Modelling Financial Markets using Methods from Network Theory

Thesis submitted in accordance
with the requirements of the
University of Liverpool for the
degree of Doctor in Philosophy

by Jenna Louisa Birch

August 2015

Abstract

This thesis discusses how properties of complex network theory can be used to study financial time series, in particular time series for stocks on the DAX 30.
First, we make a comparison between three correlation-based networks: minimum spanning trees; assets graphs and planar maximally filtered graphs. A series of each of these network types is created for the same dataset of time series' of DAX 30 stocks and we consider what information each network can provide about the relationship between the stock prices from the underlying time series. We also analyse two specific time periods in further detail – a period of crisis and a period of recovery for the German economy.

Next, we look at the structure and representations of planar maximally filtered graphs and in particular we consider the vertices that form the 3-cliques and 4-cliques [Tumminello *et al.* (2005)] state *'...normalizing quantities are* $n_s - 3$ *for 4-cliques and* $3n_s - 8$ *for 3-cliques. Although we lack a formal proof, our investigations suggest that these numbers are the maximal number of 4-cliques and 3-cliques, respectively, that can be observed in a PMFG of* n_s *elements.'* Within this thesis we provide a proof for these quantities and a different construction algorithm.

Finally, rather than correlation-based networks, we discuss two relatively new types of networks: visibility graphs and the geometrically simpler horizontal visibility graphs. We review the field's that these networks have already been applied to and consider if this is an appropriate method to apply to financial time series – specifically stock prices. We also consider using horizontal visibility graphs as a method for distinguishing between random and chaotic series within stock price time series.

Publications

The Maximum Number of 3- and 4-Cliques Within a Planar Maximally Filtered Graph. J. Birch, A.A. Pantelous and K. Zuev. (2015).

Physica A: Statistical Mechanics and its Applications. Vol. 417 (Iss. C) 221-229.

Analysis of Correlation Based Networks Representing DAX 30 Stock Price Returns. J. Birch, A.A. Pantelous and K. Soramäki. (2015).

Computational Economics. DOI 10.1007/s10614-015-9481-z.

Conference Presentations

- Forecasting Financial Markets 2013, Hannover, Germany (29/05/13 - 31/05/13),
- 60th ISI World Statistics Congress 2015, Rio de Janeiro, Brazil (26/07/15 - 31/07/15).

Conferences and Workshops Attended

- SCOR Conference, University of Nottingham, Nottingham (20/04/12 - 22/04/12),
- Financial Network Workshop: Hosted by FNA, Pompeu Fabra University, Barcelona (03/05/12 - 05/05/12),
- LMS EPSRC: Common Themes in Financial & Actuarial Mathematics, Management School, University of Liverpool, Liverpool (15/04/13 - 19/04/13),
- Inaugural Conference of Journal of Network Theory in Finance, Centre for Risk Studies, University of Cambridge, Cambridge (23/09/14),

- NetSci 2014 School and Conference, University of California, Berkeley (02/06/14 - 06/06/14),
- NetSci 2015 Conference, Zaragoza, Spain (03/06/15 - 05/06/15).

Acknowledgements

I would like to express my appreciation and thanks to my supervisor Dr. Pantelous for your advice and guidance, and for providing me with so many opportunities over these past four years. Also to my second supervisor Dr. Zuev, thank you for your contributions over the past year, and for your time, patience and dedication to my research. To those at FNA, particularly Kimmo Soramäki, thank you for the internship opportunity. Your constant support over the years has been much appreciated and I hope to continue working with you in the future.

I would not have been able to complete my thesis without the help of people within the Department of Mathematics and particularly the members of IFAM. This research would not have been possible without the funding through the GTA scheme. I would like to specifically mention Dr. Constantinescu, your advice and believe in me over the years have helped me during the times I've needed support the most and Jia Shao (Sara), I appreciate your help with proof-reading and problem solving but most of all I appreciate what a good friend you have been to me.

There are friends from each stage at university, including some of the lovely girls at LULHC and also from outside of university that have kept me sane at times with good advice, laughter and tea breaks I feel a list would certainly miss people out so I would like to thank you collectively and I know that this work would not have been completed without you all.

Most of all, I would like to thank my family (Mum, Dad and Our Mills). After so many years of helping with school projects until late into the night, reading over my work and advising me on education and 'real life jobs,' I'm sure you didn't think I'd still be needing it at 26. But you've always been there for me with love and support – know that I'm forever grateful and that this work is dedicated to you.

Contents

List of Figures

List of Tables

1 Introduction

In recent years there has been increasing interest in how we can model complex systems using network theory. These can include information, technological and biological systems [Newman (2003)], social networks [Toivonen *et al.* (2006)] and financial markets ([Allen & Gale (2000)] and [Bonanno *et al.* (2004)]). In particular, a network based approach of studying complex systems has become very popular in econophysics [Stanley & Mantegna (1999)], an interdisciplinary research field that studies economic and financial phenomena. Networks can be used to model the interactions between banks and other financial institutions. Interbank markets have been covered extensively in the literature, for example [Boss *et al.* (2004)] constructed networks to model the Austrian banking system which consists of many sectors and tiers. [Soramäki *et al.* (2007)] described the topology of the interbank payment system in the USA. [Iori *et al.* (2008)] used network topology to analyse the Italian overnight money market and the lending/borrowing that occurred between foreign banks and Italian banks of various sizes. [Li *et al.* (2010)] used data from Japan to construct a directed network model and also provided a summary of the banking systems in several other countries. As well as looking at the structure of financial systems, the literature has covered robustness and contagion in financial networks. [Allen & Gale (2000)] and [Leither (2005)] both modelled contagion in the banking networks. [Becher *et al.* (2008)] used data from CHAPS Traffic Survey 2003 to illustrate the broad network topology of the interbank payments in the UK. [Galbiati & Soramäki (2012)] modelled clearing systems as networks whose function is to transform exposures and studied how their topology affects the resulting exposures and margin requirements.

One area with significant recent developments is that of correlation based networks. These networks can be used to reduce complexity of financial dependencies and to understand and forecast the dynamics in financial markets. One of the important and fundamental problems in this approach is to filter the most relevant information from financial networks. As a result traditional algorithms from network theory have been adapted and some new methods have been introduced. In 1999, a method for finding a hierarchical arrangement for a portfolio of stocks by extracting the Minimum Spanning Tree (MST) from the complete network of correlations of daily closing price returns for US stocks was introduced by [Mantegna (1999)]. This method has been expanded using coordination numbers by [Vandewalle *et al.* (2001)] and also applied to other markets such as global stock exchange indices by [Bonanno *et al.* (2000)] and currency markets by [Mizuno *et al.* (2006)]. More recently [Brookfield *et al.* (2013)] examined the properties of the MST as applied to the book-to-market ratio and market returns. This technique was studied and further developed by [Onnela *et al.* (2002)] who considered the effect of a stock market crash on the MST, or asset tree, using the 1987 stock market crash as evidence. They concluded that there was strong shrinkage of the asset tree during the crash, with the normalised tree lengths decreasing and remaining low for the duration of the crash. [Onnela *et al.* (2003)] extended their study with the introduction of the Asset Graph (AG) – a network structure similar to the MST where a network with n vertices has $n-1$ edges; however the algorithm for the AG selects the largest $n-1$ correlations regardless of the resulting structure, i.e. unlike the MST, the AG does not have to be a connected network. Similar to this work [Tse *et al.* (2010)] created Threshold Networks (TN), by taking the cross correlations of stock prices, price returns and trading volumes and connecting vertices based on a 'winner-take-all' method, so that an edge existed between two stocks if the cross correlation was larger than a particular threshold value. This method was also used by [Qiu *et al.* (2010)] to study the dynamical behaviour of American and Chinese stock markets.

The MST and AG are methods for reducing the complete network to a basic minimum structure that contains only the most relevant information and, in the case of the MST, the general hierarchical structure. One of the more recent developments was an algorithm proposed by [Tumminello *et al.* (2005)] where the complete network can be filtered at a chosen level, by varying the genus of the resulting filtered graph. So if a graph is embedded into a surface of genus g, as g increases the resulting graph becomes more complex and so reveals more information about the clusters formed, while keeping the same hierarchical tree as the corresponding MST. The simplest form of this graph is the Planar Maximally Filtered Graph (PMFG), on a surface with $g = 0$. The PMFG has proven to be an important tool for filtering the most relevant information from a network, particularly in correlation based networks that model the correlation between stock prices. For example [Pozzi *et al.* (2013)] considered the level of risk and the returns on portfolios selected using filtered graphs, including PMFG. [Eryigit & Eryigit (2009)] used PMFGs (along with MSTs and clustering methods) to analyse the daily and weekly return correlations among indices from stock exchange markets of 59 countries. In general, the PMFG can tell us about the clusters that form within the dataset, regardless of the network nature, as a result of the underlying topological properties of the network. [Song *et al.* (2012)] introduced a technique to extract the cluster structure and detect the hierarchical organisation within a complex dataset. This method has been developed using the topological structure of the PMFG such as the separating 3-cliques which separate a graph into two disconnected parts. For the PMFG we consider the vertices that form the 3- and 4-cliques (as the maximum number of elements that can form a clique is 4). [Tumminello *et al.* (2005)] state *'...normalizing quantities are $n_s - 3$ for 4-cliques and $3n_s - 8$ for 3-cliques. Although we lack a formal proof, our investigations suggest that these numbers are the maximal number of 4-cliques and 3-cliques, respectively, that can be observed in a PMFG of n_s elements.'* One way that we use the cliques to analyse networks is to consider the ratio between the number of cliques that have formed to the maximum number of cliques that could form. For this, [Tumminello *et al.* (2005)] used the normalizing quantities that have been mentioned above, an approach that has also been used by [Eryigit & Eryigit (2009)], [Aste *et al.* (2005a)] and [Tumminello *et al.* (2007)] and used when defining the connection strength of a sector by [Coronnello *et al.* (2005)]. Within Chapter 4 of this thesis we provide a proof for these quantities and a different construction algorithm.

Each of the different methods discussed here filter various amounts of information from the complete correlation network. This can either be at a set level due to the nature of the construction algorithm (for example the MST is a severe form of data reduction leaving only the minimum number of edges for a connected network) or a level chosen during construction (for example the threshold value chosen in TNs or the genus of the surface in Filtered Graphs). Within Chapter 3 of this thesis a comparison is made between MSTs, AGs, PMFGs by creating each of these networks for the same dataset of time series'. A complete network is constructed from the data showing the pairwise correlations between stock price returns for companies on the German stock market (DAX 30). A series of MSTs, AGs and PMFGs are created and we consider what information each network can provide us with about the relationship between the stock prices from the underlying time series. We also analyse two specific time periods in further detail – a period of crisis and a period of recovery for the German economy. Our aim is to test whether or not there is a difference between the networks created from the two datasets and, if so, what information we can extract about the stocks during these time periods.

These networks determine the similarity between the different time series and we use the structural properties of the networks to show how this relationship can change over time. Another way of analysing these time series' is to consider how the individual data series' themselves change over a period of time, making use of the temporal ordering. The idea here is to map time series to networks so that the network inherits properties from the underlying time series. In their 2006 paper, Zhang and Small discussed a method for creating complex networks from pseudoperiodic time series, where each repeated cycle is a single vertex in a network [Zhang & Small (2006)]. An edge would connect two of these vertices if their underlying cycles are similar in shape and form (measured quantitatively using phase space distance or a linear correlation coefficient). [Xu *et al.* (2008)] embedded time series' to an appropriate phase space. For the edges they used a threshold for the minimum distance and also a maximum degree for the vertices. [Lacasa *et al.* (2008)] expanded upon this idea and introduced the Visibility Algorithm, a method which can be applied to different time series, not only pseudoperiodic as with [Zhang & Small (2006)]. This algorithm was further developed by [Luque *et al.* (2009)] to form the Horizontal Visibility Algorithm – a subgraph of the original Visibility Graph (VG) and as such a geometrically simpler algorithm and more analytically tractable. The authors analyse the networks created by these algorithms in terms of their structural properties such as degree distribution, average path length and clustering coefficient, and as stated above, these can reflect certain properties of the time series. For example, if we create a visibility graph for a periodic time series then the network will inherit the regularity of the time series and as such will be a regular network. By similar reasoning the algorithm also creates an exponential random network from a random times series and a scale-free network from a fractal time series [Lacasa *et al.* (2008)].

The literature in this area mainly covers theoretical results. For example [Luque *et al.* (2009)] showed that a Horizontal Visibility Graph (HVG) generated from a bi-infinite random time series will have a degree distribution of $P(k) = (\frac{1}{3})(\frac{2}{3})^{k-2}$. This means that the horizontal visibility algorithm can be used as a method to determine if a time series is random or chaotic. [Nuñez *et al.* (2012)] discussed in more details how the HVG can be used as a method of noise filtering, recognising that periodicity detection algorithms can be grouped into two categories: time domain (autocorrelation based) and frequency domain (spectral). The authors proposed HVG as a third category: graph theoretical. [Gutin *et al.* (2011)] used combinatoric properties of the networks and proved that HVGs will always be an outerplanar graph and always have a Hamilton path. The authors show that the algorithm can be used as a linear time recognition algorithm. The visibility graph can be used to estimate the Hurst exponent. The Hurst exponent provides a measure for whether a dataset is a pure white noise random process or if there is an underlying trend to the dataset. Brownian walks can be generated from a defined Hurst exponent and if it is between 0.5 and 1 then the random process will be a long memory process and the dataset is referred to as fractional Brownian Motion. [Lacasa *et al.* (2009)] covered this in more details. Further details can also be found in [Xie & Zhou (2011)]. These algorithms have been applied to time series from various fields such as physics datasets and financial time series. Modelling energy dissipation rates in turbulence using VGs, [Liu *et al.* (2009)] looked at the statistical properties and found the degree distribution to be power-law, $P(k) \sim k^{-\alpha}$ where $\alpha = -3$. [Yu *et al.* (2012)] applied the horizontal visibility algorithm to daily time series of the solar x-ray brightness from 1986 - 2007 and found that multifractality exists in both the daily time series and the corresponding HVGs. In [Yang *et al.* (2009)] VGs were constructed from six exchange rate series (US dollars to Australian dollars, Canadian dollars, euro, GB pound, Japanese Yen and NZ dollar.) The authors considered the original time series, as well as shuffled and detrended data, and found that the series converted to a scale-free and hierarchically structured network, also the original and detrended time series were multifractal. The hierarchies for the Yen and euro came across weaker compared to the others. In Chapter 5 of this thesis we present the formal construction algorithm for VG and HVGs. We discuss the properties of the graphs created from the time series and proven results from the literature. Finally, we consider whether the family of visibility algorithms is an appropriate method to apply to financial time series.

The remaining of this thesis is structured as follows. Chapter 2 contains preliminary details on network theory as well as the German economy and DAX 30 index. Chapter 3 provides a comparison between correlation based networks: MSTs (Subsection 3.2.1), AGs (Subsection 3.2.2) and PMFGs (Subsection 3.2.3). Maximally Filtered Graphs are discussed in further detail in Chapter 4, with the various representations, including standard spherical triangulation form, analysed in Section 4.2 and the main results, including the proof of the maximum 3- and 4-cliques and a different construction algorithm, given in Section 4.3. Visibility and Horizontal Visiblity Graphs are discussed in Chapter 5 (the construction algorithm for both graphs given in Section 5.1 and their properties and applications in Sections 5.2 and 5.3 respectively). Finally Chapter 6 concludes.

2 Financial Network Theory

In this chapter we introduce some of the key terminology from Network Theory that is used throughout this thesis. We also discuss the euro area, in particular details of the German economy for 2001 – 2014 as this time period covers the dataset used in later chapters. Our discussions include the introduction of the euro, Germany's imports and exports and the German stock market, the DAX 30.

2.1 Preliminaries

A *network*, also referred to in literature as a *graph*, is a set of *vertices* (or nodes) connected by *edges*. Denote the graph $G(V, E)$ where V is the set of vertices belonging to G and E is the set of edges belonging to G. Denote the number of vertices $|V| = n$ and the number of edges $|E| = m$. A *loop* is an edge whose end vertices coincide and a *multiple edge* is formed when two or more edges join the same vertices. If a network does not contain any loops or multiple edges then it is called a *simple network.* A *subgraph* H, of a graph $G(V, E)$, is a graph whose vertices are a subset of the vertex set V and whose edges are a subset of the edge set E. A subset of vertices $C \in V$ is called a *clique* if the subgraph $G(C)$ is a complete graph (a simple graph with every pair of distinct vertices connected by a distinct edge) and is denoted C_j where $|C| = j$.

Various terms are used to describe the structure of the network. A *component* is a subset of the vertices of a network such that there exists at least one path (a continuous walk along the edges) between each pair of vertices within the subset. If there is a path from every vertex in the network to every other vertex then the network is called *connected*. Without a path the network is called *non-connected* and consists of separate components. A *tree* is a connected, undirected network that contains no closed loops. A network can be *directed* or *undirected* depending on whether or not the edges show the direction of the flow between the vertices. If the flow between vertices can only be one way then the edge is called a *directed edge* (or alternatively an *arc*) and thus it follows that a network consisting of directed edges is simply called a *directed network* (or a *digraph*). An *undirected edge* has a two-way flow between vertices. Note that an undirected network can be transformed into a directed one by representing the undirected edges between the vertices as two directed edges. As well as the direction of the edges we also consider the values assigned to the edges. A *weighted network* is one in which the edges between the vertices have weights assigned to them. Depending on the subject of the network these weights can have different meanings. For example, it could show a cost of 'using' an edge or the total flow allowed to travel through the edge. In financial models it could be the values or volume traded. If no value is assigned then there is assumed to be unlimited flow throughout the network and the network itself is described as an *unweighted network*. The edges can be scaled to reflect these weights either by edge length or thickness.

Once a network has been formed there are several metrics which can be computed and the results used for comparing and making observations of the networks (or components of a network), such as how the network changes over a period of time. Over recent years, network theory research has progressed from considering the properties relating to individual vertices or edges to instead analysing the statistical properties of the network as a whole. For example centrality measures would be vertex properties as they consider which is the most central (or important) vertex in the network, whereas connectivity looks at the whole network and how dense/sparse it is. Some of these metrics are calculated for each vertex but are then averaged over the network (for example degree and mean degree).

The number of edges adjacent to a vertex is known as the *degree* of the vertex. For a vertex, v, we use the notation $deg(v)$ to show the degree, or alternatively k_i. In any network, the sum of the degrees of all the vertices is equal to twice the number of all edges, i.e. for a network $G(V, E)$, with n vertices v_i (where $i = 1, ..., n$):

$$\sum_{i=1}^{n} \deg(v_i) = 2|E|. \tag{1}$$

The mean degree for a network is calculated as:

$$\frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2|E|}{|V|}. \tag{2}$$

A vertex can be classed as either an odd or even vertex depending on whether the degree of the vertex is odd or even. In a directed network the number of edges coming into a vertex is called the *indegree* and the number of edges leaving a vertex is called the *outdegree.* The following equation holds for all directed networks:

$$\sum_{v \in V} \text{indeg}(v) = \sum_{v \in V} \text{outdeg}(v) = |E|. \tag{3}$$

In a directed network a vertex with indegree $= 0$ is called the *source* (and can be seen as the origin) whilst a vertex with outdegree $= 0$ is called the *sink.* If a vertex, v, has indegree(v) = outdegree(v) then it is called a *balanced vertex* and similarly a network with all its vertices balanced is called a *balanced network.*

When considering large scale networks an important property to consider is the degree distribution. Let $P(k)$ be the percentage of vertices with degree k in the network. The *degree distribution* is the distribution of $P(k)$ over all k, i.e. the probability that a vertex has degree k. When constructing a random network, vertices are added at random meaning that there tends to be an average degree. This results in the degree of most of the vertices within the network distributed around this average with few vertices having a much higher or lower degree. However, real life networks have been shown to have a much more skewed distribution with most vertices having only a few edges and a (proportionally) small number of vertices being highly connected with a large degree value. This leads to a power-law distribution where $P(k) \sim k^{-\gamma}$. A network that demonstrates a power-law degree distribution is known as a *scale-free network.*

Other metrics measure the connectivity and the transitivity of the network. The *connectivity* is the number of edges that actually exist between the vertices of the network compared to the number of edges that are possible. Alternatively, the connectivity can be seen as an unconditional probability (p) that two vertices share an edge. In a directed network the connectivity $(p) = \frac{|E|}{(|V|(|V|-1))}$ and for an undirected network the connectivity $(p) = \frac{2|E|}{(|V|(|V|-1))}$. The closer that this value is to 1, the more dense the network is. The *local clustering coefficient* of a vertex v is the probability that the vertices adjacent to v are also connected i.e. the ratio of the number of actual edges there are between the adjacent vertices to the number of potential edges there are between them. The *global clustering coefficient*, C, of a network measures the connectivity of the network by considering the number of closed triplets as a fraction of the total triplets within the network (three vertices joined by two edges form an open triplet and three vertices joined by three edges form a closed triplet). $C = \frac{\text{Number of closed triplets}}{\text{Number of total triplets}}$, where C = 1 implies perfect transitivity.

Further details on network theory can be found in [Albert & Barabási (2002)] and [Newman (2010)].

2.2 German Economy

In Chapter 3 we analyse the DAX 30 blue chip stocks for the time period 2001 - 2014 (see Appendix A for a list of all stock symbols and the sectors to which they belong and Section 2.3 for further details on DAX 30). The dataset, created from Thomson Reuters Datastream[a], consists of the closing prices, adjusted for dividends and splits, of the 30 stocks traded on the Frankfurt Stock Exchange, that form the DAX 30, for the time period between 2001 and 2014. This is a significant time period for the German economy as the euro area (a monetary union, originally between eleven EU members) was established on 1st January 1999 and Germany officially accepted the euro as its legal tender on 31st December 2001.

The GDP (Gross Domestic Product) can be used as a good indicator when considering the economic growth of a country. We define a recession as two consecutive periods of negative growth and there have been several periods of recession for the euro area[b] (see Figure 1). Since its establishment in 1999 the euro area has had several periods of financial crisis; however these have not always been reflected by the German economy.

[a] Thomson Reuters Datastream 5.0 (`thomsonreuters.com`).

[b] Data source: ECB statistical Data Warehouse `http://www.ecb.europa.eu/stats/keyind/html/sdds.en.html`.

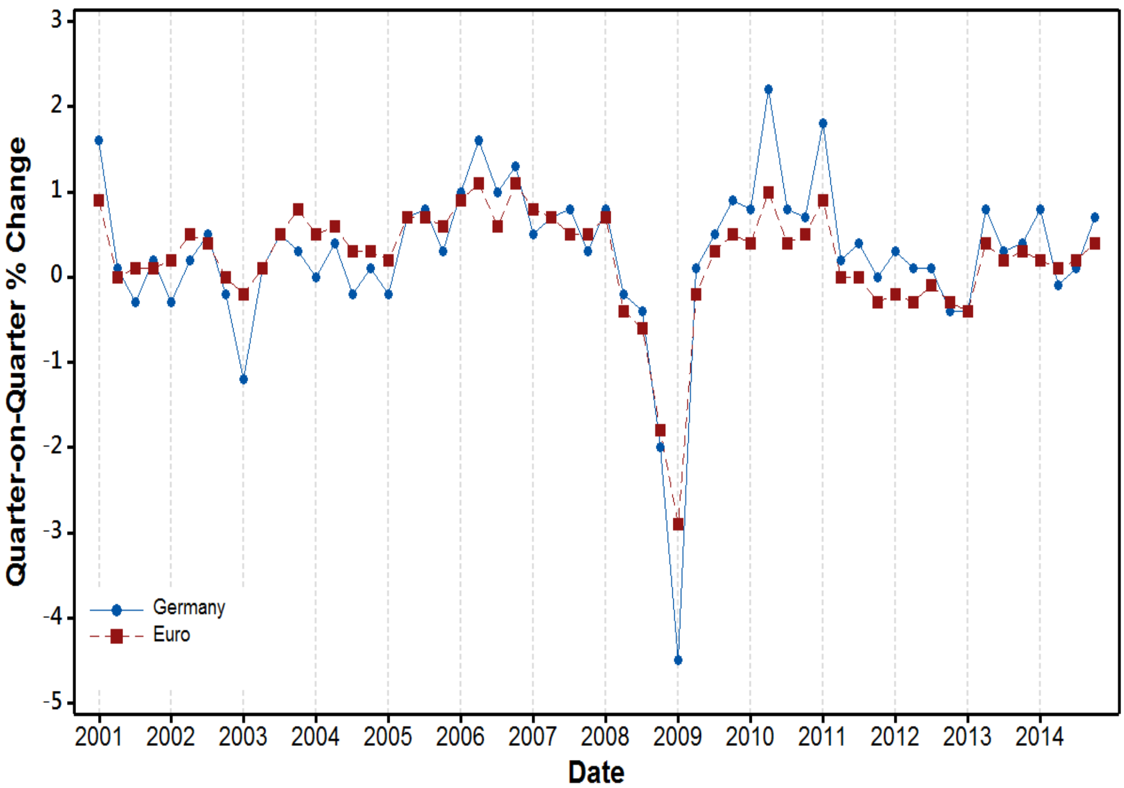

Figure 1: The time series shows the quarter-on-quarter volume growth of GDP and expenditure components for Germany (shown in blue) and the euro area (shown in red).

After the introduction of the euro certain countries within Europe (e.g. France and Germany) suffered a decline in their GDP between 2001 and 2004. After a period of recovery and economic growth, Europe was affected by the 2007 – 2009 financial crisis led by the U.S. subprime mortgage crisis. The GDP of Germany decreased by 0.2% in the 2nd quarter of 2008 and then a further 0.4% in the 3rd quarter (bigger than the economists' predicted value of 0.2%). This meant that the German economy was officially in recession as of 13th November 2008. There have been several reasons proposed for this recession, stemming from the mortgage crisis, including high inflation, a strong euro and tight monetary policies in place. Then on 15th November 2012 the euro area officially entered recession for the second time in four years, despite continuing growth in the largest economies of the area – Germany and France. The latter period is often referred to as the European sovereign-debt crisis (or the euro zone crisis). It is important to note that a financial crisis in the euro area will affect countries at different times and the rate of recovery will vary depending on the state of the country's economy prior to the crisis.

Germany is one of the most highly developed nations in the world and the German economy is the 5th largest national economy by GDP (at PPP exchange rates[c]). As such it is the largest economy within the euro area and plays a dominant role not only in the European Union but also within world economics (Germany's share of world trade (exports and imports) for 2014 was 7.2%[d]). It has invested in the emerging markets within Asia and also been influential in the expansion of the EU to include countries in Central and Eastern Europe. Its most important trading partners, based on their percentage share of overall exports, are France, USA, UK, Netherlands and in recent years PR of China (in 2014 the share of Germany's overall exports to these countries were 9.0%, 8.5%, 7.4%, 6.5% and 6.6% respectively[d].)

The German economy is predominantly based on exports, with exports accounting for 45.7% of its GDP in 2014 (Germany was the 3rd largest importer and exporter in the world in 2014; see Table 1 for exports/imports as a percentage of GDP for other years). They focus on industrially produced goods and services (in 2013 machinery accounted for 18.3% of exports, motor vehicles and parts 16.6% and chemical goods 11.6%[d]). This means that the status of the exports market can be a significant factor for growth within the German economy. The euro is a weaker currency than

[c] www.cia.gov/library/publications/the-world-factbook/geos/gm.html.

[d] Data and statistics taken from Germany's Federal Ministry for Economic Affairs and Energy.

the Deutschmark and this can be positive for the German economy as it means that German exports are cheaper to overseas consumers. However, as the value of the euro increased through 2002 the German economy once again fell into recession (see Figure 1) with a possible factor being the undesirable exchange rate between the euro and major currencies affecting the export markets with the increased price of goods produced in Germany. The financial crisis 2007 - 2009 also had an effect on the export markets when a lack of orders and sales resulted in a severe fall in German exports from 2008 Q4 (in 2008 Germany was the 3rd largest exporter in the world). However, a weak euro can have a positive effect on the export market and thus on the German economy – a record high of 2.2% GDP growth was reported for the 2nd quarter of 2010 (Figure 1). As we can see from Figure 1 the quarter-on-quarter volume growth of GDP for 2012 Q1, Q2 and Q3 were 0.3%, 0.1% and 0.1% respectively, meaning that Germany avoided a further recession at this time, unlike other countries within the euro area (e.g. Greece, Spain and Cyprus).There were two consecutive periods of decrease for Germany in 2012 Q4 and 2013 Q1 (both quarters a decrease of 0.4%) however by 2013 Q2 there was an increase of 0.8% meaning that Germany had recovered from the recession, again unlike countries such as Greece and Cyprus (both countries suffered from a negative quarter-on-quarter volume growth of GDP until 2013 Q4 and 2014 Q4 respectively.)[e]

[e]ECB statistics. Year-on-Year volume growth of GDP and expenditure components: 2.4 Exports (Q-on-Q).

	Exports			Imports		
Year	Total	Goods	Services	Total	Goods	Services
2000	30.9	26.6	4.3	30.6	23.6	7.0
2001	31.9	27.5	4.4	30.1	22.9	7.3
2002	32.6	27.7	4.9	28.2	21.3	6.9
2003	32.6	28.0	4.7	29.0	22.1	6.9
2004	35.5	30.3	5.2	30.4	23.5	6.9
2005	37.8	32.2	5.6	32.7	25.4	7.4
2006	41.2	35.2	6.0	35.9	28.5	7.5
2007	43.1	36.9	6.1	36.4	28.9	7.5
2008	43.5	37.1	6.4	37.5	29.9	7.7
2009	37.8	31.4	6.5	32.9	25.6	7.3
2010	42.3	35.6	6.6	37.1	29.4	7.7
2011	44.8	38.2	6.6	40.0	32.1	7.9
2012	45.9	39.1	6.9	40.0	31.9	8.2
2013	45.6	38.5	7.1	39.8	31.1	8.7
2014	45.7	38.6	7.1	39.1	30.6	8.5

Table 1: The table shows Germany's exports and imports as a percentage of Gross Domestic Product (GDP). The data is taken from 'German Foreign Trade in 2014', a report by Germany's Federal Ministry for Economic Affairs and Energy.

2.3 The DAX Index

This is the benchmark index for the German equity market, representing around 80% of the market capitalisation listed in Germany. Along with some general prerequisites which must be fulfilled for a company to be listed on the DAX (equities listed in the Prime Standard, continuously traded on Xetra with a widely held stock of at least 10% and a head office (or largest sales volume) in Germany) there are two main criteria that must be met based on turnover and market capitalization. Based on these two main criteria, the DAX members are reviewed annually in September for regular entry/exit and in March, June, September and December for fast entry/exit. The rules for these adjustments are outlined below, in order.[f]

Fast Exit (45/45) A company is removed from the DAX if it is no longer one of

[f] `www.dax-indices.com/EN/MediaLibrary/Document/Guide_Equity_Indices.pdf`.

the 45 largest companies according to two quantitative criteria: exchange turnover and market capitalisation, provided that an advancing company ranks 35 or above in both criteria.

Fast Entry (25/25) A company is recorded in the DAX if it is within the 25 largest companies according to both of the two quantitative criteria.

Regular Exit (40/40) A company is removed from the DAX if it is no longer one of the 40 largest companies according to the two criteria. (A non-index value but ranked at least 35 in two criteria.)

Regular Entry (30/30) A company is recorded in the DAX if it is within the 30 largest companies according to the two quantitative criteria unless there is an index value which is no longer in the 35 largest companies according to at least one criterion.

3 A Comparison of Correlation Based Networks

In this chapter we consider three methods for filtering pertinent information from a series of complex networks modelling the correlations between stock price returns of the DAX 30 blue chip stocks for the time period 2001 - 2014. The dataset, created from Thomson Reuters Datastream[g], consists of the closing prices, adjusted for dividends and splits, of the 30 stocks that form the DAX 30 for the time period between 2001 and 2014. This is a significant time period for the German economy as discussed in the previous section. Using the Thomson Reuters Datastream database and also the FNA platform[h] we create the visualisations of the correlation-based networks. These methods reduce the complete 30×30 correlation coefficient matrix to a simpler network structure consisting only of the most relevant edges. The chosen network structures include the Minimum Spanning Tree (MST), Asset Graph (AG) and the Planar Maximally Filtered Graph (PMFG). The resulting networks and the extracted information are analysed and compared, looking at the clusters, cliques and connectivity. We also consider two specific time periods: a period of crisis (October 2008 - December 2008) and also a period of recovery (May 2010 - August 2010) where we discuss the possible underlying economic reasoning for some aspects of the network structures produced.

This chapter is organised as follows. The dataset is presented in Section 3.1. The network structures are discussed in Section 3.2: MSTs (Subsection 3.2.1), AGs (Subsection 3.2.2) and PMFGs (Subsection 3.2.3). The Dax 30 is analysed in Section 3.3 (a period of crisis in 3.3.1 and a period of recovery in 3.3.3). Finally a summary is

[g] Thomson Reuters Datastream 5.0 (`thomsonreuters.com`).

[h] Financial Network Analytics - `http://www.fna.fi`.

given in Section 3.4.

3.1 Data

We begin by taking the daily adjusted closing prices of the 30 stocks that form the DAX 30 for the time period between January 2001 and December 2014. The members of the DAX 30 can change as it is reviewed quarterly (see Section 2.3 for further details) and so we take the current 30 members for each time period considered.

Denoting the adjusted closing price of stock i on day t as $P_i(t)$, we calculate the daily logarithmic returns of the stock prices Y_i, as:

$$Y_i(t) = \ln P_i(t) - \ln P_i(t - \delta t). \tag{4}$$

[Bonanno *et al.* (2004)] and [Tumminello *et al.* (2007)] considered the affect that varying the time horizon, δt, has on the hierarchical organisation of stocks. For our work we use one trading day, setting $\delta t = 1$. To look at the affiliation between the price returns of stocks i and j we calculate the pair-wise correlation coefficient using Pearson product-moment correlation for all trading days in the time period:

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} \tag{5}$$

where $\langle \cdot \rangle$ is an average over the time period (see [Stanley & Mantegna (1999)], Chapter 12 for further details). We use a moving window technique when calculating the correlation coefficient matrices – so the data is separated into annual sets and then we consider a time period of 23 observations (based on an average number of trading days per month) with an interval of 10 days chosen for simplicity. Note that the final window for each annual set will not necessarily contain 23 observations but will end with the last observation for that specific year. This technique gives us a smoother transition of the networks - although can be a compromise with the chance of error. For n stocks, this results in an $n \times n$ matrix with all entries within the interval [-1, 1]. These end values correspond to total anti-correlation between stocks i and j and complete linear correlation between stocks i and j respectively. $\rho_{ij} = 0$ represents no correlation between stocks i and j.

As discussed, the DAX 30 is reviewed quarterly so members can be removed or added to the DAX 30 during certain time windows we consider. For consistency we remove the stocks that are not present throughout the entire time window resulting in some having 28 or 29 stocks rather than 30. For example, 22nd September 2003 saw the regular exit of MLP and the entry of Continental (CON). When modelling the 2003 data we have a time window ranging from 10th September - 10th October 2003 which had 29 stocks as MLP and CON were both omitted. This is done automatically with FNA. In the following sections we consider various correlation based networks that have been presented in the literature as a way of filtering the most relevant data from the complete networks.

3.2 Network Structures

3.2.1 Minimum Spanning Tree (MST)

The first structure that we consider is the Minimum Spanning Tree (MST). As discussed in the Introduction, the MST was used by [Mantegna (1999)] to show the hierarchical arrangement of a portfolio of stocks. The MST extracts the most relevant connections from the correlation matrix and directly gives the subdominant ultrametric hierarchical arrangement of stocks. The stocks are clustered in a way that is entirely based on their correlations and Mantegna noted how this seems to be related to their economic sector.

Let $G(V, E)$ be a connected, undirected graph, where V is the set of vertices and E is the set of edges. A spanning tree $S(V, E')$ of the graph G is a subgraph that is a tree connecting all vertices of G, so if the number of vertices $|V| = n$ then the number of edges $|E'| = n - 1$. For a graph $G(V, E)$ with positively weighted edges we can select the MST – a spanning tree where the sum of the edge weights is less than or equal to that of all other spanning trees. The MST is unique if all of the edge weights are distinct. Various algorithms have been proposed to construct a MST such as [Kruskal (1956)] and [Prim (1957)]. We have applied the Kruskal's algorithm as this method is most common in the literature. To be able to construct a correaltion-based MST we need to define the distance between the vertices and the main method used in the literature is to construct the network using the Euclidean metric.

The distance between the stocks is defined so that the three axioms of a metric space are satisfied:

1. **Positive Definiteness:** For all $p, q, r \in S$ we have $d(p, q) \geq 0$ and $(p, q) = 0 \Leftrightarrow p = q$;
2. **Symmetry:** $d(p, q) = d(q, p)$;
3. **Triangular Inequality:** $d(p, r) \leq d(p, q) + d(q, r)$,

where S is a set and d is a metric on S.

We cannot construct a MST directly from the correlation coefficient matrix as using the correlations as distances would not satisfy these metric axioms – in particular, they do not satisfy the positive definiteness axiom as the correlations range from -1 to 1. Also a stock correlated with itself would give a correlation of 1 and not 0 as required by the first axiom. Furthermore, it is possible to have a high correlation between two stocks but for each of these stocks to have a low correlation with a third stock, which would thus not satisfy the third axiom. To transform the correlation matrix into a distance matrix, a metric function that incorporates the correlation coefficient and satisfies all axioms is needed. We have used a distance function used by [Mantegna (1999)] based on work by [Gower (1966)]:

$$d(i,j) = \sqrt{2(1-\rho_{ij})}, \tag{6}$$

where $d(i,j)$ is the distance between stock i and stock j and ρ_{ij} is the Pearson product-moment correlation coefficient (Eqn. 5) between stock i and stock j. With this distance function we create networks where the shorter the edge length between the vertices (i.e. stocks) the higher the correlation between them (see Appendix B for further details).

The 30×30 correlation coefficient matrix, C, is converted to a distance matrix, D, using the distance function shown in (Eqn. 6). The $n(n-1)/2 = 435$ distances from the upper triangular section of D are then placed in ascending order, so that we can apply Kruskal's algorithm.

The following shows the beginning steps of the construction of a MST for DAX 30 data, with a time window from 5th February 2001 - 7th March 2001. The first nine of these ordered distances are shown in Table 2, along with the corresponding vertices:

Distance	Vertices
0.515	SIE-IFX
0.591	EPC-IFX
0.625	SIE-EPC
0.643	DRB-DBK
0.755	EPC-DBK
0.781	SIE-DBK
0.782	EPC-DRB
0.794	SIE-DRB
0.807	DTE-DBK

Table 2: The first nine ordered distances for the MST construction from 5th February 2001 - 7th March 2001.

The first two vertices are added to the network (Siemens (SIE) and Infineon Technologies (IFX)) with an edge of length 0.515 connecting them. Next, the vertex Epcos (EPC) is added to the network, connected to IFX with an edge of length 0.591. The third edge SIE-EPC is omitted from the network as it would form a cycle between the three vertices. This process continues with SIE-DBK, EPC-DRB and SIE-DRB also being omitted and three more vertices (Dresdner Bank (DRB), Deutsche Bank (DBK), Deutsche Telekom (DTE)) being added from these first nine edges. The first section of the MST is shown in Figure 2. The complete MST for 5th February 2001 - 7th March 2001 is shown in Figure 3.

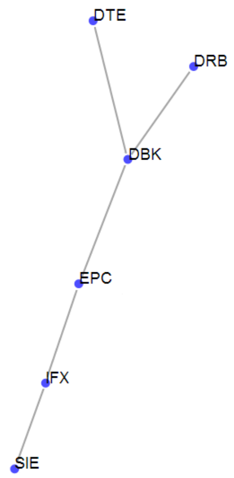

Figure 2: The first six vertices (labelled using the stock's ticker symbol) and five edges of the MST extracted from data showing the correlation between the daily returns of the closing prices for the 30 members of the DAX 30 from 5th February 2001 - 7th March 2001.

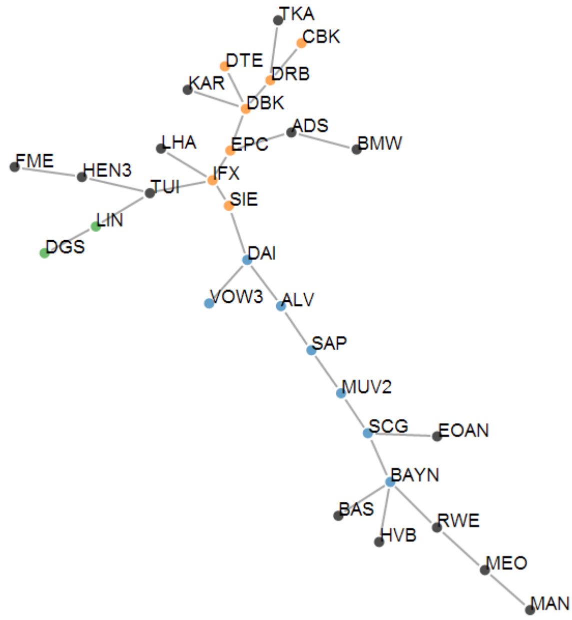

Figure 3: The complete MST extracted from data showing the correlation between the daily returns of the closing prices for the 30 members of the DAX 30 from 5th February 2001 - 7th March 2001. The vertices are labelled using the stock's ticker symbol and have been coloured to highlight the clusters so they can be compared to the AG for the same time period (see Figure 5).

The advantage of constructing this network compared to other methods (AG, PMFG) is that, when calculated in this way, the MST directly determines the subdominant ultrametric distance matrix. The axioms for an ultrametric space are similar to that of a metric space:

1. For all $p, q, \in S$ we have $u(p,q) = 0 \leftrightarrow p = q$;

2. $u(p,q) = u(q,p)$;

3. $u(p,r) \leq \max[u(p,q), u(q,r)]$,

where S is an ultrametric space and u is an associated distance function.
The subdominant ultrametric is a unique ultrametric space that satisfies these axioms and also $u(p,q) \leq d(p,q)$. The subdominant ultrametric distance matrix, $D^<$, can be calculated where the entry $d^<(i,j)$ shows the maximum value of any Euclidean distance from all edges in the shortest path connecting i and j in the MST. This means that a stock i with two different Euclidean distances between itself and two other stocks, say j and k, can have the same ultrametric distance between itself and stocks j and k. These stocks with the same ultrametric distance can then be clustered together, leading to another method for data reduction – hierarchical clustering. This can be shown using a hierarchical clustering structure (known as a *hierarchical tree* or *dendrogram*) which can also be obtained using methods such as Single Linkage Cluster Analysis (SLCA) and Average Linkage Cluster Analysis (ALCA) (see [Gower & Ross (1969)] for further details). The SLCA converts the original correlation matrix C into the subdominant ultrametric distance matrix $D^<$ by reducing C, using an algorithm that selects the maximum correlations. The ALCA reduces the correlation matrix in a similar way; however the resulting matrix and dendrogram vary slightly to that produced by the SLCA as the algorithm uses

average ultrametric distances between vertices.

To illustrate the ultrametric distance between the stocks let us consider the subdominant ultrametric distance matrix, $D^<$, for the first six vertices added to the MST from 5th February 2001 - 7th March 2001, as shown in Figure 2 and Table 2. Notice that there is an edge connecting vertices SIE-IFX, IFX-EPC and EPC-DBK with lengths 0.515, 0.591 and 0.755 respectively. Although there is no edge directly connecting vertices SIE-DBK there is a unique path connecting the two vertices ($T(V, E)$ is a tree $\Leftrightarrow$ there is exactly one path between any 2 vertices $u, v \in V$). The Euclidean distance between these vertices would be the total sum of the lengths of each edge in this path (i.e. $0.515 + 0.591 + 0.755 = 1.861$). The ultrametric distance would be the max length of all edges in the unique path between the vertices (i.e. $\text{Max}[0.515, 0.591, 0.755] = 0.755$).

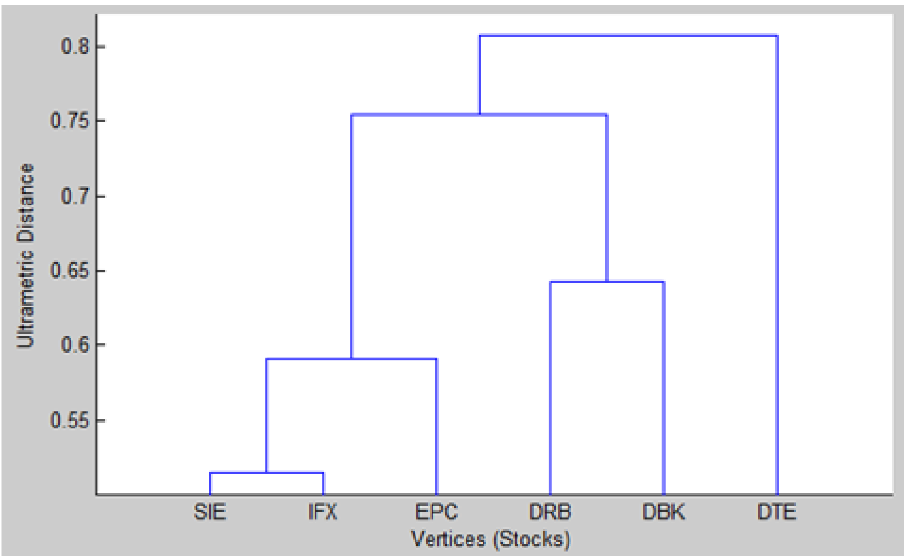

Figure 4: The hierarchical tree extracted from Table 2. The vertices are shown along the x-axis labelled by their ticker symbol. The ultrametric distance is shown on the y-axis.

	SIE	IFX	EPC	DBK	DRB	DTE
SIE	0	0.515	0.591	0.755	0.755	0.807
IFX	0.515	0	0.591	0.755	0.755	0.807
EPC	0.591	0.591	0	0.755	0.755	0.807
DBK	0.755	0.755	0.755	0	0.643	0.807
DRB	0.755	0.755	0.755	0.643	0	0.807
DTE	0.807	0.807	0.807	0.807	0.807	0

Table 3: The table shows the ultrametric distances between the first six vertices added to the MST extracted from data showing the correlation between the daily returns of the closing prices for the 30 members of the DAX 30 from 5th February 2001 - 7th March 2001, with the stocks labelled using their ticker symbols.

As stated above, stocks with the same ultrametric distances to other stocks can be clustered together. For example, SIE and IFX would form a cluster closely linked with EPC. DRB and DBK would form a second cluster. This can be shown using a hierarchical tree; Figure 4 shows a hierarchical tree for the first six stocks.

[Schaeffer (2007)] defined graph clustering as the task of grouping vertices of a graph into clusters taking into consideration the edge structure of the graph so that there are many edges within each cluster but relatively few between the clusters. As the MST does not contain cycles we consider clusters as the groups of vertices with high weighted edges between them. Possible reasons for the formation of these clusters are discussed in Subsection 3.2.2. The MST is probably the most severe form of data reduction. To satisfy the construction algorithm for the MST we may have to omit higher correlations in place of lower correlations so as to keep the resulting graph acyclic. This can be misleading, implying relationships exist between some stocks when they do not.

3.2.2 Asset Graph (AG)

The Asset Graph (AG) was introduced by [Onnela *et al.* (2003)] as a network similar to the MST but as one that includes all of the strongest correlations. The time dependent graph $G^t(V^t, E^t)$ is constructed from either the $n(n-1)/2$ entries of the upper or lower triangular section of the distance matrix, D^t. Note that the distance matrix is calculated using the distance function in Eqn. (6). The $n(n-1)/2$ distances are placed in ascending order. As with the MST, the AG has $n-1$ edges however now the set of edges chosen are the $n-1$ smallest distances from the ordered list. With this selection the set of edges E^t are the $n-1$ strongest correlations (as shorter distances correspond to stronger positive correlations) and are chosen regardless of whether or not they form cycles within the network. A similar approach to the AG is to create threshold networks. [Tse *et al.* (2010)] constructed a threshold network from closing price data on US stocks, using a winner-take-all approach. This method reduces the complete network to a less complex one by including an edge between two stock prices if their cross correlation is larger than a set threshold value. The complexity of the resulting network can be determined by varying this threshold value.

The AG is useful as it again gives us an idea of the clusters formed by the stocks. The graphs created tend to consist of some clique components with the remaining vertices forming either one or two edges with other vertices or being completely unconnected. As both the AG and the MST contain $n-1$ edges we can make comparisons between the two networks, with the AG being useful in identifying any misleading selections made by the MST construction algorithm.

We return to the earlier worked example and look at the data from 5th February 2001 - 7th March 2001. Using the construction algorithm outlined above we create an AG with the 30 members of the DAX 30 during this time period. From Figure 5 we can see that six stocks have formed a 6-clique (SIE, DBK, DRB, DTE, EPC and IFX) and there are also three 3-cliques. Only 16 of the 30 vertices have been included in the AG.

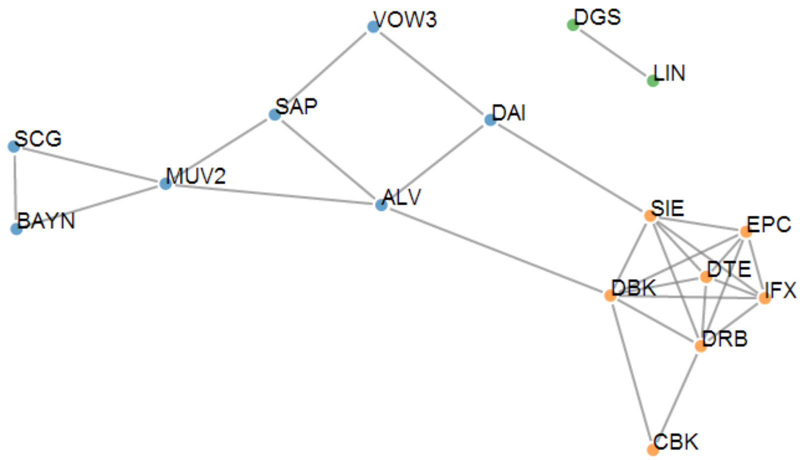

Figure 5: The complete AG extracted from data showing the correlation between the daily returns of the closing prices for the 30 members of the DAX 30 from 5th February 2001 - 7th March 2001. The vertices are labelled using the stock's ticker symbol and have been coloured to highlight the clusters so they can be compared to the MST for the same time period (see Figure 3). Note that the lengths of the edges are not to scale.

From the MST and AG we get a clear indication of the clusters that form between the stocks. These can be stocks from within the same economic sector, for example if we take the set of stocks Bayer (BAYN), BASF (BAS) and Linde (LIN), all within the chemical sector, and look at the 25 MSTs for 2007 we see that at least two of these stocks are connected in 80% of the networks and actually all three are connected in 32% of the networks. In addition we notice that ThyssenKrupp (TKA) is often connected to this set of stocks, sometimes forming the link between two of the connected stocks from the set and the third stock. TKA is in the industrial sector; LIN produces industrial gases and so is classed as being in the industrial gases subsector. Thus the four stocks would form a cluster based on their sectors and subsectors. With the AGs from the same time period we see that it is BAYN that is the central vertex in this group – with a connection between BAYN and BAS in 48% of the networks (an average correlation of 0.7107). As well as producing industrial gases, one of the largest sectors of LIN is Linde Gas Therapeutics (production of medical gases). So LIN can also be seen to form a cluster with Fresenius (FRE), Fresenius Medical Care (FME) and Henkel (HEN3) (all of which belong to the pharmaceutical and health care sector). A similar example can be seen with the set of stocks BMW, Daimler (DAI) and Volkswagen (VOW3), all within the automobile sector, and the networks for 2004 data. The MSTs show that at least two of these stocks are connected in 80% of the networks and all three are connected in 40% of the networks. The AGs for 2004 also show the strong correlations between these stocks, but they also identify that there are strong correlations between two insurance companies, Allianz (ALV) and Munich Re (MUV2), and BMW and DAI. This was something not shown with the MSTs.

Another example is between the two stocks that belong to the utilities sector, E.On (EOAN) and RWE. The 25 MSTs and AGs for 2009 show that the two stocks are connected in 72% of the MSTs and in 64% of the AGs. We can see many clusters of this form are present within the networks and we can identify them using the MST and AG. There are, however, other reasons that these clusters may form that may not be immediately clear. Companies from different sectors can form partnerships or be involved in mergers and acquisitions. For example, in January 2003 Siemens acquired majority control in Sinius GmbH, a technology service set up by Deutsche

Bank. In the 25 MSTs and AGs for 2003 SIE and DBK are connected in 56% of the MSTs and 72% of the AGs (an increase from the previous 2 years). Although we have not considered any social influences, e.g. companies having the same board members, the impact this can have on the networks has been discussed in [Halinen & Tornroos (1998)].

The disadvantage to this method is that we do not get a complete image due to the disconnected vertices. Also, as with the MST, it favours strong, positive correlations. To show this disadvantage we highlight from our data the correlations for VOW3 from 26th August 2008 - 18th December 2008. After several years of acquiring VOW3 shares, Porsche owned 42.6% of VOW3 shares outright and had derivative contracts for a further 31.5% by October 2008 (with 20% of VOW3 shares being Government owned) when they revealed plans to increase this stake to 75% during 2009. There was a rapid increase in the price of VOW3 shares, which was encouraged by Porsche buying options to purchase more shares. On 29th October 2008 Porsche announced they would settle up to 5% of VOW3 options, resulting in a fall in the price of VOW3 shares. During this time period the returns of VOW3 showed some unusual patterns and as a result the correlation matrices showed a negative correlation between the returns for VOW3 and most other stocks (in some cases with all other stocks e.g. the correlation matrix for 23rd September - 23rd October 2008). Due to the nature of the construction algorithms these returns were not highlighted by the MST or the AG.

3.2.3 Planar Maximally Filtered Graph (PMFG)

The final network that we discuss in this section is the filtered graph proposed by [Tumminello *et al.* (2005)] with particular focus on the planar filtered graph (PMFG – created when the graph is embedded into a surface with genus set equal to 0). The networks discussed so far are a severe form of data reduction, containing the minimum number of edges. The proposed filtered graphs allow us to choose how much information we filter from the complete network, so by increasing the genus of the surface we are able to construct a more complex network containing more edges. The PMFG is constructed in a similar way to the MST. For a graph $G(V, E)$ with $|V| = n$ and $|E| = m$ all edges, $e_1, e_2, \dots, e_m$, from the upper triangular section of C are placed in descending order $e_{(1)}, e_{(2)}, \dots, e_{(m)}$. Select the first edge $e_{(1)}$ and construct a graph with $e_{(1)}$ and the two vertices that it connects. Continue selecting the ordered edges and add them to the network structure only if the resulting network can be drawn on a planar surface without edges crossing. There are some tests for planarity based on Kuratowski's theorem [Kuratowski (1930)] that a graph G is planar if and only if it contains neither K_5 nor $K_{3,3}$ as a topological minor. (For more detail on these and others see [Hopcroft & Tarjan (1974)]). The algorithm ends when all vertices $v_1, v_2, \dots, v_n$ are connected, using $3(n-2)$ edges (this is the maximum number of edges in a PMFG – for further details please see Subsection 4.1.2; Eqn. (8)).

The advantage of the PMFG is that it will always contain the corresponding MST and so shows some of the clusters between stocks, but also provides additional information. Unlike the MST, the PMFG does not have a unique path between each of the vertices. This means that we cannot identify the hierarchical clustering between stocks using the subdominant ultrametric distances in the direct way that we can with the MST. However, as the construction algorithm allows the inclusion of cycles the PMFG contains cliques, as with the AG, so we can extract further information from the network by analysing these cliques.

Looking specifically at the PMFG we consider 3- and 4-cliques, as the maximum number of elements that can form a clique is four. By considering the topology of the PMFG we can see that the basic structure (or motif) of the PMFG is a series of 3-cliques. Consider a sphere, a surface with $g = 0$. The PMFG separates the sphere into a sequence of triangular faces, with each vertex of the network belonging to a 3-clique. We can say that the PMFG is the triangulation of a sphere as the network consists entirely of 3-cliques (triangulation of a surface is a partition of that surface by triangles into facets). With our dataset of 30 stocks, we have a total of $\binom{30}{3} = 4060$ possible combinations of 3-cliques from each complete graph. By constructing the maximally filtered graph we considerably reduce the connectivity of the network leaving the most relevant cliques. (The possible structures of 3-cliques are discussed further in Section 4.2). We analyse the 4-cliques by showing the sectors that the four stocks belong to as well as the average correlation coefficient inside the clique, the range between the highest and lowest correlation coefficient in the clique and the standard deviation. Note that [Tumminello *et al.* (2005)] states the maximum number of 4-cliques formed by a PMFG is $n-3$ and we also prove this in Section 4.3.

Let us consider some of the examples highlighted in the previous subsections. We have noted from the 2007 MSTs and AGs that BAS, BAYN and LIN often formed a cluster and they all belong to the chemical sector. For the PMFGs for 2007 the three stocks are connected in 60% of the networks and actually form a 3-clique in 44% of the networks. We also considered the cluster of stocks in the automobile sector for the 2004 data. These clusters are also shown in the PMFGs, with the three stocks being connected in 72% of the networks and a 3-clique forming in 44% of the networks. Finally, the two stocks in the utilities sector, RWE and EOAN, were connected in a high proportion of the MSTs and AGs for 2009 and this was also the case with the PMFGs with a connection in 84% of the PMFGs for 2009.

Cliques also allow us to identify the most connected stocks so that they can not only be clustered but also separated into two sets: core and periphery. This can be done using the AG as, due to the construction algorithm, we often have clique components and unconnected vertices. However, the benefit of the PMFG is that, as it is a connected network, we have a better understanding of the relationships between the stocks that are not identified as being within the core.

Unlike the MST and AG, the PMFG does not necessarily favour the strong, positive stocks. We highlighted VOW3 as an example of a stock that was not fairly represented in the 2008 networks due to its negative correlation. For the PMFG in 2008 we see that VOW3 is mainly connected to three other stocks (84% of the networks) and these are mostly other stocks from the automobile sector (BMW, CON and DAI). At most it is connected to 6 other stocks (this included BMW and DAI). It forms 3-cliques and in some networks a 4-clique, although this 4-clique has a lower average correlation compared to the others from the same PMFG due to the negatively correlated VOW3.

3.3 Analysis of DAX 30

So far we have made comparisons between each of the network structures and discussed their construction and the possible information we can extract. The filtered networks extract clusters of stocks from the complete networks which have high correlations between their return prices. These clusters often form between stocks that belong to the same economic sector and subsector with cross-sector clusters appearing less frequently. There may be some economic reasons for these cross-sector clusters; however they could also be due to errors with the multiple simultaneous estimates made when creating the correlation matrices, such as type I errors (i.e. false positives – identifying a correlation when one does not exist). To this end, we have included the Bonferroni correction parameter when constructing the networks with FNA. For the Bonferroni correction, the familywise error rate (the probability of making one or more type I errors among all hypotheses when performing multiple tests) is set to the chosen level of α (here $\alpha = 0.05$) and each individual test is performed at significance level $\alpha^* = \frac{\alpha}{\mathcal{M}}$ where $\mathcal{M}$ is the total number of tests performed. This method identifies edges in the network structures where the correlation may be classified as being statistically significant or insignificant.

We now discuss two specific time periods in more detail, discussing possible economic reasons for some of their features.

3.3.1 Period of Crisis

The first time period assessed is 7th October 2008 - 31st December 2008 and includes two important events; in October 2008 the German government, market regulators and other financial institutions agreed a €50 billion rescue plan (originally €35 billion, a later deal with an additional €15 billion was agreed on 5th October 2008) to prevent the collapse of Hypo Real Estate, the second largest commercial property lender. This was a sign of the economic problems in Germany – the GDP had declined 0.2% in the 2nd quarter of 2008 and a further 0.4% in the 3rd quarter of 2008 meaning as of 13th November 2008 the German economy was officially in a state of recession (see Figure 1).

(For the MSTs and AGs for this section please refer to Appendix C). The diameter of the MST increases as we move through the time period – this implies that the distances between the vertices is increasing and so the correlations are decreasing. There are some clusters that form – the two stocks from the utilities sector (RWE and EOAN) are strongly connected across the first four MSTs. Stocks in the FIRE (Finance, Insurance and Real Estate) sector (particularly the three banks Commerzbank (CBK), DBK and Deutsche Postbank (DPB)) are also strongly connected, in four of the five MSTs. However, in the final MST many of the edges that connect stocks from the FIRE sector to the tree are classified as insignificant – including CBK, DBK, DPB and MUV2. For the remaining MSTs the edges shown to be insignificant were rather predictable – mainly the edges connecting VOW3, HRE, CON and IFX to the networks for the period of crisis. The correlations that the test has found to be insignificant in the MST are the lower correlations that may have only been chosen to satisfy the construction algorithm. Some of the clusters identified in the complete data set are not present in the MSTs – such as the automobile and the chemical cluster.

We have seven stocks that are not included in any of the AGs for this time period.

VOW3 has been previously discussed. CON and Hypo Real Estate (HRE) were both excluded from the DAX 30 on the basis of the fast-exit rule in December 2008 and similarly DPB and IFX were excluded in Q1 of 2009. The final two stocks that were not included are FME and Metro (MEO). We can see from the complete dataset that for some years FME does not cluster with any other stock and is included in very few AGs between 2002 and 2004 (actually it is not included in any AG for 2003). This could be because the company is fairly unique, being the only healthcare company included in DAX 30 at this point. Let us consider the correlations of the stocks that were included in the AGs. Across the series there is a decrease in correlations – the highest correlated pair falling from 0.9607 for the first AG to 0.8409. Although this is not a significantly large decrease if we consider that for the first AG the lowest correlated pair (the 29th and therefore last edge to be included) was 0.8631 we can see that there has been a decrease in the correlation throughout the complete graph. This supports what is shown by the increase in the diameter of the MSTs.

From the PMFG we can consider the changes observed in the 4-clique analysis (average correlation within the clique, the range and the standard deviations). We also take into account the number of 4-cliques that were observed compared to the maximum total number of 4-cliques that were possible within the graph and the economic sectors that the stocks of each clique belong to. We can see from Table 4 that each PMFG had the maximum number of 4-cliques possible for the number of vertices included. At least one 4-clique formed in each PMFG containing VOW3. The average correlation within this clique was lower than the other averages (due to VOW3 being negatively correlated to all other stocks during this time period). For now we will omit the clique containing VOW3 from the following discussion as this was identified as a special case and explained above (Subsection 3.2.2).

Date	PMFG Analysis						
	Stocks	4-Cliques	Max.4-Cliques	4 Sectors	3 Sectors	2 Sectors	1 Sector
7 Oct - 6 Nov 2008	30	27	27	10	14	3	0
21 Oct - 20 Nov 2008	30	27	27	10	15	2	0
4 Nov - 4 Dec 2008	30	27	27	11	13	3	0
18 Nov - 18 Dec 2008	30	27	27	12	12	3	0
2 Dec - 31 Dec 2008	28	25	25	9	14	2	0

Table 4: The table shows the number of stocks for each time period, within the overall crisis period, the number of 4-cliques that formed, the maximum number of 4-cliques possible for that time period and how many of the cliques were made up of stocks from 4 different economic sectors, 3 different sectors, 2 different sectors or all from the same economic sector.

Overall we can see a decrease in the average correlation within the 4-cliques – the highest and lowest averages for 7th October - 6th November were 0.9044 and 0.7821 respectively, whereas for 2nd December - 31st December the highest was 0.7967 and the lowest 0.4168. When considering the economic sectors that the stocks of the 4-cliques belong to we can see from Table 4 that there are many cliques where all four stocks are from a different sector. To further analyse the 4-cliques we compute a quantity $\langle y \rangle$, as shown by [Tumminello *et al.* (2005)], to calculate the spread of the correlation among the stocks belonging to each clique (where $\rho_{ij} \geq 0$). $\langle y \rangle$ is the mean value of the disparity measure $y(i) = \sum_{j \neq i} [\frac{\rho_{ij}}{s_i}]^2$ over the clique (where i, j are elements of the clique and S_i is the strength of element i). For a clique with all correlations shared evenly between the stocks within the clique $\langle y \rangle = \frac{1}{3}$. For the cliques contained in the PMFGs for this first time period, most have the expected value $\langle y \rangle \approx 0.333$. Within each of the PMFGs for 21st October - 20th November, 4th November - 4th December and 18th November - 18th December there are three cliques that have $\langle y \rangle$ slightly higher than 0.34 (ranging from 0.341 to 0.365). Each of these cliques continued one of the seven stocks mentioned above that were omitted from all AGs for this overall time period. For the final PMFG for this time series (2nd December - 31st December) the value for $\langle y \rangle$ was greater than 0.34 for 11 cliques. The highest value was 0.474 for a 4-clique formed with DTE, FME, MEO and VOW3. This PMFG is the only one for this time period where VOW3 has non-negative correlations; however they are very small in comparison to the others which would explain the higher mean disparity value.

If we consider the edges that have been classified as insignificant within the PMFG we can see that, as with the MST, it is mainly the edges connecting the vertices VOW3, HRE and IFX for the first networks in the series. However, as each vertex in the PMFG has a degree of at least 3 there were more edges that were classified as insignificant compared with the MST. The final PMFG in the series, representing data from 2nd December - 31st December 2008, actually has a larger number of edges classified as insignificant – with vertices from a range of sectors having all the edges connecting it to the remaining network being insignificant.

3.3.2 Case Study – Hypo Real Estate Holdings AG

Hypo Real Estate Holding AG is a German holding company consisting of numerous real estate financing banks, formed in 2003 as part of a reorganisation plan of HVB Bank group. Figure 6 shows the adjusted closing price for HRE shares for the time period between 10th September 2007 and 10th September 2009. HRE had long periods of successful operation and was a member of the DAX 30 from 19th

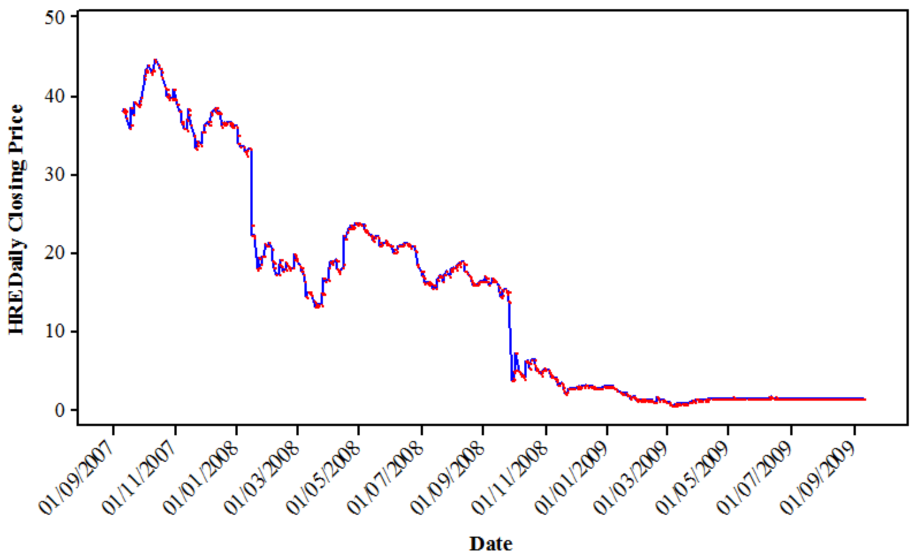

Figure 6: Times series of the daily closing price, adjusted for dividends and splits between 1st September 2007 and 1st September 2009. The red markers indicate the individual days.

December 2005 - 21st December 2008. However, on 15th September 2008 Lehman Brothers (the 4th largest investment bank in US) filed for bankruptcy due to financial problems stemming from the US subprime mortgage crisis. The bankruptcy had an affect on international financial markets and was a contributing factor (combined with its acquisition of Depfa Bank in October 2007) to the liquidity shortage facing HRE by the end of September 2008.[i] In an attempt to prevent the collapse of HRE a €35 billion Government State aid plan was announced on 29th September 2008 to provide much needed liquidity. This is highlighted by the severe dip in Figure 6 where the closing price falls from €13.96 on 26th September 2008 down to €3.8 on 29th September 2008. Plans for further State aid were announced on 6th October 2008 and February 2009. We can see some rises and falls around these time periods in both graphs. HRE, however, was replaced as a DAX 30 member by Salzgitter AG (SZG) on 22nd December 2008.

[i]The Rescue and Restructuring of Hypo Real Estate. Buder *et al.* (2011).

3.3.3 Period of Recovery

The second time period between 7th May and 3rd August 2010 is considered a time of economic success for Germany. With the country officially out of recession in August 2009, there was a significant growth in the country's exports and with that a 3.6% growth to their economy in 2010. The 2nd quarter of 2010 showed a record high in the GDP growth rate (2.2%) (see Figure 1).

(For the MSTs and AGs for this section please refer to Appendix D). For the AGs created for this time period the only stock that is not included is Merck (MRK). FME and FRE are only included in one AG where they form a separate component with only one edge between each vertex. These are the only three stocks in the pharmaceutical and healthcare supersector and, although we cannot comment on the performance of the companies based solely on the networks, we can say that their price returns do not appear to follow the same patterns as the other stocks. There are some companies and selected services that are known to be more resilient during periods of financial crisis and this includes those in the pharmaceuticals and healthcare (Merck reported a record after-tax profit in 2007 and became a member of DAX 30. In 2008 they reported a 7.1% increase in total revenue and in particular an 11% increase in revenue for their pharmaceutical business sector. There was some continued growth through 2009 and by 2010 both Merck and Fresenius Medical shares were considerable outperforming the DAX[j]). MEO is also only included in one AG and for the two time periods this is the only stock in the multiline retail subsector. The range from the largest to smallest correlation in the AGs increases across the time period and the correlations are generally not as high as in the previous time

[j]For further details on each company please refer to the annual reports published for each company.

period.

The MSTs again show that the edges classified as being insignificant are fairly predictable – mainly the edges connecting FRE, FME and MRK to the networks for the period of recovery. The MSTs show some clear clusters based on the economic sectors that the stocks belong to, particularly the automobile, chemical and FIRE sector.

	PMFG Analysis						
Date	Stocks	4-Cliques	Max.4-Cliques	4 Sectors	3 Sectors	2 Sectors	1 Sector
7 May - 8 Jun 2010	30	24	27	7	15	2	0
21 May - 22 Jun 2010	29	23	26	7	12	4	0
4 Jun - 6 Jul 2010	29	16	26	7	7	2	0
18 Jun - 20 Jul 2010	29	26	26	12	13	1	0
2 Jul - 3 Aug 2010	30	27	27	7	17	3	0

Table 5: The table shows the number of stocks for each time period, within the overall recovery period, the number of 4-cliques that formed, the maximum number of 4-cliques possible for that time period and how many of the cliques were made up of stocks from 4 different economic sectors, 3 different sectors, 2 different sectors or all from the same economic sector.

From Table 5 it can be seen that, unlike the first time period, the maximum possible number of 4-cliques did not form in the PMFG. The most significant example of this is during the time between 4th June - 6th July 2010 when only 16 from the possible 27 cliques formed. This is interesting as the AG actually included more of the 30 stocks compared to the AGs constructed for the first time period. A possible reason for this could be that only stocks from certain sectors were performing well – stocks in the FIRE sector and companies that produce goods for exports. Overall the average correlations for the 4-cliques were generally lower for the second time period when compared to those of the first. If we compare the values calculated for $\langle y \rangle$ to the values calculated in the first time period we see that there are even fewer cliques that have $\langle y \rangle$ greater than 0.333, with the highest value being 0.355. There are 13 cliques for the whole time series that have $\langle y \rangle$ higher than 0.34, and of these all but five contain one or more of MEO, FME, FRE or MRK which have been omitted from, or shown in only one, AG for this time period.

[Tumminello *et al.* (2005)] used a similar 4-clique analysis to investigate 100 US stocks from January 1995 to December 1998. The total number of 4-cliques formed was 97, and of these 31 had all four stocks in the same economic sector and 22 had three in the same economic sector. Our 4-clique analysis actually showed that it was more likely for a 4-clique to form with each stock in a different sector or at most two stocks to be in the same sector. Possible reasons for this difference could be that the time periods considered here were not 'average days', as they were a period of crisis and of recovery, and also that the length of the time period's were shorter. The German DAX 30 is also considerably smaller than the 100 US stocks considered by [Tumminello *et al.* (2005)].

Looking at the edges that have been classified as insignificant within the PMFG we see a similar pattern to the PMFGs for the period of crisis. The first networks in the series show that the edges that connect the vertices FME, FRE and MRK to the remaining network are insignificant (the same vertices identified within the MST). For the remaining PMFGs a larger number of edges are shown to be insignificant, and some vertices, such as MEO and MRK, have all of their edges classified as insignificant. This could show that they are not highly correlated with other stocks with the network and have only been included to satisfy the construction algorithm. This supports what was shown with the AGs for the period of recovery.

Some of the correlations may be driven by another factor, such as markets moving up or down in general. To control for this we apply Principal Component Analysis (PCA). PCA identifies patterns in data and expresses data in a way to highlight these similarities so we can control the effect of common factors such as the market return. As PCA needs a complete dataset some vertices were omitted if they were not present throughout the whole time period i.e. for the period of crisis Beiersdorf (BEI), CON, HRE, SZG and TUI were omitted from the networks and for the period of recovery Heidelberg Cement (HEI) and SZG were omitted. When performing PCA with all components we found that there was very little difference between the resulting networks and the original networks. However, following [Laloux *et al.* (1999)], for a second analysis we removed the first and largest component as this most likely represented the variance due to the market return and also removed components greater than component 6 as less than 1.5% of the variance was explained by these components. These networks were slightly different to our original networks but this could be due to the missing vertices. They still supported the findings from our analysis.

3.4 Summary

In summary, in this chapter we have shown three possible methods for filtering information from a complete network of the correlations of the daily adjusted closing prices for DAX 30 stocks. The minimum spanning tree reduces the complete network to the minimum connected structure and can be used to show the hierarchical clustering of the stocks. The clusters that form are likely to be between stocks in the same economic sector. The asset graph separates the complete network into components – generally complete cliques and unconnected vertices. The planar maximally filtered graph combines these two methods by showing some hierarchical clustering, as it will contain the corresponding MST and also highlight the most connected stocks, as with the AG.

We have considered two time periods in detail – a period of crisis and of recovery. Overall we can see that during the period of crisis the correlations decreased throughout the time period and they were generally lower than during the time of recovery. The AGs for the period of recovery had less unconnected stocks than the period of crisis, although the stocks not included in the AGs for the first time period seemed to show some companies that would be omitted from the DAX 30 during, or soon after, the crisis time period. There were fewer clusters for the first time period compared to the second time period – which contained clusters of stocks from the same economic sectors. We note from the 4-clique analysis that the cliques that formed in both time periods contained stocks from three or four different sectors, rather than from one sector as in the literature and from our full time period.

4 Maximally Filtered Graphs

In this chapter, we consider maximally filtered graphs in more detail and consider the construction and possible representations of planar graphs.

As discussed in Chapter 3 one of the key properties of the AG, threshold networks and PMFG is that cliques can form between the vertices in the network which can highlight relationships. [Huang *et al.* (2009)] creates threshold networks to analyse the Chinese stock market using a correlation threshold value $-1 \leq \theta \leq 1$ where θ is the correlation coefficient between two stocks. They study the relationship between the maximum clique, maximum independent set (a subset $I \subseteq V$ such that the subgraph $G(I)$ has no edges) and the threshold value θ. [Huang *et al.* (2009)] state that *'the financial interpretation of the clique in the stock correlation network is that it defines the set of stocks whose price fluctuations exhibit a similar behaviour.'*

We have already shown that the PMFG is an important tool for filtering the most relevant information from a network, particularly in correlation based networks that model the correlation between stock prices. [Aste *et al.* (2005b)] discuss the benefits of studying networks in terms of their surface embeddings. We have previously discussed how the basic structure of the PMFG is a series of 3-cliques. For a set of vertices there are various representations that this underlying series of 3-cliques can form (see Section 4.2). A set of three 3-cliques joined by the shared edges of a fourth 3-clique will form a 4-clique between a group of four vertices. [Aste *et al.* (2005b)] discusses that there must be strong relations between the properties of these 4-cliques and the ones of the system from which the cliques have been generated. [Tumminello *et al.* (2005)] state *'...normalizing quantities are* $n_s - 3$ *for 4-cliques and* $3n_s - 8$

for 3-cliques. Although we lack a formal proof, our investigations suggest that these numbers are the maximal number of 4-cliques and 3-cliques, respectively, that can be observed in a PMFG of n_s elements.' As well as looking at the average correlations within the cliques and whether the cliques are from one sector or cross-sector we also consider the ratio between the number of cliques that have formed to the maximum number of cliques that could form. For this, [Tumminello *et al.* (2005)] used the normalizing quantities that have been mentioned above. This chapter provides the formal proof that $3n - 8$ and $n - 4$ are indeed the maximum numbers of 3-cliques and 4-cliques possible in a PMFG and also an alternative construction algorithm.

4.1 Relational Definitions and Notations

Here we introduce some key terminology that is needed for the proof. Let G be a *planar graph*, i.e. a graph that can be embedded in the plane in such a way that the edges of G will only intersect at the end points (the vertices of G). The planar graph divides the plane into *faces*, with each face bound by a simple cycle of G. The number of edges in this boundary is the *degree* of the face. The *planar representations* of G are all possible isomorphic embeddings of G in the plane.

A *triangulation* of a closed surface is a simple graph, one that does not contain self- or multiple-edges, which is embedded into the surface so that each face is a triangle and that two faces meet along at most one edge. A planar graph is *maximal* if it is triangulated because if a face has more than three edges we can add a diagonal edge. A PMFG is a triangulation of a sphere. Within this thesis we shall denote P_n as a *maximal planar graph* with n vertices.

A *chord* is an edge connecting two vertices of a cycle, which is not included in the cycle itself. For a graph P_n, a cycle of length k ($k \geq 3$) is called a k-cycle, denoted $\mathcal{C}_k$. A cycle $\mathcal{C}$ is a *pure chord-cycle* if the interior of $\mathcal{C}$ contains no vertices and all of the interior faces of $\mathcal{C}$ are triangles. If each of the cycles of four or more vertices within a graph has a chord then the graph is called a *chordal graph*. A *wheel graph*, denoted W_n, is a graph with $n \geq 4$ formed by connecting a single vertex to all other vertices of an $(n-1)$-cycle.

4.1.1 Diagonal Flips

Consider two triangular faces which share a common edge and form a quadrilateral, (see Figure 7). [Negami (1994)] defines a *diagonal flip* of an edge as replacing the existing common edge with a new edge between the other two vertices. A diagonal flip is only possible if the resulting quadrilateral does not contain any multiple edges.

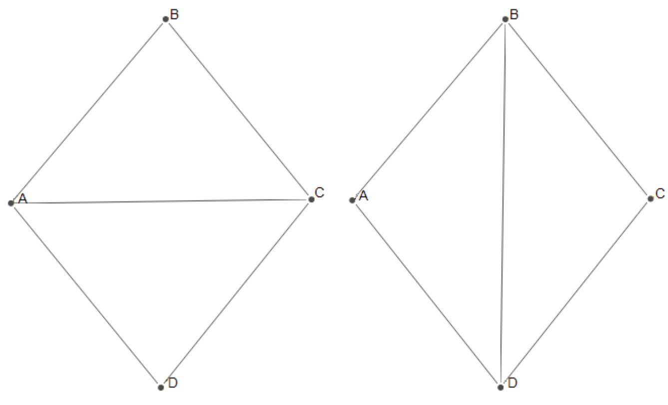

Figure 7: A quadrilateral ABCD is formed by the two adjoining triangles ABC and ACD which share a common edge (A,C). If we perform a diagonal flip the edge (A,C) is replaced by the edge (B,D).

4.1.2 Surface Triangles and Separating 3-Cycles

As a result of Kuratowski's Theorem [Kuratowski (1930)], we know that the PMFG allows cliques up to a maximum size of four vertices (the maximum number considered in this thesis). The 3-cliques can take the form of triangles on the surface (a pure chord-cycle of length 3 that forms a face of the PMFG) or a *separating 3-cycle* (a 3-cycle where both the interior and exterior of C_3 contain one or more vertices). Figure 8 shows this in more detail.

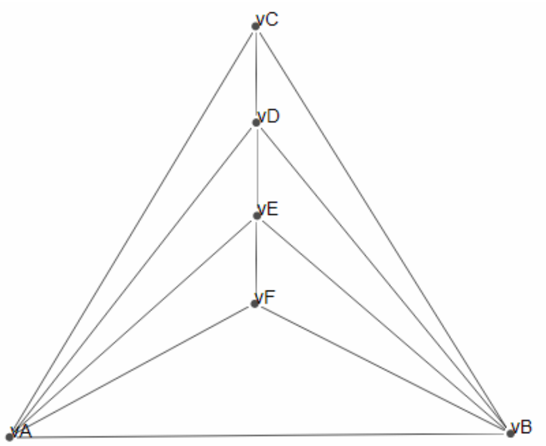

Figure 8: This PMFG with 6 vertices highlights the two possible 3-cliques. Vertices A,C,D form a 3-clique and they outline a triangle on the surface. Vertices A,B,E also form a 3-clique however they do not outline a surface triangle but rather the edges enclose 3 surface triangles which share common edges. A,B,E forms a separating 3-cycle.

The faces bounded by a cycle of edges are called *finite faces* whereas the unbounded face (ABC in Figure 8) is called the *infinite face*. As the PMFG is a triangulation of the sphere this unbounded infinite face will also form a triangle.

In Section 4.3.2 we will study the maximum number of 3-cliques possible in a PMFG, however we begin by studying the maximum number of triangles on the plane (i.e.

the maximum number of faces). To do this we use the Handshaking Lemma and Euler's formula (see Appendix E). For the remaining of this section let $G(V, E)$ be a simple, undirected, finite planar graph.

Proposition 1. *Let G be a PMFG with n vertices, e edges and f faces. Then have $e = 3n - 6$ and $f = 2n - 4$.*

Proof. Since G is planar and $deg(v_i) \geq 2$,

$$\sum_{i=1}^{f} deg(f_i) = 2e.$$

Since G is a triangulation, $deg(f_i) = 3 \Rightarrow 3f = 2e$.
We can substitute into Euler's formula and obtain the following,

$$n - \frac{3}{2}f + f = 2$$

$$f = 2n - 4 \tag{7}$$

and similarly,

$$n - e + \frac{2}{3}e = 2$$

$$e = 3n - 6. \tag{8}$$

□

So for a PMFG we see that the maximum number of these surface triangles is $2n-4$.

4.2 Representations of Each Maximal Planar Graph

In this section the various representations of planar graphs are presented, which are used to achieve the main results shown in Section 4.3.

Using the relationship between the number of edges and the sum of the vertex degree we can calculate the maximum sum of all vertex degrees for a PMFG with n vertices. By considering all combinations of the possible degrees of each vertex we can see what embeddings would be possible and, from these, which would be planar graphs. The following worked example shows this more clearly.

Example 1. Take $G(V, E)$ where $|V| = 8$ then we know $e = 3n - 6 = 18$ and so $\sum_{i=1}^{8} deg(v_i) = 2e = 36$.

Then each vertex can have a degree value from the set $\{3, 4, 5, 6, 7\}$ due to the restrictions that each vertex can be joined to all other vertices at most once (the PMFG does not allow for multiple edges) and the degree of each vertex must be at least 3. From this set there are 27 possible combinations that would give the total degree sum of 36 and from these combinations 13 would produce a planar graph. Graphs that have the same combination of vertex degrees and are isomorphic to each other are known as *planar representations* and they will have the same number of 3-cliques (denoted C_3) and 4-cliques (denoted C_4), (see Figure 9). It is possible however to have graph structures with the same combination of vertex degrees that are not isomorphic, (see Figure 10).

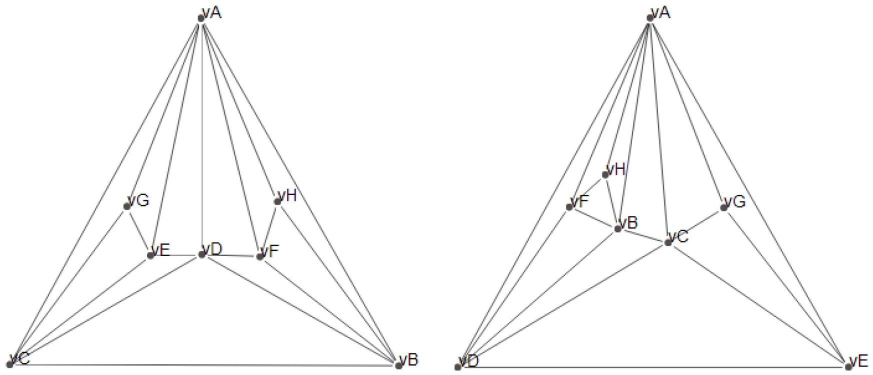

Figure 9: PMFGs with $n = 8$ and $deg[v_A, v_B, v_C, ..., v_H]$=[7, 5, 5, 5, 4, 4, 3, 3]. These graphs are isomorphic and have $C_3 = 16$ and $C_4 = 5$.

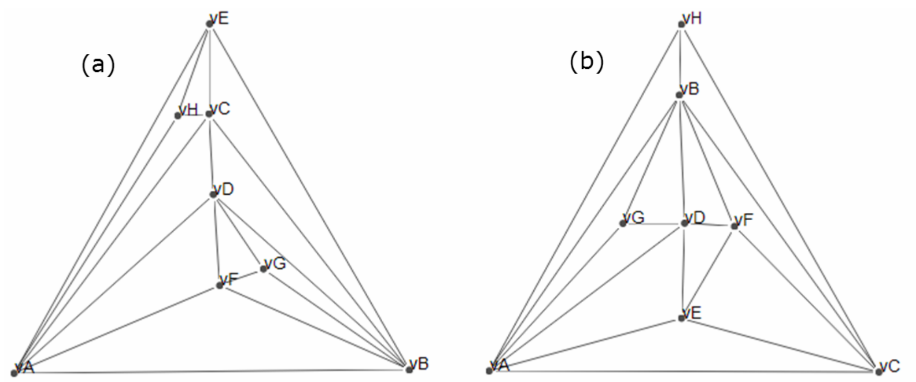

Figure 10: Both graphs have $deg[v_A, v_B, v_C, ..., v_H]$=[6, 6, 5, 5, 4, 4, 3, 3] however they are not isomorphic and so are not planar representations of a single graph. Furthermore, they have different numbers of C_3 and C_4 with the graph shown in Panel (a) having $C_3 = 16$, $C_4 = 5$ and the graph shown in Panel (b) having $C_3 = 14$, $C_4 = 2$.

4.2.1 Standard Spherical Triangulation Form

We now consider how these different embeddings relate to each other using the idea of diagonal flips, as introduced by [Negami (1994)].

In 1936 Wagner proved that any two triangulations of the sphere can be transformed into each other by a finite series of diagonal flips (see also [Bose & Verdonschot (2012)]). Although this does not hold for surfaces in general it has been shown to be true for triangulations of the torus, projective plane and Klein bottle. [Negami

(1994)] generalised the result from [Wagner (1936)],

Theorem 1 (Negami, 1994). *For any closed surface F^2, there exists a positive integer N such that two triangulations G and G' of F^2 are equivalent to each other under diagonal flips if $|V(G)| = |V(G')| \geq N$.*

In the case of the PMFG, the triangulation of the sphere, $N = 4$.

Lemma 1. *Any maximal graph with n vertices, $n \geq 4$, can be transformed to the standard spherical triangulation (or normal form), (see Figure 11), using a series of diagonal flips.*

(For proof please refer to [Ore (1967)], Chapter 1).

From $n \geq 4$, the degrees of each vertex in the standard spherical triangulation are as follows:

$deg[v_1, v_2, v_3, v_4, ..., v_i, ..., v_{n-1}, v_n] = [n-1, n-1, 4, 4, ..., 4, ..., 3, 3]$.

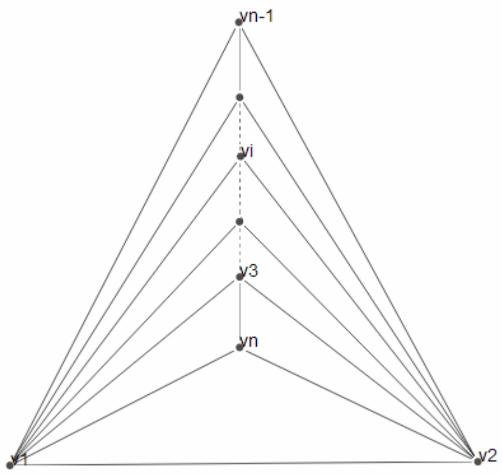

Figure 11: A maximal graph with n vertices in the standard spherical triangulation.

4.3 Main Results

4.3.1 Generating Maximal Planar Graphs

In 1891 Eberhard proposed a system in which a combination of a set of three operations could generate all possible maximally (filtered) planar graphs. We begin with the complete graph K_4 and then choose a generating operation, $\varphi_1, \varphi_2, \varphi_3$, from the operation set, Φ. Each generating operation adds a new vertex to the graph. This system is denoted as $\langle K_4; \Phi = \{\varphi_1, \varphi_2, \varphi_3\}\rangle$.

For $\varphi_1, \varphi_2, \varphi_3$ begin by deleting all of the chords of a pure chord-cycle $\mathcal{C}_k$ with length $k = (3, 4, 5)$, respectively. Then add a new vertex inside $\mathcal{C}$, which is connected to all vertices of $\mathcal{C}$ so that a wheel subgraph is created. Figure 12 shows how a pure chord-cycle transforms under each Eberhard operation.

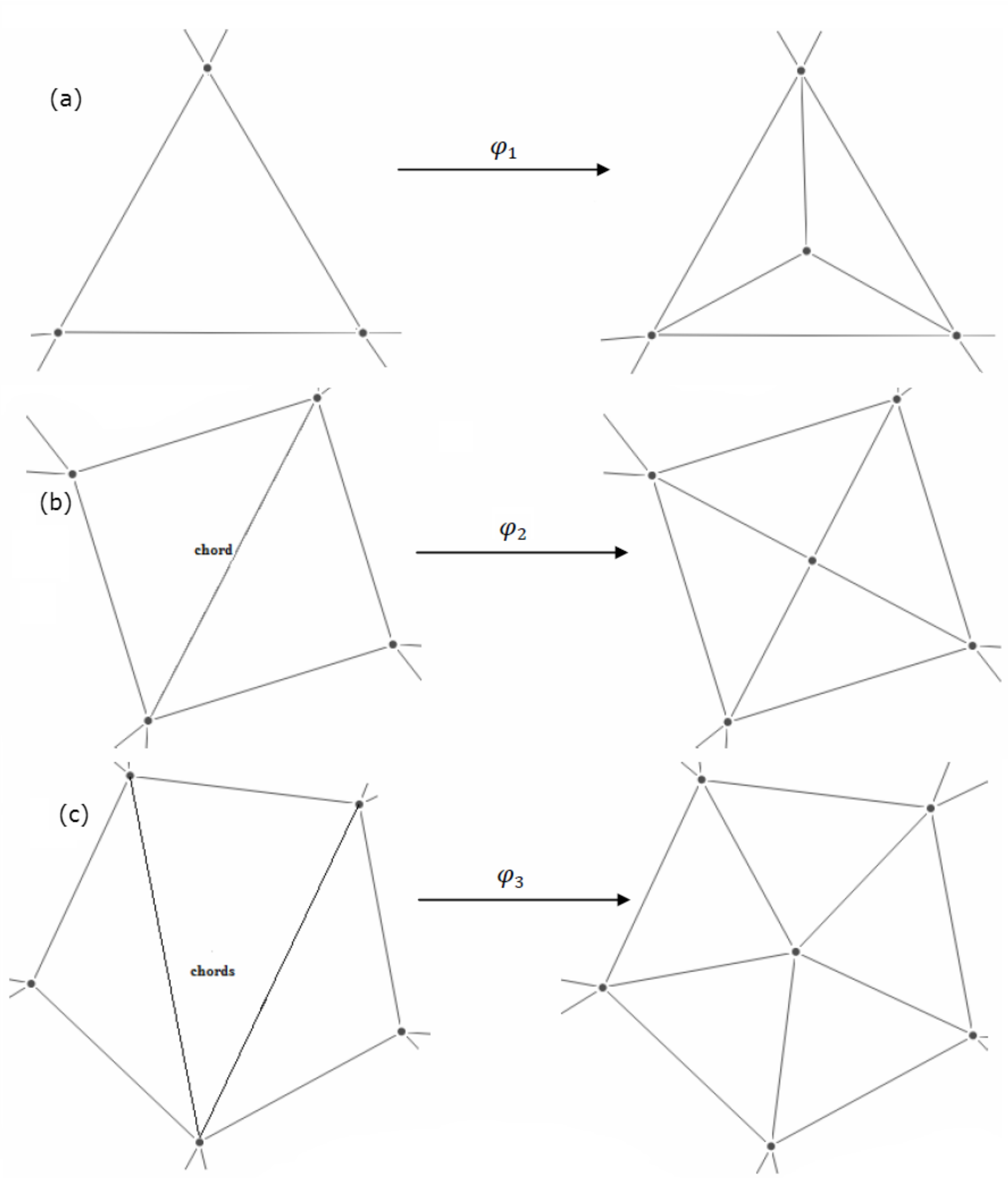

Figure 12: Panel (a) - The first Eberhard operation, φ_1, Panel (b) - The second Eberhard operation, φ_2, Panel (c) - The third Eberhard operation, φ_3. (Presented in [Eberhard (1891)]).

Example 2. For P_6 two of the five possible representations are planar:

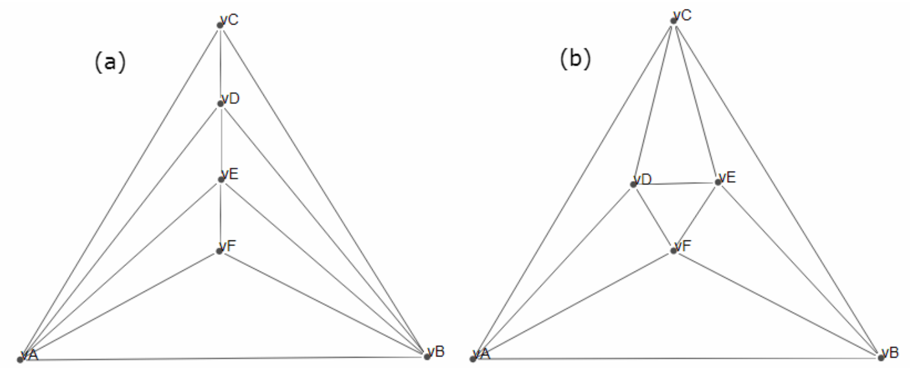

Figure 13: Panel (a) Standard spherical triangulation form, with $C_3 = 10$, $C_4 = 3$ and Panel (b) the alternative form, with $C_3 = 8$, $C_4 = 0$.

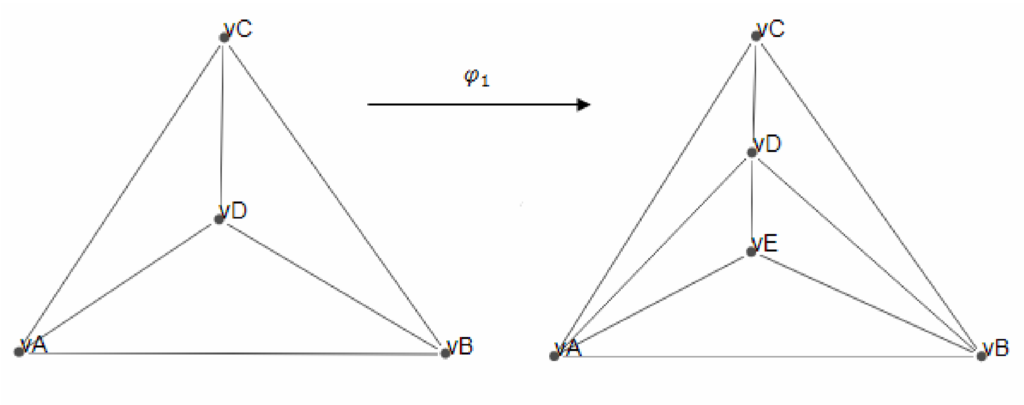

Figure 14: The transformation of K_4 to P_5 using Eberhard operation φ_1.

Both of the graphs shown in Figure 13 can be generated using a series of Eberhard's operations, starting with K_4.

We first generate P_5 from the complete graph K_4. There is only one representation of P_5 that is planar (shown in Figure 14). Then we can generate P_6 from P_5, by using any of the three operations from Φ. In P_5 there are five pure chord-cycles of length 3 which are as follows: (A, B, E), (B, D, E), (B, C, D), (A, C, D) and (A, D, E). For

φ_1 we add a vertex to the interior of one of the above pure chord-cycles which we join to the three edges of the cycle to create a wheel subgraph and a representation of P_6. Note that all $\mathcal{C}_3$ in P_5 will generate P_6 in the standard spherical triangulation form, (see Figure 15).

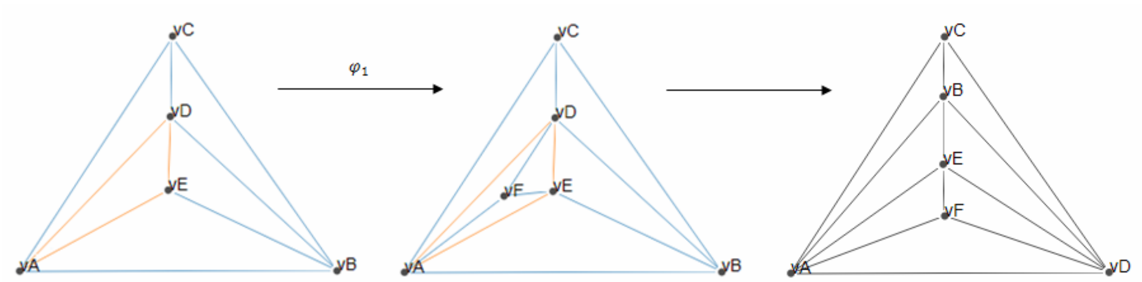

Figure 15: The transformation of P_5 to P_6 using Eberhard operation φ_1.

In P_5 there are four pure chord-cycles of length 4, all with one chord edge, which are as follows: (A, C, B, D), (A, B, D, E), (A, C, D, E) and (B, C, E, D). For φ_2 we select one of these pure chord-cycles and start by deleting the chord edge. Then we add a vertex to the interior which we join to the four edges of the cycle to create a wheel subgraph and a representation of P_6. Using φ_2 can generate P_6 in both standard spherical triangulation form and the alternative form, depending on which pure chord-cycle is chosen, (see Figure 16).

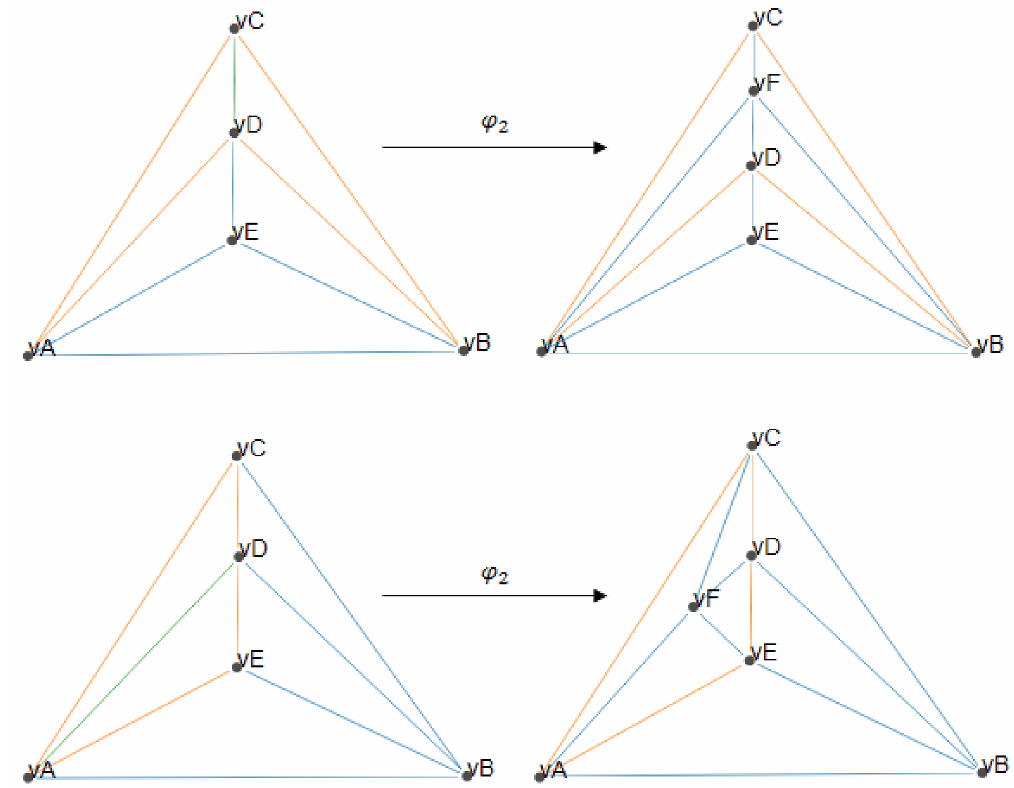

Figure 16: The transformation of P_5 to P_6 using Eberhard operation φ_2.

The final option is to use φ_3 and the one pure chord-cycle of length 5 (A, C, B, D, E). This has two chord edges (A,D) and (C,D) which will be removed before adding a new vertex to the interior. This is then joined to each of the five vertices in the cycle to produce P_6 in standard spherical triangulation form, see Figure 17.

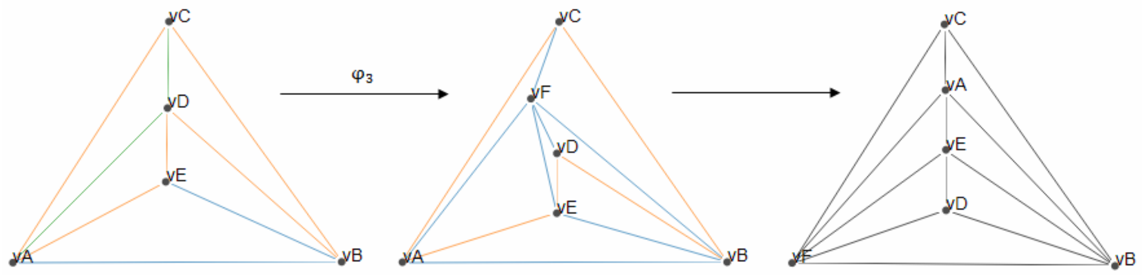

Figure 17: The transformation of P_5 to P_6 using Eberhard operation φ_3.

4.3.2 Maximum Number of 3- and 4- Cliques

Using the Eberhard operations to generate maximal planar graphs we can find the maximum number of 3-cliques that will be added during each iteration of the construction algorithm and so consequently the maximum number of 3-cliques that is possible in P_n.

Now the maximum number of 3-cliques is shown using Theorem 2.

Theorem 2. *Let P_n be a maximal planar graph with n vertices, $n \geq 3$. Then the maximum number of 3-cliques, C_3^{max}, that are possible is $C_3^{max}(P_n) = 3n - 8$.*

Proof. With each Eberhard operation there is a new vertex added and also a certain number of 3-cliques – φ_1 adds three new 3-cliques whereas φ_2 and φ_3 both add two new 3-cliques. Therefore the maximum number of 3-cliques that can be added is three and so we can say that:

$$C_3^{\max}(P_n) \leq C_3(K_4) + 3(n-4), \tag{9}$$

where $(n-4)$ is the number of Eberhard operations required to construct P_n. As we know that the number of 3-cliques in K_4 is always 4 we can simplify (9) and obtain $C_3^{\max}(P_n) \leq 4 + 3n - 12 = 3n - 8$.

So the maximum number of 3-cliques possible in $P_n, C_3^{\max} = 3n - 8$.

□

We can apply a similar argument to obtain the maximum number of 4-cliques in P_n.

Theorem 3. *Let P_n be a maximal planar graph with n vertices, $n \geq 3$. Then the maximum number of 4-cliques, C_4^{max}, that are possible is $C_4^{max} = n - 3$.*

Proof. With each Eberhard operation there is a new vertex added, however only φ_1 adds one new 4-clique, neither φ_2 and φ_3 add any new 4-cliques.

Therefore the maximum number of 4-cliques that can be added is one and so we can say that:

$$C_4^{\max}(P_n) \leq C_4(K_4) + (n-4), \tag{10}$$

where $(n-4)$ is the number of Eberhard operations required to construct P_n. As we know that the number of 4-cliques in K_4 is always 1 we can simplify (10) and obtain $C_4^{\max}(P_n) \leq 1 + n - 4 = n - 3$.

So the maximum number of 4-cliques possible in $P_n, C_4^{\max} = n - 3$. □

4.3.3 3- and 4-Cliques in the Standard Spherical Triangulation Form

As discussed in Section 4.2 there can be various representations of a graph and the number of 3-cliques that form between the vertices will depend on the structure of the graph. The minimum number of 3-cliques that will form in a PMFG with n vertices $= 2n - 4$ as it will be equal to the number of surface triangles, including the vertices that form the infinite face (as shown in Proposition 1). From Eqn. (9) we now have an expression for the maximum number of 3-cliques that can form in a PMFG. We now show that the standard spherical triangulation form always contains the maximum number of 3-cliques. When a maximal planar graph is in standard spherical triangulation form we have two types of 3-cliques – firstly those formed by surface triangles and secondly those that enclose 3 surface triangles which share common edges (as shown in Figure 8) which form between a vertex and the two vertices with degree $n-1$. We call these two forms of triangles surface triangles and enclosing triangles respectively.

Theorem 4. *For a maximal graph with n vertices in the standard spherical triangulation form the number of $C_3 = 3n - 8$ and the number of $C_4 = n - 3$.*

Proof. Number of C_3 in the standard spherical triangulation form = (number of surface triangles) + (number of enclosing triangles) - (unbounded face)
$= (2n - 4) + (n - 3) - 1 = 3n - 8$. Note that the number of enclosing triangles is equal to what we have shown to be the maximum number of C_4. □

4.4 Summary

In this chapter, we have analysed the structure and properties of maximally filtered graphs, with particular focus on Planar Maximally Filtered Graphs (PMFGs). These graphs are an important tool in filtering information from complex networks and by studying the basic structure of these graphs we can gain some insight into the underlying system which has generated the network.

As a result of Kuratowski's Theorem we know that the PMFG allows cliques up to a maximum of four vertices and so the basic structure of a PMFG, which we have considered in this chapter, is a series of 3- and 4-cliques. We have discussed the possible formation of 3-cliques: triangles on a surface and separating 3-cycles as well as possible representations of each maximal planar graph. To do this, we calculated the maximum sum of all vertex degrees for a PMFG with n vertices and considered the combinations of the possible degrees of each vertex so that we could see the possible embeddings.

We used the generating operations proposed by Eberhard to construct these maximal planar graphs and have proven that the maximum number of 3-cliques that can exist in a maximal planar graph with n vertices is $3n - 8$ and the maximum number of 4-cliques that can exist is $n - 3$, where the number of vertices $n \geq 4$. This is true for when a maximal planar graph is constructed using the PMFG algorithm.

Finally, we have shown how any maximal planar graph can be transformed to a standard spherical triangulation form retaining the original number of vertices and edges and that this structure will always contain the maximum number of 3- and 4- cliques.

5 Visibility and Horizontal Visibility Graphs

In this chapter, we consider ways of mapping between the time series and complex networks so that we may use analysis techniques well established within Network Theory ([Newman (2010)] and [Easley & Kleinberg (2010)]) as a useful tool for characterising time series. The networks created inherit properties from the time series so by studying the networks we reveal nontrivial information about the time series itself.

[Zhang & Small (2006)] discussed a method for creating complex networks from pseudoperiodic time series, where each repeated cycle is a single vertex in a network. An edge would connect two of these vertices if the phase space distance between the cycles, corresponding to the vertices is less than a chosen value D. Alternatively, a linear correlation coefficient ρ between two cycles could be used, where two cycles with a larger temporal correlation would be close in phase space. [Lacasa *et al.* (2008)] expanded upon this idea and introduced the Visibility Algorithm, a method which can be applied to different time series, not only pseudoperiodic. The algorithm creates a Visibility Graph (VG) by assigning a vertex to every datum point in the time series, keeping the same temporal order. Vertices are joined with an edge if 'visibility' exists between the vertices. This algorithm was further developed by [Luque *et al.* (2009)] to form the Horizontal Visibility Algorithm – a subgraph of the original Visibility Graph (VG) and as such a geometrically simpler algorithm and more analytically tractable.

5.1 Construction Methods and Results

Within this section we present the formal construction algorithms and also a graphical example for clarity.

5.1.1 Original Algorithm

The visibility algorithm is a fast and simple computational method that converts a time series into a *Visibility Graph*, which inherits some properties of the series in its structure [Lacasa *et al.* (2008)].

Formal Construction. Let $\{x_a\}_{a=1,\ldots,N}$ be a time series of N data. Two arbitrary data values $(t_a, x_a), (t_b, x_b)$ will have visibility (and consequently their corresponding vertices will be connected in the associated graph) if any other data (t_c, x_c) which lies between them, that is for any $t_a < t_c < t_b$, satisfies:

$$x_c < x_b + (x_a - x_b)\left(\tfrac{t_b - t_c}{t_b - t_a}\right).$$

Consider a representation of a time series as a bar chart, where each datum is represented by a bar and the height of this bar takes its value from the times series (keeping the same chromatic order). Two datum points are 'visible' if a straight line can be drawn between the bars without intersecting any other bar. Their corresponding vertices in the associated visibility graph, where every point in the time series is represented with a vertex, are then connected with an edge. Visibility of two vertices is dependent on the height of the bars (or data) and the distance between them. The example shown in Figure 18 gives further details. A bar chart is created for the sample time series (0.698, 0.269, 0.597, 0.178, 0.422, 0.881, 0.030) and two bars are connected if the data are visible e.g. datum point (3, 0.597) and datum point (6, 0.881) are connected as all data points between them ((4, 0.178) and (5, 0.422)) satisfy the above criteria. The corresponding vertices in VG are connected with an edge.

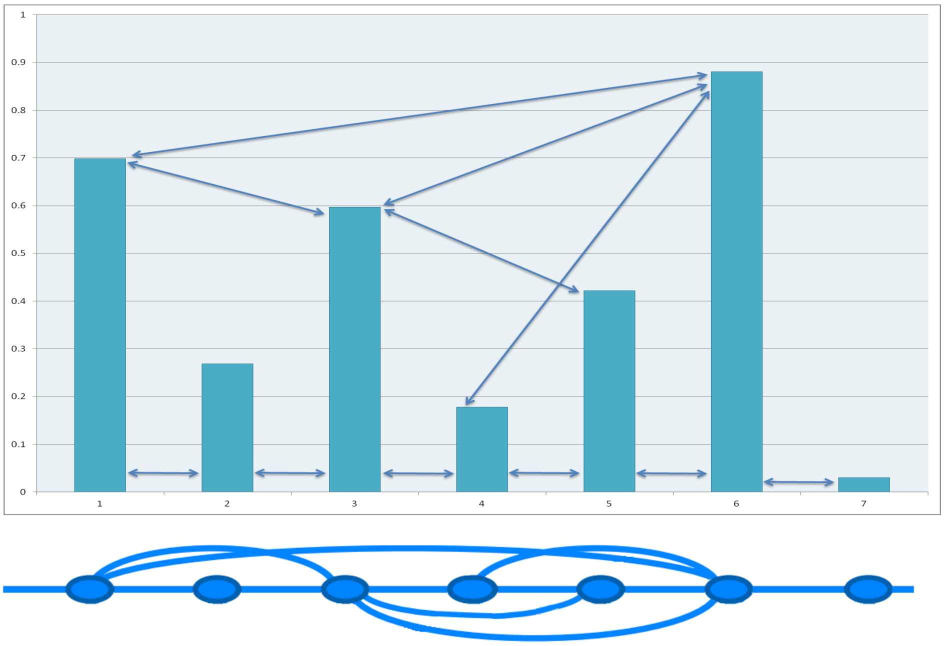

Figure 18: An example of a VG constructed from the time series: 0.698, 0.269, 0.597, 0.178, 0.422, 0.881, 0.030. A bar chart is created, the visible data are connected and the corresponding VG is shown underneath.

5.1.2 Adapted Algorithm

A modification to the above algorithm leads to the *Horizontal Visibility Graph*, a subgraph of the visibility graph with a geometrically simpler visibility criterion [Luque *et al.* (2009)].

Formal Construction. Let $\{x_a\}_{a=1,\ldots,N}$ be a time series of N data. Each datum of the series is assigned a vertex in the horizontal visibility graph (HVG). Two vertices a and b, representing data x_a and x_b, are connected if the following geometrical criterion is fulfilled:

$$x_a, x_b > x_c \text{ for all } n \text{ such that } a < c < b.$$

So again consider a representation of a time series as a bar chart, where each datum is represented by a bar and the height of this bar takes its value from the times series. If a horizontal line can be drawn between two bars of this chart, without intercepting another bar, then the two data represented are connected. An associated HVG, where every point in the time series is represented with a vertex, has two vertices that are connected if visibility exists between the corresponding data. The example shown in Figure 19 gives further details.

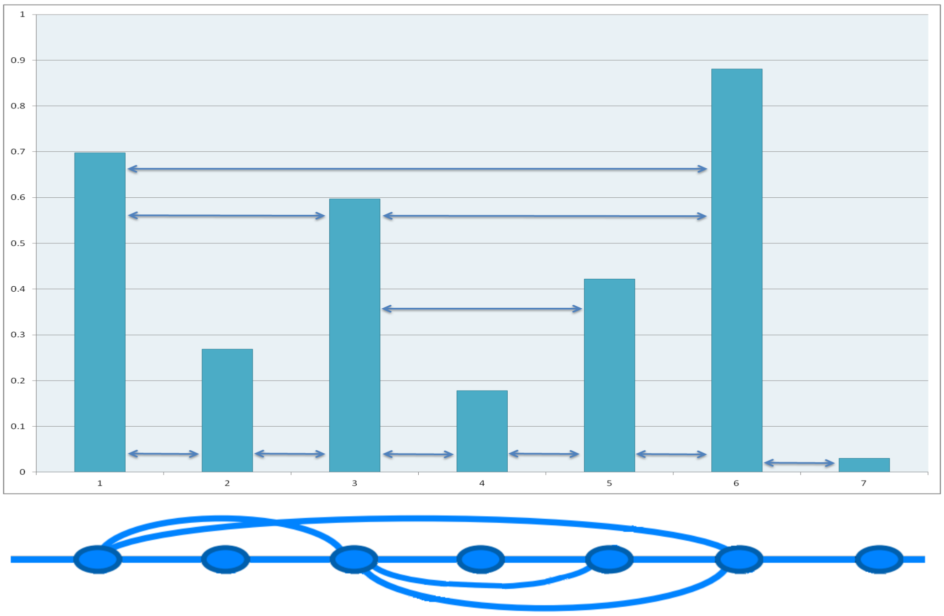

Figure 19: An example of a HVG constructed from the time series: 0.698, 0.269, 0.597, 0.178, 0.422, 0.881, 0.030. A bar chart is created, the visible data are connected using horizontal lines and the corresponding HVG is shown underneath.

5.2 Properties and Proven Results for HVGs

Both visibility and horizontally visibility graphs will always be connected (each vertex is always connected to at least its nearest neighbours) and invariant under affine transformations of the series data (if the series is rescaled horizontally (time) or vertically (values) it will not change the resulting VG or HVG. The same applies for translation of data). Due to the nature of the construction algorithm, any graphs produced will be undirected. There has been some work on directed HVGs in [Lacasa *et al.* (2012)] where the degree of the vertex i is separated into an indegree (edges linking vertex i with other past vertices) and outgoing (edges linking vertex i with other future vertices). As the algorithm absorbs transformation, two time series that differ only by the affine transformations will have the same adjacency matrix and so the same visibility graph. As a result some information from the time series is lost when it is converted to the network structure, hence the algorithm being a form of data reduction. See [Luque *et al.* (2009)] for further details, including how a weighted adjacency matrix is used so that the weights determine the height difference and the resulting graphs differ.

These networks can reflect certain properties of the time series. For example, if we create a visibility graph for a periodic time series then the network will inherit the regularity of the time series and as such will be a regular network (with repeated motifs). The degree distribution is formed by a finite number of peaks related to the series period. By similar reasoning the algorithm also creates an exponential random network from a stochastic times series. Large values refer to rare events and the time distribution of these events is exponential, therefore we expect the tail of the degree distribution to be exponential. And finally scale-free network form from a fractal time series. Examples of these are shown in [Lacasa *et al.* (2008)].

This is a relatively new area of research but some results for the visibility and horizontal visibility graphs have been proven in the literature.

Theorem 5 (**Mean Degree of Periodic Series**). *The mean degree of a HVG associated to an infinite periodic series of period T (with no repeated values within a period) is:*

$$\bar{k}(T) = 4\big(1 - \tfrac{1}{2T}\big).$$

(Please refer to [Nuñez *et al.* (2012)] for proof).

Theorem 6 (**Degree Distribution**). *For a bi-infinite sequence of independent, identically distributed random variables extracted from a continuous probability density, the degree distribution of its associated HVG is:*

$$P(k) = \tfrac{1}{3}\left(\tfrac{2}{3}\right)^{k-2} \textit{ for } k = 2, 3, 4.$$

(Please refer to [Luque *et al.* (2009)] for proof).

As a consequence of Theorem 5 and 6, it has been shown that all HVGs have a mean degree $2 \leq \bar{k} \leq 4$ where a constant series would give the lowerbound of 2 and an aperiodic series would give the upperbound 4.

Theorem 7 (**Mean Degree of Random Series**). *From the previous results we can see that:*

$$\bar{k} = \sum kP(k) = \sum_{k=2}^{\infty} \tfrac{k}{3}\left(\tfrac{2}{3}\right)^{k-2} = 4.$$

As well as the mean degree, there have been results proven for the clustering coefficient distribution for random time series.

Theorem 8 (**Local Clustering Coefficient Distribution**). *For a HVG associated to a random series, a vertex with degree k has the clustering coefficient:*

$$C(k) = \tfrac{2}{k}$$

and the clustering coefficient distribution:

$$P(C) = \tfrac{1}{3}\left(\tfrac{2}{3}\right)^{\frac{2}{C}-2}.$$

(Please refer to [Luque *et al.* (2009)] for proof).

5.3 Applications of HVGs

Reviewing the literature on the family of visibility algorithms, most papers have covered theoretical results for the algorithms and graphs but there have been some applications, as discussed in the Introduction. To follow our work in Chapter 3, the remainder of this chapter focuses on using the HVG to model time series from financial markets, in particular stock price time series.

5.3.1 Using HVG to model stock price time series

We apply the HVG algorithm to a dataset of time series' of daily closing stock prices, adjusted for dividends and splits, of 18 DAX 30 stocks taken from 01/01/1973 - 16/03/2015, so n=11011 days (the data is taken from the earliest records on Thomson Reuters Datastream[a]). The list of stocks and their ticker symbols are shown in Table 6. From these time series' we construct a HVG that will contain some useful information about the underlying phenomena which is, in this paper, the stock price fluctuations.

Table 6: 18 chosen DAX 30 stocks and their ticker symbols.

Ticker	Stock	Ticker	Stock
ALV	Allianz	HEI	HeidelbergCement
BAS	BASF	LHA	Deutsche Lufthansa
BAYN	Bayer	LIN	Linde
BEI	Beiersdorf	MAN	MAN
BMW	BMW	MUV2	Munich Re
CBK	Commerzbank	RWE	RWE
CON	Continental	SDF	K+S
DBK	Deutsche Bank	SIE	Siemens
EOAN	E.On	TKA	ThyssenKrupp

However, before we can model the time series with these graphs we must check that the HV algorithm is appropriate for this task. First we ask whether two graphs that are obtained from two series that differ only in sampling time have the same structural properties? We consider the mean degree and clustering coefficient when analysing the structural properties of a network.

We started by looking at samples of data that were not consecutive days but rather took samples of increasingly fine granularity. For each time series we separated the data into four series consisting of data taken every 4th day, with each series having a different start date ($n = 2753$ or 2752) and also into two series consisting of data from every other day: 'even' days (n=5505 days) and 'odd' days (n=5506) from the time period 01/01/1973 - 16/03/2015. We then constructed a HVG for each of the six time periods for each stock and calculated the metrics above. From these HVGs the mean degrees were all between 3.1948 - 3.8594, and the clustering coefficient 0.2644 - 0.3770. These calculations were also repeated for the HVG representing each of the full datasets (n=11011 days) and the mean degrees were all between 3.4903 and 3.7851 and the clustering coefficient 0.3175 - 0.3644 (please refer to Appendix F for the full table).

So this validates the hypothesis that the HVG approach to the considered stock price series' has some 'physical' sense and if we take coarse and fine resolutions of the same time series, then their HVGs have similar structural properties.

Next, we considered smaller samples of the time series over consecutive days. To do this we looked at the companies from the DAX 30 index (listed in Table 6) for two time periods: the first from the beginning of 2008 through the end of 2009 and the second from the beginning of 2010 through the end of 2011 as these include the dates detailed in Section 3.3 – a period of crisis (7th October 2008 - 31st December 2008) and a period of recovery (7th May 2010 - 3rd August 2010). As before the HVG was constructed for each time period and the mean degree and clustering coefficient calculated; the results for each metric are shown in Tables 7 and 8.

Stock	Crisis	Recovery	Stock	Crisis	Recovery
ALV	3.7667	3.7965	HEI	3.7094	3.8081
BAS	3.6711	3.6967	LHA	3.8738	3.7198
BAYN	3.6979	3.8964	LIN	3.7285	3.7198
BEI	3.8317	3.8733	MAN	3.5985	3.6891
BMW	3.7820	3.6430	MUV2	3.8662	3.7927
CBK	3.6673	3.7236	RWE	3.8050	3.6891
CON	3.7476	3.7159	SDF	3.5641	3.7620
DBK	3.6902	3.7812	SIE	3.7323	3.7198
EOAN	3.6902	3.7735	TKA	3.7361	3.7620

Table 7: The mean degree of the HVG for the 18 stocks (labelled using their ticker symbols for a period of crisis (2008 - 2009) and a periods of recovery (2010 - 2011)).

Stock	Crisis	Recovery	Stock	Crisis	Recovery
ALV	0.3899	0.3954	HEI	0.3772	0.3843
BAS	0.3966	0.3914	LHA	0.3886	0.3919
BAYN	0.3833	0.3802	LIN	0.3864	0.3917
BEI	0.3748	0.3744	MAN	0.3987	0.3919
BMW	0.3900	0.4003	MUV2	0.3970	0.3990
CBK	0.3993	0.3963	RWE	0.3868	0.3913
CON	0.3826	0.3957	SDF	0.3909	0.4003
DBK	0.3967	0.3955	SIE	0.3890	0.3885
EOAN	0.3838	0.3941	TKA	0.3932	0.3954

Table 8: The clustering coefficient of the HVG for the 18 stocks (labelled using their ticker symbols for a period of crisis (2008 - 2009) and a periods of recovery (2010 - 2011)).

From these tables we can see the characteristics of the index and how the HVGs of companies with similar characteristics have similar structural properties, at least mean degree and clustering coefficient. There was little difference between the results for the two different time periods, however the HVGs highlight the randomness of the DAX 30 market. From Theorem 7 we know that a time series generated from a random series will have a mean degree equal to the upperbound i.e. 4. We can see from Table 7 that all mean degrees were between 3.5 and 3.9. This led us to consider the degree distribution of the stocks in more detail.

5.3.2 Stocks: Random or Chaotic?

As discussed in the previous section, [Luque *et al.* (2009)] presented a theorem for the degree distribution of a random time series:

For a bi-infinite sequence of independent, identically distributed random variables extracted from a continuous probability density, the degree distribution of its associated HVG is:

$$P(k) = \frac{1}{3}\left(\frac{2}{3}\right)^{k-2} \text{ for } k = 2, 3, 4.$$

The authors proposed that this theorem can be used to distinguish between random series and chaotic series and used examples to demonstrate this, including a chaotic series with noise pollution and a high dimensional chaotic series. For each of the HVGs representing the full data set (n=11011 days) of 18 DAX 30 time series we have calculated the degree distribution and plotted each of them against the theoretical result given above on a semi-log plot. If the stock price follows the theoretical line then the stock prices are random; if $d(k)$ has a smaller variance then then the stock prices move in a chaotic manner; otherwise they are stochastic. Figure 20 shows the resulting plot.

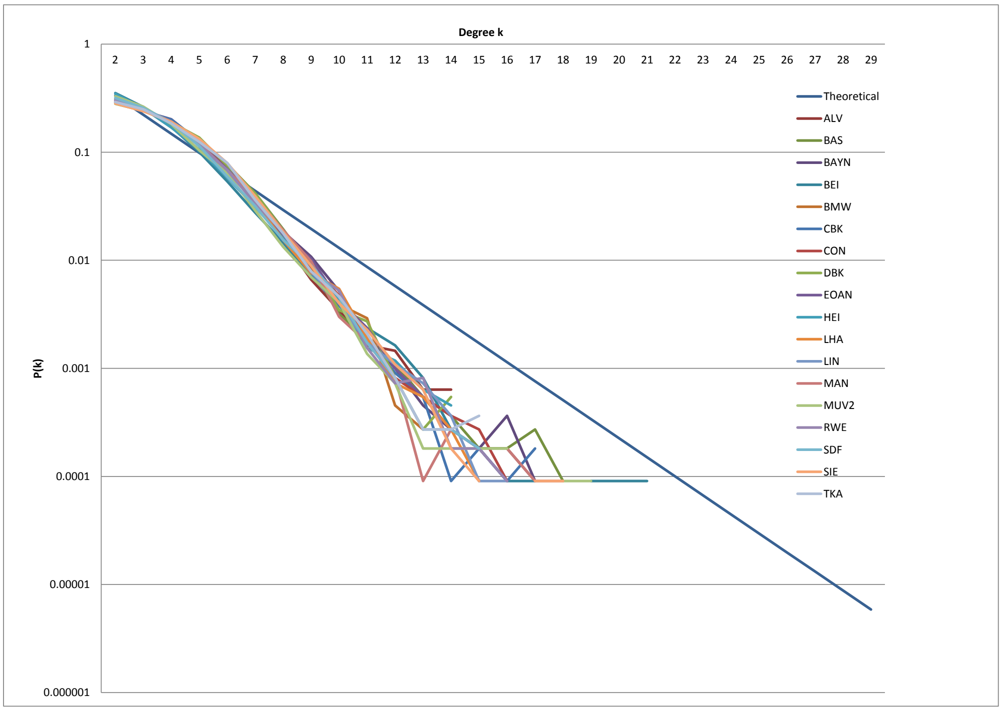

Figure 20: Graph showing the degree distribution for each stock from the full set of n=11011 days along with the theoretical probability given by $P(k) = (\frac{1}{3})(\frac{2}{3})^{k-2}$ as shown by [Luque *et al.* (2009)].

From Figure 20 we can see that, despite the noise within the tail of the distribution, the degree distribution plots for the stock prices falls below the theoretical line of

$P(k)$ meaning that the stock prices are correlated stochastic.

5.4 Summary

In this chapter we have introduced the family of visibility algorithms, namely the visibility graph algorithm and the horizontal visibility graph algorithm. We have discussed the construction algorithms for each graph and their properties such as always connected and invariant under affine transformations. We then reviewed the proven results from the literature and some applications. These theorems covered the mean degree for the HVG associated to random and periodic series, and also the distributions of the degree and clustering coefficients.

The main aim of this chapter was to show that the horizontal visibility graph algorithm is a suitable method to map stock price time series to networks. We have shown that whether we take coarse or fine samples from the time series or smaller samples of consecutive days the HVG will capture some physical sense of the underlying time series. HVGs were constructed for two time series, based on the periods of crisis and recovery presented in Chapter 3, and the mean degree and clustering coefficient calculated. Although the results did not highlight a distinction between the time series in a way that the networks from Chapter 3 did, they highlighted the correlated randomness of the stocks. This was confirmed using the degree distribution and a proven result from [Luque *et al.* (2009)] that for a bi-inifinte sequence of i.i.d random variables extracted from a continuous probability density, the degree distribution of its associated HVG is $P(k) = \frac{1}{3}\left(\frac{2}{3}\right)^{k-2}$ for $k = 2, 3, 4$.

6 Conclusion

Network theory has been used to model a variety of complex systems from research fields including biology, sociology and physics. An important area where network theory has been applied is the study of financial systems and more specifically econophysics, an interdisciplinary research field studying economics and financial phenomena. There are several reasons for studying algorithms converting time series into networks. The first is philosophical – we link two different representations of complexity: temporal (time series) and spatial (networks). The second reason is statistical – the algorithm that converts a time series into a network can be viewed as a data reduction algorithm (continuous information is compressed into discrete). Finally, the third reason is methodological – we bridge two different fields, Time Series Analysis and Network Theory.

Within the third chapter of this thesis our aim was to bring together some of the current work within the literature by creating various network models of the same dataset. By doing so we could make a comparison of the current methods available, gain a better understanding of how this field has developed and consider the advantages of the various methods. The dataset we used consisted of adjusted closing stock price returns for companies who are members of the DAX 30 from 2001 - 2014. This time period covered many economically significant dates including the start of recessions and Government State aid plan's. We created the same series of complete networks and used the methods introduced in the literature to filter the most relevant information from the complete networks using Minimum Spanning Trees (MSTs), Asset Graphs (AGs) and Planar Maximally Filtered Graphs (PMFGs). We then analysed the networks created considering the clusters and cliques created by the vertices (i.e. stocks), the edge lengths and also changes to the networks over time. The MST reduces the complete network to the minimum connected structure and can be used to show the hierarchical clustering of the stocks. The clusters that form are often between stocks in the same economic sector. The AG separates the complete network into components – generally complete cliques and unconnected vertices. As with the clusters in MSTs, the cliques that formed tend to be from the same economic sector. The PMFG combines these two methods by showing some hierarchical clustering, as it will contain the corresponding MST and also highlight the most connected stocks, as with the AG.

We have considered two time periods in detail – a period of crisis and of recovery. Overall we can see that during the period of crisis the correlations decreased throughout the time period and they were generally lower than during the time of recovery. The AGs for the period of recovery had less unconnected stocks than the period of crisis, although the stocks not included in the AGs for the first time period seemed to show some companies that would be omitted from the DAX 30 during, or soon after, the crisis time period. There were fewer clusters for the first time period compared to the second time period – which contained clusters of stocks from the same economic sectors. We note from the 4-clique analysis that the cliques that formed in both time periods contained stocks from three or four different sectors, rather than from one sector as in the literature and from our full time period.

We then focused on maximal planar graphs in Chapter 4 and have discussed possible embeddings of n-vertex triangulations, considering the various representations that are possible and from these which would be planar and which would be isomorphic representations. Within this thesis we considered cliques of 3 and 4 vertices (the maximum size within a PMFG) and discussed the types of the 3-cliques that could form (i.e. surface triangles and separating 3-cycles). We used the generating operations proposed by Eberhard to present a different construction algorithm for maximal planar graphs and have proven that the maximum number of 3-cliques that can exist in a maximal planar graph with n vertices is $3n - 8$ and the maximum number of 4-cliques that can exist is $n - 3$, where the number of vertices $n \geq 4$. This is true for when a maximal planar graph is constructed using the PMFG algorithm presented in the literature. We have shown how any maximal planar graph can be transformed to a standard spherical triangulation form retaining the original number of vertices and edges and that this structure will always contain the maximum number of 3- and 4- cliques.

Throughout the research process we were presented with possible area's of error. For example with the comparison between networks, errors can occur when creating the correlation matrices and calculating multiple simultaneous estimates. We addressed this problem by including the Bonferroni correction parameter when constructing the networks with FNA. We also applied Principal Component Analysis (PCA) to control for any factors that could be affecting all correlations i.e. the general market movements.

Finally, in Chapter 5 we have studied the family of visibility algorithms, namely the visibility graph algorithm and the horizontal visibility graph algorithm. This is a relatively new area of research and our aim was to investigate whether the visibility algorithm would be a suitable method for mapping financial time series, such as stock prices, to a HVG. We have discussed the construction algorithms for each graph, the proven results from the literature and some applications. We validated the hypothesis that when the HVG approach is applied to a stock price time series', the resulting graph has some physical sense and if we take times series of increasingly fine granularity, then their HVGs have similar structural properties.

We then considered smaller samples of consecutive days and constructed HVGs for two datasets of stock price time series for 18 stocks belonging to the DAX 30, for a time period that covered the periods of crisis and recovery presented in Chapter 3, calculating the mean degree and clustering coefficient of the graphs. Unlike the networks from our earlier work, there was no clear distinction between the graphs and the metrics calculated for the two time periods. A possible reason for this is that the algorithm captures the trends within the time series dataset as a whole, identifying changes in the individual time series (i.e. stock prices) rather than comparing the individual time series'. It would be interesting to investigate other measures, however these may need to be adapted for use with HVGs i.e. connectivity would not be the same as with standard networks since a property of HVGs is that each vertex is always connected to its neighbour.

There were also limitations with the datasets, in particular we questioned whether the datasets used when generating visibility and horizontal visibility graphs were of an adequate length. To this end, we will be considering high frequency data in our future research. We would again investigate the two time periods (crisis and recovery) and propose that there would be more variation in the calculations of the metrics, showing some differences between the two time periods.

The HVGs constructed for the two time periods did highlight the correlated randomness of the stocks. This was confirmed using the degree distribution and a proven result from [Luque *et al.* (2009)] that for a bi-inifinte sequence of i.i.d random variables extracted from a continuous probability density, the degree distribution of its associated HVG is $P(k) = \frac{1}{3}\left(\frac{2}{3}\right)^{k-2}$ for $k = 2, 3, 4$. We would like to continue with this direction and consider the clustering coefficient distribution, applying the proven result from [Luque *et al.* (2009)] that for a HVG associated to a random time series, $P(C) = \frac{1}{3}(\frac{2}{3})^{\frac{2}{C}-2}$. It has been shown that the mean degree for a HVG $k \in [2, 4]$, can a similar interval been proven for the clustering coefficient? Finally, we will use this method to consider the hidden periodicity in the stocks. We would make use of a result from [Nuñez *et al.* (2012)] who have proposed the HVG as an algorithm for detecting periodicity in time series. Current algorithms can be classified in two categories: time domain and frequency domain. Nuñez *et al.* propose a third category, graph theoretical methods, making use of HVGs.

Overall within this thesis we have investigated how the properties of complex network theory can be used to explain and better understand financial markets and used to study the economy as a whole. In particular we have concentrated on time series of stock prices for companies belonging to the DAX 30. In the future we would also extend our analysis to include other markets and compare the results with our results from the DAX 30 dataset. This could include larger markets such as FTSE 100 but also emerging markets.

References

[Albert & Barabási (2002)] R. Albert and A.-L. Barabási. *Statistical Mechanics of Complex Networks.* Reviews of Modern Physics, Vol. 74 (No. 1) 47-97. (2002)

[Allen & Gale (2000)] F. Allen and D. Gale. *Financial Contagion.* Journal of Political Economy, Vol. 108 (Iss. 1) 133. (2000)

[Aste *et al.* (2005a)] T. Aste, T. Di Matteo, M. Tumminello, R.N. Mantegna. *Correlation Filtering in Financial Time Series.* Noise and Fluctuations in Econophysics and Finance. Proceedings of SPIE, 5848: 100-109. (2005)

[Aste *et al.* (2005b)] T. Aste, T. Di Matteo, S.T. Hyde. *Complex Networks on Hyperbolic Surfaces.* Physica A, Vol. 346 (No. 1-2 Spec. Iss.) 20-26. (2005)

[Becher *et al.* (2008)] C. Becher, S. Millard, K. Soramäki. *The Network Topology of CHAPS Sterling.* Working Paper No. 355, Bank of England. (2008)

[Bonanno *et al.* (2000)] G. Bonanno, N. Vandewalle, R.N. Mantegna. *Taxonomy of Stock Market Indices.* Physical Review E, 62: R7615-R7618. (2000)

[Bonanno *et al.* (2004)] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciche, N. Vandewalle, R. N. Mantegna. *Networks of Equities in Financial Markets.* European Physical Journal B, Vol. 38 (Iss. 2) 363-371. (2004)

[Bose & Verdonschot (2012)] P. Bose, S. Verdonschot. *A History of Flips in Combinatorial Triangulations.* Computational Geometry 29 (2012)

[Boss *et al.* (2004)] M. Boss, H. Elsinger, M. Summer, S. Thurner. *Network Topology of the Interbank Market.* Quantitative Finance, Vol. 4 (No. 6) 677-684. (2004)

[Brookfield *et al.* (2013)] D. Brookfield, H. Boussabaine, C. Su. *Identifying Reference Companies using the Book-to-Market Ratio: a Minimum Spanning Tree Approach.* The European Journal of Finance. Vol. 19 (Iss. 6) 466-490. (2013)

[Coronnello *et al.* (2005)] C. Coronnello, M. Tumminello, F. Lillo, S. Micciche, R.N. Mantegna. *Sector Identification in a Set of Stock Return Time Series Traded at the London Stock Exchange.* Acta Physica Polonica B Vol. 36 (No. 9) 2653-2679. (2005)

[Easley & Kleinberg (2010)] D. Easley and J. Kleinberg. *Networks, Crowds and Markets.* Cambridge University Press. (2010)

[Eberhard (1891)] V. Eberhard. *Morphology of the Polyhedron* BG Teubner, [German Text]. (1891)

[Eryigit & Eryigit (2009)] M. Eryigit and R. Eryigit. *Network structure of cross-correlations among the world market indices.* Physica A, Vol. 388 (No. 17) 3551-3562. (2009)

[Gutin *et al.* (2011)] G. Gutin, T. Mansour and S. Severini. *A Characterization of Horizontal Visibility Graphs and Combinatorics on Words.* Physica A, Vol. 390 (No. 12) 2421-2428. (2011)

[Galbiati & Soramäki (2012)] M. Galbiati and K. Soramäki. *Clearing Networks.* Journal of Economic Behavior and Organization, 83: 609-626. (2012)

[Gower (1966)] J.C. Gower. *Some Distance Properties of Latent Root and Vector Methods used in Multivariate Analysis.* Biometrika, 53: 325-338. (1966)

[Gower & Ross (1969)] J.C. Gower and G.J.S. Ross. *Minimum Spanning Trees and Single Linkage Cluster Analysis.* Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 18 (No. 1) 54-64. (1969)

[Halinen & Tornroos (1998)] A. Halinen and J. Tornroos. *The Role of Embeddedness in the Evolution of Business.* Scandinavian Journal of Management, Vol. 14 (No. 3) 187-205. (1998)

[Hopcroft & Tarjan (1974)] J. Hopcroft and R. Tarjan. *Efficient Planarity Testing.* Journal of the Association for Computing Machinery, Vol. 21 (No. 4) 549-568. (1974)

[Huang *et al.* (2009)] W-Q. Huang, X-T. Zhuang, S. Yao. *A Network Analysis of the Chinese Stock Market.* Physica A, Vol. 388 (No. 14) 2956-2964. (2009)

[Iori *et al.* (2008)] G. Iori, G. De Masi, O.V. Precup, G. Gabbi, G. Caldarelli. *A Network Analysis of the Italian Overnight Money Market.* Journal of Economic Dynamics and Control, Vol. 32 (No. 1) 259-278. (2008)

[Kruskal (1956)] J.B. Kruskal. *On the Shortest Spanning Subtree of a Graph and the Travelling Salesman Problem.* Proceedings of the American Mathematical Society, 7: 48-50. (1956)

[Kuratowski (1930)] K. Kuratowki. *On the Problem of Space Curves Topology.* Fundamenta Mathematicae 15, 271-283. (1930)

[Lacasa *et al.* (2008)] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J.C. Nuño. *From Times Series to Complex Networks: The Visibility Graph.* Proceedings of the National Academy of Sciences. Vol.105 (No. 13) 4972-4975. (2008)

[Lacasa *et al.* (2009)] L. Lacasa, B. Luque, J. Luque and J.C. Nuño. *The Visibility Graph: A New Method for Estimating the Hurst Exponent of Fractional Brownian Motion.* Europhysics Letters 86, 30001-5. (2009)

[Lacasa *et al.* (2012)] L. Lacasa, A. Nuñez , É. Roldán, J.M.R. Parrondo and B. Luque. *Time Series Irreversibility: A Visibility Graph Approach* European Physical Journal B, 85: 217. (2012)

[Laloux *et al.* (1999)] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters. *Noise Dressing of Financial Correlation Matrices.* Physical Review Letters, 83. (1999)

[Leither (2005)] Y. Leither. *Financial Networks: Contagion, Commitment, and Private Sector Bailouts.* Journal of Finance, Vol. 6 (No. 6) 2925-2953. (2005)

[Li *et al.* (2010)] S. Li, J. He, Y. Zhuang. *A Network Model of the Interbank Market.* Physica A, Vol. 389 (No. 24) 5587-5593. (2010)

[Liu *et al.* (2009)] C. Liu, W.-X. Zhou and W.-K. Yuan. *Statistical Properties of Visibility Graph of Energy Dissipation Rates in Three-dimensional Fully Developed Turbulence.* Physica A, Vol. 389 (No. 13) 2675-2681. (2010)

[Luque *et al.* (2009)] B. Luque, L. Lacasa, F. Ballesteros and J. Luque. *Horizontal Visibility Graphs: Exact Results for Random Time Series.* Physical Review E, 80: 046103. (2009)

[Mantegna (1999)] R.N. Mantegna. *Hierarchical Structure in Financial Markets.* European Physical Journal B, Vol 11 (Iss. 1) 193-197. (1999)

[Mizuno *et al.* (2006)] T. Mizuno, H. Takayasu and M. Takayasu. *Correlation Networks Among Currencies.* Physica A, Vol. 364 (Iss. C) 336-342. (2006)

[Negami (1994)] S. Negami. *Diagonal Flips in Triangulations of Surfaces.* Discrete Mathematics Vol. 135: 225-232. (1994)

[Newman (2003)] M.E.J Newman. *The Structure and Function of Complex Networks.* Society for Industrial and Applied Mathematics Review 45(2), 167-256. (2003)

[Newman (2010)] M.E.J Newman. *Networks: An Introduction.* Oxford University Press. (2010)

[Nuñez *et al.* (2012)] A. Nuñez, L. Lacasa, E. Valero, J.P. Gomez and B. Luque. *Detecting Series Periodicity with Horizontal Visibility Graphs.* International Journal of Bifurcation and Chaos. Vol. 22 (No. 7) 12501601-10. (2012)

[Onnela *et al.* (2002)] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz. *Dynamic Asset Trees and Portfolio Analysis.* European Physical Journal B, Vol. 30 (Iss. 3) 285-288. (2002)

[Onnela *et al.* (2003)] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto. *Asset Trees and Asset Graphs in Financial Markets.* Physic Scripta, T106: 48-54. (2003)

[Ore (1967)] O. Ore. *The Four-Color Problem.* Academic Press 9. (1967)

[Pozzi *et al.* (2013)] F. Pozzi, T. Di Matteo, T. Aste. *Spread of Risk Across Financial Markets: Better to Invest in the Peripheries.* Scientific Reports 3, 1665. (2013)

[Prim (1957)] R.C. Prim. *Shortest Connection Networks and Some Generalizations.* Bell System Technical Journal, 36: 1389-1401. (1957)

[Qiu *et al.* (2010)] T. Qiu, B. Zheng, G. Chen. *Financial Networks with Static and Dynamic Thresholds.* New Journal of Physics, Vol. 12 (No. 4) 1-16. (2010)

[Schaeffer (2007)] S.E. Schaeffer. *Graph Clustering.* Computer Science Review, Vol. 1 (No. 1) 27-64. (2007)

[Song *et al.* (2012)] W-M. Song, T. Di Matteo, T. Aste. *Hierarchical Information Clustering by Means of Topologically Embedded Graphs.* Public Library of Science ONE, Vol. 7 (No. 3) (2012) e31929. (2012)

[Soramäki *et al.* (2007)] K. Soramäki, M. Bech, J. Arnold, R.J. Glass, R.J. Beyeler. *The topology of interbank payment flows.* Physica A, Vol. 379 (No. 1) 317-333. (2007)

[Stanley & Mantegna (1999)] H. E. Stanley and R. Mantegna. *Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press. (1999)

[Toivonen *et al.* (2006)] R. Toivonen, J.-P Onnela, J. Saramaki, J. Hyvonen, K. Kaski. *A Model for Social Networks* Physica A, Vol. 371 (No. 2) 851-860. (2006)

[Tse *et al.* (2010)] C.K. Tse, J. Liu, F. C. M. Lau. *A Network Perspective of Stock Market.* Journal of Empirical Finance, Vol. 17 (No. 4) 659-667. (2010)

[Tumminello *et al.* (2005)] M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna. *A Tool for Filtering Information in Complex Systems.* Proceedings of the National Academy of Sciences, Vol. 102 (No. 30) 0421-10426. (2005)

[Tumminello *et al.* (2007)] M. Tumminello, T. Di Matteo, T. Aste, R.N. Mantegna. *Correlation Based Networks of Equity Returns Sampled at Different Time Horizons.* European Physical Journal B Vol. 55 (Iss. 2) 209-217. (2007)

[Vandewalle *et al.* (2001)] N. Vandewalle, F. Brisbois, X. Tordoir. *Non-random Topology of Stock Markets.* Quantitative Finance, Vol. 1 (No. 3) 372-374. (2001)

[Wagner (1936)] K. Wagner. *Remarks About the Four Color Problem.* Journal der Deutschen Mathematiker-Vereinigung Vol. 46, 26-32. (1936)

[Xie & Zhou (2011)] W-J. Xie and W-X Zhou. *Horizontal Visibility Graphs Transformed From Fractional Brownian Motions: Topological Properties Versus Hurst Index.* Physica A, Vol. 390 (No. 20) 3592-3601. (2011)

[Xu *et al.* (2008)] X. Xu, J. Zhang and M. Small. *Superfamily Phenomena and Motifs of Networks Induced from Time Series.* Proceedings of the National Academy of Sciences, Vol. 105 (No. 50) 19601-19605. (2008)

[Yu *et al.* (2012)] Z.G. Yu, V. Anh, R. Easte and D.-L. Wang. *Multifractal Analysis of Solar Flare Indices and their Horizontal Visibility Graphs.* Nonlinear Processes in Geophysics, Vol. 19 (No. 6) 657-665. (2012)

[Yang *et al.* (2009)] Y. Yang, J. Wang, H. Yang and J. Mang. *Visibility Graph Approach to Exchange Rate Series.* Physica A, Vol. 388 (No. 20) 4431-4437. (2009)

[Zhang & Small (2006)] J. Zhang and M. Small. *Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics.* Physical Review Letters, Vol. 9 (No. 23) 238701-4. (2006)

Appendix A

Table 9: List of all stock symbols and the supersector, sector and subsector that the company belongs to. The details of the various sectors can be found in Guide to the Equity Indices of Deutsche Börse. Version 6.6, November 2008 (http://www.Deutscheboerse.com).

Ticker	Company	Supersector	Sector	Subsector
AAA	Altana	Basic Materials	Chemicals	Chemicals, Specialty
ADS	Adidas	Consumer Goods	Consumer	Clothing and Footwear
ALV	Allianz	FIRE	Insurance	Insurance
BAS	BASF	Basic Materials	Chemicals	Chemicals, Specialty
BAYN	Bayer	Basic Materials	Chemicals	Chemicals, Specialty
BEI	Beiersdorf	Consumer Goods	Consumer	Personal Products
BMW	BMW	Consumer Goods	Automobile	Automobile Manufacturers
CBK	Commerzbank	FIRE	Banks	Credit Banks
CON	Continental	Consumer Goods	Automobile	Auto Parts and Equipment
DAI	Daimler	Consumer Goods	Automobile	Automobile Manufacturers
DB1	Deutsche Börse	FIRE	Financial Services	Securities Brokers

Continued on next page

Table 9 – *Continued from previous page*

Ticker	Company	Supersector	Sector	Subsector
DBK	Deutsche Bank	FIRE	Banks	Credit Banks
DGS	Degussa Huls	Basic Materials	Chemicals	Chemicals, Specialty
DPB	Deutsche Postbank	FIRE	Banks	Credit Banks
DPW	Deutsche Post	Industrials	Transportation and Logistics	Logistics
DRB	Dresdner Bank	FIRE	Banks	Credit Banks
DTE	Deutsche Telekom	Telecomms.	Telecomms.	Fixed-Line Telecomms.
EOAN	E.On	Utilities	Utilities	Multi-Utilities
EPC	Epcos	Information Technology	Technology	Electronic Components and Hardware
FME	Fresenius Medical Care	Pharma and Healthcare	Pharma and Healthcare	Healthcare
FRE	Fresenius Care	Pharma and Healthcare	Pharma and Healthcare	Healthcare
HEI	Heidelberg Cement	Industrials	Construction	Building Materials
HEN3	Henkel	Consumer Goods	Consumer	Personal Products

Continued on next page

Table 9 – *Continued from previous page*

Ticker	Company	Supersector	Sector	Subsector
HNR1	Hannover Re	FIRE	Insurance	Re-Insurance
HRE	Hypo Real Estate	FIRE	Financial Services	Real Estate
HVB	HypoVereinsbank	FIRE	Banks	Credit Banks
IFX	Infineon Technologies	Information Technology	Technology	Semiconductors
KAR	KarstadtQuelle	Consumer Services	Retail	Retail, Multiline
LHA	Deutsche Lufthansa	Industrials	Transportation and Logistics	Airlines
LIN	Linde	Basic Materials	Chemicals	Industrial Gases
LXS	Lanxess	Basic Materials	Chemicals	Chemicals, Commodity
MAN	MAN	Industrials	Industrial	Industrial, Diversified
MEO	Metro	Consumer Services	Retail	Retail, Multiline
MLP	MLP	FIRE	Financial Services	Diversified Financial
MRK	Merck	Pharma and Healthcare	Pharma and Healthcare	Pharmaceuticals
MUV2	Munich Re	FIRE	Insurance	Re-Insurance

Continued on next page

Table 9 – *Continued from previous page*

Ticker	**Company**	**Supersector**	**Sector**	**Subsector**
RWE	RWE	Utilities	Utilities	Multi-Utilities
SAP	SAP	Information Technology	Software	Software
SCG	Schering	Pharma and Healthcare	Pharma and Healthcare	Pharmaceuticals
SDF	K + S	Basic Materials	Chemicals	Chemicals, Commodity
SIE	Siemens	Industrials	Industrial	Industrial, Diversified
SZG	Salzgitter AG	Basic Materials	Basic Resources	Steel and Other Metals
TKA	ThyssenKrupp	Industrials	Industrial	Industrial, Diversified
TUI	TUI	Industrials	Transportation and Logistics	Transportation Services
VOW3	Volkswagen Group	Consumer Goods	Automobile	Automobile Manufacturers

Appendix B

From Eqn. (6), Section 3.2, we have Y_i = ln price difference of a stock i. We normalise this value as follows:

$$\tilde{Y}_i \equiv \frac{Y_i - \langle Y_i \rangle}{\sqrt{\langle Y_i^2 \rangle - \langle Y_i \rangle^2}}. \tag{11}$$

For a time period of n days, we have an n-length time series of the daily price differences. Consider this time series as an n-dimensional vector, $\widetilde{\mathbf{Y_i}}$, consisting of components Y_{ik}. The Euclidean distance between two vectors $\widetilde{\mathbf{Y_i}}$ and $\widetilde{\mathbf{Y_j}}$ can be calculated as:

$$\begin{aligned} d_{ij}^2 = ||\widetilde{\mathbf{Y_i}} - \widetilde{\mathbf{Y_j}}||^2 &= \sum_{k=1}^{n} \left(\widetilde{Y_{ik}} - \widetilde{Y_{ik}} \right)^2 \\ &= \sum_{k=1}^{n} \widetilde{Y_{ik}}^2 + \sum_{k=1}^{n} \widetilde{Y_{jk}}^2 - 2 \sum_{k-1}^{n} \widetilde{Y_{ik}} \widetilde{Y_{jk}} \\ &= n + n - 2 \sum_{k-1}^{n} \widetilde{Y_{ik}} \widetilde{Y_{jk}} \\ &= 2 \left(n - \sum_{k-1}^{n} \widetilde{Y_{ik}} \widetilde{Y_{jk}} \right). \end{aligned} \tag{12}$$

We now consider the last expression in this equation and show that this summation is equivalent to $n\rho_{ij}$:

$$\begin{aligned} \sum_{k-1}^{n} \widetilde{Y_{ik}} \widetilde{Y_{jk}} &= \sum_{k-1}^{n} \left(\frac{Y_{ik} - \langle Y_i \rangle}{\sqrt{\langle Y_i^2 \rangle - \langle Y_i \rangle^2}} \right) \left(\frac{Y_{jk} - \langle Y_j \rangle}{\sqrt{\langle Y_j^2 \rangle - \langle Y_j \rangle^2}} \right) \\ &= \sum_{k-1}^{n} \frac{(Y_{ik} Y_{jk} - Y_{ik} \langle Y_j \rangle - Y_{jk} \langle Y_i \rangle + \langle Y_i \rangle \langle Y_j \rangle)}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}. \end{aligned} \tag{13}$$

We consider the numerator (N) and denominator (D) of equation (13) separately as follows:

$$
\begin{aligned}
N &= \sum_{k-1}^{n} Y_{ik}Y_{jk} - \sum_{k-1}^{n} Y_{ik}\langle Y_j\rangle - \sum_{k-1}^{n} Y_{jk}\langle Y_i\rangle + n\langle Y_i\rangle\langle Y_j\rangle \\
&= \sum_{k-1}^{n} Y_{ik}Y_{jk} - \sum_{k-1}^{n} Y_{ik}\sum_{k-1}^{n}\frac{Y_{jk}}{n} - \sum_{k-1}^{n} Y_{jk}\sum_{k-1}^{n}\frac{Y_{ik}}{n} + n\langle Y_i\rangle\langle Y_j\rangle \\
&= n\sum_{k-1}^{n}\frac{Y_{ik}Y_{jk}}{n} - \left(n\sum_{k-1}^{n}\frac{Y_{ik}}{n}\right)\sum_{k-1}^{n}\frac{Y_{jk}}{n} - \left(n\sum_{k-1}^{n}\frac{Y_{jk}}{n}\right)\sum_{k-1}^{n}\frac{Y_{ik}}{n} + n\langle Y_i\rangle\langle Y_j\rangle \\
&= n\langle Y_iY_j\rangle - n\langle Y_i\rangle\langle Y_j\rangle - n\langle Y_j\rangle\langle Y_i\rangle + n\langle Y_i\rangle\langle Y_j\rangle \\
&= n(\langle Y_iY_j\rangle - \langle S_i\rangle\langle S_j\rangle) \\
D &= \sqrt{\left(\langle Y_i^2\rangle - \langle Y_i\rangle^2\right)\left(\langle Y_j^2\rangle - \langle Y_j\rangle^2\right)} \\
&= \sqrt{\left(\frac{\sum_{k-1}^{n} Y_{ik}^2}{n} - \langle Y_i\rangle^2\right)\left(\frac{\sum_{k-1}^{n} Y_{jk}^2}{n} - \langle Y_j\rangle^2\right)} \\
&= \sqrt{\left(\sum_{k-1}^{n}\frac{Y_i^2}{n} - \frac{n\langle Y_i\rangle^2}{n}\right)\left(\sum_{k-1}^{n}\frac{Y_j^2}{n} - \frac{n\langle Y_j\rangle^2}{n}\right)} \\
&= \sqrt{\sum_{k-1}^{n}\left(\frac{Y_i^2 - \langle Y_i\rangle^2}{n}\right)\sum_{k-1}^{n}\left(\frac{Y_j^2 - \langle Y_j\rangle^2}{n}\right)} \\
&= \sqrt{\langle Y_i^2 - \langle Y_i\rangle^2\rangle\langle Y_j^2 - \langle Y_j\rangle^2\rangle} \\
\therefore \frac{N}{D} &= \frac{n(\langle Y_iY_j\rangle - \langle Y_i\rangle\langle Y_j\rangle)}{\sqrt{\langle Y_i^2 - \langle Y_i\rangle^2\rangle\langle Y_j^2 - \langle Y_j\rangle^2\rangle}} = n\rho_{ij}.
\end{aligned}
$$

So we can conclude that:

$$\sum_{k-1}^{n} \widetilde{Y_{ik}} \widetilde{Y_{jk}} = n\rho_{ij}$$

and substituting this back into Eqn. (12) we can show that:

$$\begin{aligned} d_{ij}^2 &= 2\left(n - \sum_{k-1}^{n} \widetilde{Y_{ik}} \widetilde{Y_{jk}}\right) = 2(n - n\rho_{ij}) \\ &= 2n(1-\rho_{ij}) \\ \therefore d_{ij} &= \sqrt{2n(1-\rho_{ij})}. \end{aligned}$$

As we are interested in relative distances the absolute values of the distances is not a concern and so for large values we can remove n from the above equation. This leaves us with:

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

Appendix C

The following are the MST figures for the period of crisis from 7th October 2008 - 31st December 2008. The vertices represent the various DAX 30 companies, labelled using their stock ticker symbol (please see Appendix A). The edge length is determined by the corr-distance so that shorter edges correspond to higher positive correlations and the edges highlighted in orange are those identified as insignificant by the Bonferroni correction. Following these figures are the tables to show the correlations of the AGs, again for the period of crisis, listed by the order of their addition and the vertices that the edge connects and then the graphically representations of the AGs. Note that the AGs here show the correlations and not the distances so that they can be compared with the correlations in the 4-clique analysis. The final figures are the PMFGs for this time period, again the orange edges highlight those identified as insignificant by the Bonferroni correction.

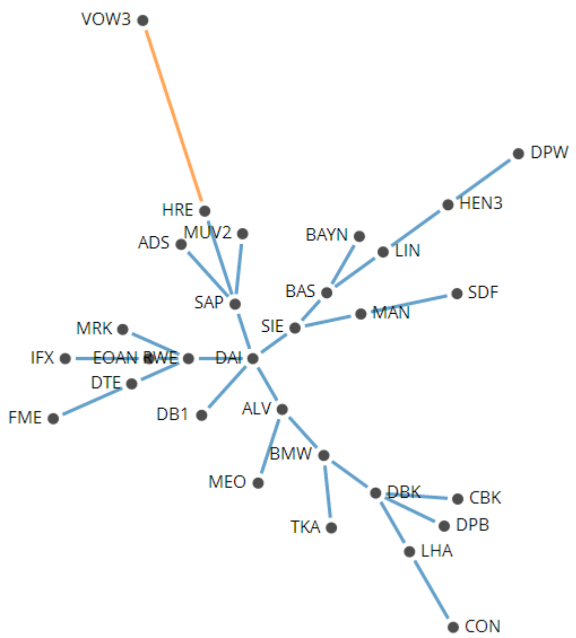

Figure 21: The minimum spanning tree for 7th October - 6th November 2008.

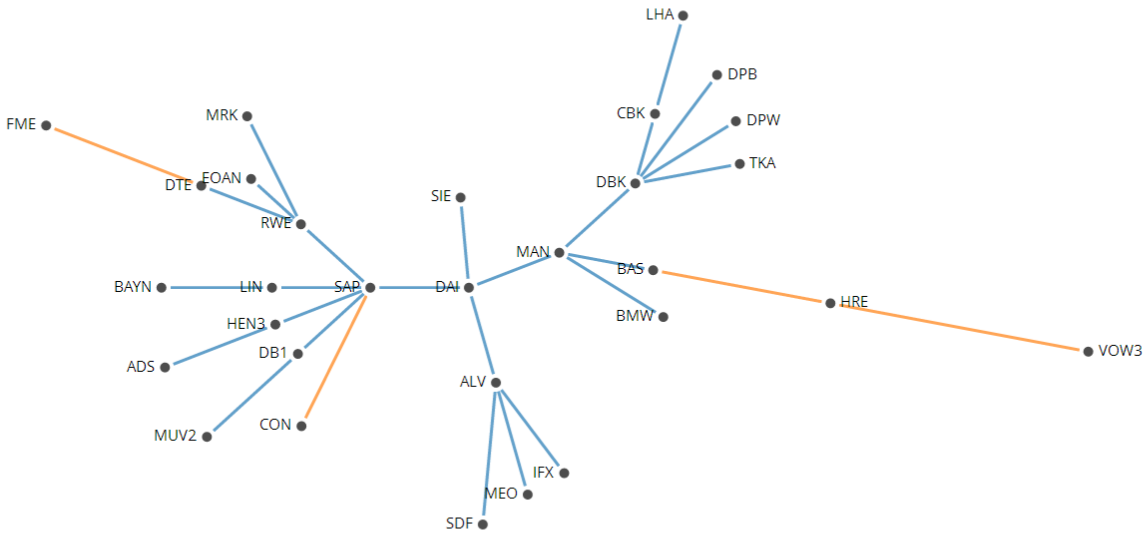

Figure 22: The minimum spanning tree for 21st October - 20th November 2008.

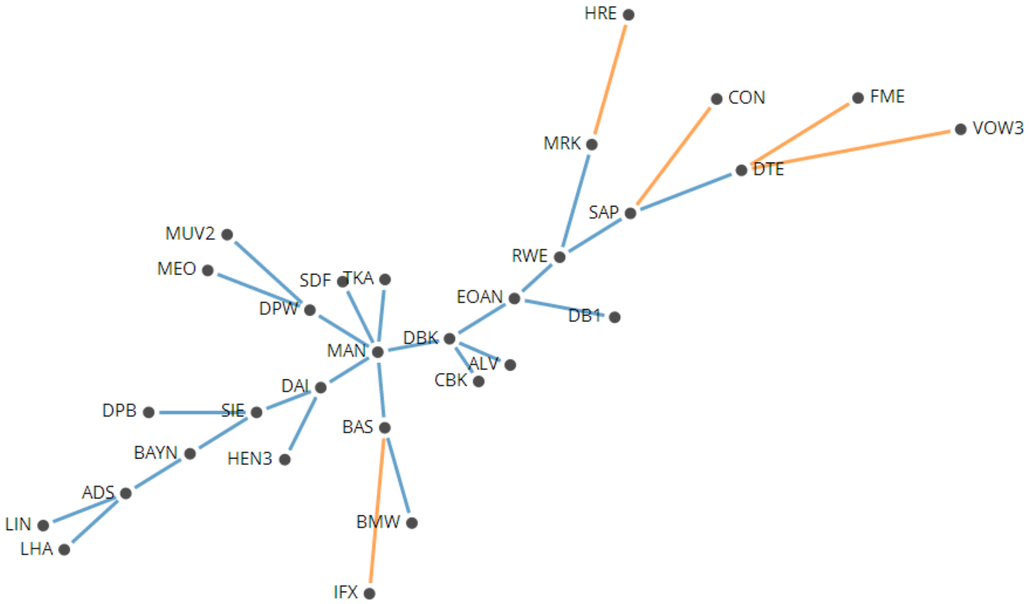

Figure 23: The minimum spanning tree for 4th November - 4th December 2008.

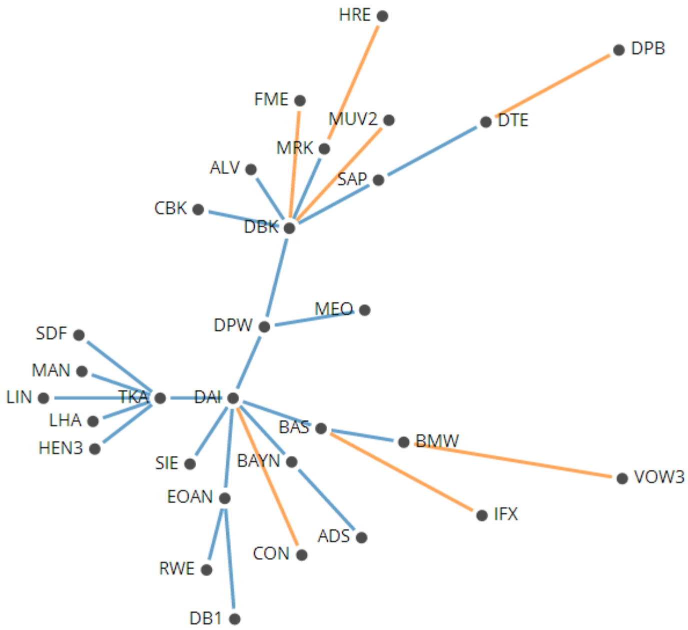

Figure 24: The minimum spanning tree for 18th November - 18th December 2008.

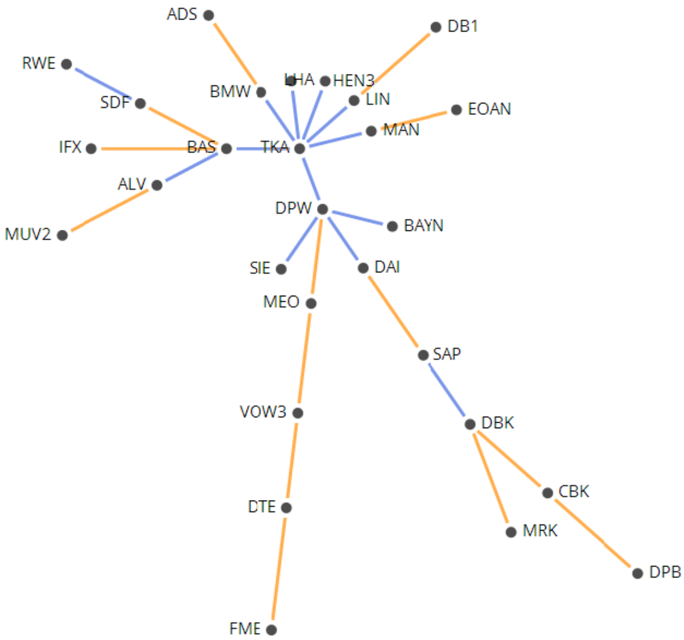

Figure 25: The minimum spanning tree for 2nd December - 31st December 2008.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.9607	RWE	EOAN	16	0.8848	RWE	BAS
2	0.9426	SIE	BAS	17	0.8847	SIE	RWE
3	0.9314	SIE	DAI	18	0.8838	LHA	DBK
4	0.9167	SAP	DAI	19	0.8777	LIN	BAS
5	0.9129	DAI	ALV	20	0.8742	EOAN	DTE
6	0.9023	RWE	DTE	21	0.8733	SAP	RWE
7	0.9023	BMW	ALV	22	0.8729	SAP	MUV2
8	0.8960	DBK	BMW	23	0.8689	RWE	MRK
9	0.8952	SIE	ALV	24	0.8653	SIE	EOAN
10	0.8949	RWE	DAI	25	0.8653	TKA	BMW
11	0.8926	BAYN	BAS	26	0.8644	TKA	DBK
12	0.8893	EOAN	DAI	27	0.8644	TKA	MAN
13	0.8861	SIE	SAP	28	0.8637	SIE	LIN
14	0.8851	DAI	BMW	29	0.8631	DAI	BAS
15	0.8848	SIE	MAN				

Table 10: The edges that form the asset graph for 7th October - 6th November 2008.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.9358	RWE	EOAN	16	0.8521	SIE	BAS
2	0.9262	DBK	CBK	17	0.8505	MAN	DBK
3	0.8841	SIE	DAI	18	0.8498	EOAN	ALV
4	0.8748	SAP	RWE	19	0.8498	RWE	LIN
5	0.8714	MAN	BAS	20	0.8487	LHA	DBK
6	0.8670	MAN	DAI	21	0.8431	LIN	HEN3
7	0.8638	SAP	DB1	22	0.8400	TKA	DBK
8	0.8632	SIE	MAN	23	0.8393	RWE	DAI
9	0.8629	SAP	DAI	24	0.8387	LIN	BAS
10	0.8629	SAP	LIN	25	0.8385	RWE	DTE
11	0.8624	DAI	ALV	26	0.8382	SAP	EOAN
12	0.8598	LIN	EOAN	27	0.8379	SAP	ALV
13	0.8596	EOAN	DAI	28	0.8360	RWE	ALV
14	0.8534	SAP	HEN3	29	0.8289	EOAN	DTE
15	0.8529	LHA	CBK				

Table 11: The edges that form the asset graph for 21st October - 20th November 2008.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.9458	DBK	CBK	16	0.8903	SIE	BAYN
2	0.9326	RWE	EOAN	17	0.8894	EOAN	ALV
3	0.9230	DBK	ALV	18	0.8892	SDF	MAN
4	0.9194	MAN	DAI	19	0.8861	MAN	DPW
5	0.9154	SIE	DAI	20	0.8847	MAN	CBK
6	0.9077	TKA	MAN	21	0.8847	DAI	BAYN
7	0.9046	MAN	DBK	22	0.8843	HEN3	DAI
8	0.9026	SIE	MAN	23	0.8816	DAI	ADS
9	0.8981	BAYN	ADS	24	0.8766	TKA	CBK
10	0.8977	TKA	DAI	25	0.8766	SAP	RWE
11	0.8972	MAN	BAS	26	0.8758	LHA	ADS
12	0.8961	EOAN	DBK	27	0.8744	DAI	BAS
13	0.8949	DBK	DAI	28	0.8741	EOAN	DAI
14	0.8939	CBK	ALV	29	0.8734	SIE	ADS
15	0.8903	TKA	SIE				

Table 12: The edges that form the asset graph for 4th November - 4th December 2008.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.9099	DBK	ALV	16	0.8536	HEN3	DAI
2	0.9078	TKA	LHA	17	0.8535	LHA	DAI
3	0.9032	TKA	DAI	18	0.8511	MAN	DPW
4	0.9003	RWE	EOAN	19	0.8494	SIE	HEN3
5	0.8891	DPW	DAI	20	0.8475	SIE	LHA
6	0.8880	TKA	DPW	21	0.8475	SIE	DPW
7	0.8874	SIE	DAI	22	0.8452	SIE	BAYN
8	0.8752	TKA	HEN3	23	0.8437	SIE	MAN
9	0.8750	TKA	MAN	24	0.8422	DAI	BAS
10	0.8746	DBK	CBK	25	0.8411	TKA	SIE
11	0.8734	MAN	DAI	26	0.8398	LHA	DPW
12	0.8714	BMW	BAS	27	0.8162	EOAN	DAI
13	0.8635	DAI	BAYN	28	0.8146	MAN	LHA
14	0.8625	MRK	DBK	29	0.8131	CBK	ALV
15	0.8567	HEN3	DPW				

Table 13: The edges that form the asset graph for 18th November - 18th December 2008.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.8409	TKA	DPW	15	0.7654	BAS	ALV
2	0.8230	TKA	HEN3	16	0.7596	TKA	SIE
3	0.8207	TKA	BMW	17	0.7596	BMW	BAS
4	0.8053	DPW	DAI	18	0.7576	SIE	HEN3
5	0.8039	DPW	BAYN	19	0.7555	LIN	DPW
6	0.8030	DAI	BMW	20	0.7507	LHA	DPW
7	0.8015	TKA	LHA	21	0.7351	HEN3	DAI
8	0.7991	TKA	LIN	22	0.7339	MAN	DAI
9	0.7981	SIE	DPW	23	0.7311	SAP	DBK
10	0.7933	TKA	MAN	24	0.7275	SDF	RWE
11	0.7909	TKA	BAS	25	0.7259	DAI	BAS
12	0.7890	SIE	DAI	26	0.7258	TKA	BAYN
13	0.7868	TKA	DAI	27	0.7183	LHA	ALV
14	0.7801	DAI	BAYN				

Table 14: The edges that form the asset graph for 2nd December - 31st December 2008.

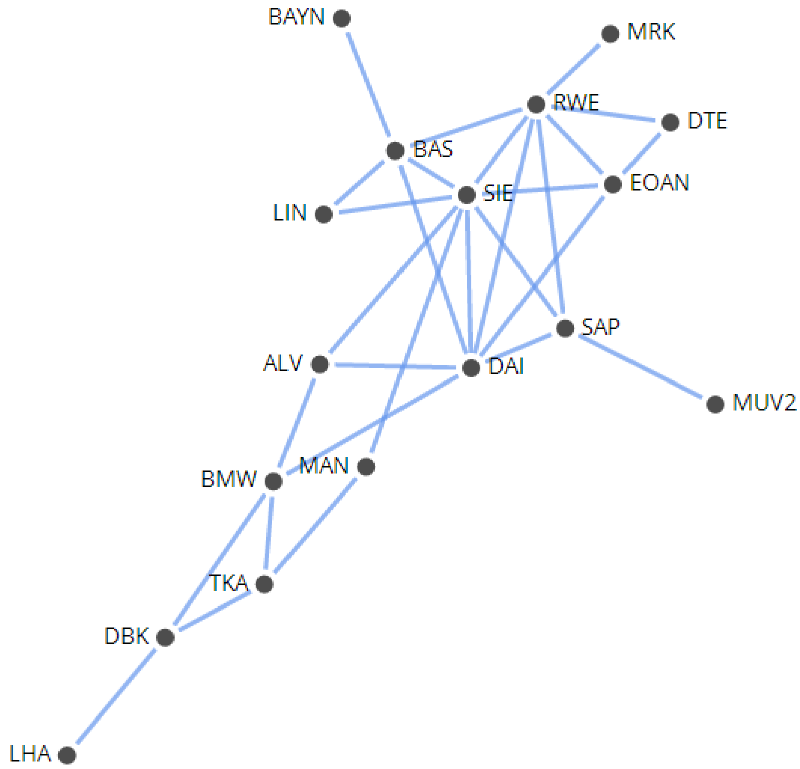

Figure 26: The asset graph for 7th October - 6th November 2008.

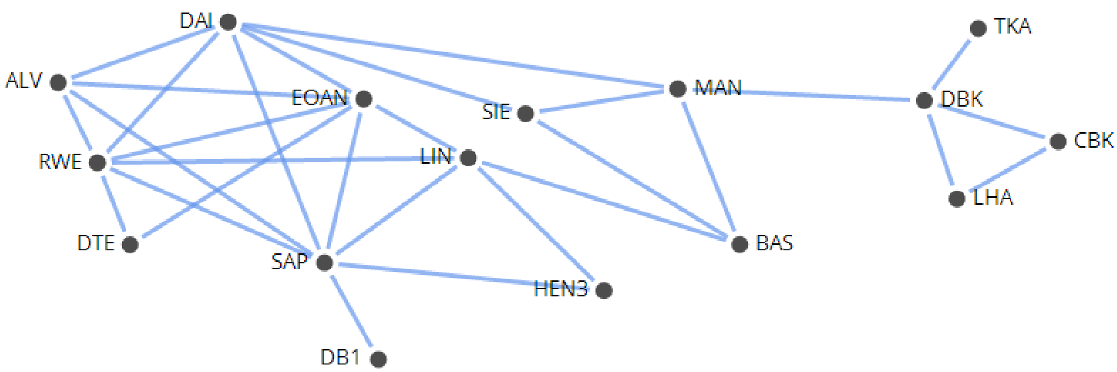

Figure 27: The asset graph for 21st October - 20th November 2008.

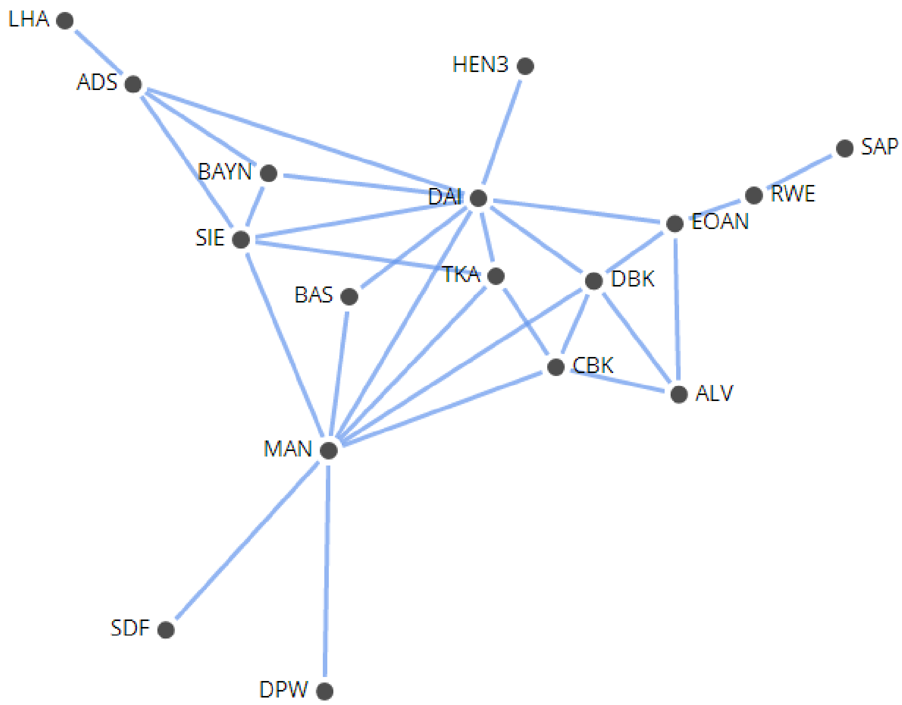

Figure 28: The asset graph for 4th November - 4th December 2008.

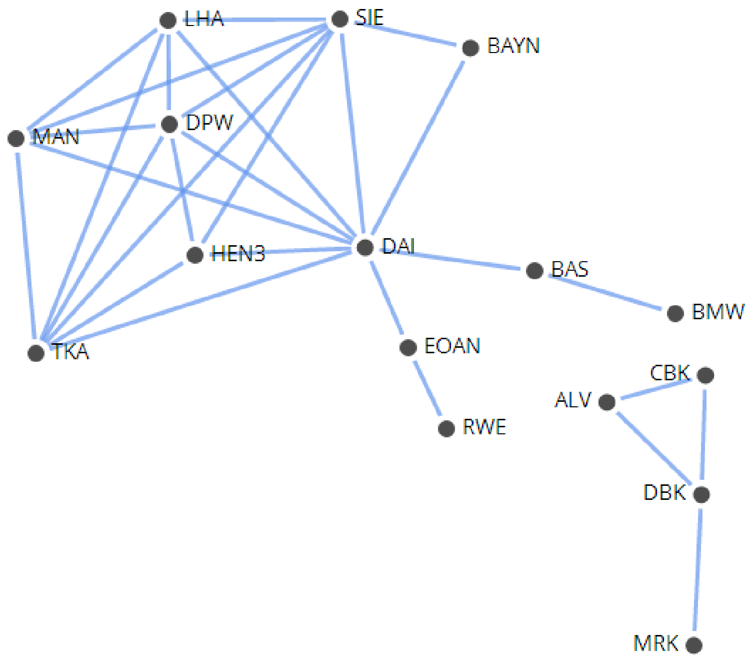

Figure 29: The asset graph for 18th November - 18th December 2008.

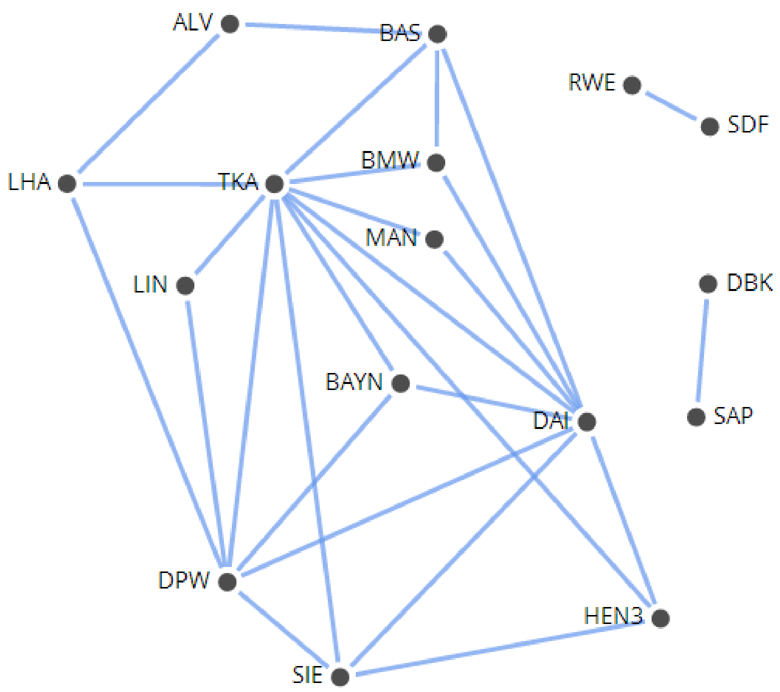

Figure 30: The asset graph for 2nd December - 31st December 2008.

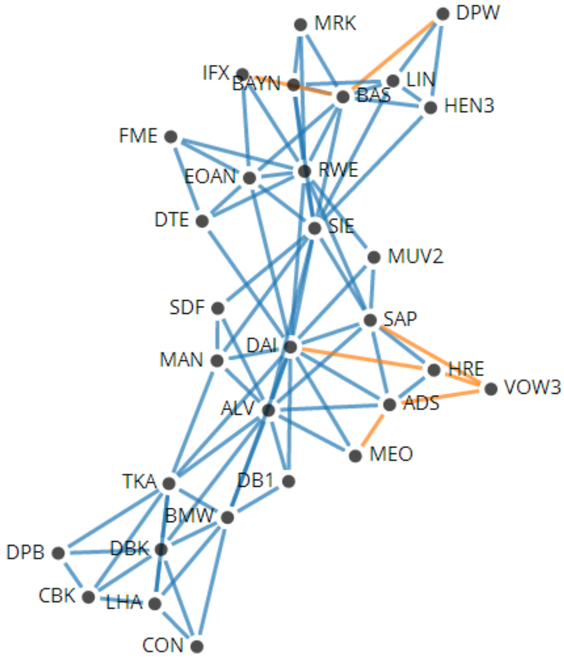

Figure 31: The planar maximally filtered graph for 7th October - 6th November 2008.

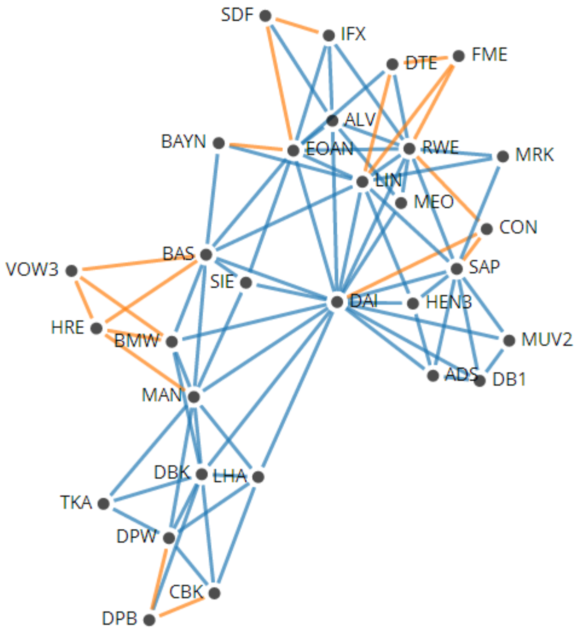

Figure 32: The planar maximally filtered graph for 21st October - 20th November 2008.

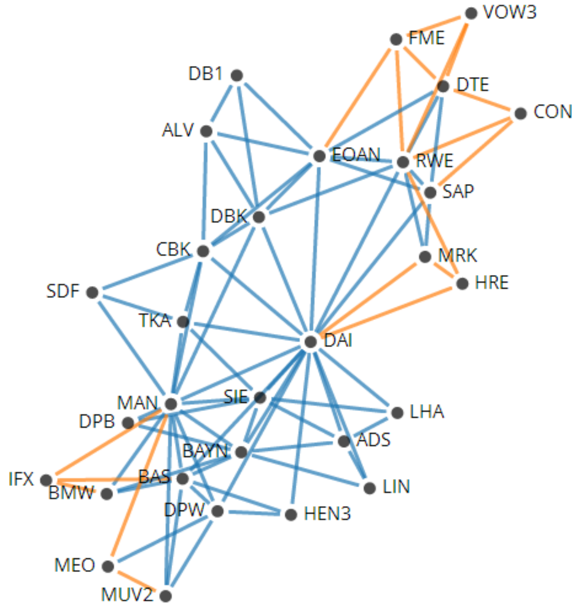

Figure 33: The planar maximally filtered graph for 4th November - 4th December 2008.

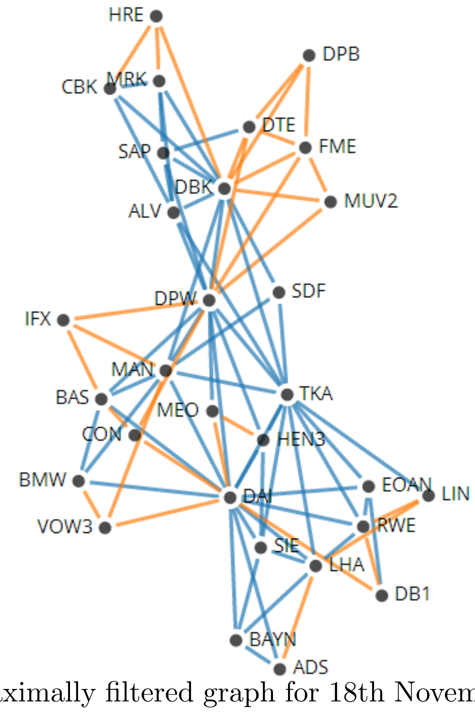

Figure 34: The planar maximally filtered graph for 18th November - 18th December 2008.

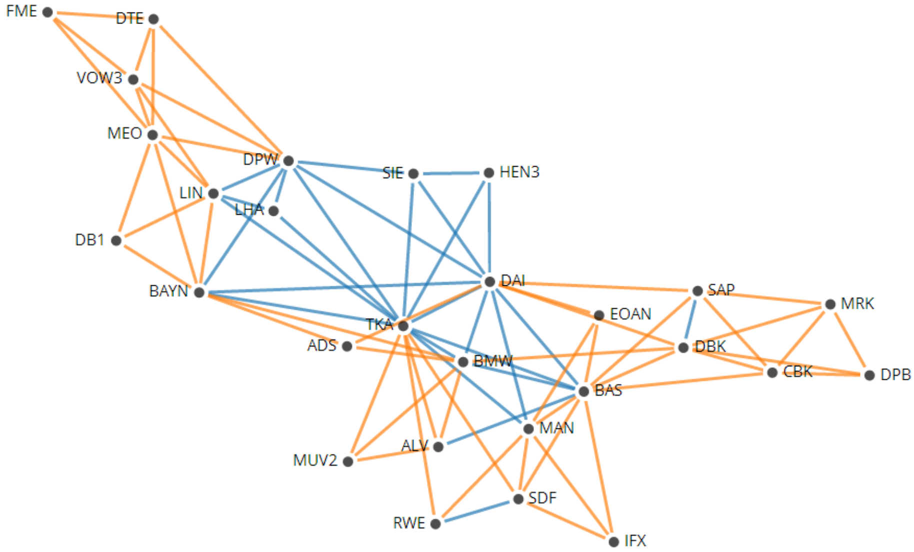

Figure 35: The planar maximally filtered graph for 2nd December - 31st December 2008.

Appendix D

The following are the MST figures for the period of recovery from 7th May 2010 - 3rd August 2010. The vertices represent the various DAX 30 companies, labelled using their stock ticker symbol (please see Appendix A). The edge length is determined by the corr-distance so that shorter edges correspond to higher positive correlations and the edges highlighted in orange are those identified as insignificant by the Bonferroni correction. Following these figures are the tables to show the correlations of the AGs, again for the period of recovery, listed by the order of their addition and the vertices that the edge connects and then the graphically representations of the AGs. Note that the AGs here show the correlations and not the distances so that they can be compared with the correlations in the 4-clique analysis. The final figures are the PMFGs for this time period, again the orange edges highlight those identified as insignificant by the Bonferroni correction.

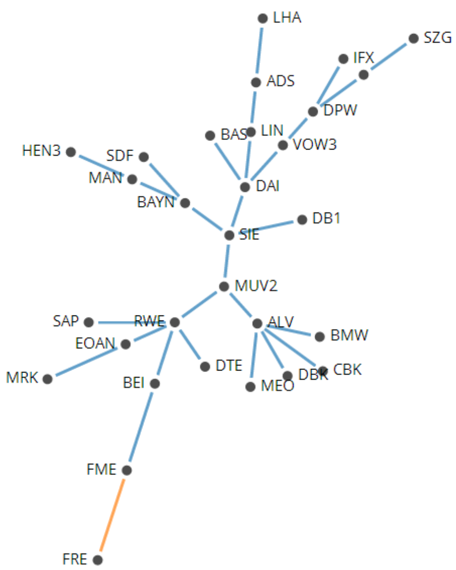

Figure 36: The minimum spanning tree for 7th May - 8th June 2010.

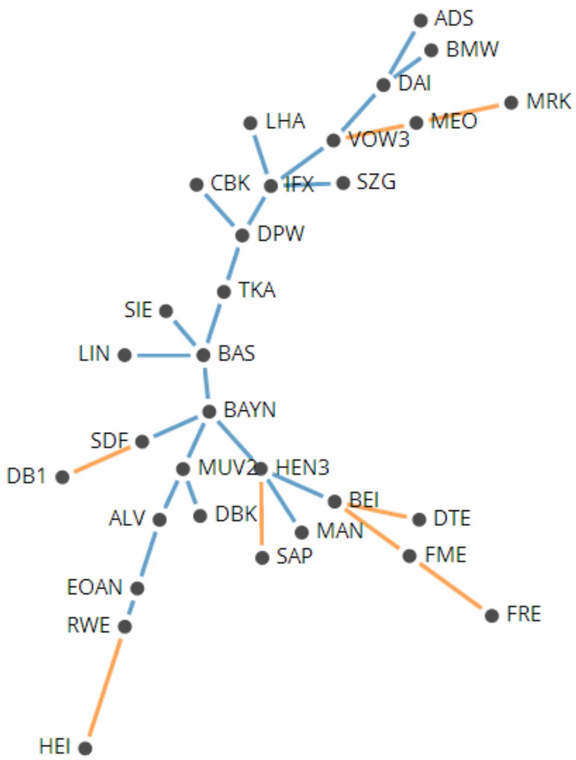

Figure 37: The minimum spanning tree for 21st May - 22nd June 2010.

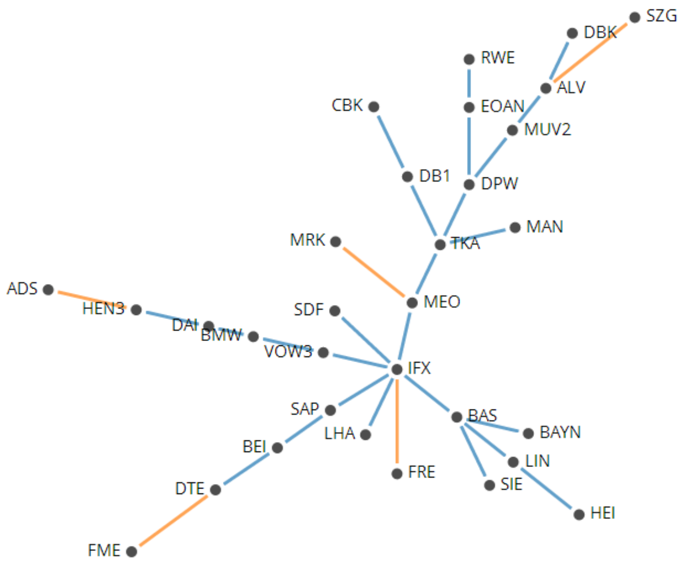

Figure 38: The minimum spanning tree for 4th June - 6th July 2010.

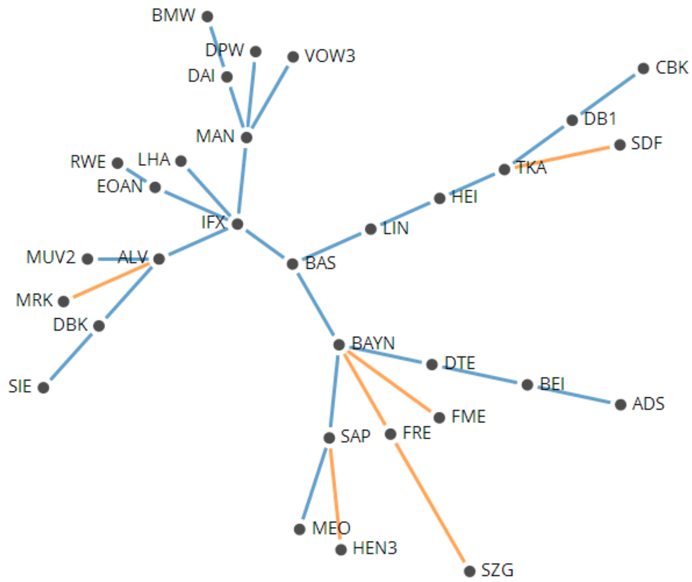

Figure 39: The minimum spanning tree for 18th June - 20th July 2010.

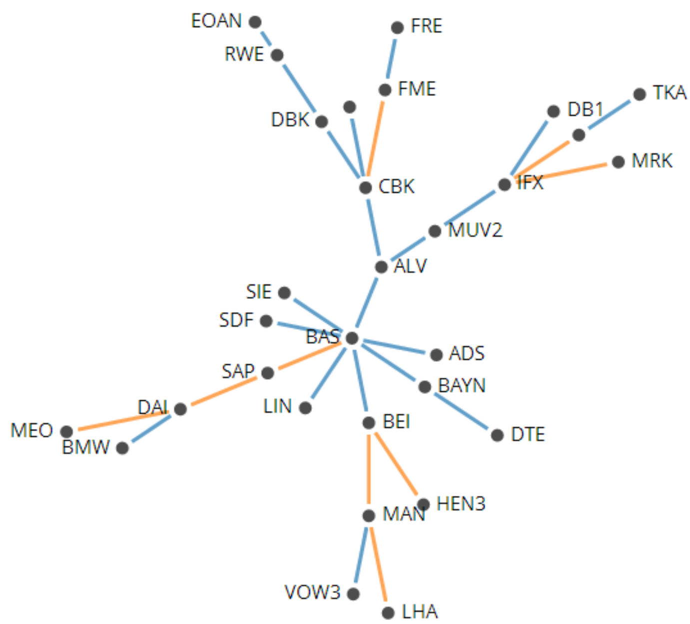

Figure 40: The minimum spanning tree for 2nd July - 3rd August 2010.

Edges	Correlation	Vertex	Vertex	Edges	Correlation	Vertex	Vertex
1	0.9280	VOW3	DPW	16	0.8702	VOW3	SIE
2	0.9119	MUV2	ALV	17	0.8701	SIE	MAN
3	0.9119	LIN	ADS	18	0.8697	DBK	ALV
4	0.9092	SIE	DAI	19	0.8681	RWE	MUV2
5	0.9055	SIE	MUV2	20	0.8681	LIN	DPW
6	0.8967	RWE	EOAN	21	0.8678	IFX	DPW
7	0.8929	SIE	BAYN	22	0.8674	VOW3	IFX
8	0.8910	RWE	DTE	23	0.8671	DPW	ADS
9	0.8905	LIN	DAI	24	0.8657	BAYN	ADS
10	0.8875	DAI	BAS	25	0.8651	VOW3	MAN
11	0.8860	VOW3	DAI	26	0.8638	SDF	BAYN
12	0.8841	LIN	BAS	27	0.8637	TKA	SZG
13	0.8801	MAN	BAYN	28	0.8631	SIE	DPW
14	0.8783	MUV2	BAYN	29	0.8622	SIE	ADS
15	0.8707	SIE	ALV				

Table 15: The edges that form the asset graph for 7th May - 8th June 2010.

Edges	Correlation	Vertex	Vertex	Edges	Correlation	Vertex	Vertex
1	0.8833	MUV2	ALV	15	0.8027	TKA	SIE
2	0.8667	SIE	BAS	16	0.7994	TKA	IFX
3	0.8641	BAYN	BAS	17	0.7985	MAN	HEN3
4	0.8634	IFX	DPW	18	0.7920	DBK	BAS
5	0.8539	DAI	BMW	19	0.7852	LHA	DPW
6	0.8520	TKA	DPW	20	0.7814	EOAN	ALV
7	0.8450	MUV2	DBK	21	0.7807	LHA	DBK
8	0.8274	DBK	ALV	22	0.7804	IFX	BAS
9	0.8165	LHA	IFX	23	0.7801	BAS	ALV
10	0.8147	TKA	BAS	24	0.7755	LHA	BAYN
11	0.8126	MUV2	BAYN	25	0.7728	SDF	BAYN
12	0.8120	MUV2	BAS	26	0.7676	DAI	ADS
13	0.8047	DPW	CBK	27	0.7662	VOW3	DAI
14	0.8027	IFX	CBK	28	0.7595	DPW	BAS

Table 16: The edges that form the asset graph for 21st May - 22nd June 2010.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.9099	RWE	EOAN	15	0.7843	HEN3	DAI
2	0.8878	MUV2	ALV	16	0.7807	SAP	IFX
3	0.8537	DBK	ALV	17	0.7799	LHA	DBK
4	0.8371	TKA	MEO	18	0.7771	TKA	DB1
5	0.8258	TKA	DPW	19	0.7768	SDF	IFX
6	0.8176	MEO	IFX	20	0.7766	SIE	BAS
7	0.8107	MUV2	DPW	21	0.7757	VOW3	IFX
8	0.8084	DTE	BEI	22	0.7756	IFX	DPW
9	0.7988	VOW3	BMW	23	0.7704	IFX	BAS
10	0.7968	LIN	BAS	24	0.7703	SAP	BEI
11	0.7940	LHA	IFX	25	0.7697	DBK	BAYN
12	0.7897	BAYN	BAS	26	0.7684	TKA	MAN
13	0.7890	DPW	ALV	27	0.7668	EOAN	DPW
14	0.7883	IFX	DBK	28	0.7649	DB1	CBK

Table 17: The edges that form the asset graph for 4th June - 6th July 2010.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.8839	DAI	BMW	15	0.7749	MUV2	IFX
2	0.8834	MAN	DAI	16	0.7675	TKA	IFX
3	0.8667	IFX	BAS	17	0.7663	DBK	ALV
4	0.8522	MUV2	ALV	18	0.7657	CBK	ALV
5	0.8012	SIE	DBK	19	0.7656	BAS	ALV
6	0.7976	TKA	DB1	20	0.7652	IFX	EOAN
7	0.7932	LHA	IFX	21	0.7629	RWE	DB1
8	0.7879	LIN	BAS	22	0.7609	IFX	DPW
9	0.7870	IFX	ALV	23	0.7589	DPW	DB1
10	0.7868	DB1	ALV	24	0.7507	IFX	DB1
11	0.7851	MAN	DPW	25	0.7504	VOW3	MAN
12	0.7830	LHA	BAS	26	0.7496	BAYN	BAS
13	0.7820	MAN	IFX	27	0.7490	SIE	ALV
14	0.7766	DB1	CBK	28	0.7472	SAP	BAYN

Table 18: The edges that form the asset graph for 18th June - 20th July 2010.

Edges	**Correlation**	**Vertex**	**Vertex**	**Edges**	**Correlation**	**Vertex**	**Vertex**
1	0.9396	RWE	EOAN	16	0.7225	EOAN	DBK
2	0.8462	FRE	FME	17	0.7156	BAS	ADS
3	0.8417	MUV2	ALV	18	0.7154	BEI	BAS
4	0.8134	DAI	BMW	19	0.7103	RWE	IFX
5	0.7954	TKA	HEI	20	0.7099	DTE	BAYN
6	0.7736	BAS	ALV	21	0.7098	RWE	MUV2
7	0.7591	DBK	CBK	22	0.7087	MUV2	BAS
8	0.7568	VOW3	MAN	23	0.7087	EOAN	ALV
9	0.7553	RWE	DBK	24	0.7074	SDF	BAS
10	0.7504	CBK	ALV	25	0.7057	BAYN	BAS
11	0.7478	RWE	ALV	26	0.7039	IFX	DB1
12	0.7467	SIE	BAS	27	0.7034	IFX	ALV
13	0.7406	DPW	CBK	28	0.6978	SIE	ALV
14	0.7308	LIN	BAS	29	0.6967	EOAN	CBK
15	0.7308	MUV2	IFX				

Table 19: The edges that form the asset graph for 2nd July - 3rd August 2010.

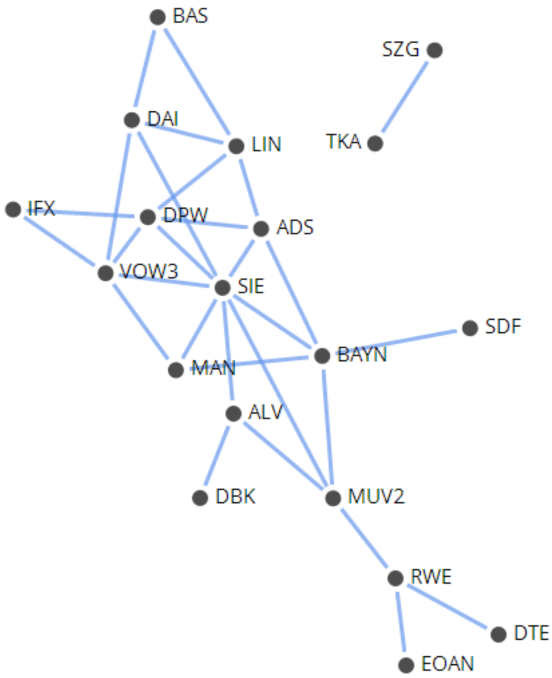

Figure 41: The asset graph for 7th May - 8th June 2010.

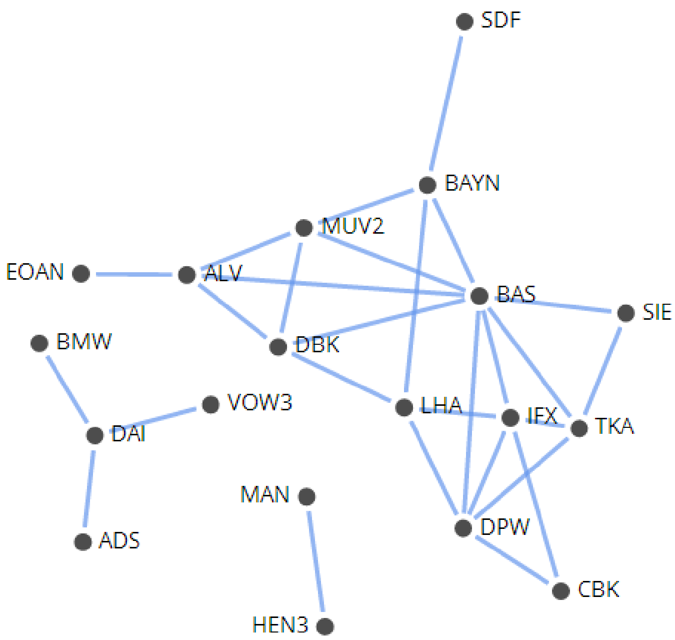

Figure 42: The asset graph for 21st May - 22nd June 2010.

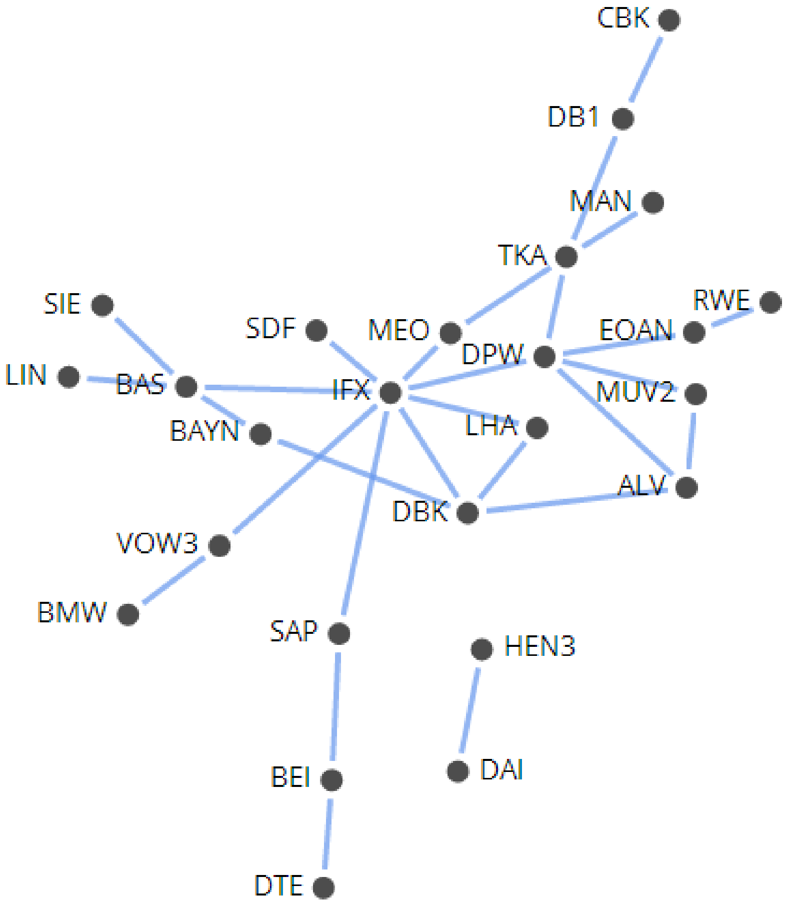

Figure 43: The asset graph for 4th June - 6th July 2010.

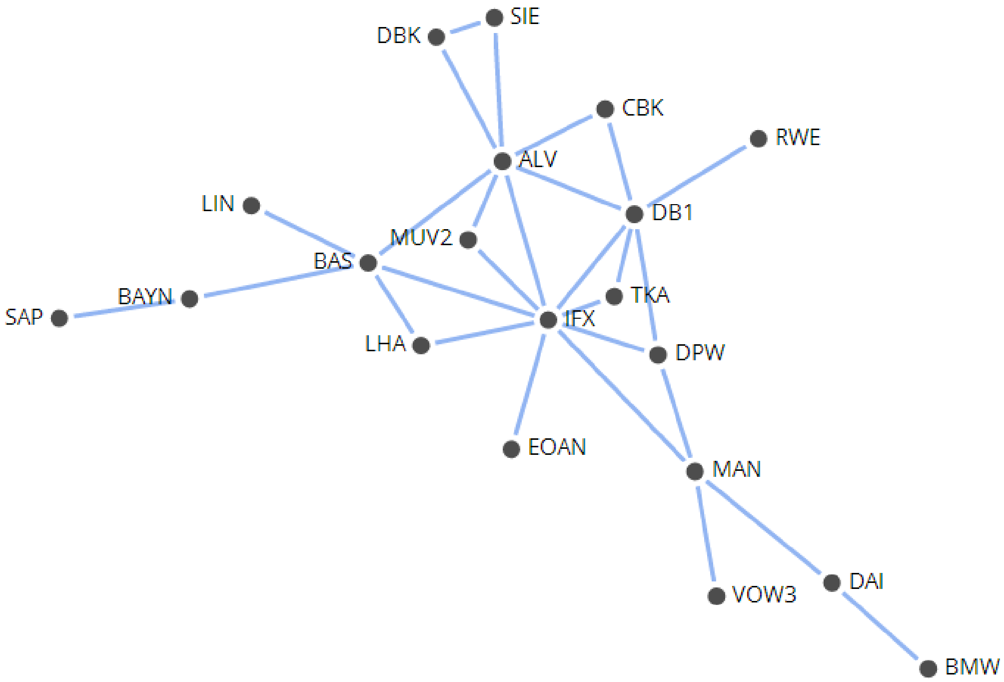

Figure 44: The asset graph for 18th June - 20th July 2010.

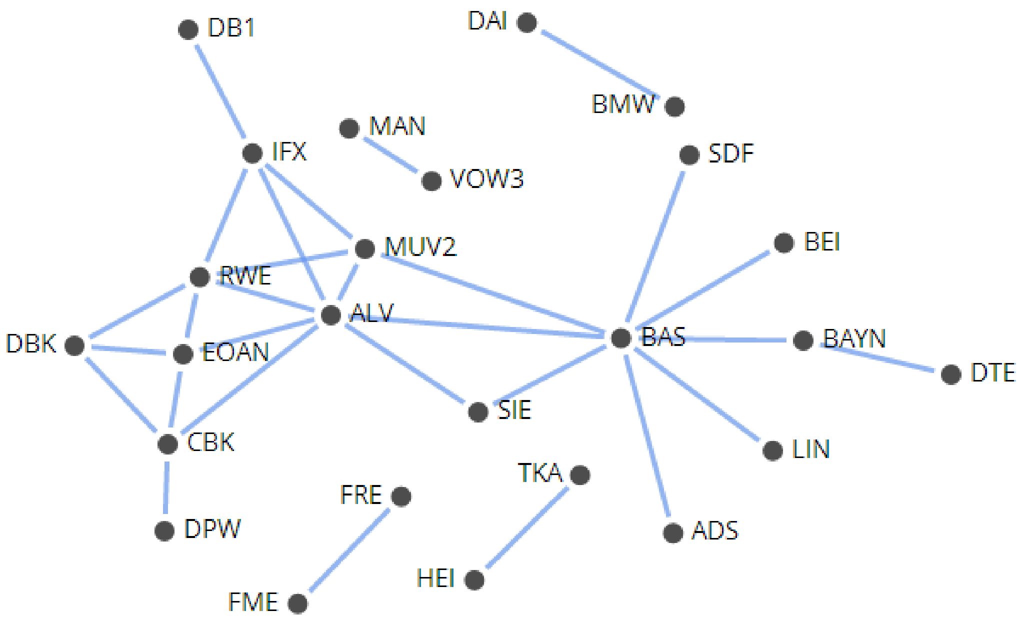

Figure 45: The asset graph for 2nd July - 3rd August 2010.

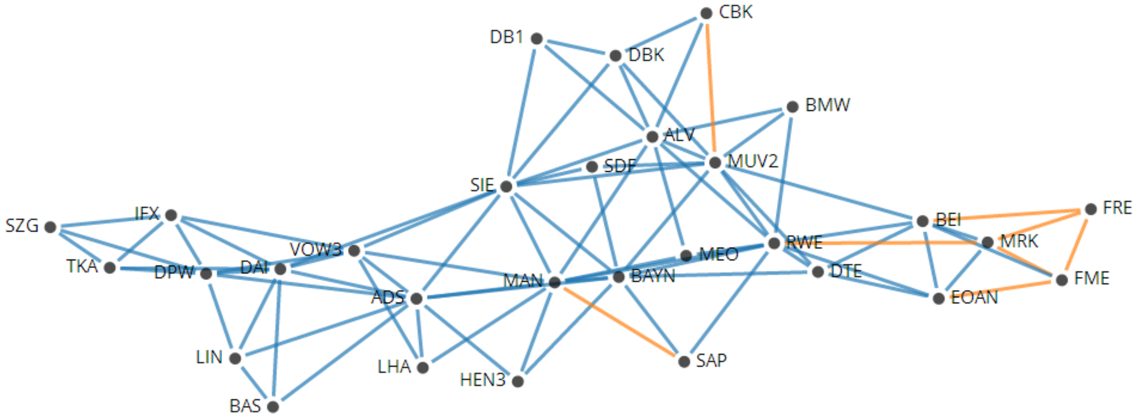

Figure 46: The planar maximally filtered graph for 7th May - 8th June 2010.

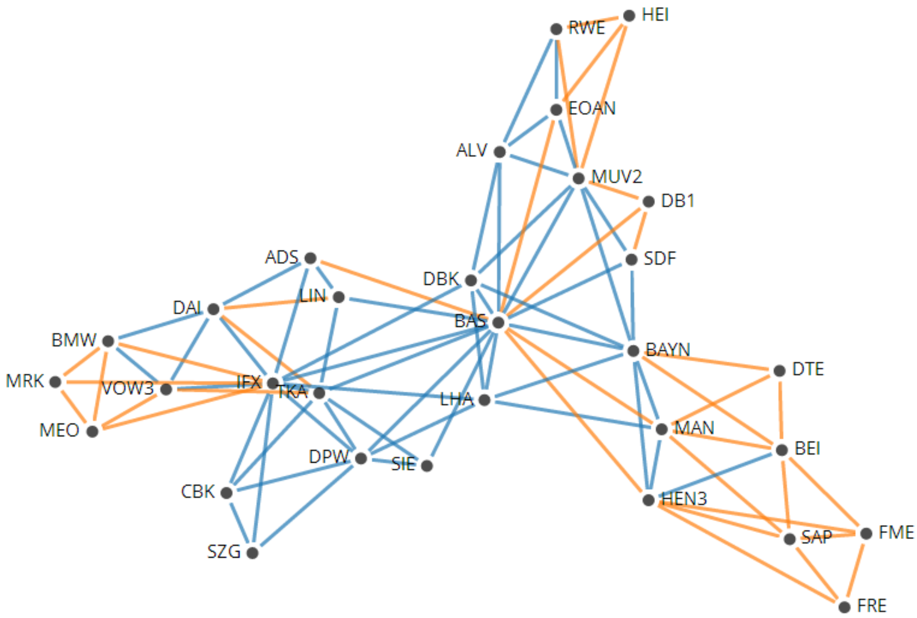

Figure 47: The planar maximally filtered graph for 21st May - 22nd June 2010.

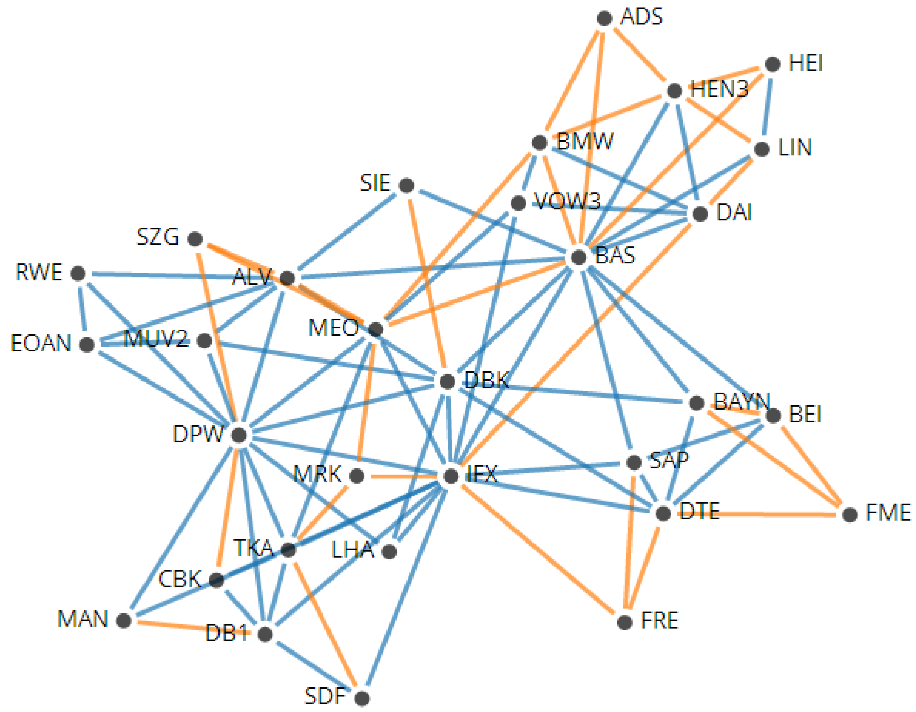

Figure 48: The planar maximally filtered graph for 4th June - 6th July 2010.

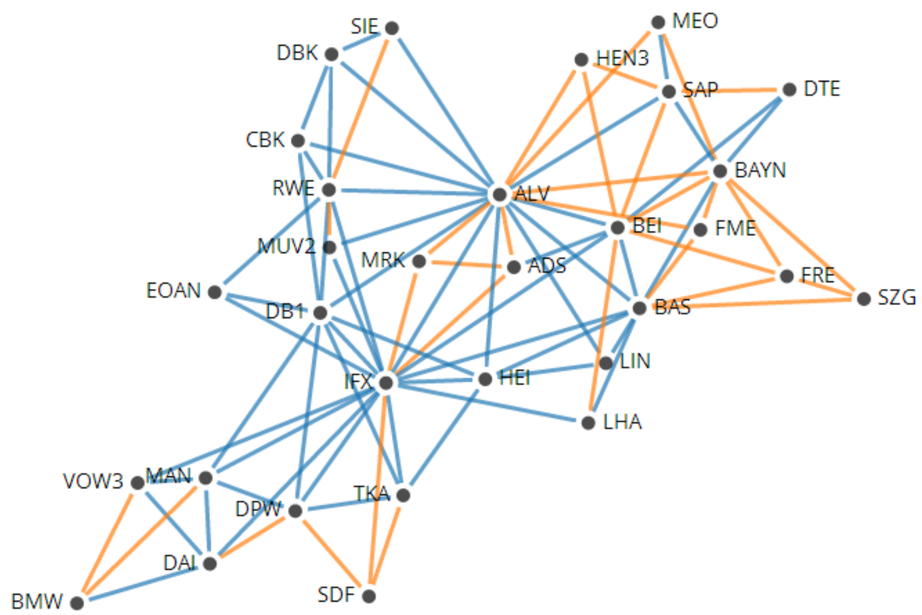

Figure 49: The planar maximally filtered graph for 18th June - 20th July 2010.

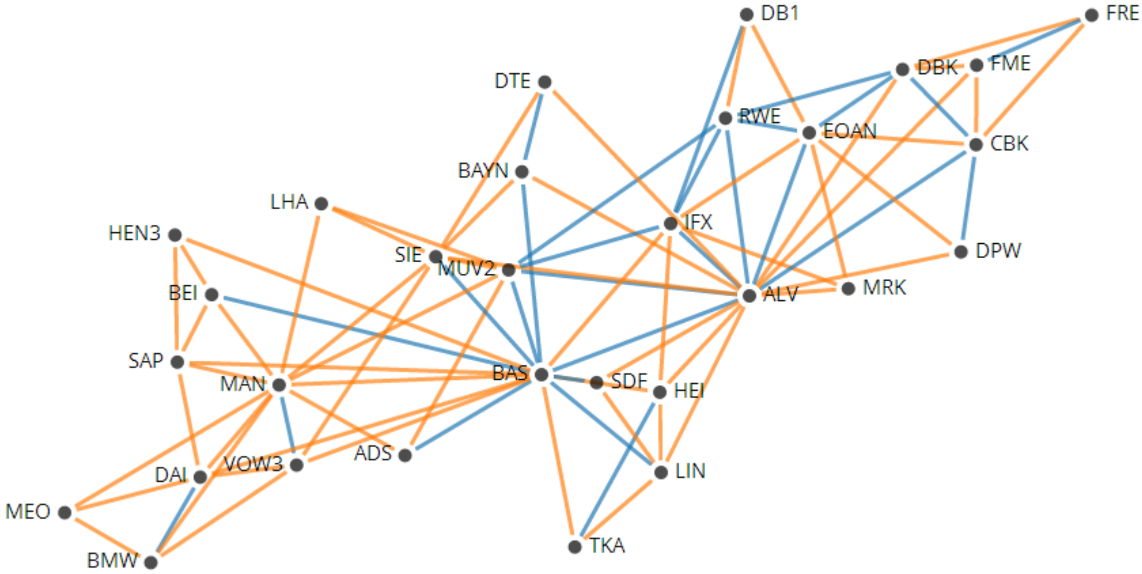

Figure 50: The planar maximally filtered graph for 2nd July - 3rd August 2010.

Appendix E

Handshaking Lemma

Theorem 9 (Handshaking Lemma). *Every finite, undirected graph has an even number of vertices of odd degree i.e.*
Let $G(V, E)$ be a simple, non-directed finite graph. Then:

$$\sum_{v \in V} deg(v) = 2|E|.$$

Proof. Each edge is incident to exactly two vertices.
The degree of each vertex = number of edges to which it is incident.
$\therefore$ when we sum up the degrees of all vertices, we are counting all of the edges of the graph twice. $\square$

Euler's Formula

Theorem 10. *Let G(V,E) be a finite, connected planar graph (drawn in its planar representation) with $|V| = n, |E| = e$ and f faces (including the unbounded face). Then:*

$$n - e + f = 2$$

Proof. Mathematical induction. If $e = 0$ then $n = 1$ and $f = 1$, so it is obvious the formula is as required.
Assume true for all $G(V, E)$ with $|E| < e$, where $e \geq 1$. Suppose $G(V, E)$ has e edges.

There are two options:

- $G(V, E)$ is a tree. Then $n = e + 1$ and $f = 1 \therefore (e + 1) - e + 1 = 2$ so formula is as required.

- $G(V, E)$ is not a tree. Let m be an edge from a cycle $\subset G$ and consider the graph formed by removing edge m, graph $G - m$. The connected, planar graph has n vertices, $e - 1$ edges and $f - 1$ faces $\therefore$ by the induction hypothesis $n - (e - 1) + (f - 1) = 2 \implies n - e + f = 2$.

$\square$

Appendix F

Table 20: Mean degree (Mean Deg.) and Clustering Coefficient (Cluster.) for the HVG created for each stock for four time periods taking data on every 4th day with n=2753 or n=2752 for Set 4 (Set 1 beginning 01/01/1973, Set 2 beginning 02/01/1973, Set 3 beginning 03/01/1973 and Set 4 beginning 04/01/1973) and also the set of 'even' days (n=5505 days) and 'odd' days (n=5506) for the time period 01/01/1973 - 16/03/2015.

Ticker	**Set**	n	**Mean Deg.**	**Cluster.**
ALV	Full Set	11011	3.6767	0.3515
	Set 1	2753	3.4844	0.3195
	Set 2	2753	3.638	0.3497
	Set 3	2753	3.7472	0.3687
	Set 4	2752	3.8296	0.3743
	Even Days	5505	3.7624	0.3639
	Odd Days	5506	3.7588	0.3643
BAS	Full Set	11011	3.7555	0.3578
	Set 1	2753	3.7072	0.3472
	Set 2	2753	3.722	0.3561
	Set 3	2753	3.7468	0.3604
	Set 4	2752	3.768	0.3695
	Even Days	5505	3.8158	0.3714
	Odd Days	5506	3.8111	0.3709
BAYN	Full Set	11011	3.7635	0.3586
	Set 1	2753	3.684	0.3466
	Set 2	2753	3.724	0.3579
	Set 3	2753	3.7964	0.3644
	Set 4	2752	3.8164	0.3693

Continued on next page

Table 20 – *Continued from previous page*

Ticker	Set	n	Mean Deg.	Cluster.
BAYN	Even Days	5505	3.8223	0.3701
	Odd Days	5506	3.8198	0.3698
BEI	Full Set	11011	3.4903	0.3175
	Set 1	2753	3.1948	0.2644
	Set 2	2753	3.4512	0.3088
	Set 3	2753	3.6252	0.3336
	Set 4	2752	3.76	0.3525
	Even Days	5505	3.6338	0.3411
	Odd Days	5506	3.6211	0.3376
BMW	Full Set	11011	3.7125	0.3549
	Set 1	2753	3.604	0.3361
	Set 2	2753	3.6468	0.3518
	Set 3	2753	3.7376	0.3599
	Set 4	2752	3.794	0.3700
	Even Days	5505	3.7791	0.3661
	Odd Days	5506	3.7915	0.3682
CBK	Full Set	11011	3.7317	0.3549
	Set 1	2753	3.6776	0.3447
	Set 2	2753	3.712	0.3499
	Set 3	2753	3.7352	0.3527
	Set 4	2752	3.76	0.3662
	Even Days	5505	3.8027	0.3674

Continued on next page

Table 20 – *Continued from previous page*

Ticker	Set	n	Mean Deg.	Cluster.
CBK	Odd Days	5506	3.814	0.3661
CON	Full Set	11011	3.6967	0.3496
	Set 1	2753	3.6324	0.3359
	Set 2	2753	3.6592	0.3462
	Set 3	2753	3.672	0.3496
	Set 4	2752	3.7188	0.3642
	Even Days	5505	3.7762	0.3604
	Odd Days	5506	3.7715	0.3614
DBK	Full Set	11011	3.7851	0.3644
	Set 1	2753	3.7136	0.3520
	Set 2	2753	3.7564	0.3609
	Set 3	2753	3.8004	0.3686
	Set 4	2752	3.8208	0.3762
	Even Days	5505	3.8325	0.3730
	Odd Days	5506	3.8482	0.3751
EOAN	Full Set	11011	3.7186	0.3565
	Set 1	2753	3.6528	0.3394
	Set 2	2753	3.6876	0.3535
	Set 3	2753	3.734	0.3621
	Set 4	2752	3.7648	0.3710
	Even Days	5505	3.7795	0.3688
	Odd Days	5506	3.773	0.3715

Continued on next page

Table 20 – *Continued from previous page*

Ticker	Set	n	Mean Deg.	Cluster.
HEI	Full Set	11011	3.5288	0.3190
	Set 1	2753	3.2832	0.2729
	Set 2	2753	3.3612	0.2887
	Set 3	2753	3.5684	0.3300
	Set 4	2752	3.7608	0.3582
	Even Days	5505	3.6621	0.3374
	Odd Days	5506	3.6756	0.3383
LHA	Full Set	11011	3.6418	0.3368
	Set 1	2753	3.466	0.3052
	Set 2	2753	3.5688	0.3274
	Set 3	2753	3.6904	0.3454
	Set 4	2752	3.7792	0.3619
	Even Days	5505	3.7475	0.3518
	Odd Days	5506	3.7348	0.3527
LIN	Full Set	11011	3.6676	0.3428
	Set 1	2753	3.5436	0.3228
	Set 2	2753	3.6248	0.3348
	Set 3	2753	3.7004	0.3462
	Set 4	2752	3.7684	0.3610
	Even Days	5505	3.7544	0.3617
	Odd Days	5506	3.7497	0.3619

Continued on next page

Table 20 – *Continued from previous page*

Ticker	Set	n	Mean Deg.	Cluster.
MAN	Full Set	11011	3.6999	0.3488
	Set 1	2753	3.5728	0.3267
	Set 2	2753	3.662	0.3446
	Set 3	2753	3.7112	0.3607
	Set 4	2752	3.7908	0.3669
	Even Days	5505	3.7718	0.3593
	Odd Days	5506	3.7792	0.3641
MUV2	Full Set	11011	3.5372	0.3321
	Set 1	2753	3.2528	0.2793
	Set 2	2753	3.4288	0.3191
	Set 3	2753	3.6228	0.3518
	Set 4	2752	3.8164	0.3687
	Even Days	5505	3.6603	0.3497
	Odd Days	5506	3.6669	0.3527
RWE	Full Set	11011	3.6711	0.3484
	Set 1	2753	3.5512	0.3257
	Set 2	2753	3.6464	0.3427
	Set 3	2753	3.718	0.3556
	Set 4	2752	3.7576	0.3666
	Even Days	5505	3.7504	0.3637
	Odd Days	5506	3.7544	0.3622

Continued on next page

Table 20 – *Continued from previous page*

Ticker	Set	n	Mean Deg.	Cluster.
SDF	Full Set	11011	3.5979	0.3346
	Set 1	2753	3.4884	0.3150
	Set 2	2753	3.5684	0.3198
	Set 3	2753	3.5752	0.3316
	Set 4	2752	3.6656	0.3593
	Even Days	5505	3.6923	0.3507
	Odd Days	5506	3.6971	0.3523
SIE	Full Set	11011	3.7822	0.3632
	Set 1	2753	3.6976	0.3545
	Set 2	2753	3.7316	0.3588
	Set 3	2753	3.7884	0.3686
	Set 4	2752	3.838	0.3716
	Even Days	5505	3.8594	0.3744
	Odd Days	5506	3.858	0.3770
TKA	Full Set	11011	3.7305	0.3529
	Set 1	2753	3.6112	0.3380
	Set 2	2753	3.6808	0.3496
	Set 3	2753	3.748	0.3565
	Set 4	2752	3.8148	0.3678
	Even Days	5505	3.806	0.3659
	Odd Days	5506	3.8057	0.3657

Appendix G

Here we present the FNA and Matlab codes used within the thesis.

FNA code for series of Minimum Spanning Trees. Creates a series of MSTs for DAX 30 data for a one year time period from 1st January - 31st December. There are 23 observations with a 10 day overlap:

```
resetdb
buildbycorrelation -table <file name.csv> -filter tree:gower:true -significance 0.05:bonferroni -window 23:10
corrdistance -p correlation -method gower -savep distance
radialtreelayout -p distance
viz -vlabel vertex_id -vfontsize :::15 -vsize :::5 -awidth :::3 -acolor significant -arrow :false
```

FNA code for series of Planar Maximally Filtered Graphs. Creates a series PMFGs for DAX 30 data for a one year time period from 1st January - 31st December. There are 23 observations with a 10 day overlap:

```
resetdb
buildbycorrelation -table DAX30_2008.csv -filter pmfg:gower:true -significance 0.05:bonferroni -window 23:10
viz -vlabel vertex_id -vfontsize :::15 -vsize :::5 -awidth :::3 -acolor significant -arrow :false
```

Matlab code for constructing a Horizontal Visibility Graph from an Excel file containing two columns: the 'Time' column and 'Price' column. We acknowldege and thank Ashley Brereton for his work on this code.

```
clear all
```

```
clc; format compact; format long e;

buffer=12;
Time = xlsread(<file name.xlsx'>, 'Sheet1', 'C2:C');
Price=xlsread(<file name.xlsx'>, 'Sheet1', 'C2:C');

length_of_time=length(Price);
self_marker=zeros(length_of_time,4);

for i=1:length_of_time

\%look to the right

places_to_right=find(Price(i+1:end)>=Price(i),1,'first');

\% look to left

places_to_left=i-find(Price(1:i-1)>=Price(i),1,'last');
if isempty(places_to_left)
    places_to_left=0;
end
if isempty(places_to_right)
    places_to_right=0;
end

```

```
self_marker(i,1)=i-places_to_left;
self_marker(i,2)=places_to_right+i;
self_marker(i,3)=i-1;
self_marker(i,4)=i+1;

if self_marker(i,4)>length_of_time
self_marker(i,4)=0;
end

if self_marker(i,3)==0
self_marker(i,3)=0;
end

if (self_marker(i,2)-i)==0
      self_marker(i,2)=0;
end

if (self_marker(i,1)-i)==0
      self_marker(i,1)=0;
end

end

r_right=zeros(length_of_time,buffer);

```

```
r_left=zeros(length_of_time, buffer);

for i=1:length_of_time

    if i<length_of_time-1
 [r1_right,~] = find(self_marker(i+2:end,:)==i);
 r1_right=r1_right+i+1;
 if isempty(r1_right)
 else
        r_right(i,1:length(r1_right))=r1_right;
 end
    end

    if i>2
 [r1_left,~] = find(self_marker(1:i-2,:)==i);

 if isempty(r1_left)
 else
        r_left(i,1:length(r1_left))=r1_left;
 end

    end

end

```

```

    full_network1 = [r_left self_marker r_right];

for i=1:length_of_time

    connections(i)=length(find(unique(full_network1(i,:))));
end

full_network=zeros(length_of_time,max(connections));

for i=1:length_of_time
    dummy1=unique(full_network1(i,:));
    dummy=dummy1(dummy1>0);
    full_network(i,1:length(dummy))=dummy;
end

clearvars -except full_network connections length_of_time <
    file>_price Time

p=plot(Time,connections);
set(p,'LineWidth',3)
set(p,'Color','r')
xlabel('Time')
ylabel('Connections')
```

The following code is used to transform the HVG codes generated by Matlab into a format that can be uploaded to FNA.

```
    N=length_of_time*x \% x = \#columns in 'full_network'
c=Time

cc=c(:,ones(10,1))
from_id=cc(:)

FNA=zeros(N, 2)
FNA(:,1)=from_id
FNA(:,2)=full_network(:)

A = FNA(:,2)==0
FNA(A,:) = []
```