DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

An Application of Correlation Clustering to Portfolio Diversification

Hannah Cheng Juan Zhan[1], William Rea[2], Alethea Rea[3]

WORKING PAPER

No.11/2014

ISSN 1179-3228
Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

An Application of Correlation Clustering to Portfolio Diversification

Hannah Cheng Juan Zhan[1], William Rea[2], Alethea Rea[3]

5/May/2014

Abstract: This paper presents a novel application of software developed for constructing a phylogenetic network to the correlation matrix for 126 stocks listed on the Shanghai A Stock Market. We show that by visualizing the correlation matrix using a Neighbor-Net network and using the circular ordering produced during the construction of the network we can reduce the risk of a diversified portfolio compared with random or industry group based selection methods in times of market increase.

Keywords: Visualization, Neighbour-Nets, Correlation Matrix, Diversification

JEL Classifications: G11

Acknowledgements:

1. Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand
2. Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand
3. Data Analysis Australia, Perth

*Corresponding Author: bill.rea@canterbury.ac.nz,

University of Canterbury WORKING PAPER No. 11/2014

An Application of Correlation Clustering to Portfolio Diversification

Hannah Cheng Juan Zhan, William Rea, and Alethea Rea

1 Introduction

Portfolio diversification is critical for risk management because it aims to reduce the variance in returns compared with a portfolio of a single stock or similarly undiversified portfolio. The academic literature on diversification is vast, stretching back at least as far as Lowenfeld (1909). The modern science of diversification is usually traced to Markowtiz (1952) which is expanded upon in great detail in Markowitz (1991).

The literature covers a wide range of approaches to portfolio diversification, such as; the number of stocks required to form a well diversified portfolio, which has increased from eight stocks in the late 1960's (Evans and Archer, 1968) to over 100 stocks in the late 2000's (Domian et al., 2007), what types of risks should be considered, (Cont, 2001; Goyal and Santa-Clara, 2003; Bali et al., 2005), factors intrinsic to each stock (Fama and French, 1992; French and Fama, 1993), the age of the investor, (Benzoni et al., 2007), whether international diversification is beneficial, (Jorion, 1985; Bai and Green, 2010), among others.

Despite the recommendation of authorites like Domian et al. (2007), Barber and Odean (2008) reported that in a large sample of American private investors the average portfolio size of individual stocks was only 4.3. While comparable data does not appear to be available for private Chinese investors, it seems unlikely that they hold substantially larger portfolios.

The mean returns and variances of the individual contributing stocks are insufficient for making informed decision on selecting a suite of stocks because selecting a portfolio requires an understanding of the correlations between each of the stocks available for consideration for inclusion in the portfolio. The number of correlations between stocks rises in proportion to the square of the number of stocks meaning that for all but the smallest of stock markets the very large number of correlations were beyond the human ability to comprehend them. Rea and Rea (2014) presented a method to visualise the correlation matrix, yielding insights into the relationships between the stocks.

Traditional investing wisdom has suggested that investors should select investment opportunities from a range of industries because returns within an industry would be more highly correlated than those between industries. While that may hold true there are some instances (such as companies with operations in several industries) in which a stock exchange industry classification alone is insufficient. Furthermore with some authors (including Domian et al. (2007)) recommending over 100 investments, the number of investments may exceed the number of industries meaning there is a need to select a diverse range of stocks even within industries.

Another key aspect of stock correlation is the potential change in the correlations with a significant change in market conditions (say comparing times of general market increase with recession and post-recession periods).

In this paper we explore investment opportunities in China using data from the Shanghai Stock Exchange. We compare the correlation structure reported in four periods (a period of market calm 2005/2006, a boom period of 2006/2007, market decline (2008), and a post crash period 2009/2010).

Our primary motivation is to investigate four portfolio selection strategies. The four strategies are;

1. picking stocks at random,
2. forming portfolios by picking stocks from different industry groups,
3. forming portfolios by picking stocks from different correlation clusters
4. forming portfolios by picking stocks from industry groups within correlation clusters.

Our results show that knowledge of correlations clusters can reduce the portfolio risk.

The outline of this paper is as follows; Section (2) discusses the data, Section (3) the methods used in this paper, Section (4) discusses identifying the correlation clusters, Section (5) discusses the movement of stocks in the Neighbor-Net splits graphs between study periods, Section (6) applies the results of the previous two sections to the problem of forming a diversified portfolio of stocks, and Section (7) contains the discussion and our conclusions.

2 Data

The data used in this study was downloaded from Datasteam. We obtained daily closing prices and dividend data for 126 stocks from the Shanghai A Index. The data listed the stock name, a six digit identification number, and assigned the stock to one of five industry groups. These groups were (1) Energy (12 stocks), (2) Finance (17 stocks), (3) Health Care (18 stocks), (4) Industrial (33 stocks), and (5) Materials (36 stocks). To make the identification of the stocks and their exchange-assigned industry groups simpler we generated four letter stock codes and to this code appended a single letter indicating its industry group. A list of these can be found in Table (5) in A. To estimate stock return correlations we calculated weekly returns from the daily price and dividend data. To obtain the period returns we calculated the total return for each period and treated the dividends as being reinvested into the stock that issued them.

A graph of the index and the boundaries of our study periods can be found in Figure (1). We defined the study periods so that they represented as different market conditons as we could make them, though it could be argued that our study periods one and four are similar.

Study period one was 13 May 2005 until 13 June 2006 and was a period in which the market underwent a slow rise. Study period two was 13 June 2006 until 16 October 2007 and is a cconsidered a boom or market bubble period. Study period the was 16 October 2007 until 28 October 2008 representing a sharp decline or crash. The final study period was from 29 October 2008 until 19 October 2010 was a time of initial market recovery and then a largely flat returns.

With four study periods, for the portfolio selection methods which require a model building, or estimatation, period we can form models in periods one through three and use the periods two through four for out-of-sample testing. Such extremely different market conditions represent a very severe test of portfolio diversification strategies, especially forming portfolios based on period two and testing them against period three data.

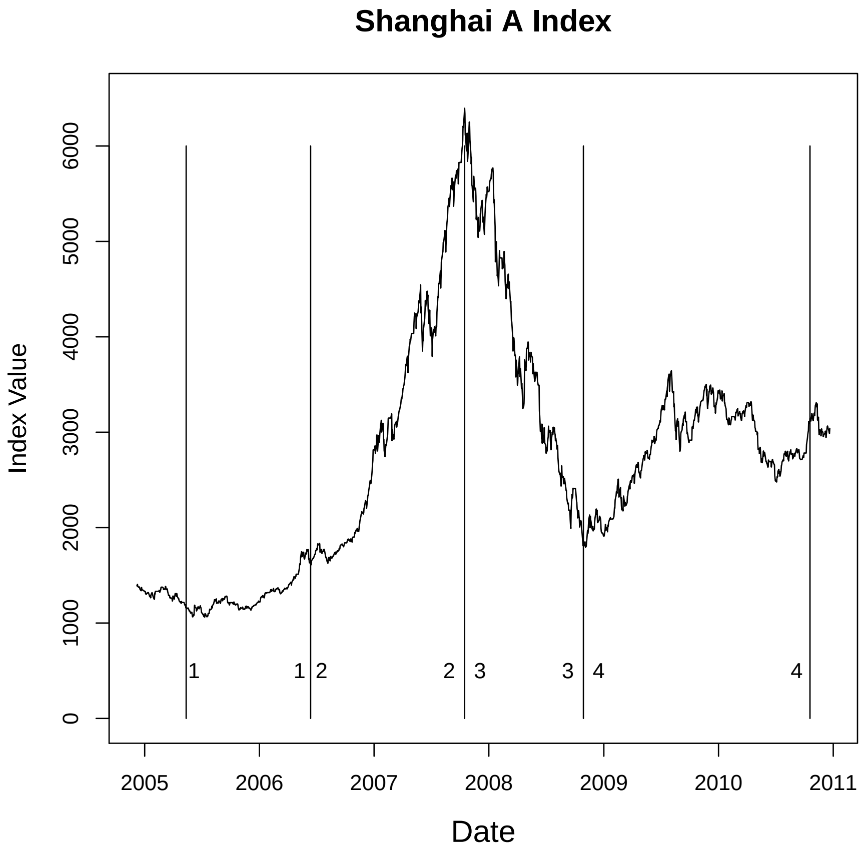

Figure 1: A plot of the Shanghai Stock Exchange A Index with the boundaries of the four study periods marked. The dates are 13-May-2005, 13-June-2006, 16-Oct-2007, 28-Oct-2008, and 19-Oct-2010 respectively.

3 Methods

3.1 Neighbor-Net Splits Graphs

A typical stock market correlation matrix for n stocks is of full rank which means that it can only be represented fully in an $(n-1)$-dimensional space. Some basic statistics on the correlations are presented in Table (1). In visualization, the high dimensional data space is collapsed to a much lower dimensional space so that the data can be represented on 2-dimensional surface such as a page or computer screen.

We need to convert the numerical values in the correlation matrix to a measure which can be construed to be a distance. In the literature the most common way to do the conversion is by using the so-called ultra-metric,

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{1}$$

where d_{ij} is the estimated distance and ρ_{ij} is the estimated correlation between stocks i and j, see Mantegna (1999) for details.

Using the conversion in Equation (1) we formatted the converted correlation matrix and augmented it with the appropriate stock codes for reading into the Neighbor-Net software, SplitsTree, available from `http://www.splitstree.org`. Using the SplitsTree software we generated the Neighbor-Nets splits graphs. Because the splits graphs are intended to be used for visualization we defer the discussion of the identification of correlation clusters and their uses to Sections (4) and (5) below.

3.2 Simulated Portfolios

Recently Lee (2011) discussed so-called risk-based asset allocation. In contrast to strategies which require both expected risk and expected returns for each investment opportunity as inputs to the portfolio selection process, risk-based allocation considers only expected risk. The four methods of portfolio selection we present below can be considered to be risk-based allocation methods. This probably reflects private investor behaviour in that often they have nothing more than broker buy, hold, or sell recommendations to assess likely returns.

The four portfolio methods were compared using simulations. For each of 1,000 iterations a portfolio was sampled based on the rules governing the portfolio type. We recorded at the mean and standard deviation of the returns for the 1,000 portfolios.

As mentioned in the introduction the primary motivation is to investigate four portfolio strategies. These are:

1. Selecting stocks at random;
2. Selecting stocks based on industry groupings;
3. Selecting stocks based on correlation clusters; and
4. Selecting stocks based on industry groups within correlation clusters.

We describe each of these in turn.

Random Selection: The stocks were selected at random using a uniform distribution with out replacement. In other words each stock was given equal chance of being selected according but with no stock being selected twice within a single portfolio.

By Industry Groups: There were five industry groups. If the portfolio size was five or less, the industries were chosen at random using a uniform distribution without replacement. From each of the selected industry groups one stock was selected. If the desired portfolio size was more than five then each group had at least s stocks selected, where s is the quotient of the portfolio size divided by five. Some (the remainder of the portfolio size divided by five) industry groups will have $s + 1$ stocks selected and the industry groups this applied to were chosen using a uniform distribution without replacement. Within each industry group stocks were selected using a uniform distribution, again without replacement.

By Correlation Clusters: The correlation clusters were determined by examining the Neighbor-Net network for the relevant periods (period one, two and three). Each stock was assigned to exactly one cluster and each cluster can be defined by a single split (or bipartition) of the circular ordering of the Neighbor-Net of the relevant period. The clusters determined in periods one, two and three were used to generate the portfolios for out-of-sample testing in periods two, thee and four respectively. Because the goal of portfolio building is to reduce risk each cluster was paired with another cluster which was considered most distant from it. This method is discussed in detail below.

As with the industry groups, if there were fewer clusters than the desired portfolio size, cluster pairs were selected at random and a stock selected from within each correlation cluster pair. If the desired portfolio size was larger than the number of correlation cluster then we apply the method described above for the industry clusters.

As indicated above each cluster was paired with the one most distant from it. Because we identified an even number of clusters in period two, cluster one was paired with cluster five, two with six and so on. In periods with an odd number of clusters the pairing may not be so straight-forward. For example, in period two (see Figure 15) we identified five clusters and cluster one was paired with four, both clusters two and three were paired with five, four was paired with one and five with two.

By Industry Group within Correlation Clusters: The final method was selecting stocks stocks from industry groups within correlation clusters. Each stock within each cluster has an associated industry group. Therefore each correlation cluster can be subdivided into up to five sub-clusters based on industry.

As indicated above each cluster was paired with the one most distant from it. Once a cluster was selected for inclusion, so was the paired cluster, however this time we did not allow any of the paired stocks to be from the same industry. This was the method used for determining the set of stocks for the fourth portfolio strategy.

4 Identifying Correlation Clusters

As Bryant and Moulton (2004) point out "the splits graphs generated by Neighbor-Net are *always* planar, an important advantage over other network methods when it comes to visualization" (emphasis original). Thus one method of identifying a group of stocks clustered by correlation is to examine the splits graph for the stocks (see, for example, Figure 2) and look for natural breaks in the structure of the network. Because this is a visual approach the results are subjective and result from the researcher or financial analyst balancing whatever competing requirements they may have. It is possible for two researchers to interpret the splits graph in different ways and identify different clusters. Here we know that in the simulations to follow the sizes of the portfolios we will generate will be two, four, eight or 16 stocks. Consequently we do not need large numbers of clusters and we would like them to have a sufficiently large number of stocks that when selecting stocks at random from within the cluster that there are a sufficiently large number of combinations available to make the simulations meaningful. These requirements are, as indicated above, subjective. For period one we chose eight clusters, which was the maximum number of clusters in any period. The smallest cluster had nine stocks giving $\binom{9}{2} = 36$ distinct ways of choosing two stocks from this cluster in the 16 stock portfolio simulation.

Figure (2) shows the clusters we identified for period one. The stocks in each cluster are listed in Appendix B, Section (B.1). Cluster one is at the bottom in black and the clusters are sequentially numbered moving counter-clockwise around the splits graph. Cluster one can be recognised by the small, but clear, gaps in the network structure between it and clusters two and eight. Similar small gaps can be seen between the other clusters.

This grouping of eight clusters is not the only division of the stocks into clusters which could have been made. If the researcher or financial analyst had other requirements some of the clusters could be further subdivided or combined. For example if small clusters were acceptible then Cluster 2 could be further split into two clusters, as could Cluster 8. In both cases there is a clear gap in the network structure where the split could be made. Conversely, if the number of clusters desired was reduced then there are some reasonably clear combinations which could be made. For example, if only two clusters were required, then, perhaps, Clusters 1, 2, 7, and 8 could be combined to form one cluster while Clusters 3, 4, 5, and 6 would form the other.

5 Movements of Stocks in the Splits Graphs between Periods

In Figures (3) through (8) we show the movement of industry groups both within a cluster and generally between study periods. We compare this with the movements of the materials industry group in the splits graph.

In Figure (3) we have selected Cluster 1 in study period 1 and assigned a colour to each industry group within the cluster. While all five industry groups are represented in the cluster it is clear that the materials group of stocks represent the largest such group within this correlation cluster. Figures (4) through (6) shows locations in the splits graph of the stocks from Cluster 1 of Period 1 in Periods 2 through 4. As can be seen the stocks in this initial cluster do not remain clustered together in subsequent periods.

However, the materials group has remained together as a block not only in study period two but also in study periods three and four. During period two (Figure 4) the materials group from Cluster 1 is now in what we identified as Cluster 3. In study period three (Figure 5) they have split into two groups and are in what we identified as Clusters 1 and 6, which are adjacent clusters in that study period. Finally in study period four they are in what we identified as Clusters 1 and 2, again, these are adjacent clusters in that study period.

In diversification one seeks groups of stocks which will tend to move together in the future but relatively independently of other so-identified groups of stocks. Then an investors spreads their investments across these groups. This is the basis for previous studies which have grouped stocks my industry assuming that stocks in the same industry will tend to have price movements more similar than stocks in different industries, see Section (5.1) below. Thus the evidence presented here is that the stocks within cluster one period one from the materials group form a financially useful grouping when considering forming a diversified portfolio.

Because of this we would not expect portfolios selected from stocks within correlation clusters alone to be significantly less risky than those chosen from industry groups. However, considering both a stock's industry group and its correlation cluster has potential to result in greater risk reduction than either method on its own.

5.1 Clustering by Industry Group

In previous studies a number of authors have included in their studies of forming diversified stock portfolios at least one method in which they dividied the stocks into industry groups and then selected portfolios by spreading the investments across the groups, see Domian et al. (2007) for example. Neighbor-Nets splits graphs give us a direct method of assessing the likely success of such a strategy. To illustrate this we have selected the energy and materials groups because they had the smallest and largest number of stocks, 12 and 36 respectively. Figures (7) through (10) show the locations of the materials stocks, while Figures (11) through (14) show the locations of the energy stocks.

Clustering of the materials stocks is clearly visible in each of the four study periods. This gives a direct visual confirmation of previous studies which have reported that selecting stocks by spreading them across industry groups gives a greater reduction in portfolio risk than randomly selecting stocks. In the smaller energy sector stocks this clustering is even clearer.

6 Example

This examples uses 126 stocks from the Shanghai exchange, for which we calculated the weekly returns from price and dividend data and we divided the data into four periods based on market behaviour as discussed in Section (2) above. Some basic statistics on the correlations are presented in Table (1). As can be seen the highest average correlation occurred in period 3, a time of a sharp market decline or crash.

For all the periods, as the portfolio size was increased the standard deviation of the returns decreased across all four portfolio selection methods. In early empirical studies of portfolio diversification focused on the number of stocks in a portfolio, see Evans and Archer (1968). A larger portfolio was reported to be less risky with the lower risk being a result of the lower level of variation in the returns. However, the benefit of reduced risk rapidly dimished with increasing portfolio size.

An ANOVA test was used to compare the means, because the variances were within a small range the ANOVA test remains valid even though the Levene test detects statistically significant differences. The Levene test was applied using the `lawstat` package in `R` (Gastwirth et al., 2013).

Period	Mean	Std. Dev.	Min	Max	Negative
1	0.266	0.170	-0.642	0.864	438/7875
2	0.328	0.196	-0.413	0.855	480/7875
3	0.441	0.191	-0.168	0.908	132/7875
4	0.437	0.192	-0.158	0.906	143/7875

Table 1: Basic statistics on the correlations. There are $n(n-1)/2 = (126 \times 125)/2 = 7875$ correlations between the 126 stocks. The final column gives the count of the number of correlations which were estimated to be negative. The highest proportion of negative correlations occurred in period 2 when approximately 6% of estimated correlations were negative.

Number of Stocks in Portfolios	Random Selection	Industry Grouping	Correlation Clusters	Industry and Correlation Clusters	ANOVA (Levene) Test p-value
2	464	449	467	457	0.0783
	(234)	(227)	(220)	(2.8)	(0.281)
4	468	459	463	4.71	0.248
	(169)	(161)	(154)	(158)	(0.041)
8	466	459	454	4.64	0.484
	(119)	(115)	(102)	(105)	(<0.001)
16	466	462	463	466	0.023
	(78)	(78)	(68)	(50)	(<0.001)

Table 2: Returns in percent under the four different portfolio selection methods for period two using period one data for the estimation of the correlations. Underneath each set of returns, in brackets, is the standard deviation of the returns. The final column reports the p-value of the ANOVA analysis which tests for differences in the means or the Levene test which tests whether the standard deviations of all four methods are equal as appropriate for each line.

Period two was a period of general market increase and the returns were good during this period. Table (2) presents the mean and standard deviations of returns together with some statistical testing of the results. The returns were statistically significantly different for portfolios of size 16 and weakly significant for portfolios of size 2. For the smallest portfolios the correlation cluster method performed best and for portfolios of size 4 and 16 the industry and correlation clusters method performed best.

For the all the portfolios the variation in the returns decreased as the portfolio size increased. The Levene test showed that there was statistically significant differences in the standard deviations for portfolios of size 4, 8 and 16. For portfolios of size 4 and 8 the correlation cluster method produced the lowest variation in the returns. For portfolios of size 16 it was the industry and correlation cluster method that produces the lowest variation, by a substantial margin.

Number of Stocks in Portfolios	Random Selection	Industry Grouping	Correlation Clusters	Industry and Correlation Clusters	ANOVA (Levene) Test p-value
2	-57	-54	-52	-53	0.007
	(25)	(27)	(29)	(27)	(0.265)
4	-58	-56	-53	-54	0.001
	(0.16)	(17)	(19)	(18)	(0.001)
8	-57	-55	-53	-54	$<$0.001
	(0.11)	(12)	(14)	(13)	(0.004)
16	-57	-54	-55	-54	$<$0.001
	(8)	(8)	(8)	(7)	($<$0.001)

Table 3: Returns in percent under the four different portfolio selection methods for period three using period two data for the estimation of the correlations. Underneath each set of returns, in brackets, is the standard deviation of the returns. The final column reports the p-value of the ANOVA analysis which tests for differences in the means or the Levene test which tests whether the standard deviations of all four methods are equal as appropriate.

Table (3) presents the mean andstandard deviations of returns together with some statistical testing of the results for period three. This was a period of general market decline. In these circumstances a widely used risk/return measure such as the Sharpe ratio is negative. In such circumstances a private investor would regard a portfolio which minimised the losses as be the most desirable. While we should not over interpret the results, the correlation clusters have slightly better returns for portfolios of sizes 2, 4 and 8. The industry and correlation clusters and industry based groupings have slightly better returns for portfolios of size 16.

As with period two out of sample testing, the variation decreased as the portfolio size was increased, regardless of the method used to select the portfolio. The Levene test showed that there was statistically significant differences in the variances in the standard deviations for portfolios of size 4, 8, and 16. Typically the correlation cluster method showed the largest standard deviations and random selection method the lowest standard deviations. For portfolios of size 16 the industry and correlation clusters method reported the smallest variation.

Table (4) presents the mean andstandard deviations of returns together with some statistical testing of the results for period four. This period showed modest

Number of Stocks in Portfolios	Random Selection	Industry Grouping	Correlation Clusters	Industry and Correlation Clusters	ANOVA (Levene) test p-value
2	2.2	211	241	237	<0.001
	(154)	(164)	(173)	(166)	(0.227)
4	229	200	235	233	<0.001
	(118)	(105)	(113)	(118)	(<0.001)
8	218	210	233	234	<0.001
	(75)	(74)	(82)	(84)	(0.003)
16	219	207	232	234	<0.001
	(53)	(50)	(53)	(41)	(<0.001)

Table 4: Returns in percent under the four different portfolio selection methods for period four using period three data for the estimation of the correlations. Underneath each set of returns, in brackets, is the standard deviation of the returns. The final column reports the p-value of the ANOVA analysis which tests for differences in the means or the Levene test which tests whether the standard deviations of all four methods are equal as appropriate.

returns. While, again, we should not over-interpret the results, the returns were lower for random and industry grouping selection methods for all four portfolios sizes tested. The highest returns were for the correlation clusters portfolio selection method for the two smaller portfolios, and for portfolios of size 8 and 16 the industry and correlation clusters method reports slightly higher returns.

As with period two and three out of sample testing, the variation decreased as the portfolio size was increased, regardless of the method used to select the portfolio. The Levene test showed that there was statistically significant differences in the variances for the portfolios of sizes 4, 8 and 16. The industry based selection method offered the greatest reduction in the variation in the returns for portfolios of size 4 and 8. For the largest portfolio size (portfolios of size 16) the industry and correlation clusters had the lowest standard deviations (the same outcome as periods two and three).

Therefore this suggests that the correlation clusters (or industry and correlation clusters) are particularly effective in times of general market increase, with the benefit being either a reduction in the variation or an increase in the return.

This study shows that combining industry and correlation clusters is particularly effective at lowering the variation for the larger portfolios, with all three period showing a much lower variation for portfolios of size 16, as well as reasonable returns. This is in line with general advice to investors to hold larger portfolios and to ensure the holdings are diversified.

7 Discussion

An earlier paper (Rea and Rea, 2014) introduced Neighbor-Net networks as a method for visualising correlations in stock markets. The method has the advantage of being able to represent a lot of the key features of the correlation matrix in a planar graphic. The paper noted that such a diagram could assist with creating diversified portfolios. This paper has highlighted the effectiveness of using correlation clusters to investigate diversified portfolios.

In this paper four risk budgeting methods of portfolio selection were compared; randomly selected portfolios, industry clusters, correlation clusters and industry and correlation clusters. Traditionally selecting stocks by industry was considered an appropriate method to diversify a portfolio. While this may be the case in some markets and under some market conditions, this investigation demonstrated that industry based clusters was generally outperformed by portfolios selected at random, however the portfolios selected using industry grouping may have lower variance in times of market increase compared with random selection.

Of the four, the most restrictive method of selecting portfolios was the industry and correlation cluster selection method. With the random selection method all possible combinations of n stocks from the 126 stocks are allowable but for the industry and correlation cluster selection method, there are many portfolios that are not admissible because they do not meet the rules of this portfolio selection method. The industry grouping and correlation cluster methods are also restrictive but less so than the industry and correlation clusters method.

The main concern was whether the rules of portfolio selection presented here offer significant benefits. If a difference in mean was detected, the correlation clusters or industry and correlation clusters method may outperform the other methods on mean return. This effect was most pronounced in the period four out of sample testing where the returns for the correlations clusters and industry and correlation clusters method always exceeded random portfolio selection. Therefore the knowledge of the circular ordering can be used to enhance portfolio returns.

The variation in the returns for portfolios of size 16 was always lowest if the method of portfolio selection was Industry and Correlation Cluster selection. For the other portfolio sizes the variation with a method decrease as the portfolio size increases, but no one method consistently outperforms the others. This suggests portfolio size has a greater impact on the variation of the returns than the method used to select of the portfolio.

Rea and Rea (2014) discussed how stocks from opposite side of the Neighbor-Net network did not necessarily create a portfolio with high returns because some stocks maybe giving negative returns while one on the opposite side of the network may be giving positive returns. The division of the data into four periods in the manner we did, represents a particularly severe test of diversifiction, particularly since no account was taken of either historical or expected returns of the stocks. It is our expectation that investor knowledge and analysis alongside correlation cluster based portfolio selection has the potential to improve the return of the portfolio, as well as reduce the variance (or equivalently, the standard deviation).But this awaits further research.

We note that the correlation clusters were determined by eye in this analysis. This is a valid method of determining clusters, however it is also subjective. It does offer the advantage of being able to assign each stock to exactly one cluster. Future work could focus on methods to automate the selection of the correlation clusters to see if this further enhances the portfolio performance.

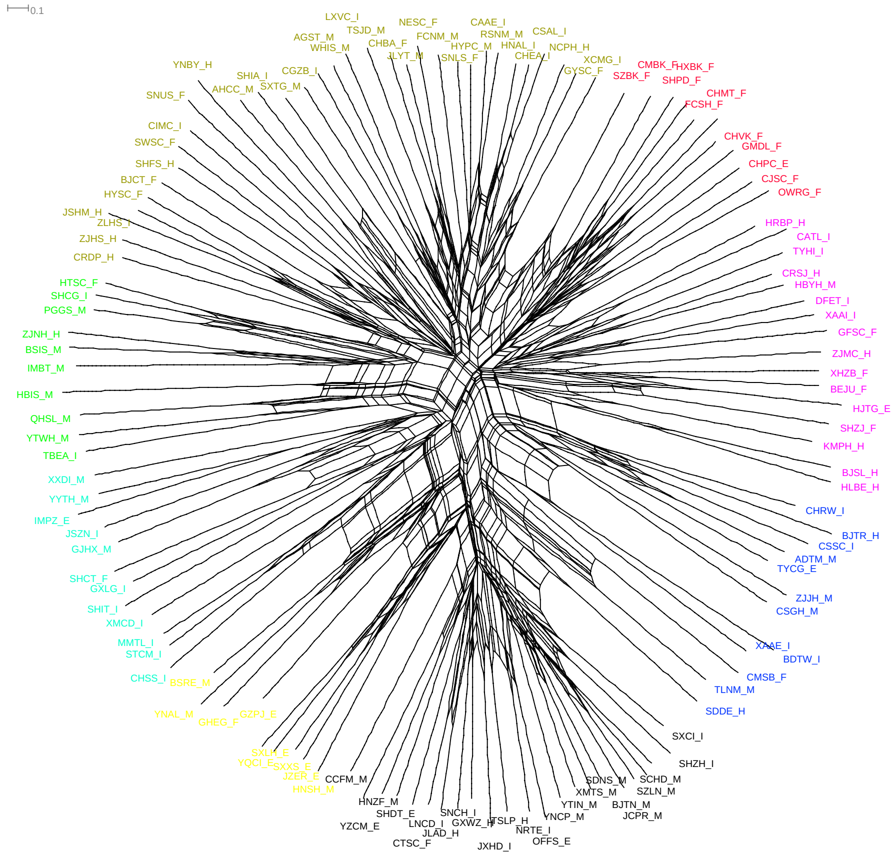

Figure 2: SplitsTree network for 126 stocks from the Shanghai A Stock Exchange for period one using five trading day returns to estimate correlations and hence distances with the stocks in cluster one colour coded. The eight correlation clusters each have different colours. In the discussion the clusters are coded as follows; Cluster 1 – Black, Cluster 2 – Blue, Cluster 3 – Purple, Cluster 4 – Red, Cluster 5 - Khaki, Cluster 6 – Green, Cluster 7 – Aqua, Cluster 8 – Yellow.

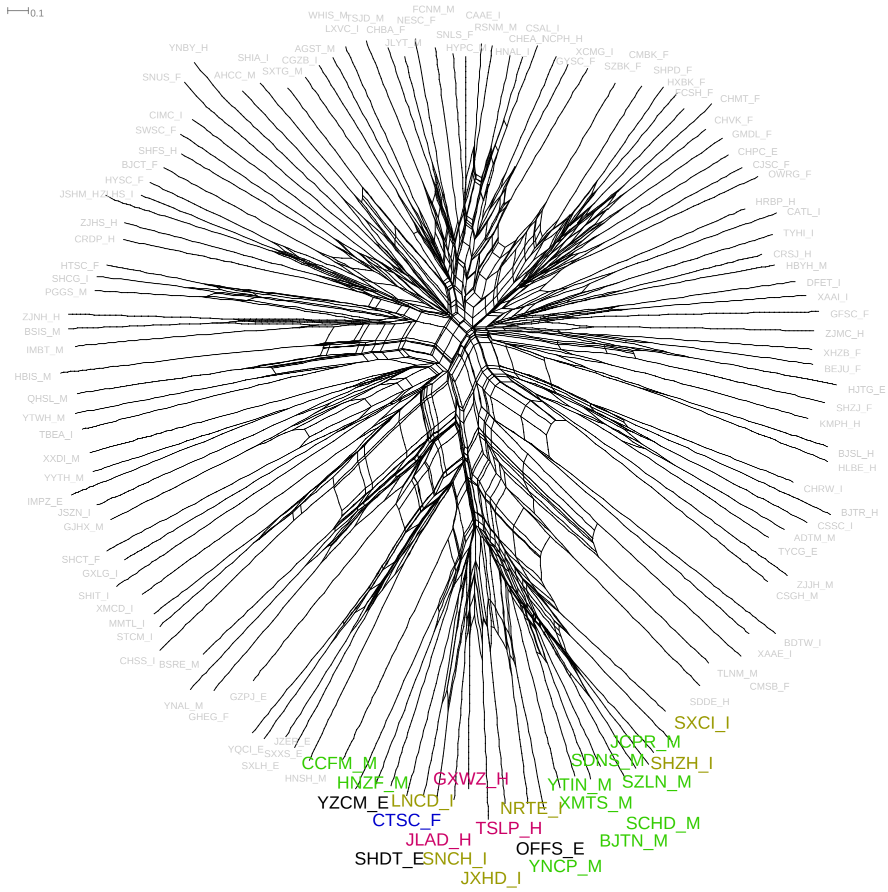

Figure 3: The SplitsTree network for the Shanghai A Stock Exchange for period one with the stocks in cluster one colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

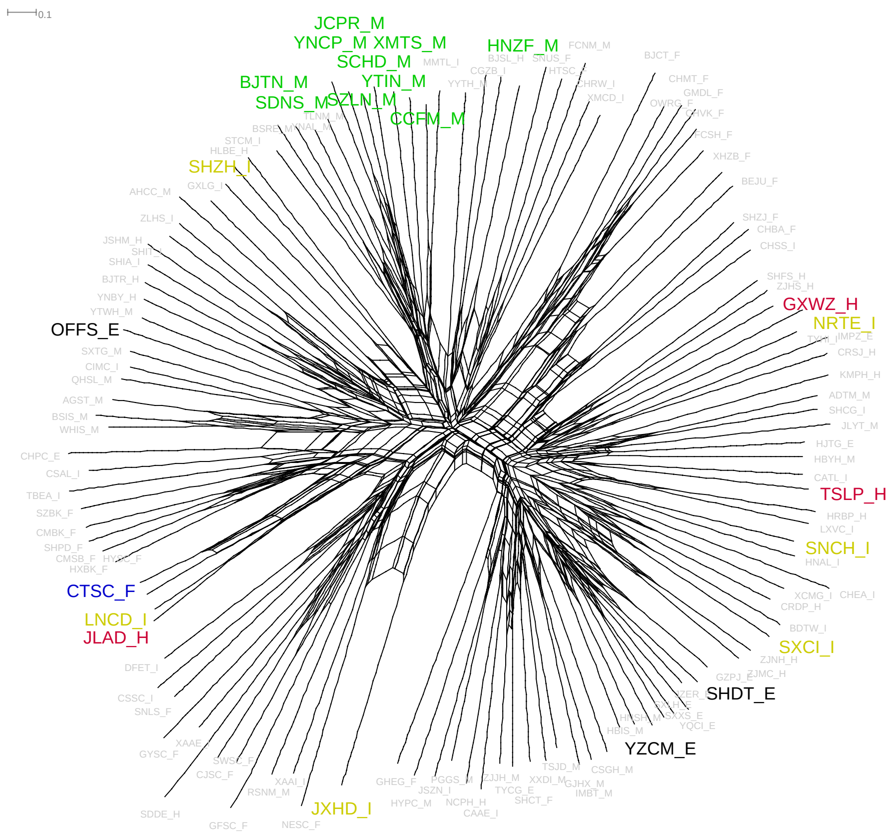

Figure 4: SplitsTree network for study period two with the stocks from cluster one, period one coloured. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

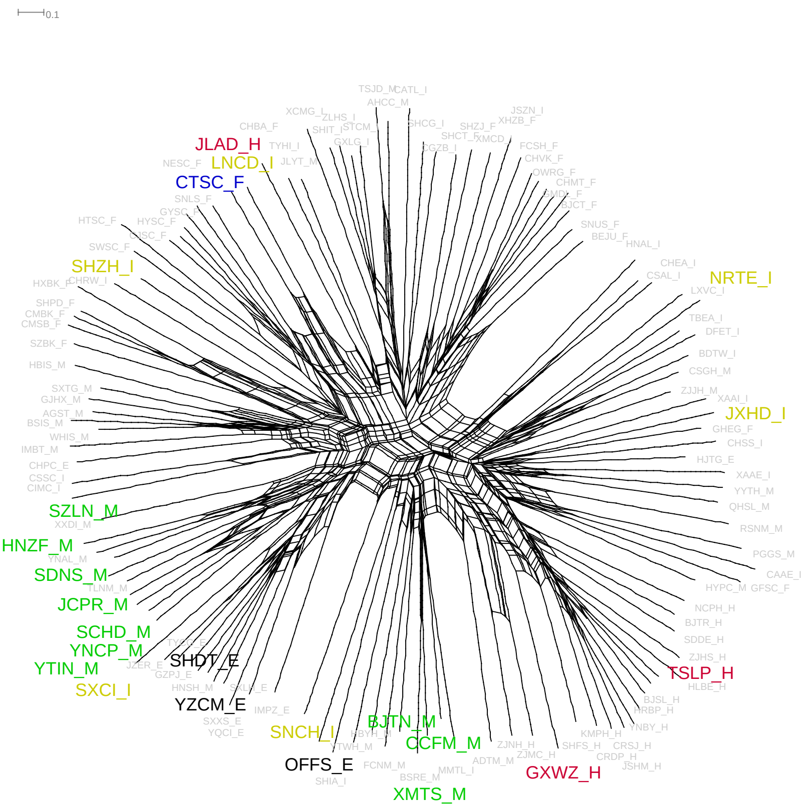

Figure 5: SplitsTree network for study period three with the stocks in cluster one, period one coloured. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

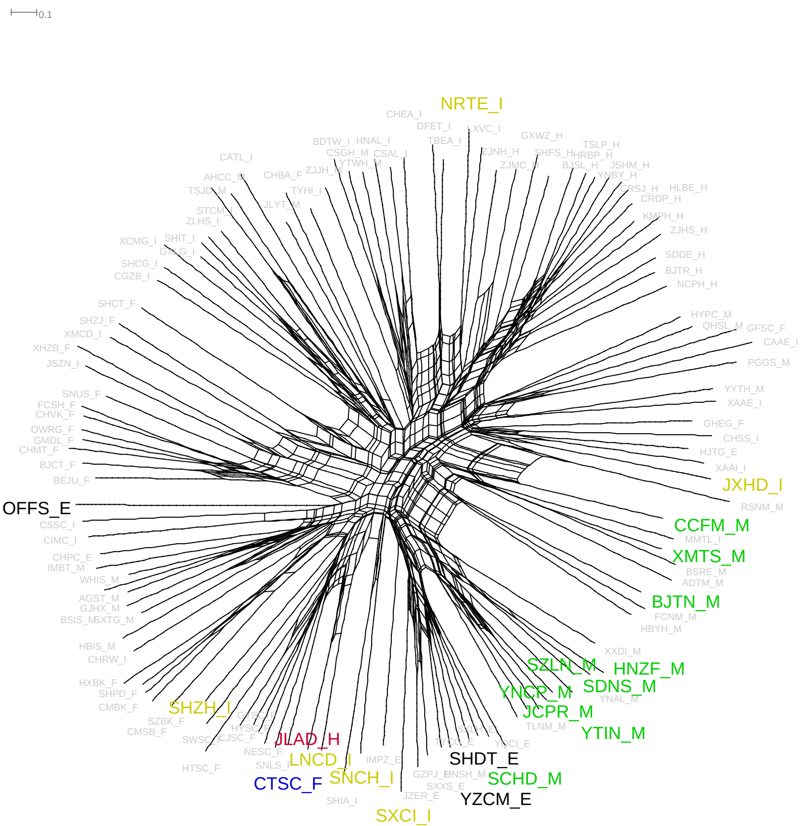

Figure 6: SplitsTree network for study period four with the stocks in cluster one, period one coloured. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

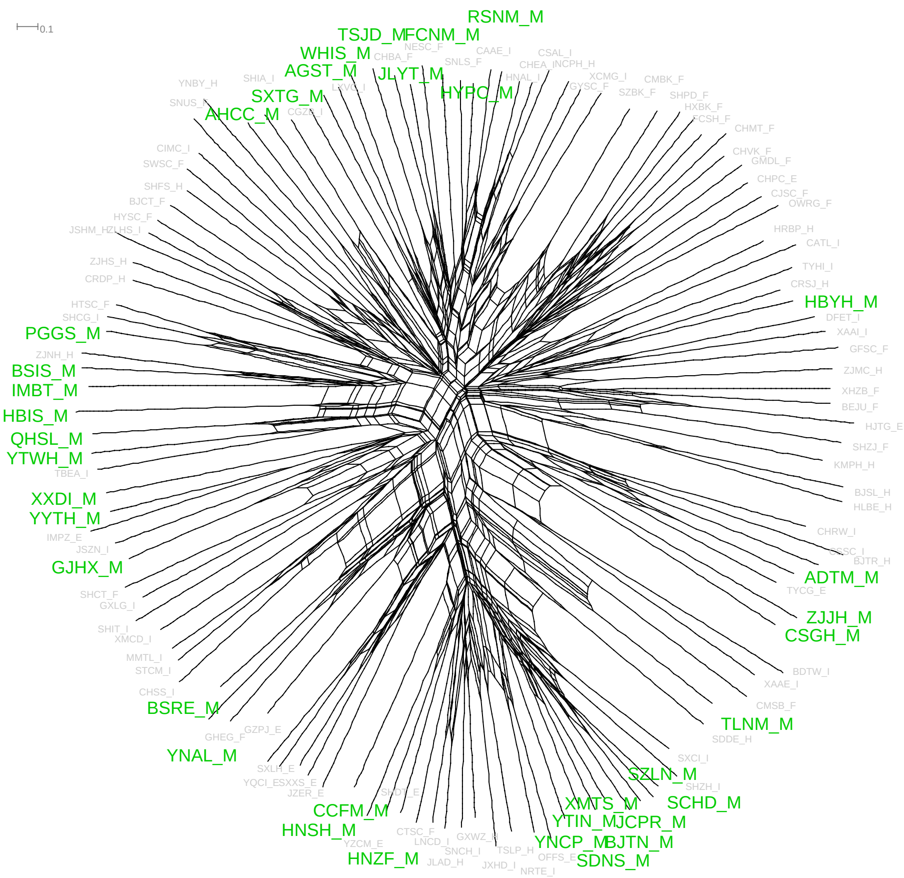

Figure 7: SplitsTree network for study period one with the stocks in the materials sector coloured green.

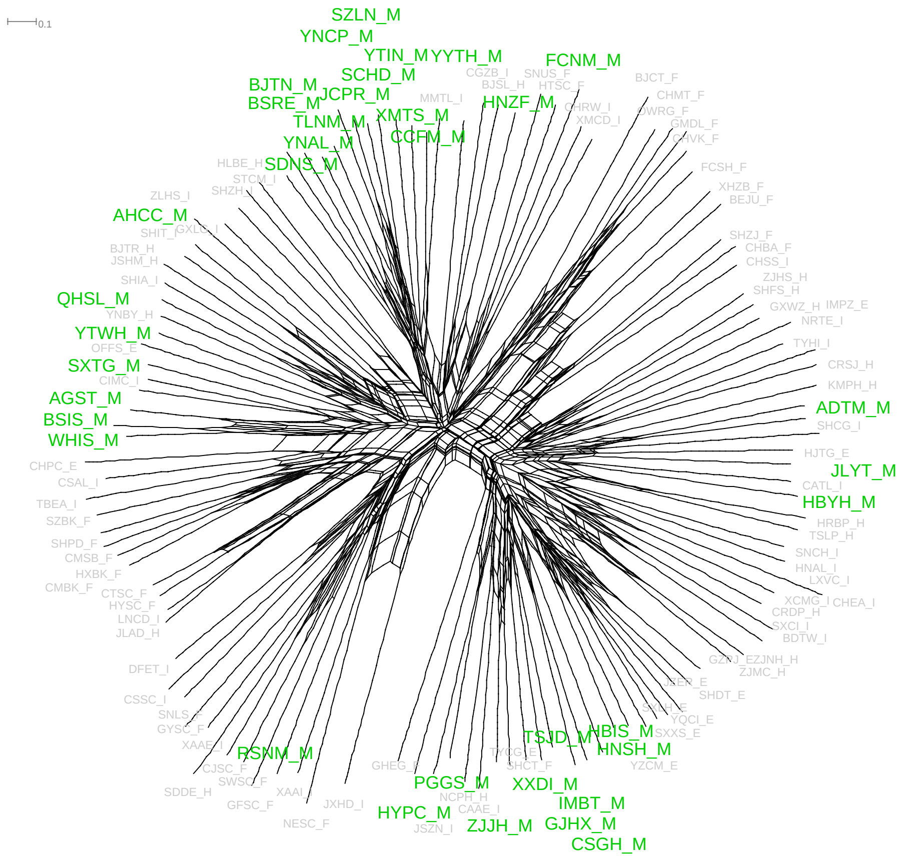

Figure 8: SplitsTree network for study period two with the stocks in the materials sector coloured green.

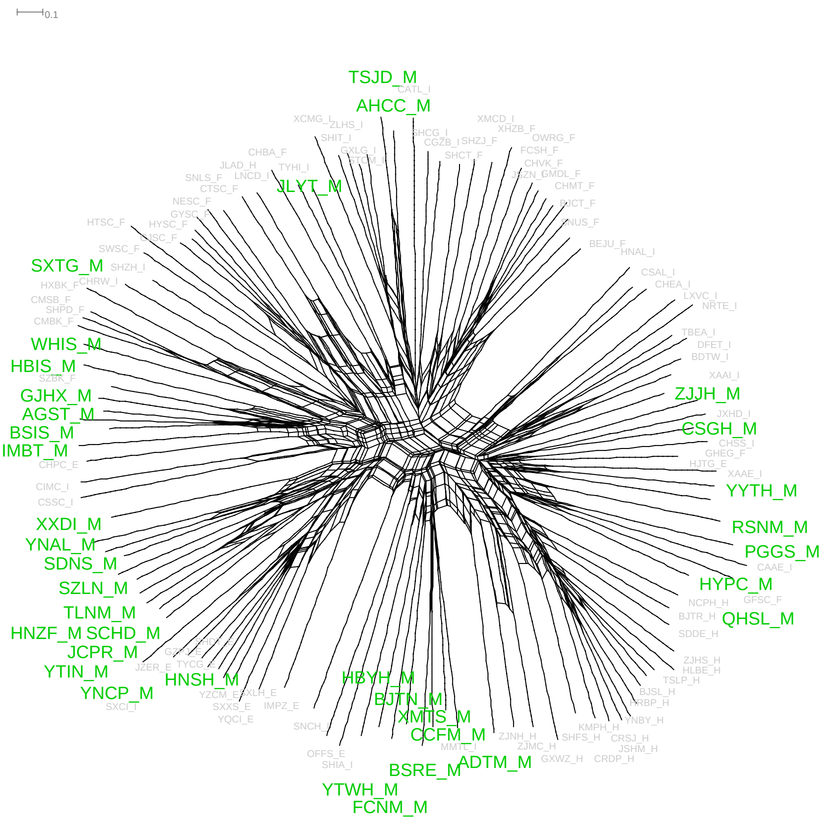

Figure 9: SplitsTree network for study period three with the stocks in the materials sector coloured green.

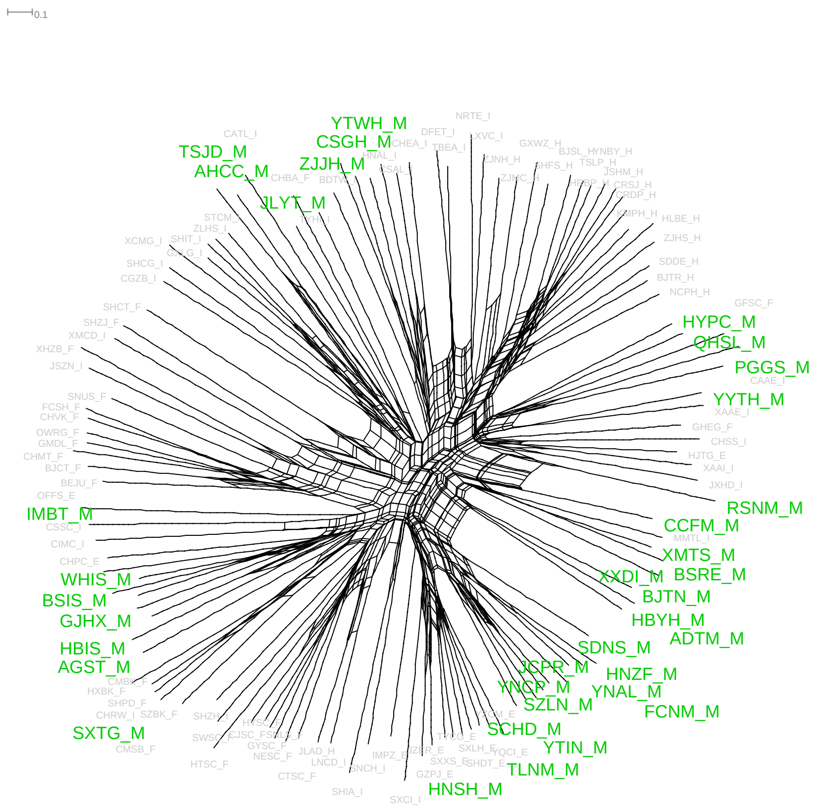

Figure 10: SplitsTree network for study period four with the stocks in the materials sector coloured green.

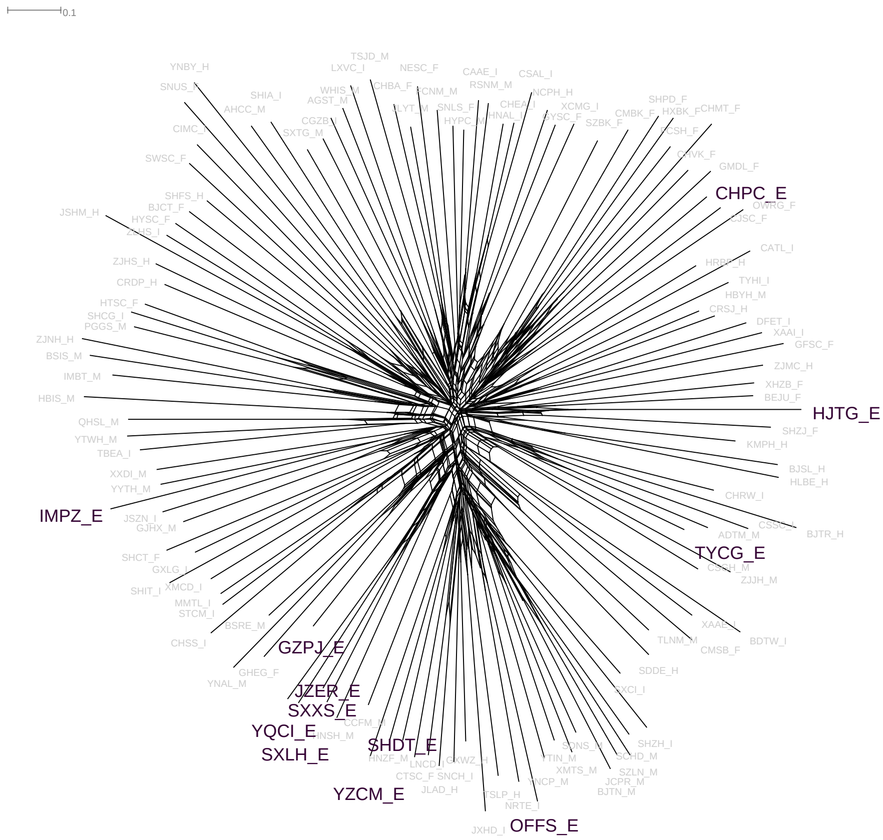

Figure 11: SplitsTree network for study period one with the stocks in the energy sector coloured black.

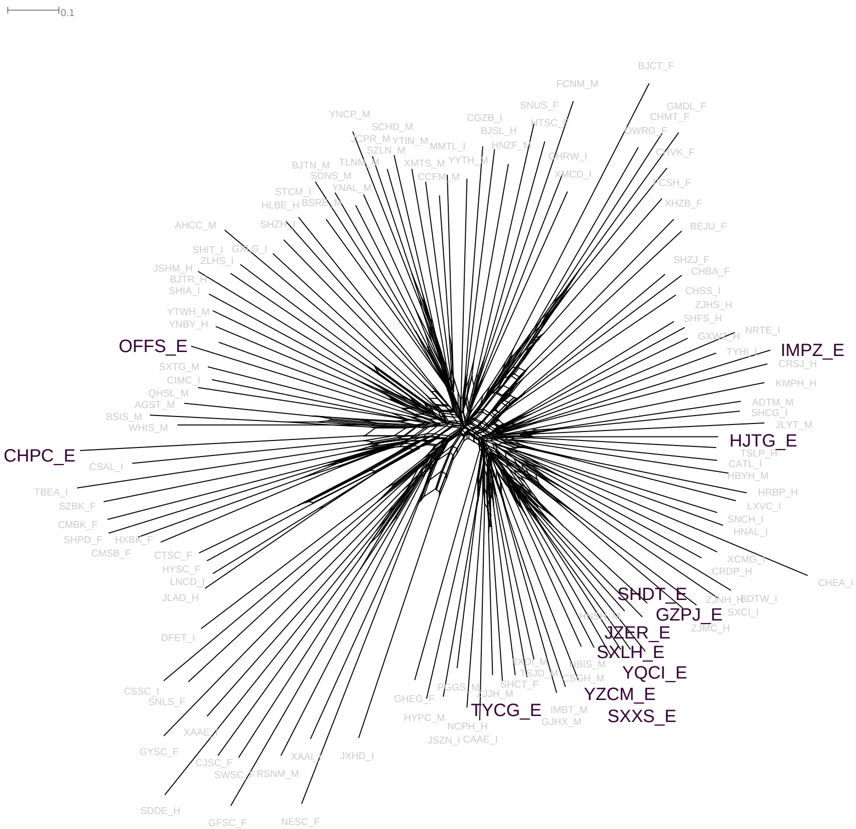

Figure 12: SplitsTree network for study period two with the stocks in the energy sector coloured black.

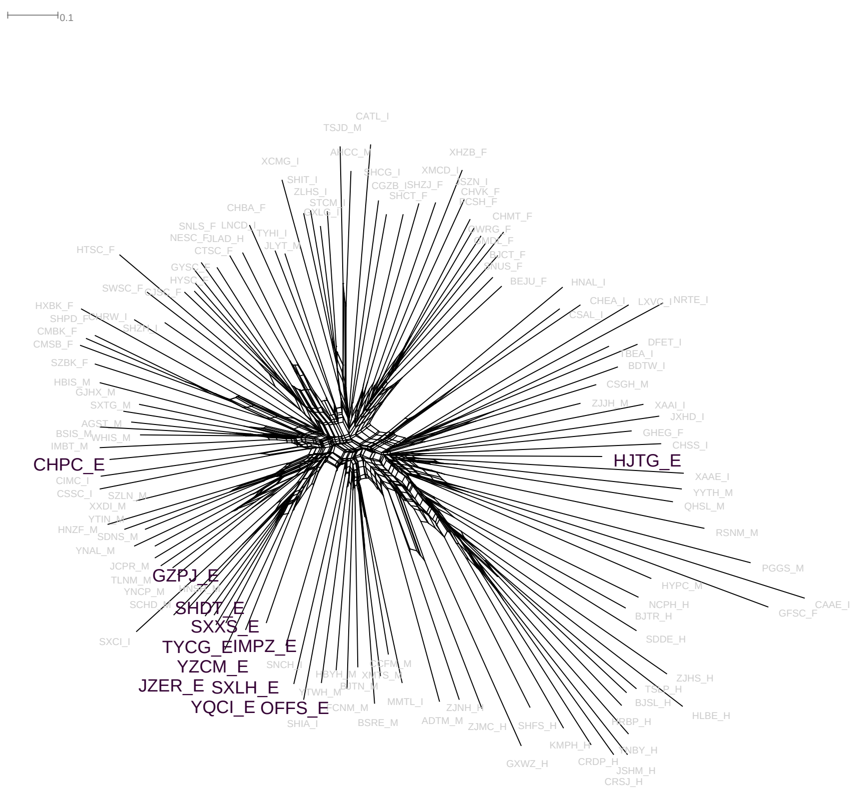

Figure 13: SplitsTree network for study period three with the stocks in the energy sector coloured black.

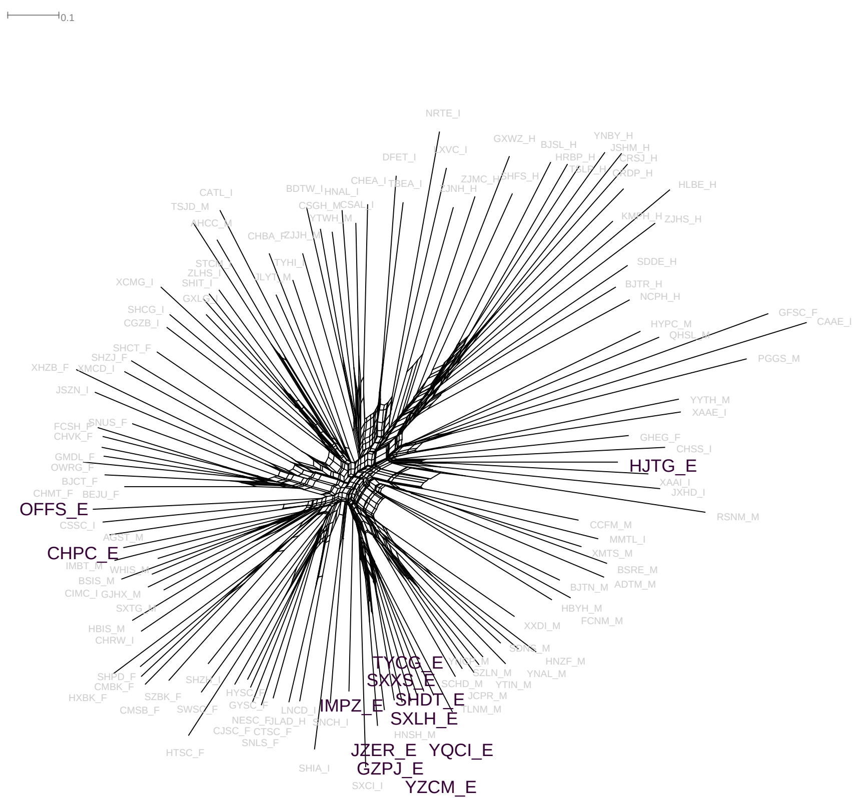

Figure 14: SplitsTree network for study period four with the stocks in the energy sector coloured black.

References

Bai, Y. and C. J. Green (2010). International Diversification Strategies: Revisited from the Risk Perspective. *The Journal of Banking and Finance 34*, 236–245.

Bali, T. G., N. Cakici, X. Yan, and Z. Zhang (2005). Does Idiosyncratic Risk Really Matter? *The Journal of Finance 60*(2), 905–929.

Barber, B. M. and T. Odean (2008). All That Glitters: The Effect of Attention and News on the Buying Behaviour of Individual and Institutional Investors. *The Review of Financial Studies 21*(2), 785–818.

Benzoni, L., P. Collin-Dufresne, and R. S. Goldstein (2007). Portfolio Choice over the Life-Cycle when the Stock and Labor Markets are Cointegrated. *The Journal of Finance 62*(5), 2123–2167.

Bryant, D. and V. Moulton (2004). Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution 21*(2), 255–265.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1:2*, 223–236.

Domian, D. L., D. A. Louton, and M. D. Racine (2007). Diversification in Portfolios of Individual Stocks: 100 Stocks Are Not Enough. *The Financial Review 42*, 557–570.

Evans, J. L. and S. H. Archer (1968). Diversification and the Reduction of Dispersion: An Empirical Analysis. *The Journal of Finance 23*(5), 761–767.

Fama, E. F. and K. R. French (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance 47*(2), 427–465.

French, K. R. and E. F. Fama (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financal Economics 33*, 3–56.

Gastwirth, J. L., Y. R. Gel, W. L. W. Hui, V. Lyubchich, W. Miao, and K. Noguchi (2013). *lawstat: An R package for biostatistics, public policy, and law.* R package version 2.4.1.

Goyal, A. and P. Santa-Clara (2003). Idiosyncratic Risk Matters! *The Journal of Finance 58*(3), 975–1007.

Jorion, P. (1985). International Portfolio Diversification with Estimation Risk. *The Journal of Business 58*(3), 259–278.

Lee, W. (2011). Risk-Based Asset Allocation: A New Answer to an Old Question? *Journal of Portfolio Management 37*(4), 11–28.

Lowenfeld, H. (1909). *Investment, an Exact Science*. Financial Review of Reviews.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B 11*, 193–197.

Markowitz, H. M. (1991). *Portfolio Selection: Efficient Diversification of Investments 2nd Edition*. Wiley.

Markowtiz, H. (1952). Portfolio Selection. *The Journal of Finance 7*(1), 77–91.

Rea, A. and W. Rea (2014). Visualization of a stock market correlation matrix. *Physica A 400*, 109–123.

A Stock Codes and Industry Segments

Table 5: Stock market codes and company names and Industrial sector of stocks in the study.

Company Name	Company Code	Industry Group
China Ptl. & Chm.	CHPC-E	Energy
Guizhou Panjiang Coal	GZPJ-E	Energy
Inner Mongolia Pingzhuang En. Rso.	IMPZ-E	Energy
Jizhong Energy Res.	JZER-E	Energy
Liaoning Hjtg. Chems.	HJTG-E	Energy
Offs. Oil Engr.	OFFS-E	Energy
Shai Datun Energy Res.	SHDT-E	Energy
Shanxi Lanhua Sci-Tech Venture	SXLH-E	Energy
Shanxi Xishan	SXXS-E	Energy
Taiyuan Coal Gasification	TYCG-E	Energy
Yangquan Coal	YQCI-E	Energy
Yanzhou Coal Mining	YZCM-E	Energy
Beijing Capital Dev.	BJCT-F	Finance
Bej. Urban Con. Inv. Dev.	BEJU-F	Finance
Changjiang Securities	CJSC-F	Finance
China Baoan Gp.	CHBA-F	Finance
China Merchants Bank	CMBK-F	Finance
China Merchants Pr. Dev.	CHMT-F	Finance
China Minsheng Banking	CMSB-F	Finance
China Vanke	CHVK-F	Finance
Citic Securities	CTSC-F	Finance
Financial Str. Sldg.	FCSH-F	Finance
Gemdale	GMDL-F	Finance
GF Securities	GFSC-F	Finance
Guanghui Energy	GHEG-F	Finance
Guoyuan Securities	GYSC-F	Finance
Haitong Securities	HTSC-F	Finance
Hong Yuan Secs.	HYSC-F	Finance
Huaxia Bank	HXBK-F	Finance
Northeast Securities	NESC-F	Finance
Oceanwide Rlst. Group	OWRG-F	Finance
Shai. Chengtou Hldg.	SHCT-F	Finance
Shai. Pudong Dev. Bk.	SHPD-F	Finance
Shai. Zhangjiang	SHZJ-F	Finance
Shenzhen Dev. Bank	SZBK-F	Finance
Sinolink Securities	SNLS-F	Finance

Table 5: Stock market codes and company names and Industrial sector of stocks in the study.

Company Name	Company Code	Industry Group
Southwest Securities	SWSC-F	Finance
Suning Universal	SNUS-F	Finance
Xinhu Zhongbao	XHZB-F	Finance
Beijing Sl Pharmaceutical	BJSL-H	Health Care
Beijing Tongrentang	BJTR-H	Health Care
China Res. Dble. Crane Pharm.	CRDP-H	Health Care
China Res. Sanjiu Med.& pharm.	CRSJ-H	Health Care
Guangxi Wuzhou Zhongheng	GXWZ-H	Health Care
Harbin Pharms. Gp.	HRBP-H	Health Care
Hualan Biological Engr.	HLBE-H	Health Care
Jiangsu Hengrui Medicine	JSHM-H	Health Care
Jilin Aodong Pharm. Gp.	JLAD-H	Health Care
Kangmei Pharm.	KMPH-H	Health Care
North China Pharm.	NCPH-H	Health Care
Shai. Fosun Pharm. Group	SHFS-H	Health Care
Shan Dong Dong E-Jiao	SDDE-H	Health Care
Tasly Pharmaceutical	TSLP-H	Health Care
Yunnan Baiyao Gp.	YNBY-H	Health Care
Zhejiang Hisun Pharm.	ZJHS-H	Health Care
Zhejiang Medicine	ZJMC-H	Health Care
Zhejiang Nhu	ZJNH-H	Health Care
Baoding Tianwei Baobian Elec.	BDTW-I	Industrial
China Avic Avionics Equ.	CAAE-I	Industrial
China Cssc Hdg.	CSSC-I	Industrial
China Eastern Airl.	CHEA-I	Industrial
China Gezhouba Group	CGZB-I	Industrial
China Intl.Mar.Ctrs.	CIMC-I	Industrial
China Railway Erju	CHRW-I	Industrial
China Railway Tielong Container Logistic	CATL-I	Industrial
China Southern Airlines	CSAL-I	Industrial
China Spacesat	CHAA-I	Industrial
Dongfang Electric	DFET-I	Industrial
Guangxi Liugong Mch	GXLG-I	Industrial
Hainan Airlines	HNAL-I	Industrial
Jiangsu Zhongnan Con.	JSZN-I	Industrial
Jiangxi Hongdu Aviation	JXHD-I	Industrial
Liaoning Chengda	LNCD-I	Industrial
Luxin Venture Cap. Gp.	LXVC-I	Industrial
Minmetals Dev.	MMTL-I	Industrial

Table 5: Stock market codes and company names and Industrial sector of stocks in the study.

Company Name	Company Code	Industry Group
Nari Tech. Dev.	NRTE-I	Industrial
Sany Heavy Industry	SHIT-I	Industrial
Shai. Shenhua Heavy Ind.	SHZH-I	Industrial
Shanghai Con. Group	SHCG-I	Industrial
Shanghai Intl. Arpt.	SHIA-I	Industrial
Shantui Con. Mch.	STCM-I	Industrial
Shanxi Coal Intl.	SXCI-I	Industrial
Sinochem Intl.	SNCH-I	Industrial
Taiyuan Hvy. Ind.	TYHI-I	Industrial
Tbea	TBEA-I	Industrial
Xcmg Con. Machinery	XCMG-I	Industrial
Xi'an Aero-Engine	XAAE-I	Industrial
Xi'an Air.Intl.	XAAI-I	Industrial
Xiamen C & D	XCMD-I	Industrial
Zoomlion Hdy. Sctc.	ZLHS-I	Industrial
Advd. Tech.& Mats.	ADTM-M	Materials
Angang Steel	AGST-M	Materials
Anhui Conch Cmt.	AHCC-M	Materials
Baoji Titanium Ind.	BJTN-M	Materials
Baoshan Iron & Stl.	BSIS-M	Materials
China Nonferrous Mtl.	CCFM-M	Materials
Csg Holding	CSGH-M	Materials
Fangda Cbn. New Mra.	FCNM=M	Materials
Gan Jiu Stl. Gp. Hongxing	GJHX-M	Materials
Ginghai Salt Lake Ind.	QHSL-M	Materials
Industrial Sichuan Hongda	SCHD-M	Materials
Inmong. Baotou Stl. Rare Earth	BSRE-M	Materials
Hebei Iron & Steel	HBIS-M	Materials
Henan Shenhuo Caa. & Pwr.	HNSH-M	Materials
Henan Zhongfu Indl.	HNZF-M	Materials
Hengyi Petrochemical	HYPC-M	Materials
Hubei Yihua Chm. Ind.	HBYH-M	Materials
Inner Mongolia Baotou Steel Union	IMBT-M	Materials
Jiangxi Cpr.	JCPR-M	Materials
Jilin Yatai Group	JLYT-M	Materials
Pangang Gp. Stl. Vmtm.	PGGS-M	Materials
Rising Nonfr. Mtls	RSNM-M	Materials
Shandong Nanshan Almn.	SDNS-M	Materials
Shanxi Taigang Stl.	SXTG-M	Materials

Table 5: Stock market codes and company names and Industrial sector of stocks in the study.

Company Name	Company Code	Industry Group
Shn. Zhongjin Lingnan Nonfemet	SZLN-M	Materials
Tangshan Jidong Cmt.	TSJD-M	Materials
Tongling Nonfr. Mtls. Gp.	TLNM-M	Materials
Xiamen Tungsten	XMTS-M	Materials
Xinxing Ductile Iron	XXDI-M	Materials
Yantai Wanhua Polyuretha	YTWH-M	Materials
Yunnan Alum.	YNAL-M	Materials
Yunnan Copper	YNCP-M	Materials
Yunnan Tin	YTIN-M	Materials
Yunnan Yuntianhua	YYTH-M	Materials
Wuhan Iron and Steel	WHIS-M	Materials
Zhejiang Juhua	ZJJH-M	Materials

B Stocks in Each Cluster

B.1 Period 1

Cluster1: YZCM_E, SXCI_I, OFFS_E, SCHD_M, JCPR_M, SHZH_I, YNCP_M, JXHD_I, CCFM_M, BJTN_M, GXWZ_H, SNCH_I, XMTS_M, YTIN_M, HNZF_M, JLAD_H, TSLP_H, LNCD_I, NRTE_I, CTSC_F, SZLN_M, SHDT_E, SDNS_M

Cluster2: XAAE_I, CSSC_I, TYCG_E, CSGH_M, ADTM_M, SDDE_H, CMSB_F, BDTW_I, CHRW_I, BJTR_H, TLNM_M, ZJJH_M

Cluster3: BEJU_F, KMPH_H, ZJMC_H, HJTG_E, SHZJ_F, XAAI_I, TYHI_I, CATL_I, BJSL_H, CRSJ_H, DFET_I, HLBE_H, XHZB_F, HBYH_M, HRBP_H, GFSC_F

Cluster4: CMBK_F, CJSC_F, OWRG_F, HXBK_F, CHMT_F, FCSH_F, SZBK_F, GMDL_F, CHPC_E, SHPD_F, CHVK_F

Cluster5: JLYT_M, CSAL_I, SNUS_F, NESC_F, AHCC_M, ZLHS_I, HNAL_I, XCMG_I, GYSC_F, SXTG_M, BJCT_F, ZJHS_H, JSHM_H, CRDP_H, CGZB_I, FCNM_M, SNLS_F, TSJD_M, YNBY_H, WHIS_M, SHFS_H, CHEA_I, CAAE_I, HYPC_M, CHBA_F, SWSC_F, HYSC_F, CIMC_I, AGST_M, RSNM_M, SHIA_I, NCPH_H, LXVC_I

Cluster6: BSIS_M, HTSC_F, HBIS_M, SHCG_I, PGGS_M, IMBT_M, TBEA_I, QHSL_M, ZJNH_H, YTWH_M

Cluster7: YYTH_M, XMCD_I, CHSS_I, JSZN_I, XXDI_M, GXLG_I, SHIT_I, SHCT_F, STCM_I, GJHX_M, IMPZ_E, MMTL_I

Cluster8: YNAL_M, SXLH_E, BSRE_M, JZER_E, SXXS_E, YQCI_E, GHEG_F, HNSH_M, GZPJ_E

B.2 Period 2

Cluster1 XXDI_M, SXCI_I, KMPH_H, ZJMC_H, JLYT_M, SHCT_F, IMPZ_E, ZJNH_H, HJTG_E, BDTW_I, GHEG_F, TYHI_I, SXLH_E, TYCG_E, HNAL_I, IMBT_M, TSLP_H, SXXS_E, NRTE_I, XCMG_I, ZJHS_H, HBYH_M, HRBP_H, HNSH_M, CRDP_H, JSZN_I, YZCM_E, SHCG_I, TSJD_M, CSGH_M, ADTM_M, PGGS_M, SHFS_H, CHEA_I, CAAE_I, HYPC_M, ZJJH_M, HBIS_M, GXWZ_H, CATL_I, SNCH_I, GZPJ_E, CRSJ_H, JZER_E, NCPH_H, LXVC_I, GJHX_M, SHDT_E

Cluster2 CHSS_I, OWRG_F, BEJU_F, CHMT_F, FCSH_F, GMDL_F, XHZB_F, BJCT_F, CHBA_F, SHZJ_F, CHVK_F

Cluster3 CGZB_I, FCNM_M, JCPR_M, SCHD_M, YNCP_M, SNUS_F, CHRW_I, TLNM_M, CCFM_M, BJTN_M, HTSC_F, XMCD_I, YNAL_M, YYTH_M, BJSL_H, XMTS_M, BSRE_M, YTIN_M, HNZF_M, MMTL_I, SZLN_M, SDNS_M

Cluster4 CMBK_F, GXLG_I, SHZH_I, CSAL_I, TBEA_I, QHSL_M, BJTR_H, BSIS_M, AHCC_M, SHIT_I, ZLHS_I, STCM_I, LNCD_I, SXTG_M, YTWH_M, JSHM_H, HXBK_F, OFFS_E, SZBK_F, YNBY_H, WHIS_M, CMSB_F, SHPD_F, HYSC_F, CIMC_I, AGST_M, SHIA_I, JLAD_H, HLBE_H, CHPC_E, CTSC_F

Cluster5, CJSC_F, SNLS_F, SDDE_H, JXHD_I, SWSC_F, XAAI_I, NESC_F, XAAE_I, CSSC_I, RSNM_M, DFET_I, GYSC_F, GFSC_F

B.3 Period, 3

Cluster1 FCNM_M, SNCH_I, XMTS_M, SHIA_I, BSRE_M, OFFS_E, MMTL_I, YTWH_M, CCFM_M, HBYH_M, BJTN_M

Cluster2 JSHM_H, CRDP_H, KMPH_H, ADTM_M, ZJMC_H, SDDE_H, YNBY_H, SHFS_H, ZJNH_H, BJTR_H, GXWZ_H, BJSL_H, CRSJ_H, NCPH_H, TSLP_H, HLBE_H, ZJHS_H, HRBP_H

Cluster3 CSGH_M, PGGS_M, CSAL_I, TBEA_I, QHSL_M, CHEA_I, CAAE_I, HYPC_M, HJTG_E, BDTW_I, GHEG_F, JXHD_I, XAAI_I, ZJJH_M, YYTH_M, CHSS_I, XAAE_I, RSNM_M, HNAL_I, LXVC_I, DFET_I, NRTE_I, GFSC_F

Cluster4 CGZB_I, JSZN_I, BEJU_F, SHCG_I, GXLG_I, CHMT_F, TSJD_M, FCSH_F, SHCT_F, SNUS_F, GMDL_F, SHZJ_F, XMCD_I, OWRG_F, CATL_I, AHCC_M, SHIT_I, ZLHS_I, STCM_I, XHZB_F, XCMG_I, BJCT_F, CHVK_F

Cluster5 CMBK_F, CJSC_F, JLYT_M, SHZH_I, TYHI_I, NESC_F, BSIS_M, HTSC_F, IMBT_M, LNCD_I, GYSC_F, SXTG_M, HXBK_F, SNLS_F, SZBK_F, WHIS_M, CMSB_F, CHRW_I, SHPD_F, CHBA_F, SWSC_F, HYSC_F, HBIS_M, CIMC_I, CSSC_I, AGST_M, JLAD_H, GJHX_M, CHPC_E, CTSC_F

Cluster6 YZCM_E, XXDI_M, SXCI_I, JCPR_M, SCHD_M, YNCP_M, IMPZ_E, YQCI_E, TLNM_M, YNAL_M, SXLH_E, TYCG_E, GZPJ_E, YTIN_M, JZER_E, HNZF_M, SXXS_E, SZLN_M, SDNS_M, HNSH_M, SHDT_E

B.4 Period 4

Cluster1 YZCM_E, XXDI_M, SXCI_I, JCPR_M, SCHD_M, YNCP_M, IMPZ_E, YQCI_E, TLNM_M, YNAL_M, SXLH_E, TYCG_E, GZPJ_E, YTIN_M, JZER_E, HNZF_M, SXXS_E, SZLN_M, SDNS_M, HNSH_M, SHDT_E

Cluster2 FCNM_M, ADTM_M, PGGS_M, QHSL_M, CAAE_I, HJTG_E, HYPC_M, GHEG_F, JXHD_I, XAAI_I, CCFM_M, BJTN_M, YYTH_M, CHSS_I, XAAE_I, RSNM_M, XMTS_M, BSRE_M, MMTL_I, HBYH_M, GFSC_F

Cluster3 JSHM_H, CRDP_H, KMPH_H, ZJMC_H, SDDE_H, YNBY_H, SHFS_H, ZJNH_H, BJTR_H, GXWZ_H, BJSL_H, CRSJ_H, NCPH_H, LXVC_I, TSLP_H, NRTE_I, HLBE_H, ZJHS_H, HRBP_H

Cluster4 CSGH_M, HNAL_I, CSAL_I, DFET_I, TBEA_I, CHEA_I, BDTW_I, YTWH_M, ZJJH_M

Cluster5 CGZB_I, GXLG_I, SHCG_I, TSJD_M, JLYT_M, , CHBA_F, TYHI_I, CATL_I, AHCC_M, SHIT_I, ZLHS_I, STCM_I, XCMG_I

Cluster6 JSZN_I, BEJU_F, CHMT_F, FCSH_F, SHCT_F, SNUS_F, GMDL_F, SHZJ_F, XMCD_I, OWRG_F, XHZB_F, BJCT_F, CHVK_F

Cluster7 CMBK_F, CJSC_F, SHZH_I, NESC_F, BSIS_M, HTSC_F, IMBT_M, LNCD_I, GYSC_F, SXTG_M, HXBK_F, SNLS_F, OFFS_E, SZBK_F, WHIS_M, CMSB_F, CHRW_I, SHPD_F, SWSC_F, HYSC_F, HBIS_M, CIMC_I, CSSC_I, SNCH_I, AGST_M, SHIA_I, JLAD_H, GJHX_M, CHPC_E, CTSC_F

C Extra Neighbor-Nets Splits Graphs

C.1 Period 2 Clusters in Period 3

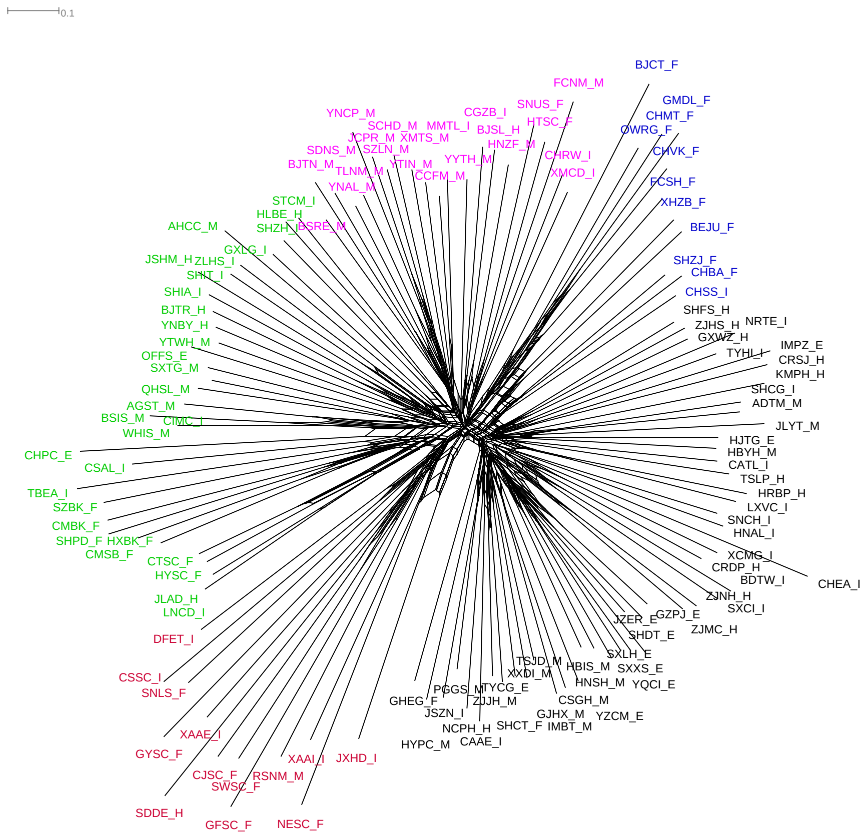

Figure 15: SplitsTree network for 126 stocks from the Shanghai A Stock Exchange for period two showing the five identified clusters. The colours are Cluster 1 – Black, Cluster 2 – Blue, Cluster 3 – Pink, Cluster 4 – Green, and Cluster 5 – Red.

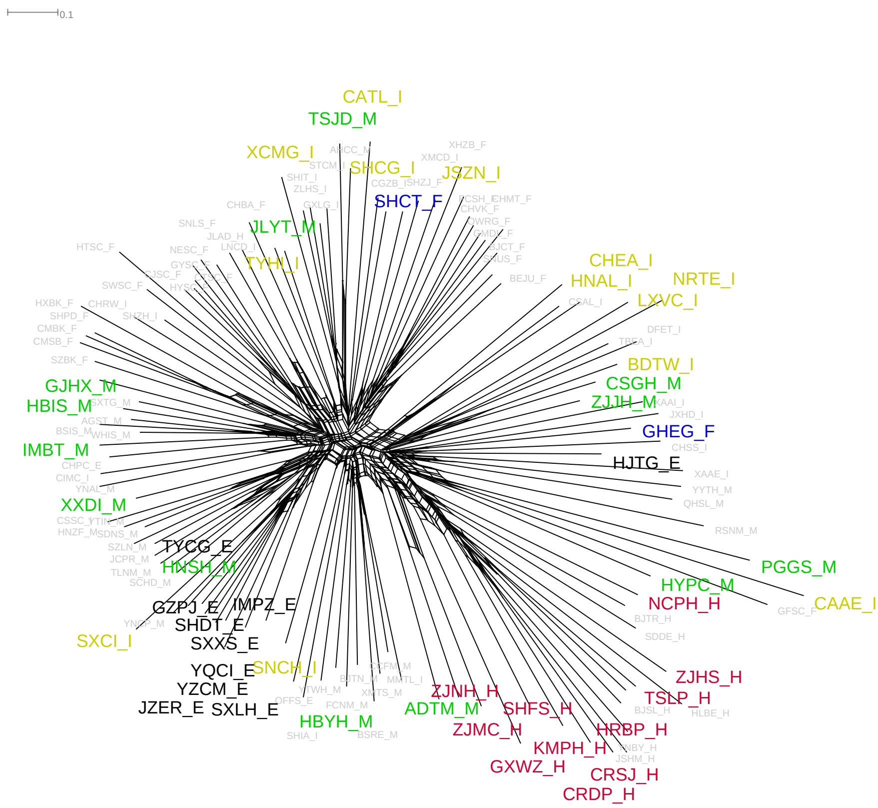

Figure 16: SplitsTree network for 126 stocks for period 3 showing the period two cluster 1 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

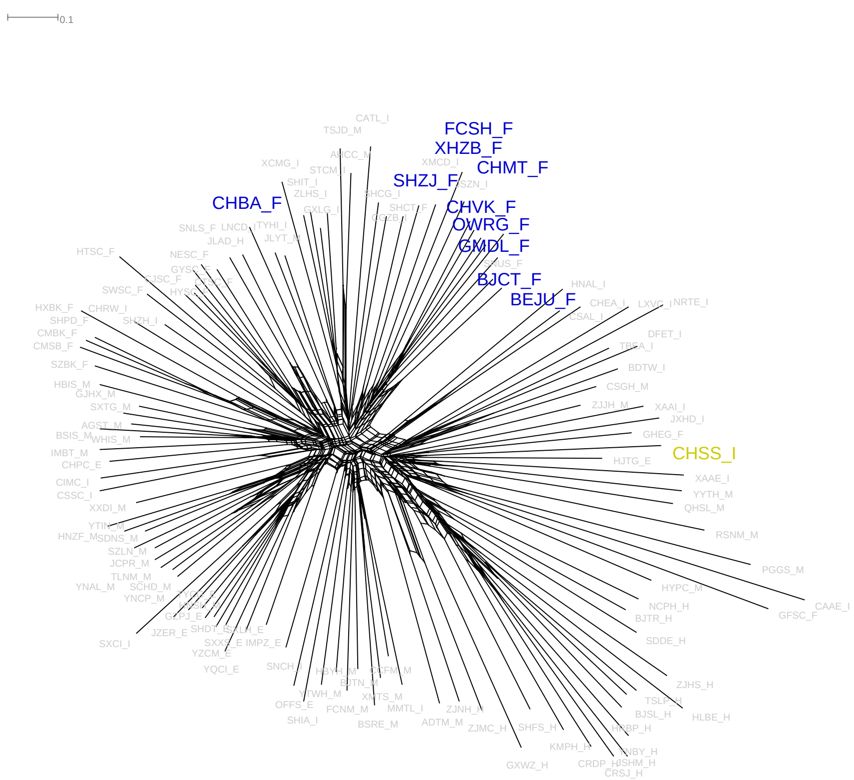

Figure 17: SplitsTree network for 126 stocks for period 3 showing the period two cluster 2 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

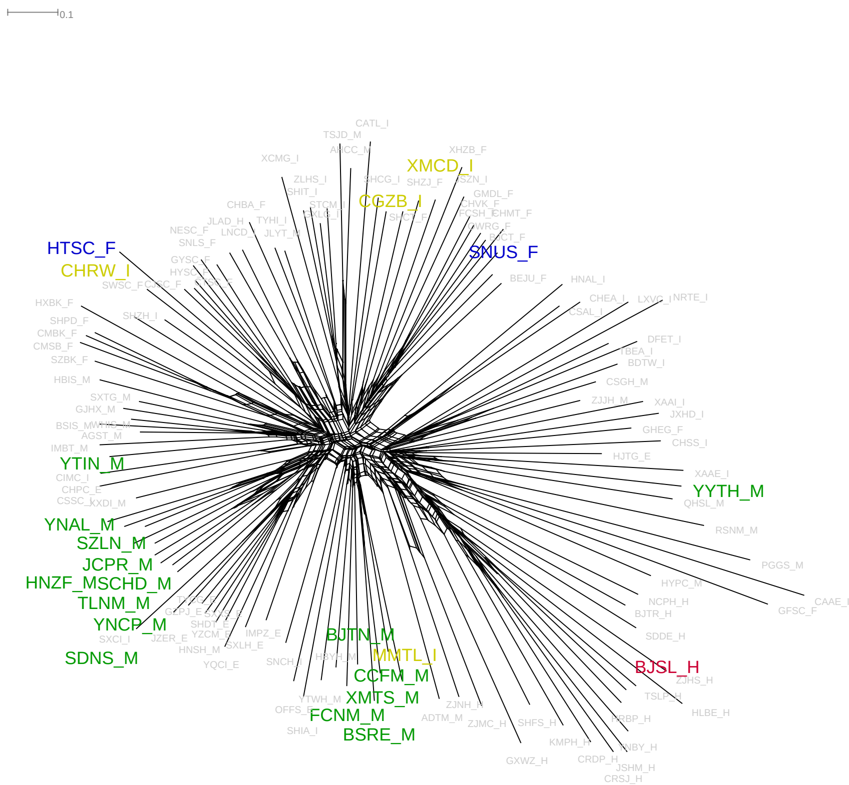

Figure 18: SplitsTree network for 126 stocks for period 3 showing the period two cluster 3 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

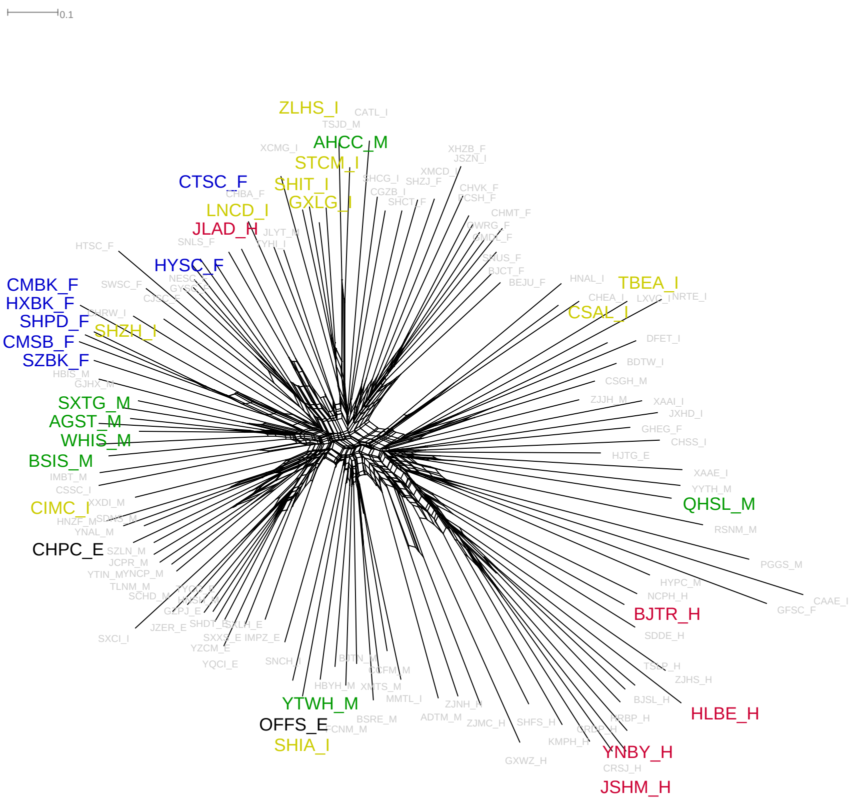

Figure 19: SplitsTree network for 126 stocks for period 3 showing the period two cluster 4 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

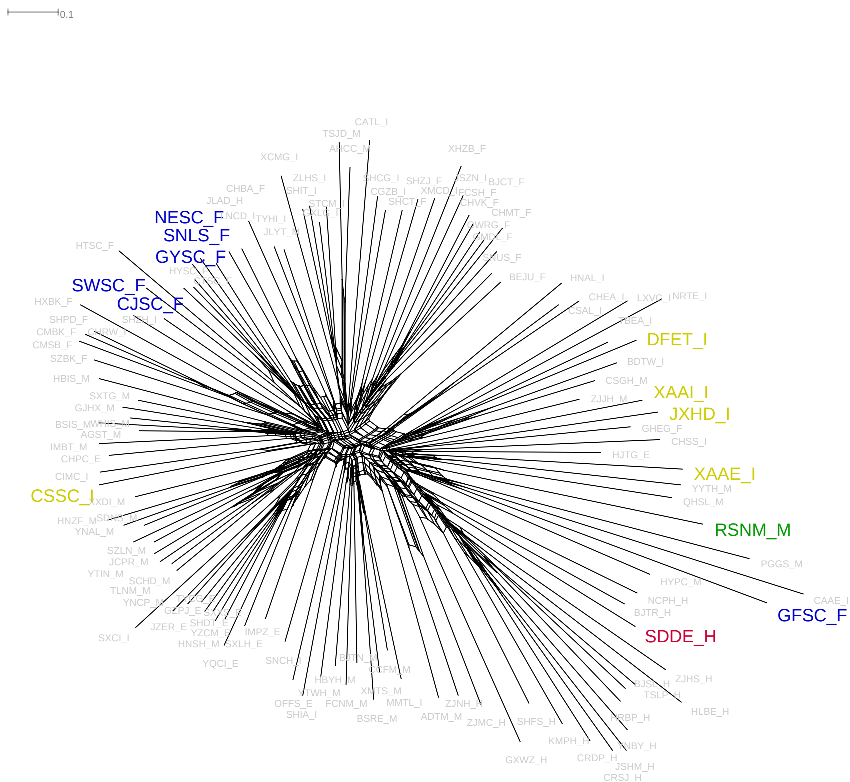

Figure 20: SplitsTree network for 126 stocks for period 3 showing the period two cluster 5 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

C.2 Period 3 Clusters in Period 4

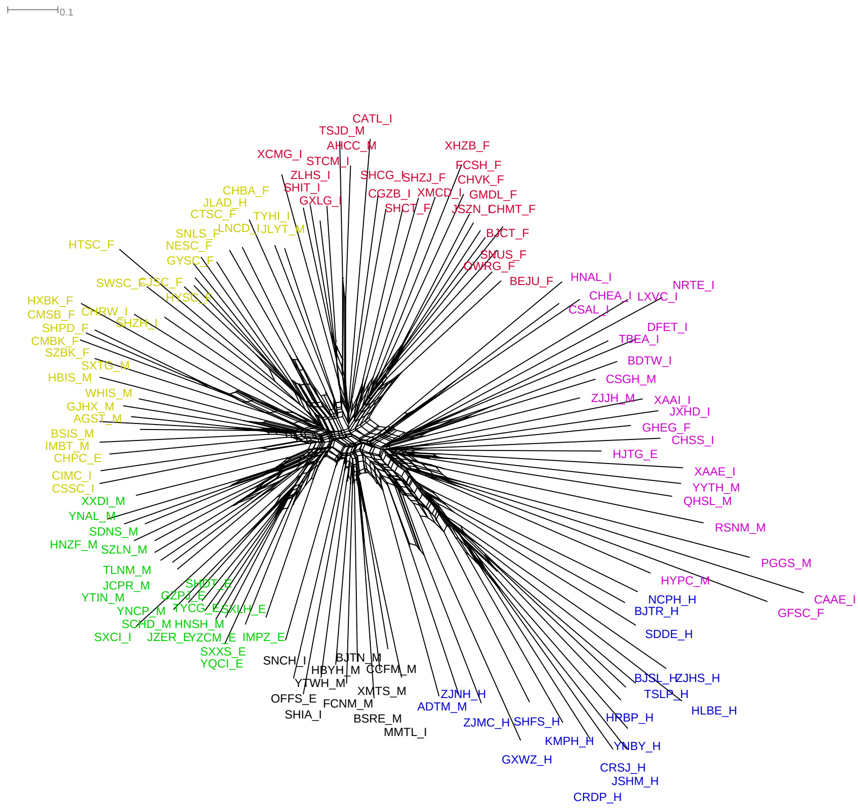

Figure 21: SplitsTree network for 126 stocks for period three showing the six identified clusters. The colours are Cluster 1 – Black, Cluster 2 – Blue, Cluster 3 – Pink, Cluster 4 – Red , Cluster 5 – Khaki, and Cluster 6 – Green.

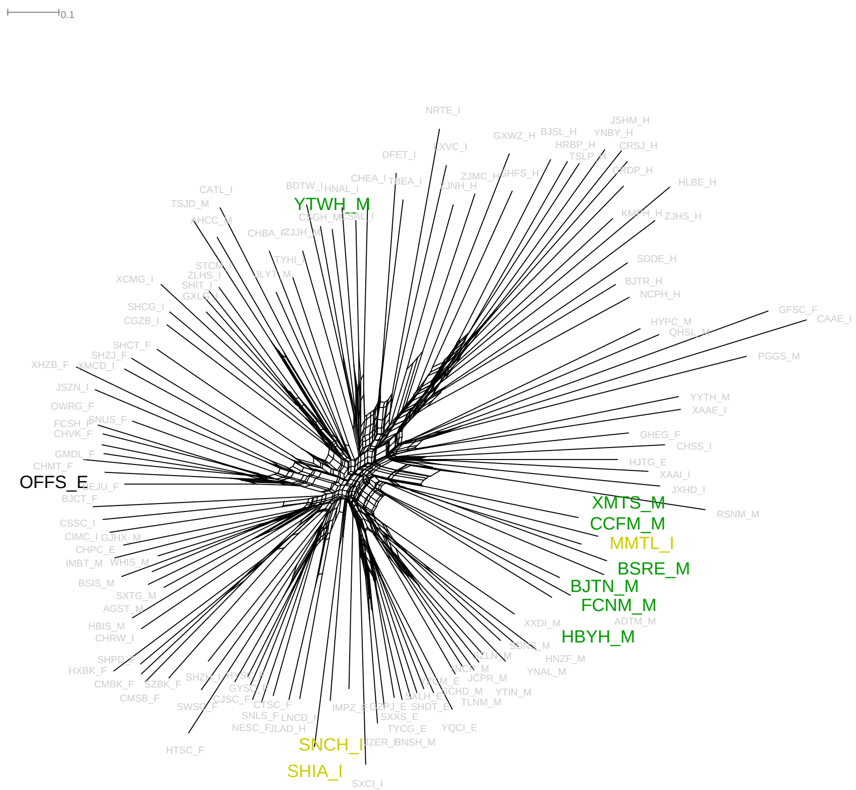

Figure 22: SplitsTree network for 126 stocks for period four showing the period three cluster 1 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

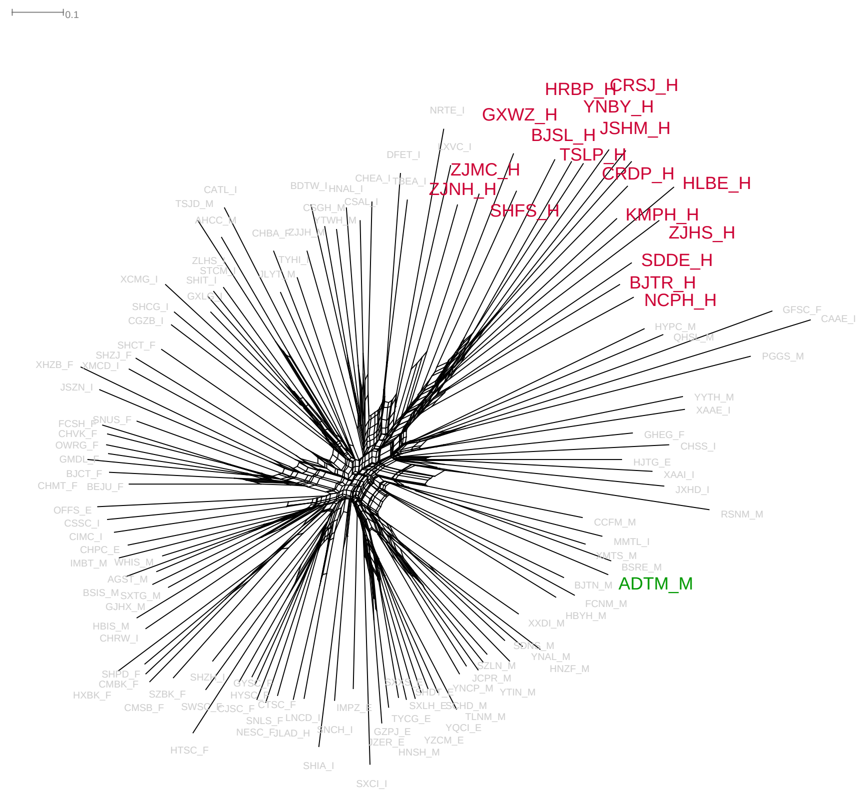

Figure 23: SplitsTree network for 126 stocks for period four showing the period three cluster 2 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

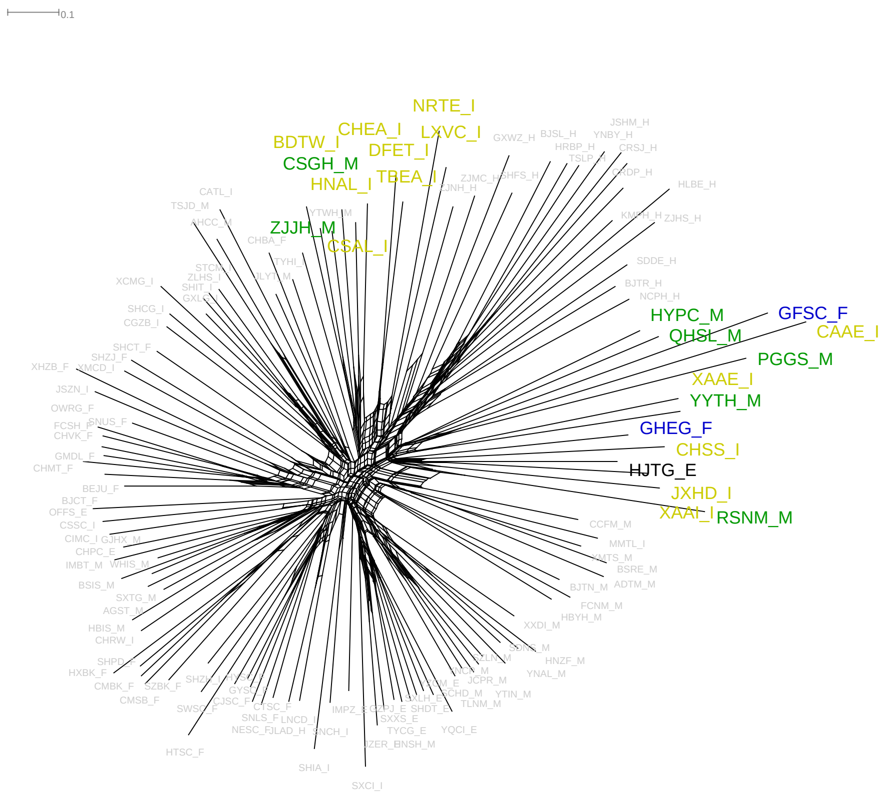

Figure 24: SplitsTree network for 126 stocks for period four showing the period three cluster 3 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

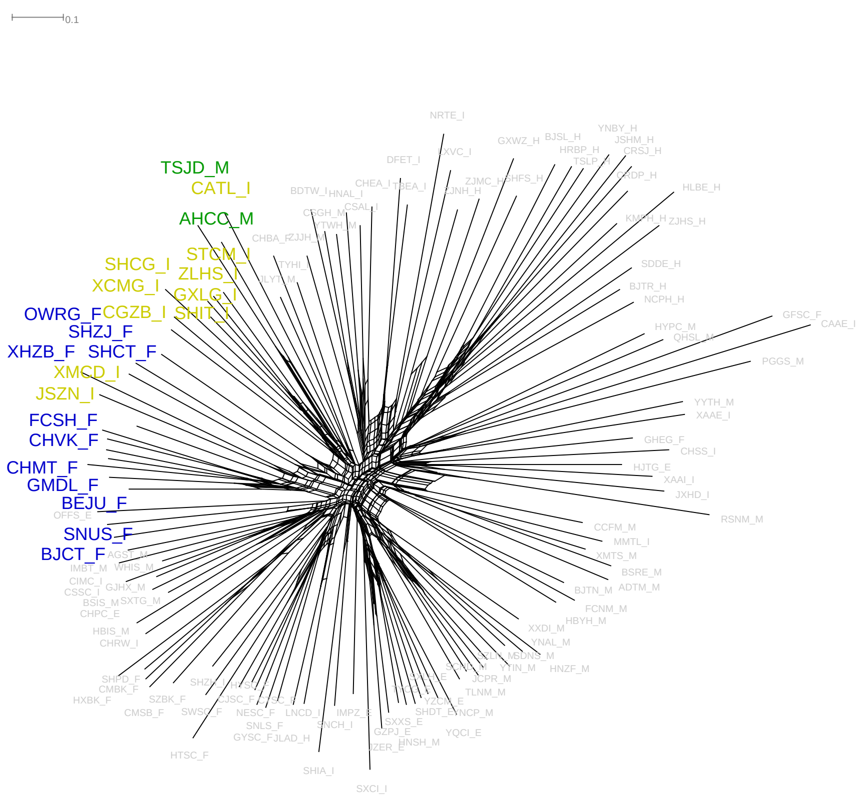

Figure 25: SplitsTree network for 126 stocks for period four showing the period three cluster 4 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

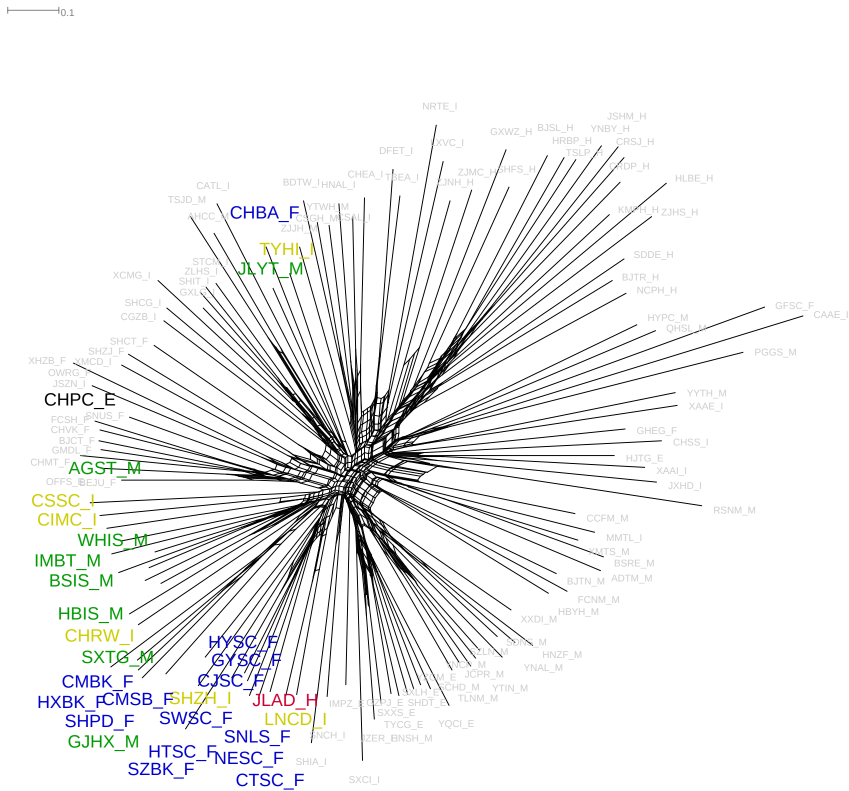

Figure 26: SplitsTree network for 126 stocks for period four showing the period three cluster 5 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.

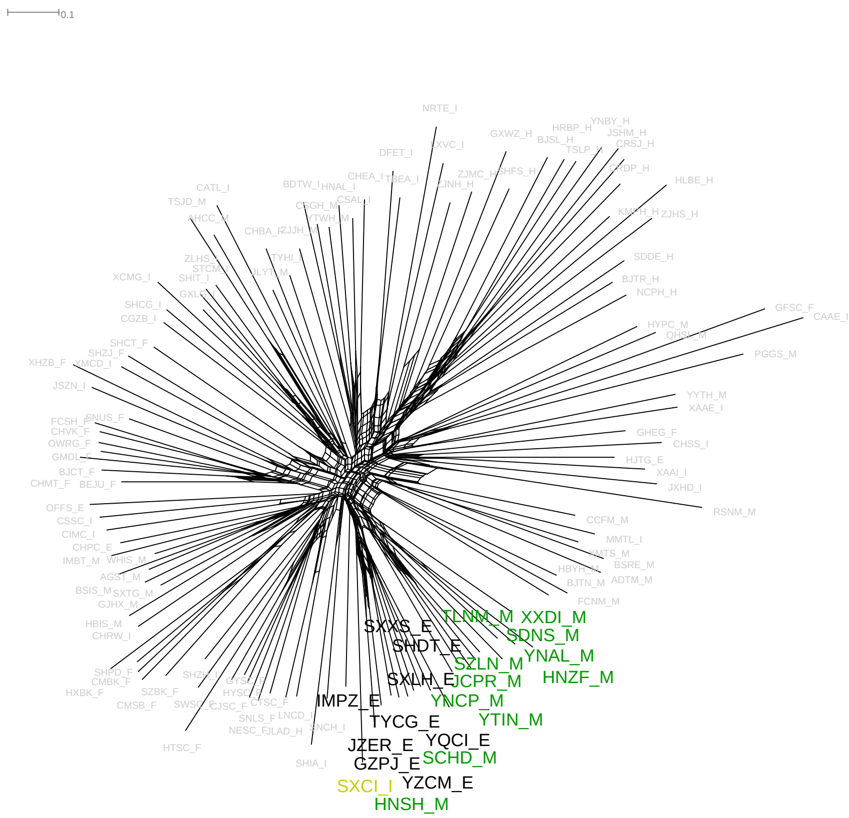

Figure 27: SplitsTree network for 126 stocks for period four showing the period three cluster 6 colour coded by industry group. The colours are Energy - Black, Finance – Blue, Health Care – Red, Industrials – Khaki, Materials – Green.