Asset Clusters and Asset Networks in Financial Risk Management and Portfolio Optimization

Zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaften

von der Fakultät für Wirtschaftswissenschaften des

Karlsruher Instituts für Technologie (KIT)

genehmigte Dissertation

von

Jochen Papenbrock

geb. in Minden

Hauptreferent: Prof. Dr. S.T. Rachev
Korreferent: Prof. Dr. G. Bol
Tag der mündlichen Prüfung: 05.12.2011

to my family

Table of Contents

1 Abstract of the Dissertation 4

2 Introduction and Background of the Dissertation 6

3 Presentation of the Methods 14

3.1 Approaches to Clustering 14
3.2 Hierarchical Clustering of Financial Time Series 16
3.3 Properties of Distance Measures 27
3.4 Selection of Distance Measures 29
3.5 Networks in Finance 31

4 New Approaches to Equity Portfolio Management 42

4.1 Overview of Clustering in Portfolio Management 43
4.2 The Cluster Based Waterfall Approach 47
4.3 The Asset Network Approach 60
4.3.1 Significance of the Correlation Coefficients and the MST 61
4.3.2 The Fall of the German Banks in the Financial Crisis: A Dynamic Analysis 64
4.3.3 Portfolio Optimization Based on Asset Networks 73

5 New Approaches to Credit Portfolio Risk Management 79

5.1 From Concentrations to Risk Concentrations 83
5.2 Practical Application: Detection of Credit Risk Concentrations 85
5.3 The Effect of Eliminating Risk Concentration Drivers 92
5.4 Scenario Based and Dynamic Analysis of Risk Concentrations 94

6 Further Potential of Clustering and Networks in Finance 96

7 Bibliography 98

1 Abstract of the Dissertation

In this thesis we use explorative empirical procedures for grouping financial assets into homogeneous groups or clusters. The assets are solely represented by their return time series. The clustering is developed from the data without any background model so it differentiates from other models of asset classification (like defining industry sectors a priori) which are often subjective and exhibit model risk. The explored structures are called Asset Clusters and they are very helpful in risk management and portfolio optimization. For example, the Asset Clustering can be used to identify assets with very different return structures, compared to the bulk of assets. Asset Networks can be designed in a similar way as Asset Clusters and provide an even richer set of applications to optimize and risk manage financial portfolios. Although these networks are called complex due to their non-regular and non-random topology, they reduce complexity of financial portfolios in a way that is favourable for risk management and portfolio optimization.

In this dissertation, we present new potential approaches based on the clustering and network procedures. There are portfolio management applications to an equity portfolio consisting of the stocks of the German DAX. Also, there are risk management applications to a large mixed portfolio of a commercial bank that can be embedded into the risk management process and into the internal capital adequacy assessment process (ICAAP). The applications are enriched by powerful visualization techniques of the portfolio structures based on the cluster and network technology so it is possible to intuitively retrace the mechanisms of the approaches. Also, the visualization techniques take advantage of human cognitive strengths and amplify intelligence of decision-makers like portfolio and risk managers. We have developed several software prototypes that visualize important portfolio interactions, indicate optimization and risk management potential, and even recommend concrete portfolio and risk management initiatives. As a byproduct, we analyse the network based time evolution of the banking sector of the DAX index (year 2005 to

2010) and show that it has for some time lost its central position in the German economy since the financial crisis. Finally, we show further potential of our approaches in other financial applications and illustrate future promising research directions. To the best knowledge, our pioneering approaches to risk management and portfolio optimization have never been presented in scientific literature or used in practice before. This empirical work contributes to the current stream of research in the direction of new economic thinking viewing economy and financial markets as complex systems and networks.

2 Introduction and Background of the Dissertation

The latest and most dramatic financial crisis has once more highlighted that current concepts in financial economics like efficient markets, rational expectations and general equilibrium may be too dogmatic and axiomatic. It is often suggested that the focus should be more on data and empirical approaches. In July 2010 for example, “The Economist” writes that conventional economic models failed to foresee the financial crisis and that The Institute for New Economic Thinking has attacked many of the assumptions, including efficient financial markets and rational expectations, on which these models are predicated. [1]

In this context, an excerpt from a speech titled “Reflections on the nature of monetary policy: non-standard measures and finance theory” by Jean-Claude Trichet, President of the European Central Bank (ECB), on November 18, 2010 is very remarkable: “I would very much welcome inspiration from other disciplines: physics, engineering, psychology, biology. Bringing experts from these fields together with economists and central bankers is potentially very creative and valuable."[2]

Popular science magazines like Nature and Science have come up with contributions to this topic. Authors propose that “there should be a change of mindset in economics and financial engineering, that should move away from dogmatic axioms and focus more on data, orders of magnitudes, and plausible, albeit non rigorous, arguments” (Bouchaud 2008). Schweitzer et al. (2009) suggest that “the current economic crisis illustrates a critical need for new and fundamental understanding of the structure and dynamics of economic net-

[1] http://www.economist.com/node/16636121, „Agents of Change, Jul 22nd 2010
[2] http://www.ecb.int/press/key/date/2010/html/sp101118.en.html

works. Economic systems are increasingly built on interdependencies, implemented through trans-national credit and investment networks, trade relations, or supply chains that have proven difficult to predict and control. We need, therefore, an approach that stresses the systemic complexity of economic networks and that can be used to revise and extend established paradigms in economic theory."

Major central banks have recently propagated interdisciplinary approaches between Economics and Biology/Ecology, Physics, and Mathematics/Engineering. Examples are the conference "New directions for understanding systemic risk" organised by the Federal Reserve Bank of New York (NY Fed) and the US National Academy of Sciences in May 2006, and the conference "Alternative Approaches to Modeling Systemic Risk" organized by ECB, NY Fed and the Center of Financial Financial Studies (CFS). Also, there has been an ECB report in 2010 that shows recent advances in modelling systemic risk using network analysis".[3]

However, it is important that researchers from natural science with non-economical background do not overshoot the mark and that for every problem there must be chosen the right tool as Farmer/Geanakoplos (2009) point out: "On the one hand, we worry that physicists often misunderstand the equilibrium framework in economics, and fail to appreciate the very good reasons for its emergence.[…] In other cases where the cognitive task is extraordinarily complex, such as the pricing of a new firm, or where estimation problems are severe, such as portfolio formation, human models may diverge significantly from rational models, and the equilibrium framework may be a poor approximation. For good science one must choose the right tool for the job, and in this case the good scientist must use an assortment of different tools. Close-mindedness in either direction is not likely to be productive. As we have stressed, equilibrium theory is an elegant attempt to find a parsimonious mod-

[3] European Central Bank Report (2010): "Recent Advances in Modelling Systemic Risk Using Network Analysis",
http://www.ecb.europa.eu/pub/pdf/other/modellingsystemicrisk012010en.pdf?d216f976f3587224bcc087cc8149ed49

el of human behaviour in economic settings. It can be criticized, though, as a quick and dirty method, a heroic attempt to simplify a complex problem. Now that we have begun to understand its limitations, we must begin the hard work of laying new foundations that can potentially go beyond it. "

Our contribution in this thesis is to propose new methods like data-driven complex networks in risk management and portfolio optimization and integrate them into the real world context of risk managers and portfolio optimizers in financial institutions. We are now going to present the main idea of the thesis.

Diversification is fundamental in finance

Diversification in finance is a risk management technique that mixes a wide variety of investments within a portfolio. It is the spreading out dissimilar investments to reduce risk. The standard mathematical formulation of this problem is the well known Markowitz mean-variance approach where the (co-)variances of assets determine the degree of portfolio diversification. Markowitz (1959) comments in his fundamental work about portfolio selection: "Like most economic quantities, the returns on securities tend to move up and down together. This correlation is not perfect: individual securities and entire industries have at times moved against the general flow of prosperity. […] To reduce risk it is necessary to avoid a portfolio whose securities are all highly correlated with each other."

This idea can already be observed when plotting the wealth function of some stocks from the German index DAX from 2008 to 2011. The following figure shows the paths followed by stocks from the automotive sector in green and by the financial sector in red.

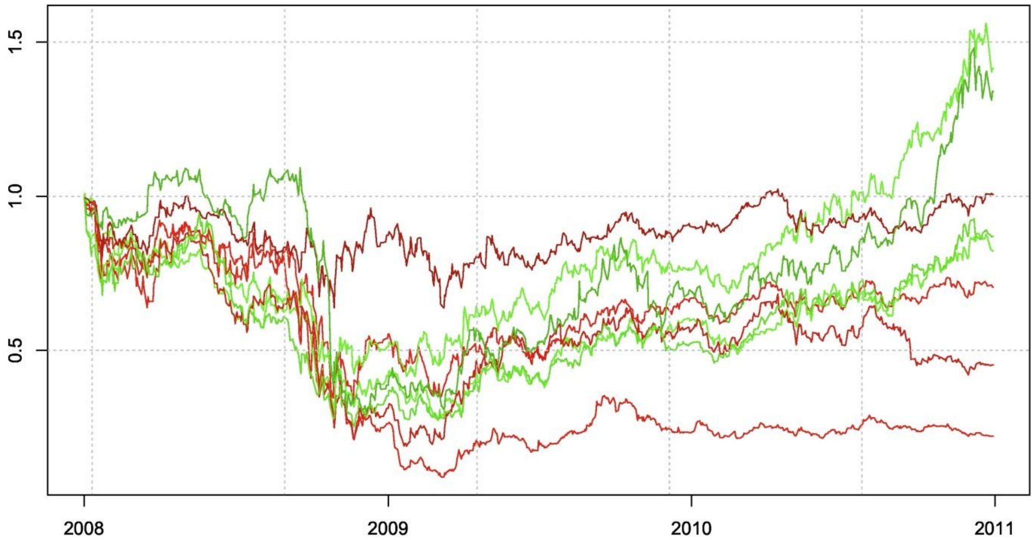

Figure 2-1: Automotive stocks from the DAX in green and stocks from the financial sector in red.

It is interesting to shed some light on this famous quotation from a new and very different perspective and we would like to contribute to his statement "individual securities and entire industries have at times moved against the general flow of prosperity" from a data-driven and explorative point of view. It is obvious that for risk management and portfolio optimization it is advantageous to know which assets move individually or in certain groups like industries or in other words: which assets show collective behaviour? However, we argue that clustering is a proper method to find those groups as objects in a cluster are similar to each other and also very dissimilar to objects outside the cluster, particularly objects in other clusters.

What drives the collective behavior of assets?

Farrell (1974) is one of the first to use cluster analysis in portfolio management. He builds indices of stocks that are homogeneous in the sense that they are significantly correlated within their own grouping and, at the same time, generally independent of other groups. He stresses the importance of clustering in portfolio management and gives article references to the relevance of homogeneous groups in different topics in finance. His insight is that grouping stocks in multi-index approaches based on industry sectors often failed due to possible collinearity of industries.

According to Onnela et al. (2003), there might be several reasons why an industry classification fails: for example, the industry classifications might be outdated or there are companies or conglomerates that are engaged in multiple industries (like Siemens or General Electric). Also, there are industries dependent on other industries or there are sectors like banks and insurance companies can be connected to all sectors.

Farrel's method can be characterized as subsuming industries into broader categories, just as King (1966) shows that companies can be grouped into broader categories than the method of industry classification. The evidence of significant co-movement among industry groupings implies that there is another factor, broader than the industry factor and in co-existence to the general market and company specific risk drivers. In particular, he uses a broader classification to see if the price action of stocks conforms to this classification and finds that his stock groupings were homogeneous, stocks within each group were highly correlated and the inter-group relationship showed near independence. These procedures were further developed to a multiple factor risk model by Arnott (1980) who also describes applications of this approach to stock classification, portfolio optimization and performance measurement.[4]

The approaches presented by Farrell and Arnott require hypothesized stock categories like growth, cyclical and energy. The classification procedure is supervised as there is some apriori economical classification structure imposed which implies some model risk and is subject to personal opinion and subjectivity. Also, as financial markets are socio-economical systems, there are psychological factors at work (see for example Onnela et al. 2003). Final-

[4] We would like to thank Frank J. Fabozzi for drawing our attention to this part of literature (Farrell, King, Arnott). He also gave other very helpful and valuable advice.

ly, financial markets can be hit by exogenous shocks like news impact or by endogenous shocks with no obvious economic explanation. Also, there are institutionalized reasons for collective behaviour of asset returns. Examples are similar trading patterns of hedge funds according to the use of similar models like coordinated deleveraging of similarly constructed portfolios (see for example Khandani/Lo 2007) or there are market constellations that trigger action of numerous investment management firms in order to follow their legal and institutional investment constraints.

Being confronted with this bulk of explanations why asset prices exhibit collective behaviour there must be another way of consistently modelling these phenomena. A solution might be data-driven explorative methods like clustering. Theses methods do not require any economical, causal, institutional, or psychological explanation (or a mixture thereof) which is subject to potentially unsafe assumptions. In contrast, machine learning tasks like clustering are specialized in finding and describing complex structures in high-dimensional and often noisy data. Various machine learning algorithms have been developed for data analysis and decomposition problems such as regression, ranking, classification, clustering, factor modelling, dimension reduction, feature selection and de-noising. They are applicable in several areas of financial data analysis including, e.g. risk analysis, portfolio construction and optimization, hedging, pricing, and trading strategies, where the identification and quantification of hidden relationships within and between many time series forms an essential part. However, it has to be stated that these methods will always find patterns in data no matter if these are really existing or not. So after the machine learning step there has to be a validation step. Fortunately, there are numerous mathematical validation techniques which are able to provide significance estimations concerning the patterns and structures found and some of them will be presented and applied in this work.

Another drawback of time series based explorative methods is the historical focus. It should be clear that prediction based on past data is very difficult if not impossible. However, there are many applications based on historical data which are quite successful in finding some elementary structures in data which are valid and persistent for some time in the future. Having addressed some drawbacks of time series clustering, it has to be stated once more that there is no a priori knowledge of the structures required as a model input so there is a potential outperformance in the application of explorative approaches compared to classical approaches.

Recently, there are more and more approaches to financial applications from the fields of statistical physics, network theory, econophysics, complex systems, and agent based models. Since the end of the 1990s, correlation based clusters and networks have become present in academic finance literature and publications whereas one of the most cited works is Mantegna (1999). We believe that there is even more potential for those kinds of methods in finance, especially after the observation of some classical models in finance to fail in times of crisis. The approaches presented in the latest academic works show a high degree of methodological maturity.

Another insightful aspect of explorative methods is their assessment against their counterparts from the assumptions-based modelling world, i.e., mathematically validated outcomes of explorative methods can be compared to economical models. An example is the clustering of market taxonomies and explorative sector constructions compared to economically defined industry sectors. There is often large correspondence between the two models which indicates that the coincidental part of both very different approaches is properly modelled. A matter of particular interest is the non-coincidental part which shows that at least one of the models is less precise. If there is reason to believe that the results of the explorative method are more reliable this non-coincidental part carries a lot of additional information.

Another advantage of using correlation based clustering and networks is an effective filtering of the correlation matrix. Estimates of correlations are often noisy and unreliable as estimation horizons are always finite. Optimization programs like the Markowitz approach suffer from theses estimation errors and clustering/networks may act as a filtering procedure whose outcomes are more reliable. These aspects are well documented in literature (see for example Mantegna/Stanley 2000 and Tumminello et al. 2010).

Many scientific works about clusters and networks in finance deliver spectacular new insights to financial markets and asset return modelling. However, only few address the classical problems arising both in academical and practical finance. So it is important to integrate the available technologies like clustering and networks into the practical requirements of financial risk management and portfolio optimization. Having shown the modelling advantages of clustering and network techniques, we would now like to fill some gap between the technical aspects of these methods and concrete applications in finance.

3 Presentation of the Methods

In this chapter we will describe the methods used throughout this work. There will be a categorization of different clustering methods and a way to find a proper clustering method for the task at hand. Clustering is often based on some distance measure between the objects. It is very important to analyse the combination of distance measures and clustering methods as both together form the unsupervised classification task. Finally, we will show how filtered networks can be generated from certain types of cluster algorithms.

3.1 Approaches to Clustering

Custer analysis is a multivariate statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, and bioinformatics. It is a so-called unsupervised classification method whereas in contrast, discriminant analysis is supervised classification (see Jain at al. 1999). Cluster analysis is the formal study of methods and algorithms for grouping, or clustering, objects according to measured or perceived intrinsic characteristics or similarity (see Jain 2010). The goal of clustering is to objectively organize data into homogeneous groups where the within-group-object similarity is minimized and the between-group-object dissimilarity is maximized.

Clustering algorithms can be broadly divided into two groups: hierarchical and partitional. Accoring to Jain (2010), hierarchical clustering algorithms recursively find nested clusters either in agglomerative mode (starting with each data point in its own cluster and merging the most similar pair of clusters successively to form a cluster hierarchy) or in divisive (top-down) mode (starting with all the data points in one cluster and recursively dividing each cluster into smaller clusters). Compared to hierarchical clustering algorithms, partitional clustering algorithms find all the clusters simultaneously as a partition of the data and do not impose a hierarchical structure. Hierarchical clustering builds a cluster hierarchy or a tree of clusters, also known as a dendrogram. Such an approach allows exploring data on different levels of granularity.

Representing the data by fewer clusters necessarily loses certain fine details, but achieves simplification. This simplification can of course be done more ore less efficient and one success factor is the number of parameters. According to Keogh et al. (2004), two main dangers of working with parameter-laden algorithms are the following: First, incorrect settings may cause an algorithm to fail in finding the true patterns. Second, a perhaps more insidious problem is that the algorithm may report spurious patterns that do not really exist, or greatly overestimate the significance of the reported patterns. This is especially likely when the user fails to understand the role of parameters in the data mining process. In their opinion, data mining algorithms should have as few parameters as possible, ideally none. A parameter-free algorithm would limit the ability to impose prejudices, expectations, and presumptions on the problem at hand, and would let the data speak for themselves.

It is obvious that there is a very important parameter in partitional clustering: the number of clusters. If practitioners are required to deliver such an important input parameter it seems to contradict the paradigm of letting the data speak for themselves. For this reason, there have been researchers who developed clustering algorithms that do not require a priori knowledge of the number of clusters making it real “cluster-mining” algorithms (an example is the U*C clustering algorithm in Ultsch 2007).

However, it is not clear if real cluster mining approaches are “best” since each clustering algorithm imposes a structure on the data either explicitly or implicitly. Since the structure of the data is not known a priori, one needs to try competing and diverse approaches to determine an appropriate algorithm for the clustering task at hand. This idea of no best clustering algorithm is partially captured by the impossibility theorem (Kleinberg 2002), which states that no single clustering algorithm simultaneously satisfies a set of basic axioms of data clustering (see Jain 2010). Kleinberg identifies three properties that one may expect all clustering functions to satisfy, but then proves no function can satisfy all three properties. Zadeh/Ben-David (2009) circumvent Kleinberg's impossibility theorem by relaxing one of his axioms and restricting the attention to clustering algorithms that take the number of clusters to be created as part of their input. It is known that such a restriction suffices to render the resulting set of axioms consistent and one interpretation of Kleinberg's impossibility result is that if one does not give algorithms the number of clusters they are to return, then the algorithm must be performing some unintuitive operations. Carlsson/Mémoli (2010) show in a similar spirit to Kleinberg's theorem, that in the context of hierarchical methods, one obtains uniqueness instead of non-existence. They emphasise that their result can be interpreted as a relaxation of the theorem proved by Kleinberg, by allowing the output of clustering methods to be hierarchical so hierarchical clustering seems to be a good choice for complex data. In 1962, Nobel laureate Simon wrote: “the central theme that runs through my remarks is that complexity frequently takes the form of hierarchy, and that hierarchic systems have some common properties that are independent of their specific content. Hierarchy, I shall argue, is one of the central structural schemes that the architect of complexity uses”.

According to Murtagh (2004), ultrametricity is a natural property of sparse, high-dimensional spaces and it emerges as a consequence of randomness and the law of large numbers.

3.2 Hierarchical Clustering of Financial Time Series

The explorative risk management and portfolio optimization applications presented in this work require time series of asset returns as input. Hierarchical clustering is a collection of procedures for organizing objects into a nested sequence of partitions on the basis of data on the similarity or respectively dissimilarity among the objects. It is the fitting of a high dimensional space into a tree-like structure which is depicted in dendrograms. The dissimilarity between objects is measured by a distance matrix D whose components d_{ij} resemble the distance between two points x_i and x_j. The hierarchical clustering procedure is a two-stage process:

1. choice of a distance measure and
2. choice of the cluster algorithm,

whereas both choices together define the whole clustering outcome. Distance measures of asset return time series focus on the dissimilarity between the synchronous time evolutions of a pair of assets. The matrix of pairwise distances will be the input of the hierarchical cluster algorithm that uses some linkage rule to determine a hierarchical structure. The choice of clustering procedure, also in combination with the distance measure of assets has to be carefully made as it is a critical part of our approaches.

Agglomerative hierarchical clustering algorithms produce nested series of partitions based on merging criterions. Each partition is nested into the next partition of the sequence. After the proximity index has been defined and a distance matrix has been calculated, the hierarchical clustering can be carried out by a suitable clustering algorithm. The clustering algorithm specifies how the distance matrix is processed in order to merge two elements/clusters until a single cluster containing all elements is created Jain et al. (1999).

In hierarchical clustering a bijection is defined between a rooted, binary, ranked, indexed tree, called a dendrogram, and a set of ultrametric distances (Murtagh (20004). The "strong triangular inequality" or ultrametric inequality is $d(x,z) \leq \max \{d(x,y), d(y,z)\}$ for any triplet of points x, y, z . Deriving the dendrogram from the raw data involves several steps (Jain/Dubes 1988, Jain at al. 1999):

1. Data collection (e.g. returns of daily closing prices)
2. Representation (proximity index in the form of a distance matrix)

3. Clustering the data set
4. Validation (validating the quality of the dendrogram).

The structure that was imposed on the distance matrix by the clustering algorithm is captured in the cophenetic/ultrametric matrix. The cophenetic matrix records the value at which a clustering is formed - or more precisely: the cophenetic proximity matrix indicates at which level (distance) two objects first appear in the same cluster. It therefore usually contains many ties. It has perfect hierarchical structure. The higher the degree of agreement between the cophenetic matrix and the distance matrix, the better does the hierarchical structure fit the data. The goal of a clustering algorithm is to find a perfect hierarchical structure that is as close to the distance matrix as possible. This insight will play a crucial role when determining the Cophenetic Correlation Coefficient (CPCC) that helps us determine the quality of the clustering Jain et al. (1999). As clustering algorithms will always find a clustering structure one has to determine to which extent the clustering could have evolved from a random structure or is itself random (Jain et al. 1999, Jain and Dubes 1988). The CPCC is defined as

$$CPCC = \frac{\sum_{i<j}(d_{ij} - \overline{d})(c_{ij} - \overline{c})}{\sqrt{[\sum_{i<j}(d_{ij} - \overline{d})^2][\sum_{i<j}(c_{ij} - \overline{c})^2]}},$$

letting $\overline{d}$ be the average of the d_{ij} and letting $\overline{c}$ be the average of the c_{ij}.

Four clustering algorithms will be used in the analysis: single-linkage, average-linkage, complete-linkage and Ward's method. The single-link and complete-link algorithms follow two very basic concepts that are oftentimes used to derive different algorithms. The idea behind single-linkage is to form groups of elements, which have the smallest distance to each other (nearest neighbouring clustering). This oftentimes leads to large groups/chaining. The complete-linkage algorithm tries to avoid those large groups by considering the largest distances between elements. It is thus called the farthest neighbour clustering. The average-linkage algorithm is a compromise between the single-linkage and complete-linkage algorithm (Jain and Dubes 1988). Ward's method joins elements/clusters that do not increase a given measure of heterogeneity too much thus tries to create groups within clusters that are as homogenous as possible. A basic agglomerative algorithm is presented in Tan et al. (2005), Jain et al. (1999):

Compute proximity matrix

repeat

Find most similar pair of clusters and merge those clusters

Update proximity matrix to reflect changes

until *One cluster remains*

It becomes clear, that the fundamental difference in many hierarchical clustering algorithms is the definition of "closest clusters". A more detailed description of the most important clustering algorithms will shed some light on their basic idea and understanding of "what is similar".

Single-Linkage

The single-linkage clusters are characterized by maximally connected subgraphs. The algorithm is clustering the elements, which are nearest to each other first, thus is often referred to as the "nearest neighbour" or "minimum algorithm". Its basic idea can also be used to construct minimal spanning trees to which the single-linkage algorithm is closely related as will be shown later. The single-linkage takes the minimum distance between two elements/clusters of the current (updated) proximity matrix to merge the next elements/clusters. It can thus be described as pseudo code in the following form

Compute proximity matrix

repeat

    Merge clusters for which the distance $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$

    Update proximity matrix to reflect changes

until *One cluster remains*

The following example shows a correlation matrix of four German stocks. The correlation matrix is transformed to a distance matrix and this in turn is transformed into the ultrametric or cophenetic matrix by the single linkage hierarchical clustering. The height of the dendrogram indicates the distances at which clusters were agglomeratively merged together.

	ADS.DE	BEI.DE	LHA.DE
BEI.DE	0.29		
LHA.DE	0.43	0.30	
LIN.DE	0.44	0.31	0.45

	ADS.DE	BEI.DE	LHA.DE
BEI.DE	1.188		
LHA.DE	1.063	1.186	
LIN.DE	1.060	1.175	1.046

	ADS.DE	BEI.DE	LHA.DE
BEI.DE	1.175		
LHA.DE	1.060	1.175	
LIN.DE	1.060	1.175	1.046

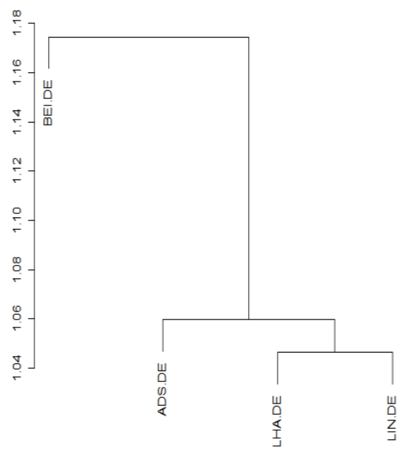

Figure 3-1: Example of single linkage hierarchical clustering.

The ultrametric distance c resulting from the single link method is such that $c_{ij} \leq d_{ij}$ always. It is also unique with the exception of ties. It is also termed the subdominant or maximal inferior ultrametric.

In the single-linkage clustering algorithm, objects to be merged just need to be neighbours thus they tend to experience "chaining".

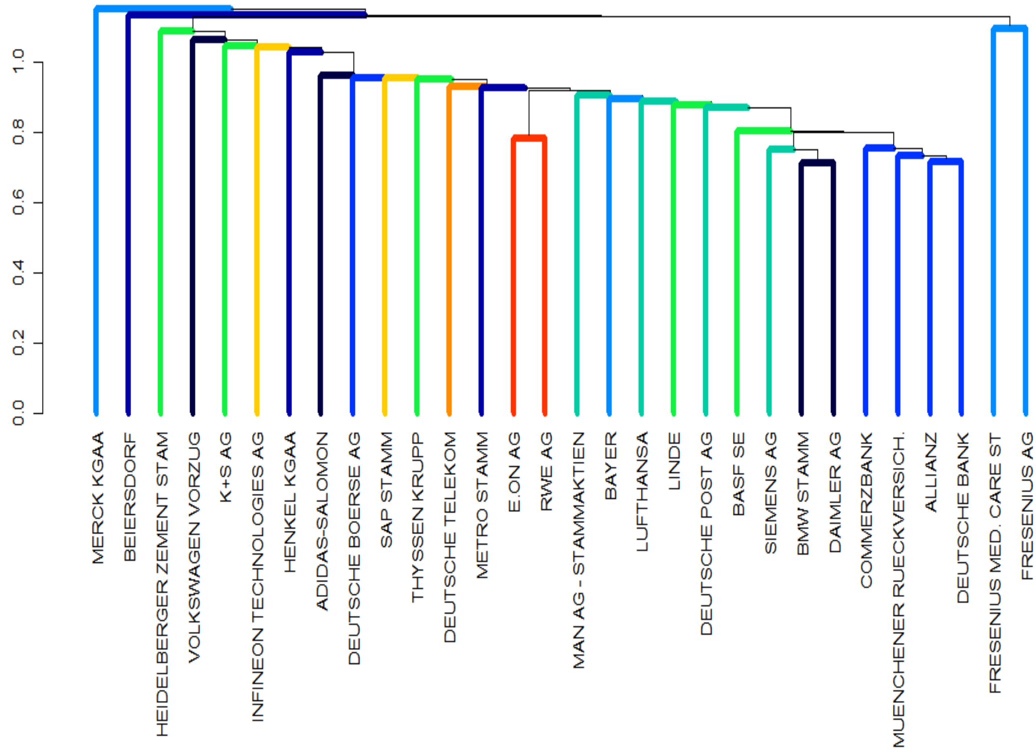

Figure 3-2: Single linkage clustering of the DAX data from 2003 to 2011. Colour codes are according to industry classification.

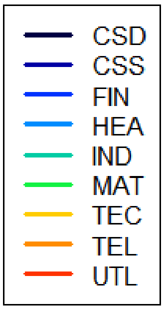

Figure 3-3: Color codes of an industry classification for DAX stocks.

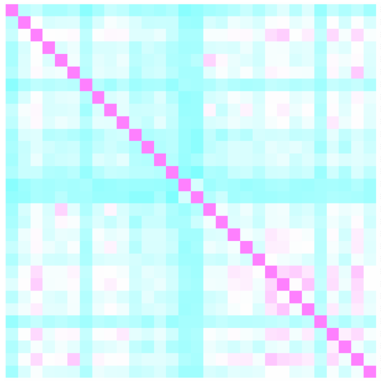

Figure 3-4: Unordered correlation matrix as input to the hierarchical clustering algorithm. High correlations are purple.

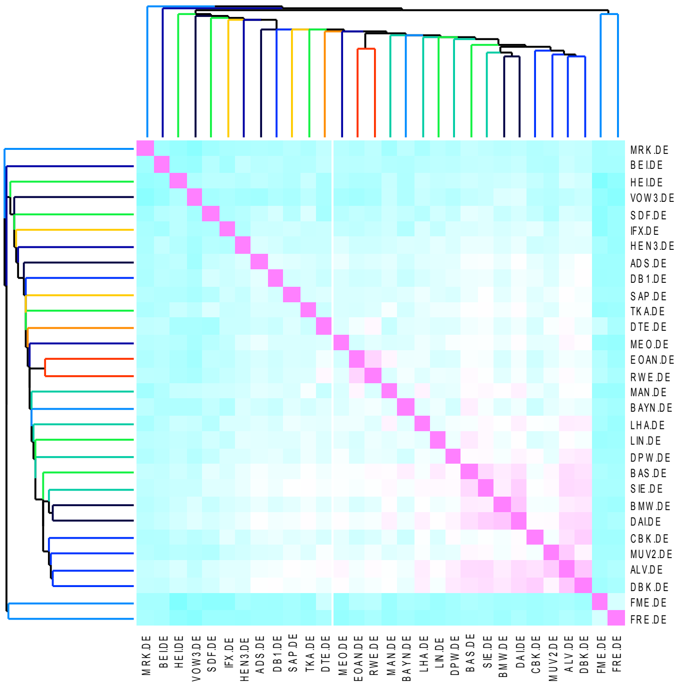

Figure 3-5: Ordered correlation matrix according to the single linkage hierarchical clustering algorithm. High correlations are purple. Clusters of high correlations can already be observed visually.

Complete-linkage

The complete-link clusters are more restrictive with respect to the pairs of clusters that are merged in a round. All pairs of objects are related before the cluster is formed. The minimum of those distances indicates, which clusters or objects to merge next. It is thus less vulnerable with respect to noise and outliers. However, it can break large clusters and lead to globular shapes. It is furthermore usually more compact than the single-link algorithm. For many practical applications, the complete-link clustering provided better results than single-link (Jain 1999). The clustering algorithm is in its design very similar to the single link, with the exception of the merging operation:

Compute proximity matrix

repeat

Merge clusters for which the distance $d(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y)$

Update proximity matrix to reflect changes

until *One cluster remains*

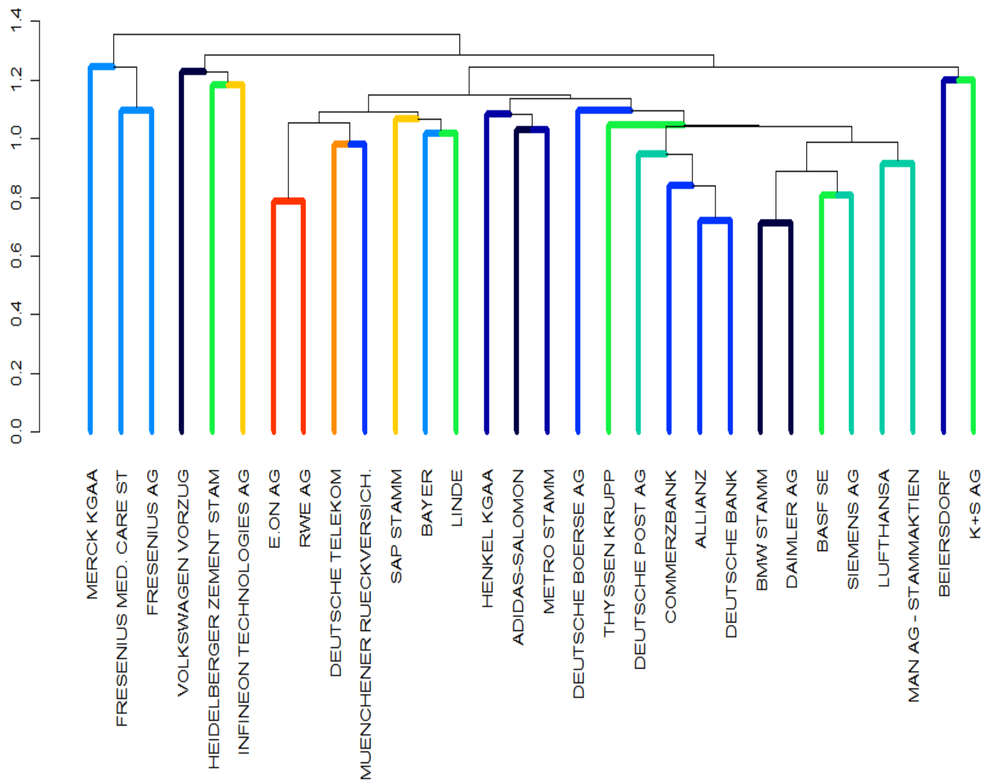

Figure 3-6: Complete linkage clustering of the DAX data from 2003 to 2011. Colour codes are according to industry classification.

Average-linkage

The average-link clustering algorithm is a combination between the complete-link and single-link as it does not take the minimum or maximum distance between pairs of clusters but the group average. The distance used to determine, which clusters are to be merged next is thus defined as:

$$d(C_i, C_j) = \frac{\sum_{x \in C_i, y \in C_j} d(x,y)}{|C_i||C_j|}$$

The clustering algorithm is the same for the average linkage as for single link or complete link with the only difference of the definition of "most similar pair of clusters".

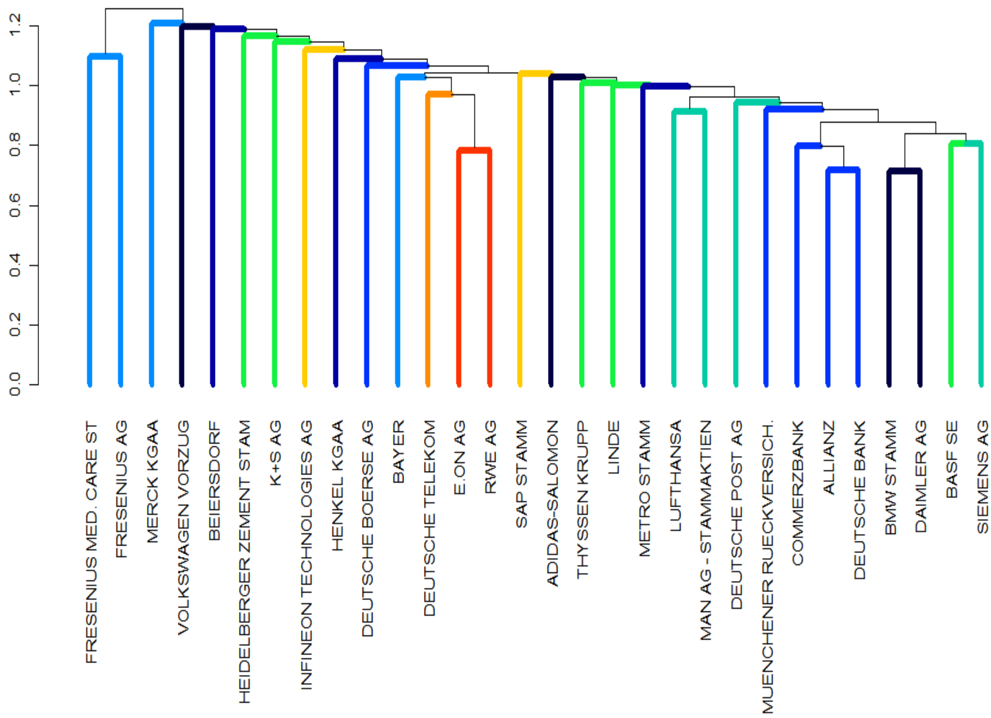

Figure 3-7: Average linkage clustering of the DAX data from 2003 to 2011. Colour codes are according to industry classification.

Ward's method

Whereas single-linkage, complete-linkage and average-linkage can be classified as graph-based clustering algorithms, Ward's method has a prototype-based view in which the clusters are represented by a centroid. For this reason, the proximity between clusters is usually defined as the distance between cluster centroids. Whereas in the clustering approaches discussed earlier, the "farest", "closest", etc. distances between clusters or elements was used to derive the next merging operation, in Ward's method the increase of the "sum of the squares error" (SSE) is determined. The SSE is the sum of errors of every data point. The error of every data point is its distance from its closest centroid. The SSE can be calculated as (Tan et al. 2005):

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(c_i, x)^2$$

The centroid (mean) of any cluster i is defined as:

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x$$

Just like the K-Means (partitioning clustering algorithm), Ward's method tries to minimize the squared errors from the mean (objective function is similar). However, it differs in the way that Ward's method is a hierarchical algorithm, where elements are merged together. The element or cluster, which is merged next is determined by the change of the SSE. Even though it may appear based on the objective function and goal of Ward's method, that it is closer related to K-Means than to other hierarchical methods, it can mathematically be shown, that Ward's is very similar to the average-linkage when the proximity measure is the Euclidean distance (Tan et al. 2005). Ward's method can nevertheless be understood as the hierarchical representation of K-Means.

An often-cited downside of the centroid methods is that the clusters that are formed at each step of the algorithm do not represent local minima with respect to total SSE. Ward's has, unlike the other hierarchical clustering algorithms mentioned earlier, furthermore the (oftentimes cited undesirable) property of possible inversions. That means that clusters merged at a later step may in fact be more similar than clusters merged in an earlier step (Jain/Dubes 1988).

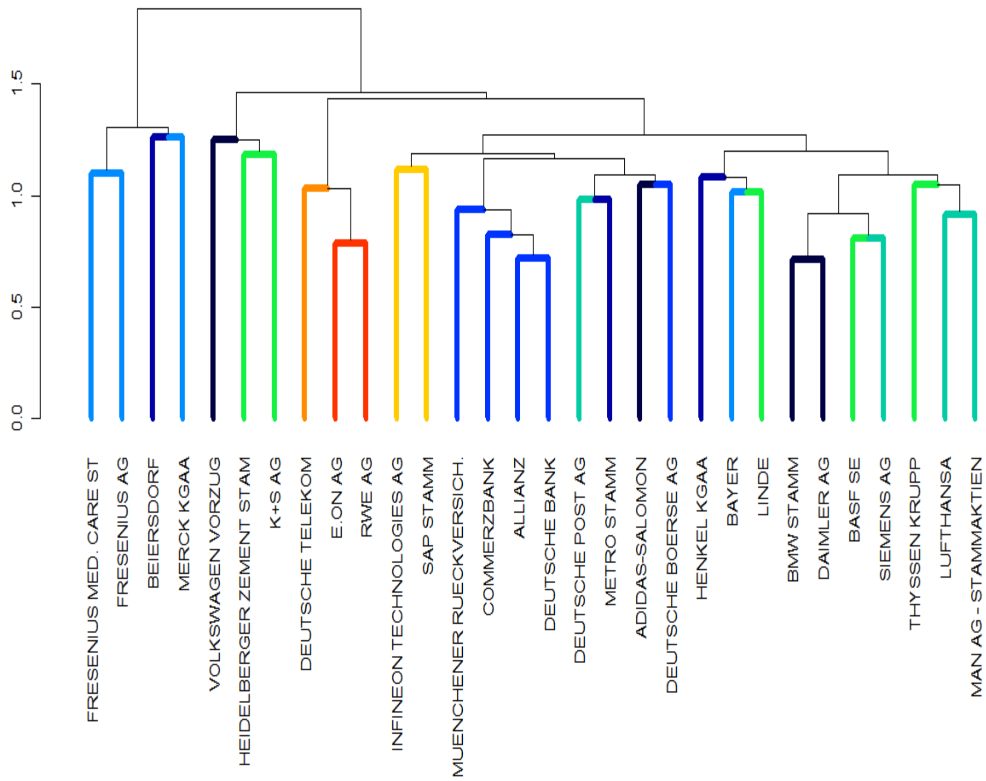

Figure 3-8: Ward linkage clustering of the DAX data from 2003 to 2011. Colour codes are according to industry classification.

Summary of the clustering methods

A theoretical comparison of clustering algorithms is not feasible, as it is practically impossible to describe the different approaches mathematically in a way that can be compared (Jain/Dubes 1988). Different distance measures further complicate the problem. Instead, a rough comparison in terms of "what is intended" with the algorithms is given and thus leaves the ultimate choice to the user.

The single-link algorithm oftentimes forms clusters that are chained together and leaves large clusters. It can probably be best understood as a way to give a "more robust" estimation of the distance matrix and furthermore preserves the original structure as much as possible. Elements departing early from the tree can be interpreted as "different" from the overall dataset. In terms of application, the single-link clustering algorithm is very useful to gain insights in the correlation structure between assets and separates assets that were very different from the rest. If this separation is preferred and high weights should be put on "outliers" the single link certainly is a good choice.

The complete-link algorithm has a different idea: elements should be grouped together in a way that they are not too different from each other when merged in a cluster. It thus has a much stronger definition of "similar pair of clusters". The complete-link algorithm therefore seems suitable for investors interested in grouping stocks that are similar in one cluster.

The average-linkage algorithm offers a trade-off between the complete-link and single-linkage algorithm. With respect to Ward's method, the clustering of stocks according to industries or sectors is even clearer and more meaningful than with the complete-link algorithm.

3.3 Properties of Distance Measures

According to Keogh, there are four categories of similarity measures for numeric time series:

- shape-based methods compare the overall appearance of the time series,
- feature-based methods extract time independent aspects of the series that are compared with static distance functions,
- model-based methods require a model of the data and measure the similarity by comparing model-based distances, and

- compression based methods analyze how well two time series can be compressed alone and together.

The Euclidean distance is the most widely used shape-based distance for the comparison of numeric time series (Keogh/Kasetty 2002). Feature-based methods and model-based methods require rather long time series in order to calibrate the features and models. However, they can handle different time series' lengths. One example of rather general features is the use of the first four moments of the empirical probability distribution of the data and the first order differences (Nanopoulos et al., 2001). Additional features based on trend, seasonality, and self-similarity is used in (Wang et al. 2004). Since the method clusters using extracted global measures, it reduces the dimensionality of the time series and is much less sensitive to missing or noisy data.

The model-based representations assume that the time series have been produced by a certain model. A popular choice for numeric time series is statistical modelling with ARMA models, e.g., (Kalpakis et al. 2001, Xiong/Yeung 2003). The statistical features extracted in Nanopoulos et al. (2001) and Wang et al. (2006) can also be interpreted as a model of the process generating the time series. Other models for time series could be regime switching, change point analysis, hidden Markov models and structural breaks.

The compression-based similarity of time series is inspired by computational theory. Keogh et al. (2007) define a distance measure based on the conditional Kolmogorov complexity called Compression-Based Dissimilarity Measure (CDM). The Kolmogorov complexity is the length of the shortest program that is able to generate the given data. The basic idea is that concatenating and compressing similar data should give higher compression ratios than doing so with very different data. Similar to model-based methods, they are allowed to have different lengths.

The proximity matrix is the single input to most clustering algorithms and contains all the information relevant for the clustering procedure. It can either measure similarity of dissimilarity (Jain et al. 1999, Jain/Dubes 1988). Since it is most common to refer to the similarity between patterns through their dissimilarity we will refer to the proximity measure more precisely with the term distance measure, especially when dissimilarity is a metric (Jain et al. 1999). The information of all asset return time-series is therefore reduced to a matrix with $n(n-1)/2$ distinct entries. The usual properties of a metric or distance measure are:

(i) $d(i,j) \geq 0$ (non-negativity)
(ii) $d(i,j) = 0$ if and only if i = j (identity)
(iii) $d(i,j) = d(j,i)$ (symmetry)
(iv) $d(i,k) \leq d(i,j) + d(j,k)$ (subadditivity / triangle inequality).

The second property shows that if there are two assets are completely correlated $(\rho_{ig} = 1)$ they are not separated by any distance $(d_{ig} = 0)$. The symmetry property reflects the symmetry of the correlation matrix. The triangular inequality expresses the Pythagorean equation in the Euclidean space.

3.4 Selection of Distance Measures

This section will explain some basic distance metrics and a related correlation based metric. We denote the price of a stock i at time t as $p_i(t)$. The logarithmic return can be written as $r_i(t,\Delta t) = \ln p_i(t+\Delta t) - \ln p_i(t)$. In this work we use daily returns.

The Minkowski metric satisfies all conditions of a distance metric. The general form of the Minkowski metric is defined as:

$$\left(\sum_T \left| r_i - r_j \right|^p \right)^{1/p}$$

If $p = 1$ the Minkowski metric becomes the Manhattan distance, for $p = 2$ the Euclidean distance. For $p \to \infty$ the metric becomes the "sup" distance. This work will focus on the most common case, the Euclidean distance. It has an intuitive appeal and is oftentimes used when the proximity of objects has to be determined in two- or three dimensional spaces. A commonly cited drawback of the Minkowski metric is that largest-scale features tend to dominate other features. Since only one feature (the return) is analyzed in this work, the stated problem does not really apply in this context. However, it is true that the Euclidean distance attributes weight to outliers of a data set. The Euclidean distance is used in several papers for financial cluster analysis (e.g. Lisi/Corazza 2008, or, Zhang/Maringer 2010).

Correlation Based Distance

The correlation based metric has found widespread use among practitioners of clustering in financial applications (see for example Lisi/Corazza 2008, Mantegna 1999, Tola et al. 2008, Tumminello et al. 2010, Dose/Cincotti 2005). The Pearson correlation coefficient is widely used as a measure of strength of linear dependence between two variables:

$$\rho_{ij} = \frac{\left\langle r_i(t,\Delta t) r_j(t,\Delta t) \right\rangle_T - \left\langle r_i(t,\Delta t) \right\rangle_T \left\langle r_j(t,\Delta t) \right\rangle_T}{\sqrt{\left\langle r_i^2(t,\Delta t) \right\rangle_T - \left\langle r_i(t,\Delta t) \right\rangle_T^2} \sqrt{\left\langle r_j^2(t,\Delta t) \right\rangle_T - \left\langle r_j(t,\Delta t) \right\rangle_T^2}}$$

$$\rho_{ij} = \begin{cases} 1, & \textit{completely correlated} \\ 0, & \textit{uncorrelated} \\ -1, & \textit{completely anticorrelated} \end{cases}$$

However, the correlation coefficient of a pair of asset returns cannot be used as a distance because it does not fulfil the axioms that form a metric. A real metric can be designed using a function of the correlation coefficient ρ. It can be rigorously determined by a transformation of the correlation coefficient so that the distance between variables is directly proportional to the correlation between them (Gower 1966):

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

It can be shown that this distance fulfils the usual metric properties including the triangle relation (see Mantegna 1999).

3.5 Networks in Finance

There exist numerous applications of networks in science and business such as social networks, internet traffic and biology with methods and algorithms drawn from statistical physics, computer science, complexity theory and others. Most real world networks display non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. An example is a scale-free network whose node degree distribution asymptotically follows a power law.

According to Porter et al. (2009), a network's structure may exhibit a complicated set of hierarchical and modular components. The term module or community is typically used to refer to a single "cluster" of nodes. Intuitively, a community is a cohesive group of nodes that are connected "more densely" in comparison to other nodes of the network. Porter et al. (2009) give an overview of the current research and applications of communities in networks and Lancichinetti/Fortunato (2009) develop benchmarks for testing community detection algorithms.

The collective behaviour of assets returns can be analysed by means of the construction of topologically constrained graphs from cross-correlation matrices. Given a connected, undirected graph, a spanning tree of that graph is a subgraph that is a tree and connects all the nodes/vertices together. So a spanning tree is the maximal set of edges of a graph that contains no cycle, or as a minimal set of edges that connect all vertices.

A single graph can have many different spanning trees. An attractive candidate is the minimal spanning tree (MST) because it provides an arrangement of assets which selects the most relevant connections of each element of the set resulting in n-1 edges. In our case these are the highest correlations (or lowest correlations distances, respectively, which leads to the expression minimal spanning tree). Finally, the minimal spanning tree gives the subdominant ultrametric hierarchical organization of the assets of the investigated portfolio so a weighted graph's MST closely corresponds to the single linkage clustering (see Mantegna 1999).

According to Aste et al. (2010) the following intuitive algorithm can be applied to construct the MST:

Step 1*: Make an ordered list of all edges i,j in a fully connected network ranking them by decreasing correlation* ρ_{ij} *(first the largest and last the smallest).*

Step 2*: Take the first element in the list and add the edge to the graph.*

Step 3*: Take the next element and add the edge if the resulting graph is still a forest or a tree; otherwise discard it.*

Step 4*: Iterate the process from step 3 until all pairs have been exhausted.*

The following example uses the same data as in the single linkage clustering in order to show the relation to MSTs.

	ADS.DE	BEI.DE	LHA.DE
BEI.DE	0.29		
LHA.DE	0.43	0.30	
LIN.DE	0.44	0.31	0.45

	ADS.DE	BEI.DE	LHA.DE
BEI.DE	1.188		
LHA.DE	1.063	1.186	
LIN.DE	1.060	1.175	1.046

	ADS.DE	BEI.DE	LHA.DE
BEI.DE	1.175		
LHA.DE	1.060	1.175	
LIN.DE	1.060	1.175	1.046

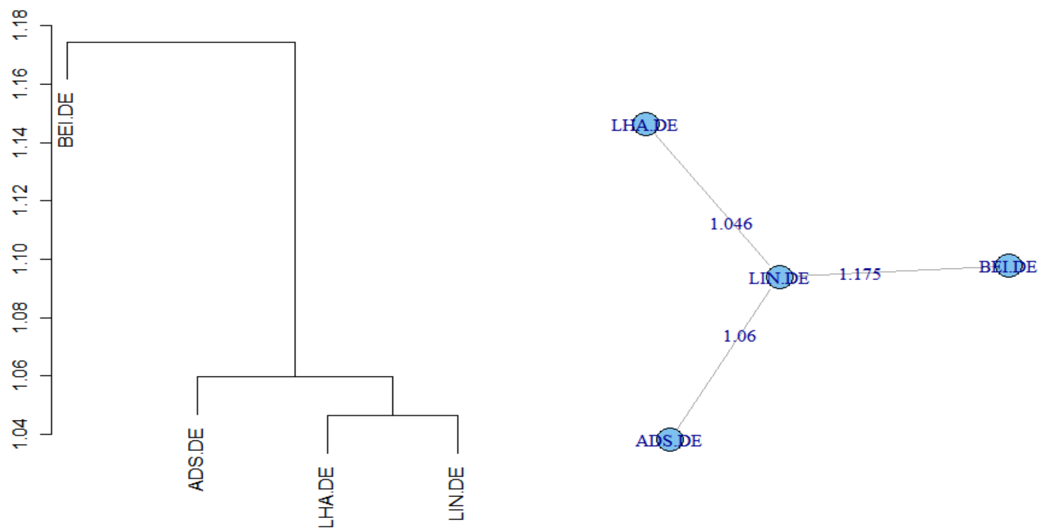

Figure 3-9: Building the MST showing the relation to the single linkage clustering.

Using the single linkage clustering on the left hand side of the figure the dendrogram helps to construct the MST stepwise according to the following procedure:

At each step (defined by a mark on the height of the dendrogram) the linkage algorithm merges together two clusters (or two points or a point and a cluster). A link is then added to the emerging tree that connects the new node(s) to where the distance is the smallest (minimum distance; single linkage). The procedure stops at the topmost level of the dendrogram when all nodes/links have been added.

Properties and metrics of MSTs

Often related methods to the MST are spectral analysis and Random Matrix Theory (see for example Heimo et al. 2007). The MST is full of information concerning the backbone dependence structure of the financial assets. There are numerous graph-based measures quantifying and visualizing the backbone dependence structure of the assets in static and dynamic context. The fundamentals of these metrics are the edge lengths or paths and the distribution of edge degrees.

Real life networks for example form hard or social sciences often exhibit a scale-free structure meaning that their edge degree distribution of the network nodes has power-law tail behaviour with tail index between 2 and 3. This indicates that there is a huge amount of nodes with very few degrees and a small amount of nodes with high edge degree making the edge degree distribution fat-tailed. The following power law distribution describes this phenomenon

$$P(k) \approx ck^{-\gamma},$$

where $P(k)$ is the fraction of nodes in the network having k connections, γ is the tail exponent, and c some constant.

Also, the moments of the asset correlations can be compared to the edge weights of the MST. The mean correlation is computed as

$$\bar{\rho} = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij} .$$

The average edge length L of the MST which is also called normalized tree length is defined as:

$$L(MST) = \frac{1}{N-1} \sum_{d_{ij} \in E(MST)} d_{ij} ,$$

where only those d_{ij} of the amount of MST edges $E(MST)$ are averaged. The average edge length of the MST can be transformed into the average of the MST correlations transforming it back to a correlation by the correlation distance measure:

$$Avg\_Corr(MST) = 1 - L(MST)^2 / 2 \,.$$

Each node or vertex can be characterized by its relative position in the network. For example, nodes can be rather central or rather eccentric concerning the network topology. An example measure for centrality is the “betweenness centrality” of the network:

$$C_B(v) = \sum_{i \neq v \neq j \in V(MST)} nsp_{ij}(v) / nsp_{ij}$$

with nsp_{ij} (v) as the number of shortest paths from node i to j going through node v and with $V(MST)$ as the quantity of nodes of the MST. Other examples for centrality and eccentricity measure definitions can be found in financial applications in Aste et al. (2010), for example.

Visualization

There are forced-based algorithms like the one by Fruchterman/Reingold (1991) or Kamada/Kawai (1989) which draw graphs in an aesthetically pleasing way. Their purpose is to position the nodes with nearly equal length and as few crossing edges as possible. These algorithms assign physical forces among the edges and nodes: edges could be springs and nodes could exhibit electric repulsions, for example. The layout algorithm simulates the different forces at work and it is terminated when some mechanical equilibrium step is reached.

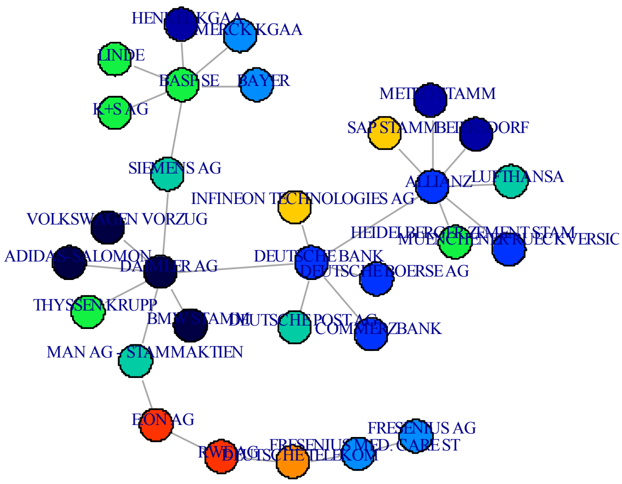

Figure 3-10: The MST of the DAX (2003-2010) with industry colour codes.

Community detection

In the network it is interesting to identify dense connections between the nodes within modules but sparse connections between nodes in different modules. For this reason there are community/module detection algorithms that reflect the concentration of nodes within modules compared with a random distribution of links between all nodes regardless of modules (random null model as a randomized realization of a particular network which is used as a reference and often exhibits the same edge degree distribution). An example is the community detection algorithm by Newman/Girvan (2004) which optimizes the measure of modularity. It is defined as the fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random. It is defined between [-1,1] whereas it is positive if the number of edges within groups exceeds the number expected on the basis of chance. Newman/Girvan (2004) iteratively remove edges (esp. weak links) from the original graph and measure modularity at each step. The partition with the highest modularity is chosen and defines the community structure. In graph visualization it seems a good idea to combine force-based layout algorithms with colour coded nodes referring to the community membership.

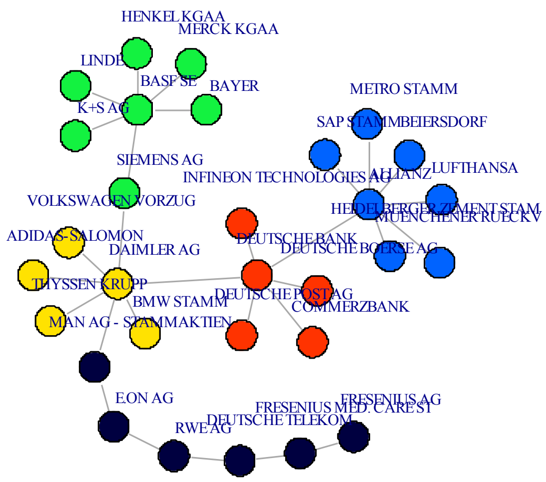

Figure 3-11: The MST of the DAX (2003-2010) with community colour codes.

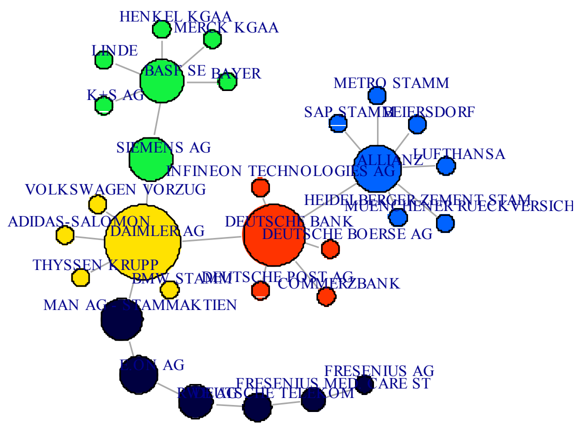

Figure 3-12: The MST of the DAX (2003-2010) with community colour codes and node sizes corresponding to edge degrees.

Dynamic MSTs

Observing the MST topology at different time snap shots by using a fixed moving calibration window reveals certain dynamics of the backbone dependence structure. Measures like average tree correlation or network centrality evolving in time as well as time dynamic merges of communities deliver new insight to the market dynamics. For example, measuring the relative centrality of one sector in comparison to the others and plotting this measure in time shows how industry sectors change their role in the structure in time.

There are special measures like edge survival or T1-distance (used for example in Aste et al. 2010) that focus on the edge rewiring of the snapshot MSTs in time. The one step edge survival is defined as

$$ES(1,t) = \frac{1}{N-1}\left|E^{t} \bigcap E^{t-1}\right|,$$

where E^{t} refers to the set of edges of the MST at time t and the $\bigcap$ operation intersects the two MSTs. So it is simply the fraction of surviving edges to all edges. This measure plotted in time indicates if there are major edge rewirings due to market disruptions.

The T1 measure compares formerly first neighbours in time t-1 with the path length between them in time t. For example, formerly first neighbours in t-1 are second neighbours in t, then T1 = 1 for this single edge. We average the T1 distances across all edges and use it as a dynamic measure of structural breaks.

Applications of networks in portfolio context

Correlation structures in complex systems can alternatively be analysed by means of threshold methods which will keep only the strong interactions and disconnecting the part that is less correlated than the threshold from the system. A graph-filtering method can instead describe the entire structure by also keeping some weaker correlations but simultaneously filtering out redundancies in the highly correlated part (Onnela et al. 2003). However, it has to be stated that the number of weak correlations is not too high when forcing a correlation matrix into an MST. This aspect is closely connected to the robustness, significance and validity analysis of MSTs which we will show below.

The MST of a portfolio of assets exhibits properties which are very valuable for portfolio optimization and risk management. First, there are graph measures characterising the global function of each node in the network. For example, there are nodes in the MST which are more central or more peripheral than others, or few nodes exhibit a large number of connections whereas others are very isolated. Second, there are numerous algorithms to detect communities in graphs. Combining both methods conveys important relations in portfolios and the graph measures and communities can be functionally visualized.

One of the most popular quality functions for community detection algorithms is modularity which measures how well partitions of a network are. For example, in a first step, communities are detected by progressively removing edges from the original graph (Newman/Girvan 2004). The end result of the Girvan–Newman algorithm is a dendrogram. In a second step, the dendrogram is cut at a position to optimize the modularity measure.

Correlation based networks have been found very useful in the elucidation of economic properties of stock returns traded in a financial market. In finance literature there are already some approaches to portfolio management based on correlation based networks. For example, there are approaches to analyze the location of Markowitz optimal portfolios on MST-layouts and to determine the general diversification potential of a market (Onnela et al. (2004). These approaches define a central node based on some centrality measure like edge degree and each asset then exhibits a distance level from this “centre of market gravity”. The levels of each asset can then be weighted either by the same weights in order to compute the mean occupation layer or they can be weighted by Markowitz’ or Minimum Variance weights in order to receive the weighted portfolio layer. An important observation of these tree measures is that the different branches of the tree often belong to different business sectors and it can be conjectured that the location of assets in the tree and in a tree branch are important for risk management and diversification.

A non-local measure of centrality can be inferred by computing the number of shortest paths that pass through a given node: the larger the number of paths, the more central the node. It is called edge betweenness. An opposite kind of measure is the eccentricity, which is the largest distance between a node and any other node in the graph. Some papers also focus on the dynamics of market correlations and analyze structural changes in the market by means of filtered graphs (see overview of these papers below). For example, the time evolution of the system is continuously monitored by repeatedly constructing MSTs from moving fixed time windows. It can be observed that some edge connections remain stable in time and some are loose. Also, the community structure changes over time as well do the centrality and peripherality measures. This can be explained when correlation is understood as a dynamical concept.

An example is Onnela at al. (2002) who constructs MSTs sampled at different points in time and find topological shrinking and strong reconfiguration of the trees during stock market crises. They demonstrate that the assets of the optimal Markowitz portfolio lie practically at all times on the outskirts of the tree and that different tree branches reflect diversification potential. They also show that the tree topology and the general investment diversification potential coincide. Tumminello et al. (2007) investigate correlation based networks of equity returns sampled at different time horizons ranging from 5 minutes up to one trading day. Their analysis confirms that the selected stocks compose a hierarchical system progressively structuring as the sampling time horizon increases and that the cluster formation can be quantitatively associated to economic sectors. Di Matteo et al. (2010) use dynamical networks to detect the hierarchical organization of financial market sectors. Their analysis is based on measures for centrality and peripherality like edge degree, betweenness, eccentricity and closeness. Aste et al. (2010) construct topologically constrained graphs from cross-correlation matrices and report significant statistical signatures of the 'credit crunch' financial crisis that unfolded between 2008 and 2009. They test the stability, statistical significance and economic meaningfulness of these graphs evolving in time. The results show an intriguing trend that highlights a consistently decreasing centrality of the financial sector over the last 10 years. Song et al. (2011) analyse the evolution of worldwide stock markets by means of correlation based graphs. They discover that the correlation among market indices presents both a fast and a slow dynamics. The slow dynamics reflects the development and consolidation of globalization. The fast dynamics is associated with critical events that originate in a specific country or region of the world and rapidly affect the global system. They define a measure for link co-occurance in order to detect the graph dynamics in a quantitative way.

4 New Approaches to Equity Portfolio Management

We now introduce a new approach to portfolio optimization that addresses the diversification problem by means of clustering. Stocks of the German index DAX are grouped according to a nested hierarchical structure that is solely explored from the daily stock return data without any a priori model assumptions.[5] The automatic classification can be verified and tested based on hypotheses. Therefore, it is not necessary to employ a model to group stocks according to industry sectors, factors or other classification approaches which are often subject to model risk.

This high quality classification is then used to derive a diversified portfolio in a way that portfolio weights are balanced across the clusters. We test the risk-adjusted performance of the diversification scheme against other well-known portfolio construction approaches like mean variance optimization. The result is that the cluster diversification outperforms even though it has no information about the stocks' individual risk/return profile. The comparisons consider investment constraints, trading costs and individual investor preference. Based on the good performance level of the diversification scheme, arbitrary estimation approaches of the stocks' risk/return characteristics can still enter the model in order to further increase risk-adjusted performance. This application has important potential implications on portfolio construction. Our approach is unique in how the properties of hierarchical clustering are used for diversification purposes. It is very straight forward in solving the diversification problem, is very tractable and transparent from an investment and visualization perspective, and is economically viable due to the hierarchical clustering representing data-driven economic sectors.

---

[5] Data source: Yahoo! Finance - Business Finance, Stock, Market, Quotes, News, www.yahoo.com/finance

4.1 Overview of Clustering in Portfolio Management

Mean-variance efficient portfolios are engineered to deliver the highest return for a given the level of risk. Modern portfolio theory (MPT) argues that the mean-variance optimization (MVO) portfolio is engineered to achieve the best diversification for a given investment universe subject to the returns and risk forecasts as inputs. The traditional mean-variance approach uses expected returns and covariance estimates as inputs, while weights and the resulting risk contributions of the optimal portfolio are the output; risk contributions simply reflect the inputs to the optimization.

However, these desirable properties in portfolio optimization cause some practical problems. Estimation of expected returns and covariance is rather difficult due to the lack of data and according to short estimation horizons the covariance matrix is often very noisy (Laloux et al. 1999). Also, it is even more difficult to estimate means than covariances of asset returns (e.g. see Merton 1980) and MVO techniques are known to be very sensitive to differences in expected returns so they actually “maximize"” estimation error.

A portfolio construction that completely circumvents the parameter estimation problem is the equally weighted (EW) portfolio. It is probably the simplest portfolio construction approach in an attempt to achieve diversification and it does not need any optimization. In terms of MVO, these portfolios would imply assets with identical expected returns and volatilities as well as zero correlations. However, these rather unrealistic assumptions may achieve an astonishingly good out-of-sample performance when simply used as an “investment strategy”'. In their extensive study of different asset allocation models applied to several datasets of the global equities universe, De Miguel et al. (2009) find that none of the theoretically sound asset allocation approaches (MVO and modifications) are consistently better out of sample than the heuristic 1/N equally weighted rule. They explain that “allocation mistakes” caused by using the 1/N weights can turn out to be smaller than the error caused by using the weights from an optimizing model with inputs that have been estimated with error. Kritzman (2010), however, point out that the dataset chosen may be beneficial to the 1/N strategy. Also, its diversification effects are portfolio dependent and not at all guaranteed. Finally they conclude that the longer the estimation horizons the more potential for the MPT to beat the 1/N strategy on a range of datasets. Also, Martellini (2008) finds evidence that improved moment estimators may outperform equally weighted schemes. Consequently, the EW portfolio is more of a performance benchmark rather than an investment strategy.

A recent academic research has focused on minimum-variance portfolios, which rely solely on estimates of covariances and thus are less vulnerable to estimation. Examples are the latest approaches in risk parity and diversification from Choueifaty/Coignard (2008) and Maillard et al. (2010). However, some authors raise several conceptual and practical concerns. An example is Lee (2011) who concludes that there is no theory to predict, ex-ante, that any of the risk-based approaches should outperform. Also, these approaches are only optimal if and only if one assumes that all stocks have the same Sharpe ratios (and the same pairwise correlations) which seems rather unrealistic from an empirical point of view.

At present, many authors attempt to put the focus back on the only truly optimal weighting scheme consistent with modern portfolio theory (see for example Martellini 2008) and they ague that using improved robust estimators for the variance-covariance matrix of stock returns, as well as for their expected returns, might rehabilitate MPT. However, strong assumptions such as normally distributed asset returns need to be made before one can conclude that they are truly optimal portfolios. As a result, portfolio managers need to consider whether assumptions and objectives behind each concept are compatible with their views and needs before employing MPT to construct portfolios.

Criticism of the normality of asset return assumption was first empirically challenged by Mandelbrot (1963) and subsequently supported by Fama (1963)

Also, risk in general is a rather abstract term for financial loss potential and (co-)variance in the MVO may not be an adequate measure for risk and diversification. For example, variance is not a downside risk measure and correlation implies only linear dependence due to the normal distribution assumption. For these reasons, models incorporating stylized facts of empirical asset returns such as excess kurtosis, skewness, volatility and copula tail dependence effects were developed by Rachev/Mittnik (2000) and Rachev et al. (2007 and 2008). Although the parameters of these models are difficult to estimate, efficient parameterization methods have been recently developed so as to make implementation of these models practical. An example of such an approach and its application to credit risk management in general and for price calibration and hedging of correlation dependent credit derivatives is Papenbrock et al. (2009).

Existing literature on clustering in portfolio management often categorize assets by means of clustering and using these clusters to create portfolios. Examples are Zhang/Maringer (2010) who propose a clustering criterion which groups market assets to maximize the Sharpe ratio of portfolios and Dose/Cincotti (2005) who limit the number of stocks in a portfolio by considering a subset derived from clusters and then setting the weight of each stock as a result of an optimization process.

Other authors think of clustering as a means of filtering and improving parameter estimation. An example is Tola et al. (2008) where the number of correlation coefficients in the matrix is reduced by a hierarchical clustering approach. This can be seen as a filtering procedure in which the number of distinctive elements in the correlation matrix is reduced. The resulting ultrametric and original metric correlation matrices are then used to build the portfolio in a Markowitz approach. The authors find that the cluster-based filtering shows better results than RMT filtering.

Since the beginning of this millennium, hierarchical clustering is shown in many studies about financial markets to be able to explicatively reproduce some economic structure (see e.g. Mantegna 1999 and Tumminello et al. 2005) and is able to deliver some theoretical description of financial markets. The gathered hierarchically embedded set of clusters can be visualized according to the tree layout and this renders possible meaningful economic interpretations of the clusters as well as economic plausibility checks of the unsupervised learning of hierarchies. Tumminello et al. (2010) for example, find that hierarchical clustering is able to detect clusters of stocks belonging to the same sectors or sub-sectors of activities without the need of any supervision of the clustering procedure and they develop a nested factor model based on hierarchical clustering.

We pursue a similar approach but without the necessity to derive nested factors. Rather, the hierarchical clustering is directly used for diversification by a mechanism that distributes capital weights evenly across the hierarchically nested clusters. The necessary hierarchically nested structure is automatically explored by standard clustering algorithms so it is not necessary to construct a model for the hierarchical structure of sectors which may be subject to additional model risk. The straightforward portfolio construction based on the explorative clustering can be seen as a benchmark for diversification and this information. It can be combined with knowledge of the stocks' individual risk/return profiles whereas the single stock risk and return prospects can be derived from any arbitrary model. For example, more coherent risk measures than variance can be used here.

Stocks with favourable risk/return profiles can be picked and at the same time the diversification scheme indicates which stocks are very isolated far away from the bulk of stocks. These should ideally receive a higher weight as they are contributors to diversification. The whole process can be visualized based on the hierarchical tree layout. We analyze in detail different hierarchical cluster algorithms in order to find adequate set-ups for portfolio management. For convenience, we focus on the correlation–based distance but it has to be remarked that a variety of linear and non-linear (dis-)similarity measures could be used here. The variety of combinations from similarity measures and hierarchical clustering procedures not only clear the way for adequate diversification processes but also for meeting investment constraints and individual investors' preferences.

It has to be stated that explorative clustering is basically an intelligent reduction of information and complexity. Unfortunately, it seems that in many papers the validation of the clustering structures is either non-existent or still leaves many questions unanswered. The difficulties arising from processing large datasets and deriving clustering structures without proper validation may explain the limited use in practice. As Jain/Dubes (1988) and Tan et al. (2005) note, clustering algorithms will always find clustering structures - no matter whether they are existent or not so there is always a need for cluster validation and cluster stability tests.

4.2 The Cluster Based Waterfall Approach

We proposed a weighting scheme that distributes capital weights according to the tree layout and its nested structure as can be seen in the following graph:

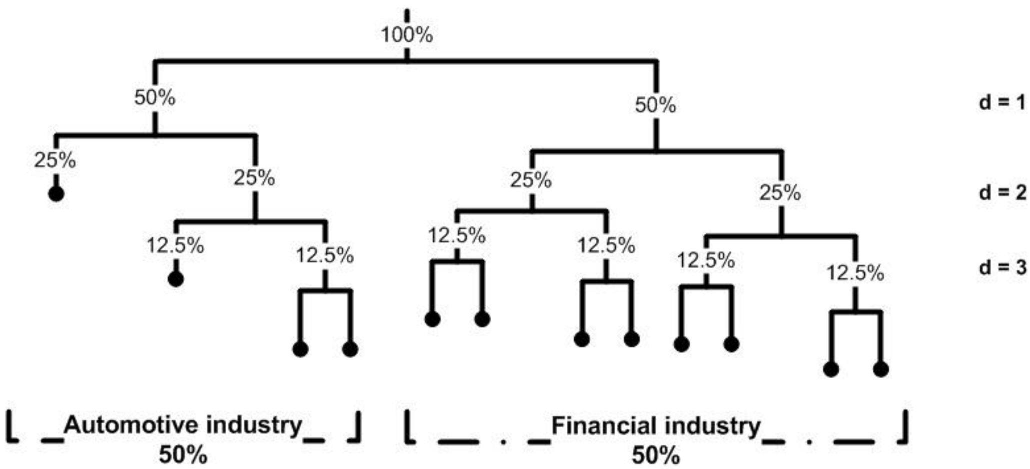

Figure 4-1: The cluster-based waterfall weighting scheme.

Step 1: split the capital invested into two equal halves at the first bisection in the tree at d=1.

Step 2: split the remaining 50% capital of each branch at its next bisection.

Step 3: continue step 2 for each branch until there are no more bisections.

The hierarchically nested structure is automatically explored by standard clustering algorithms so it is not necessary to construct a model for the hierarchical structure of sectors which may be subject to additional model risk. The diversification scheme identifies stocks which are very isolated and just surrounded by small clusters far away from the bulk of stocks. These should ideally receive a higher weight as they are contributors to diversification.

The chaining effect in the single linkage clustering may result in very high weights of single stocks. In contrast, the naïve strategy gives weights with the same investment fraction. In between these two extremes there are several hierarchical clustering procedures with respective waterfall weights:

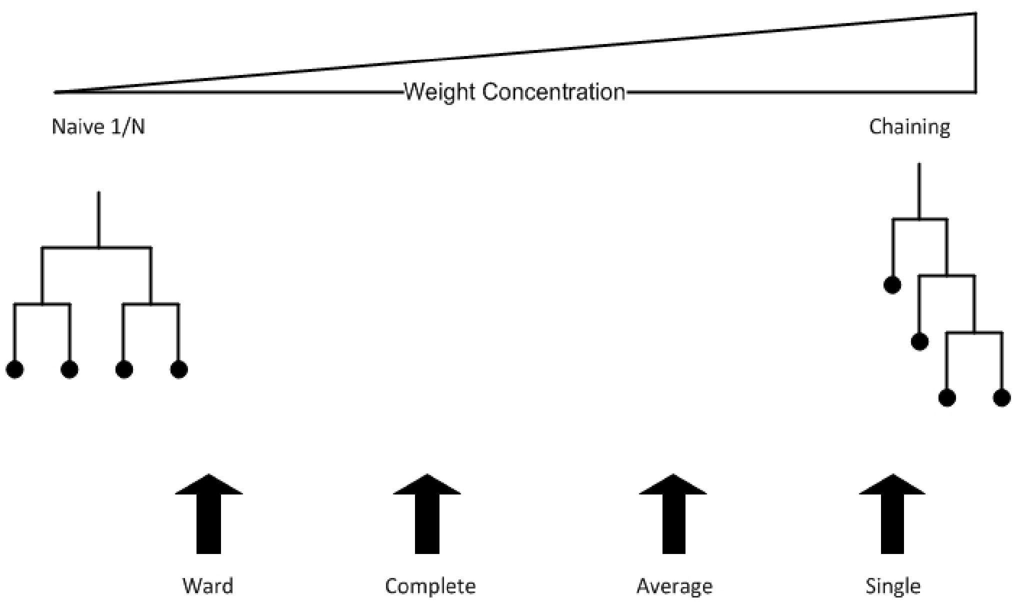

Figure 4-2: Varying the weight concentration by using trees with different symmetry properties in the cluster-based waterfall approach.

Ex-post Analysis of Risk and Performance

An "ideal risk metric" is not used in the portfolio construction process of the hierarchical weights strategy. The aim is to "diversify" investments. However, as investors are interested in the return they get and are either bound to certain risk thresholds or may need to measure an investment strategy in terms of risk and ex-post risk-adjusted performance. The following section shall thus give a short overview of the risk-metric and performance measures that were used to evaluate the strategies ex-post.

There is vast literature on the topic of risk measures and risk metrics. The properties, advantages and disadvantages of those shall not be part of the discussion of this work. Several measures of risk that have become popular in the financial industry will be used for the analysis of return series.

β -S**ystemic risk:** The systematic risk of a strategy is expressed with β . For a sufficiently large set of assets, the idiosyncratic risks are diversified away and only the systematic risk remains. It is the basis of the first asset-pricing model derived from economic theory and was subsequently used in the one-factor CAPM-model. It can be empirically determined:

$$\beta_i = \frac{cov(R_i, R_M)}{\sigma_M^2}$$

and

$$\beta_P = \beta_i \times w_i$$

In this context, $\text{cov}(R_i, R_M)$ denotes the covariance between an asset return R_i and the market index return R_M . The term σ_M^2 is the variance of the market index return and w_i the weight of the respective asset in the investment strategy. The market index will be the DAX.

Variance: As already mentioned, using the variance as a measure to assess risk can be quite problematic. It is nevertheless often used, especially because it is closely tied to the MVO-approach of Markowitz. The annualized portfolio standard deviation can thus easily be calculated. It is assumed that one year has 260 trading days:

$$\sigma_P = \sqrt{\sigma_P^2} * \sqrt{260}$$

Value-at-Risk: Value at Risk (VaR) measures the worst loss that can be expected to a certain confidence level α. The confidence level for the daily loss is usually set to 95% or 99%. In this case, VaR will only be calculated ex-post. There are generally two ways to determine the VaR from a given dataset: Either the 95th or 99th percentile of a sufficiently large ordered dataset is selected, or a distribution is determined for which the VaR is determined accordingly. In this work the 99th percentile is selected from historical data, as more than 1,300 data-points exist for the 6-year ex-post analysis.

CVaR: As VaR and variance were under harsh criticism for their shortcomings with respect to risk measurement in portfolio management, the more meaningful Conditional Value at Risk (CVaR) is oftentimes used for this task.[6] CVaR is a significant improvement over VaR, since the expected loss beyond the confidence level is also included in this metric. Furthermore it satisfies all properties of a coherent risk measure (including sub-additivity,

[6] In a section below dealing with credit risk, the CVaR will be termed the VaR for credits and is not be mixed up with Conditional Value-at-Risk which is also named Expected Shortfall (ES) or Average VaR (AVaR).

which is violated by VaR). CVaR can thus better capture diversification effects in a portfolio. It is defined as:

$$CVaR_{\epsilon} = \frac{1}{\epsilon}\int_0^{\epsilon} VaR_p dp$$

Even though two time series may have the same VaR at a given confidence level, the one may be much riskier than the other because of the heavier tail of its distribution (or lack of diversification with respect to heavy tailed assets).The advantages of the CVaR risk measure over the VaR become apparent when the underlying distributional model is skewed. This also holds true for the advantages of CVaR over the variance or standard deviation.

Maximum DrawDown: The maximum drawdown measures the decline from a historical peak of a time series. The analysis furthermore includes the Maximum DrawDown per month as an indicator of what an investor had to deal with in terms of maximum loss during any given month of the respective investment strategies.

Number of negative months: The number of negative months gives an idea, in how many months the strategies would have destroyed value.

Average loss per month: The average incurred loss at the end of every month bearing a loss is calculated.

Measuring Performance

The importance of the ex-post valuation of the performance of an investment strategy was already pointed out. Several performance measures exist for this task. The goal of this work should not be to find a suitable performance ratio or risk-adjusted performance measure, but shortly describe the most widespread measures, which are then subsequently used to evaluate the respective investment strategies.

Annual Geometric Return: The return of the investment strategies on a yearly basis are not computed using the arithmetic mean, as the returns are compounding, but with the geometric return. The annual geometric return of a strategy is computed as:

$$Return = (\frac{TerminalValue}{InitialValue})^{\frac{1}{Years}}$$

Sharpe Ratio: The Sharpe ratio is used to relate the excess return of an investment strategy measured against its total risk exposure. The total risk is measures in terms of standard deviation. Higher Sharpe ratios are usually preferable to lower Sharpe ratios. It is often used in the MVO-optimization, where the maximization of the Sharpe ratio may be the goal of the optimization. The resulting portfolio would tangent the MV-efficient frontier. The Sharpe ratio is calculated as:

$$SR = \frac{r_p - r_F}{\sigma_P}$$

In this context, r_P denotes the annual return (geometric mean), r_F denotes the risk free rate and σ_P is the standard deviation of the portfolio.

Modigliani Risk-Adjusted Performance: Whereas the Sharpe Ratio is not to be considered a risk-adjusted return measure, the Modigliani Risk-Adjusted Performance is measuring the risk-adjusted returns of a portfolio against a benchmark. Whereas the Sharpe Ratio is dimensionless, the M^2 is measured in units of return. It is more intuitive to interpret. It is computed in the following way:

$$M^2 = r_p + SR \times (\sigma_M - \sigma_P)$$

From the equation, it immediately becomes obvious, why the Sharpe Ratio is only a ratio, whereas the M^2 is a true risk-adjusted performance measure: The relative performances of portfolios are calculated with respect to the market risk. The return of the portfolios that are to be compared is thus calculated on the assumption that the portfolios would carry the same risk as the market. It is therefore relatively easy to infer, which portfolio is superior.

Treynor Ratio: It is very similar to the Sharpe Ratio with one important difference: Whereas the SR measures risk in terms of standard deviation (total risk), the Treynor Ratio measures risk in terms of systemic risk or market risk β It thus ignores specific risk and is only useful as a performance ratio, if the idiosyncratic risks of a portfolio are "diversified away". It is computed in the following way:

$$TR = \frac{r_P - r_F}{\beta_P}.$$

In this context, r_P denotes the annual portfolio return (geometric mean), r_F denotes the risk free rate and β_P is the systemic risk or market risk of the portfolio.

Jensen's α : The economic interpretation of Jensen's α is the intercept of the portfolio to the CAPM-line. Jensen's α thus measures the over- or underperformance of a portfolio with respect to its systematic risk compared to the market. It is oftentimes used by fund managers to express their superior performance over the market - potentially totally neglecting the idiosyncratic risks of their investments. Ignoring any error terms, Jensen's α is calculated as:

$$\alpha = r_p - r_F - \beta_P \times (r_M - r_F)$$

Results of the cluster-based waterfall investment strategy

In order to test this simple strategy, we implement a rolling window investment strategy with monthly rebalancing and estimation windows ranging from 130 to 260 and 390 trading days. We operationalize the waterfall strategy for the four clustering procedures and compare their performance to the DAX benchmark, to the Markowitz portfolio (with no short-selling) and to the naïve 1/n strategy. The following table compares the different strategies for the 130 days estimation window whereas performance, risk, and turnover are expressed in terms of the usual realized measures such as return per annum, standard deviation, maximum draw down, Jensen's α , Sharpe ratio and average monthly weight change:[7]

	Return p.a.	σ	Max DD	Jensens α	Sharpe	Aver. mon. weight change
Sing-Corr	20.2	25.5	66.3	13.4	0.67	34.2
Aver-Corr	16.9	20.9	62.8	9.9	0.65	26.9
Compl-Corr	16.9	22.5	64.8	9.8	0.61	25.3
Ward-Corr	13.2	21.9	63.2	5.8	0.45	34.6
Markowitz	8.7	18.5	56.7	2.4	0.18	30.5
1/N	13.5	23.5	61.5	5.5	0.43	14.9
DAX	8.3	23.2	58.7	0	0.22	-

Figure 4-3: The result of the different investment strategies based on 130 days of in-sample-data for the period of 2005-2010.

	Return p.a.	σ	VaR 99	CVaR 99	Max DD	Jensens α	Sharpe	Weights change (mo.)
Sing-Corr	20.3%	26.7%	-2.6%	-6.7%	67.2%	13.5%	0.64	23.6%
Aver-Corr	14.6	25.3	-2.5	-6.4	71.1	7.8	0.45	26.9
Compl-Corr	15.5	23.3	-2.3	-6.4	68.7	8.4	0.53	31.5
Ward-Corr	18.5	21.4	-1.9	-6.1	56.3	11.2	0.71	26.9
Markowitz	11.9	18.2	-1.6	-6.1	56.9	5.6	0.48	19.7
1/N	13.5	23.5	-2.0	-6.9	61.5	5.5	0.43	14.9
DAX	8.3	23.2	-2.0	-6.4	58.7	0	21.7	-

Figure 4-4: The result of the different investment strategies based on 390 days of in-sample-data for the period of 2005-2010.

[7] Results and especially rankings are similar when using different sampling horizons, transaction costs and weight constraints.

Performance	Single-Corr	Ward-Corr	Markowitz	DAX	1/N
Return p.a.	20.3%	18.5	11.9	8.3	13.5
Pos. Months	45	48	44	42	44
Aver. Monthly Gain	3.9%	2.9	2.3	2.4	2.8
Best Month	39.1%	18.8	11.6	15.5	20.9
Risk	Single-Corr	Ward-Corr	Markowitz	DAX	1/N
σ p.a.	26.7%	21.4	18.2	23.2	23.5
VaR 99	-2.6%	-1.9	-1.6	-2.0	-2.0
CVaR 99	-6.7%	-6.1	-6.1	-6.4	-6.9
Max DD	67.2%	56.3	56.9	58.7	61.5
Max DD in a Month	23.9%	30.5	29.2	27.6	34.9
Number of Neg. Months	26	23	27	29	27
Aver. Monthly Loss	2.0%	1.5	1.4	1.8	1.9
Worst Monthly Loss	17.1%	17.1	16.9	15.1	20.8
β	0.72	0.80	0.61	1.00	0.92
Risk-adjusted Perf.	Single-Corr	Ward-Corr	Markowitz	DAX	1/N
Sharpe Ratio	0.64	0.71	0.48	0.22	0.43
Treynor Ratio	0.24	0.19	0.14	0.05	0.11
Jensen's α	13.5%	11.2	5.6	0.0	5.5

Figure 4-5: A more detailed overview of several performance, risk and risk-adjusted performance related numbers and ratios. No Trading costs are assumed, no additional asset constraints are imposed. The in sample estimation of parameters has a length of 390 days of trading data.

It can be stated that the Single-Linkage and Ward-Linkage hierarchical weighting strategies show superior performance with respect to the dimensions return and risk-adjusted performance. The concentration of asset weights is very different for both strategies. Ward-Linkage will be most helpful for investors seeking high risk-adjusted performance and well balanced portfolios with little concentration on individual assets. The Ward- Linkage strategy is to some extent comparable to the 1/N approach but has the advantage that weight concentrations in sectors are generally avoided. The Single-Linkage strategy shows the highest performance and favours assets showing little correlation to the other assets in the asset universe. This strategy is only helpful for investors willing to take high stakes in single assets. The following graph shows a typical performance of the Single-Linkage clustering (blue) in comparison to Markowitz (red), 1/n (green) and the DAX (black):

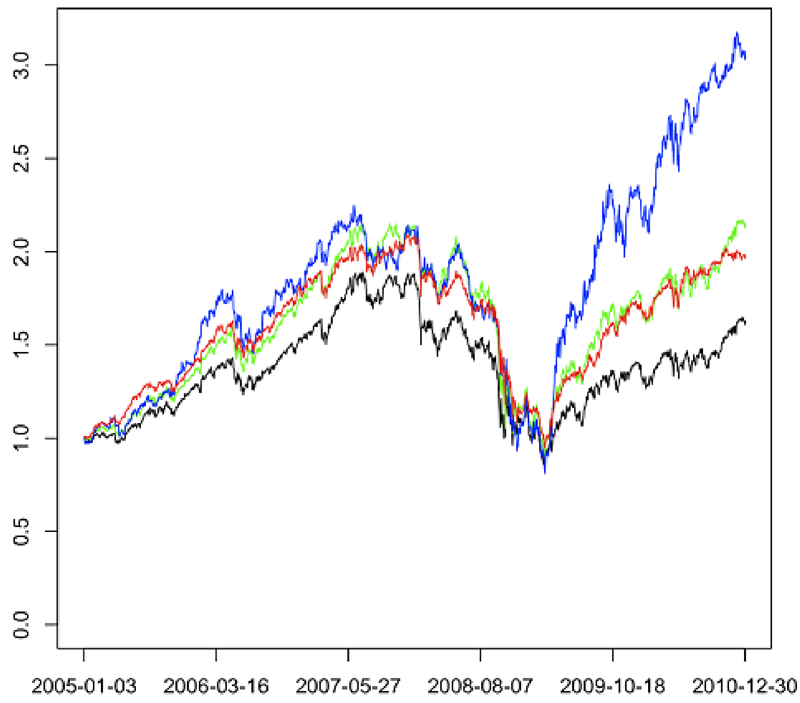

Figure 4-6: The Single-Corr approach (blue) does show better risk-adjusted performance than 1/N (green), Markowitz (red) or a simple investment in the Benchmark (DAX, black).

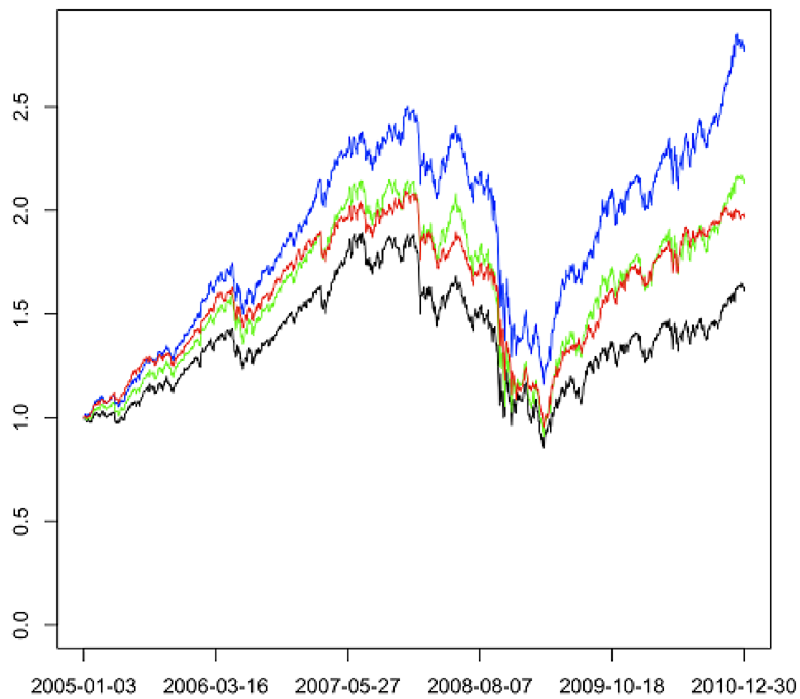

Figure 4-7: The Ward-Corr approach (blue) does show better risk-adjusted performance than 1/N (green), Markowitz (red) or a simple investment in the Benchmark (DAX, black).

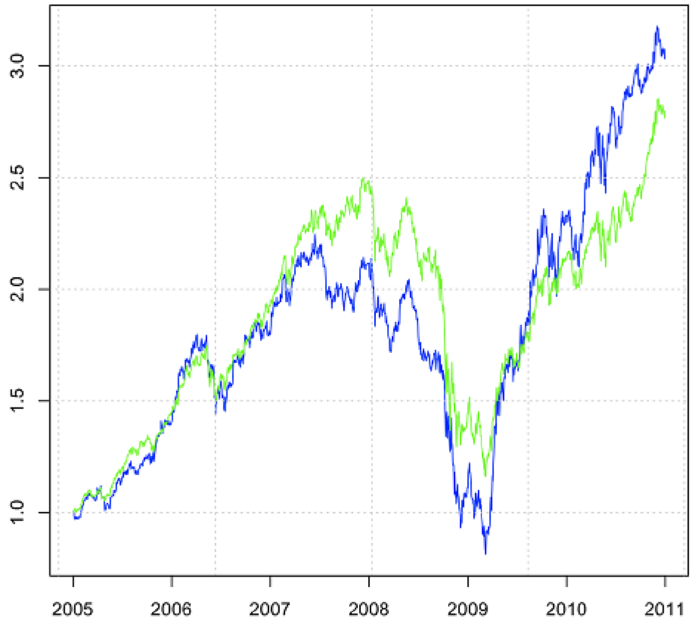

Figure 4-8: The two strategies over time. Sing-Corr is depicted in blue and Ward-Corr in green.

Finally the next three figures show the weight distributions in time of the single linkage, the Ward clustering, and the Markowitz strategy.

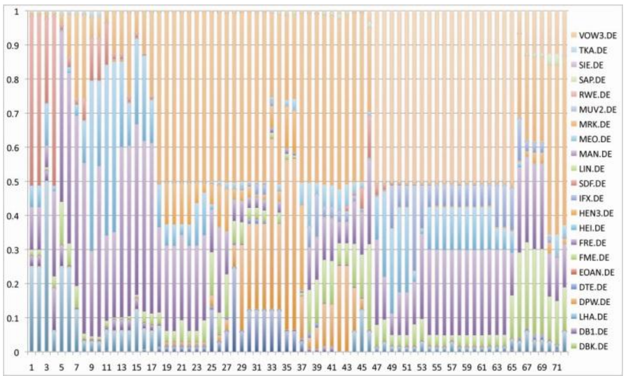

Figure 4-9: Visualization of the weights changing in time using single linkage clustering.

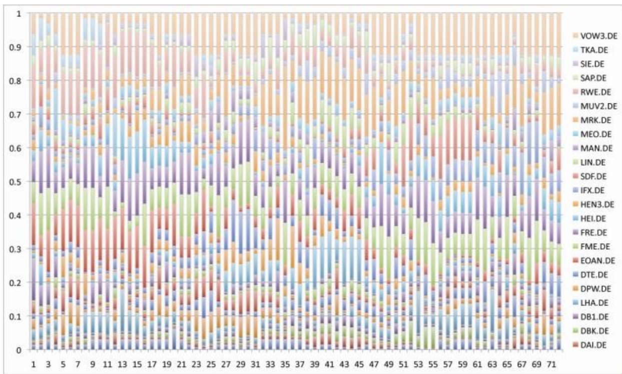

Figure 4-10: Visualization of the weights changing in time using Ward clustering.

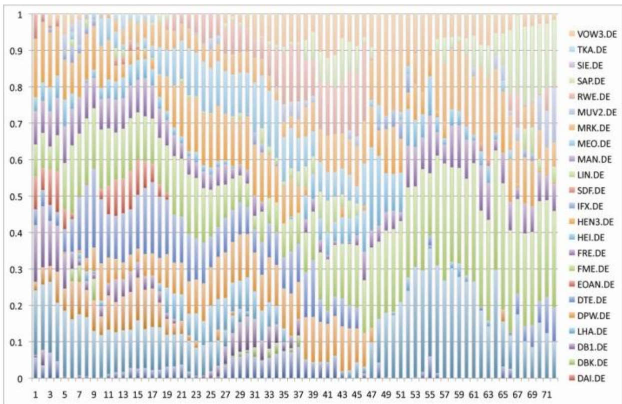

Figure 4-11: Visualization of the weights changing in time using Markowitz.

Conclusions of the Asset Clustering Approach

The following advantages can be attributed to the clustering approach:

- Proof of concept for data-driven diversification strategy
- Superior return and risk-adjusted performance based on a simple and weighting heuristic
- Implementation complexity is low
- Different investor choices among the linkage types of the clustering algorithms are possible in order to implement different strategies
- The results are just based on the correlation coefficient so that individual risk measures like variance or reward measures like expected return can still enter the model

The following disadvantages can be observed:

- The waterfall rule is static and there is not much room for weight adjustments (e.g. investment constraints)
- Some clustering algorithms tend to suggest extreme weighting schemes and especially the single linkage clustering tends to emphasize extreme market phases
- Turnover rates can be higher than in benchmark models
- Dendrogram heights have not yet been incorporated

Finally it has to be stated that the parameter estimation horizon is just chosen arbitrarily and that the validation of clustering has not been carried out. In the next step we will show a similar approach based on networks which exactly deduces an ideal estimation window size and delivers statistically validated results. It is related to the clustering approach and uses networks for several data mining tasks.

4.3 The Asset Network Approach

In this section we will start with an analysis of the correlation dynamics of the DAX with the help of evolving MSTs. In the time period under investigation, the DAX exhibits slow and fast correlation dynamics as can be seen from the following upper figure where average DAX correlations are plotted in time (x-axis) using different parameter estimation windows (y-axis), whereas high average correlation is red and low average correlation is yellow:

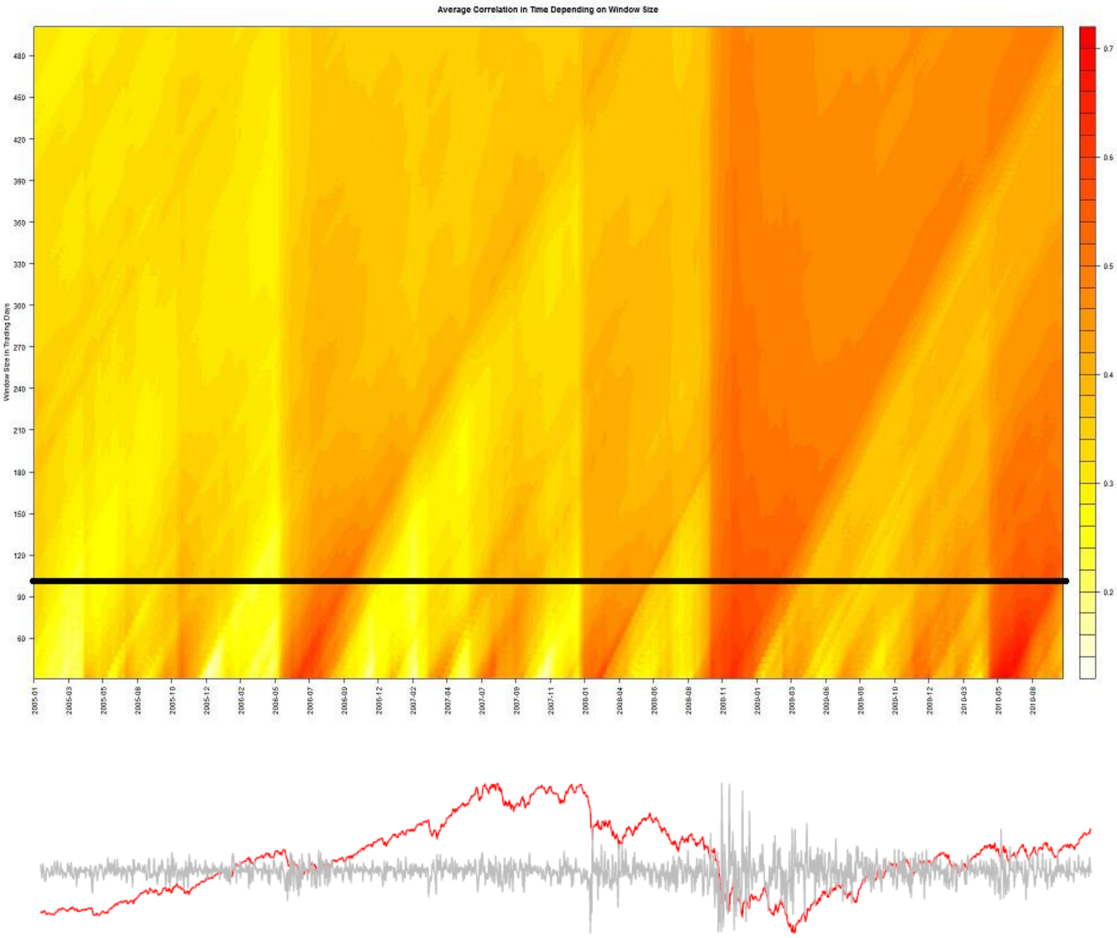

Figure 4-12: Average correlation of the DAX from 2005 to 2010. High correlations are red and the y-axis represents different parameter estimation windows ranging from short to long. Below is the DAX index curve (red) with index return (grey).

The figure below shows the corresponding DAX curve (red) from beginning of 2005 to mid 2010 and the DAX's daily return in grey. It can be observed that in crash times return volatility increases and that the corresponding changes in average correlation are best captured when using shorter estimation windows. When a long estimation interval is used, successive estimations of average correlations are not independent and therefore localized jumps are smeared out over a long time period (see also Song et al. 2011). However, it has to be stated that the shorter the choice of the time window the noisier the correlation estimations. Therefore it is necessary to test the significance of the correlation coefficients as well as the significance of the MST against the null hypothesis which we will do in the following section. The black line in the upper figure shows our choice of estimation window (100 trading days) for the subsequent analyses which resembles a good trade off between capturing large parts of the short correlation dynamics on the one hand and significance of the correlation coefficient and the MST on the other hand.

4.3.1 Significance of the Correlation Coefficients and the MST

For the dynamic temporal analysis of the DAX universe we define 40 equally spaced observation points in the time range from the beginning of 2005 to mid 2010. This is a good choice as the subsequent analyses can be clearly presented with this number of observation points and more or less observation points don't fundamentally change the results. Just before each observation point we define a time windows dating backwards 100 trading days in order to construct a frequency distribution of pair wise asset return correlations and an MST of this period. Also, for each time window, we construct a set of random time series by using the original time series with shuffled order eliminating in this way any real temporal correlation and preserving the distribution parameters of each time series (see also Aste et al. 2010). Random, uncorrelated data should result in zero correlation coefficients but for return series of finite window sizes there will be some residual finite non-zero correlations. The residual correlations between the shuffled series act as a threshold value to construct a "null hypothesis" as we assess how often a pair of real correlation coefficients is larger in absolute value than the corresponding correlation form the set of shuffled coefficients. The following figure shows the bars of a frequency distribution of the real pair wise correlations at one of the 40 time observation points:

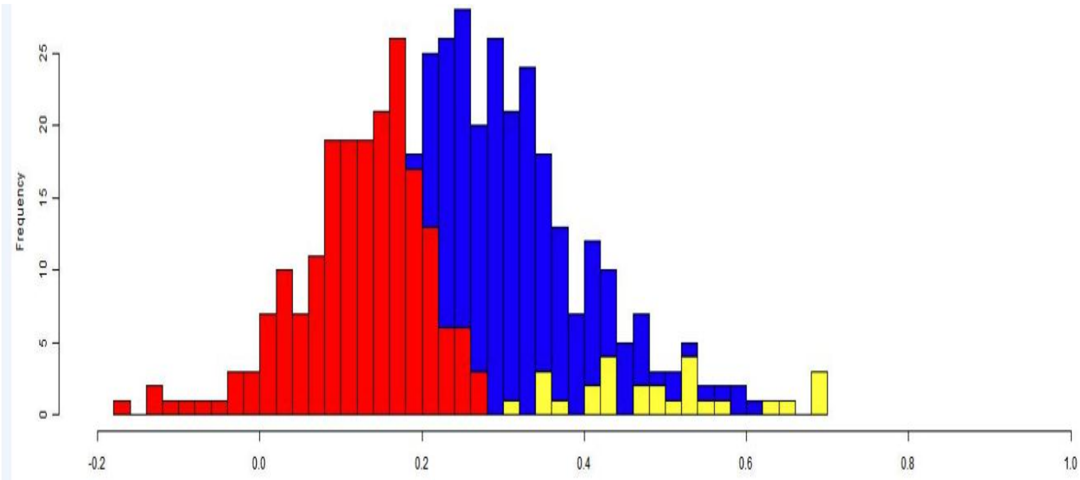

Figure 4-13: Histogram of the correlations in blue. Red bars are fraction of non-significant correlations and yellow is the MST correlation fraction.

Originally, the frequency distribution of the real correlations is completely blue whereas the red bars show the fraction of the non-significant correlations of the blue bars at 99% confidence level. It can be seen that the higher correlations are significant. If the estimation time window was longer than 100 trading days, the fraction of red bars could be expected to be smaller. The yellow bars show the fraction n-1 associated with the MST as a reduction of the n*(n-1)/2 correlations. It can be seen that a large part of the highest correlations was chosen by the MST algorithm and that all MST correlations are significant at 99% confidence level (there are in the blue area and not in the red one).

Finally, it has to be mentioned that the same figure was produced for each of the 40 time observation points and that the one presented corresponds to the histogram where the non-significant bars were nearest to the MST correlations (red bars very near to the yellow bars). From this analysis we conclude that for the dynamic MST analysis it is sufficient to use time windows of 100 trading days – at 99% confidence level.

This analysis could be confirmed by construction the “ultrametric hypothesis” based on the close relation of the MST to single linkage hierarchical clustering. For this analysis we use the same set of shuffled time series for the same time window as in the analysis just before and build the frequency distribution of the cophenetic correlation between the correlation-based distance matrix and the ultrametric distance matrix based on single linkage clustering. Both distance matrices are based on the same set of 100 shuffled series and the blue bars in the following figure show the distribution of the corresponding 100 cophenetic correlation coefficients:

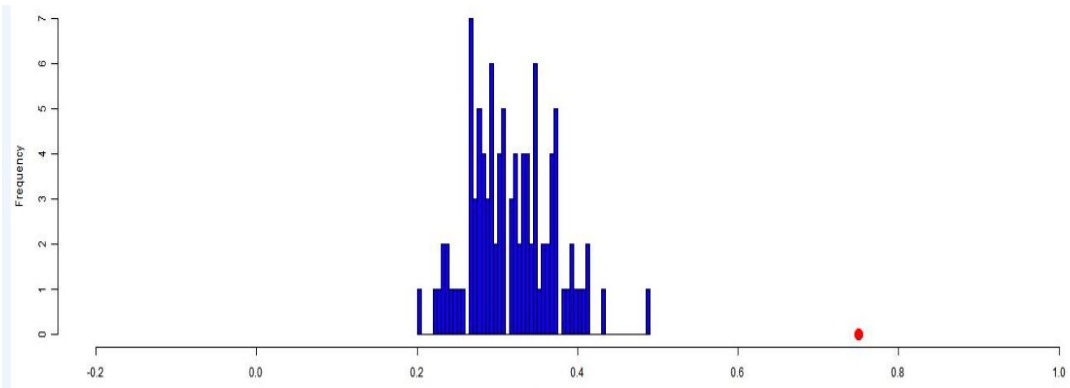

Figure 4-14: Histogram of the CPCC based on the shuffled data (blue) and the CPCC based on the real empirical matrix (red).

If the cophenetic correlation coefficient between the real correlation-based distance matrix and the real ultrametric distance matrix based on single linkage clustering (red dot in the figure above) is larger than a certain significance threshold (quantile of the blue frequency distribution) then the hierarchical clustering is significant. This was not only the case for this observation point being shown in the figure but the case for all observation points so the single linkage hierarchical structure is significant throughout the whole analysis. This implicates that the MSTs for all observation points are plausible graph filterings.

4.3.2 The Fall of the German Banks in the Financial Crisis: A Dynamic Analysis

After the validation and significance tests we now proceed to the core of the dynamical analysis of the DAX universe. The evolution of the MSTs is studied in order to outline the anatomy of the market crash in conjunction with the financial crisis 2008/2009. As developed in the section before, we use 40 observation points and fixed estimation windows of 100 trading days.

Average correlation of the DAX stocks evolves in accordance with market eruptions as can be seen in the following figure:

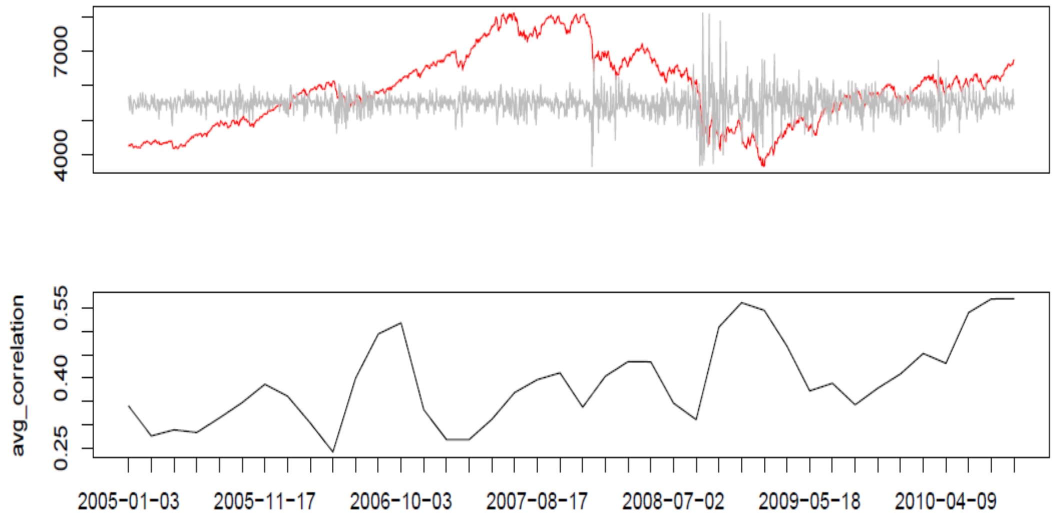

Figure 4-15: Average correlation of the DAX stocks evolving in time. It is never above 0.6 and never below 0.25.

Average correlation resembled by the sum of the MSTs' edge weights exhibits a very similar shape whereas average correlation of the MSTs' is on a generally higher level as can be seen in the scales of the y-axes in both plots:

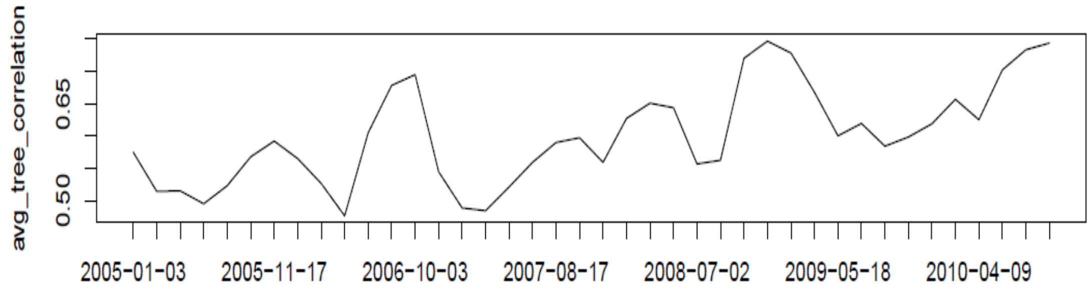

Figure 4-16: Average tree correlation of the filtered DAX network evolving in time. It is never below 0.5.

This shows that the MST captures the correlation dynamics quite well as was expected after the validation and significance tests. The evolving cophenetic correlation coefficient (CPCC) shows a consistently high level above 0.7 which is another indicator of the hierarchical organization of the DAX, especially during/after the crisis:

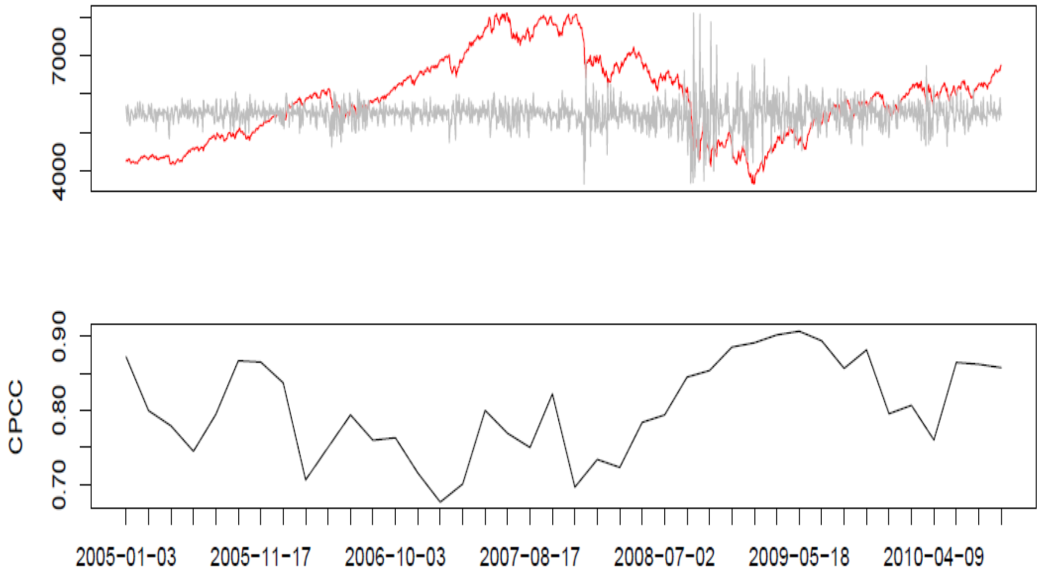

Figure 4-17: The CPCC of the filtered DAX network evolves in time at a high level (correlation of > 0.7)

The dynamic MSTs show the typical behaviour of asset trees in stock markets which exhibit scale-free behaviour of edge degrees (the power law exponent is in the range 2.05 to 2.30):

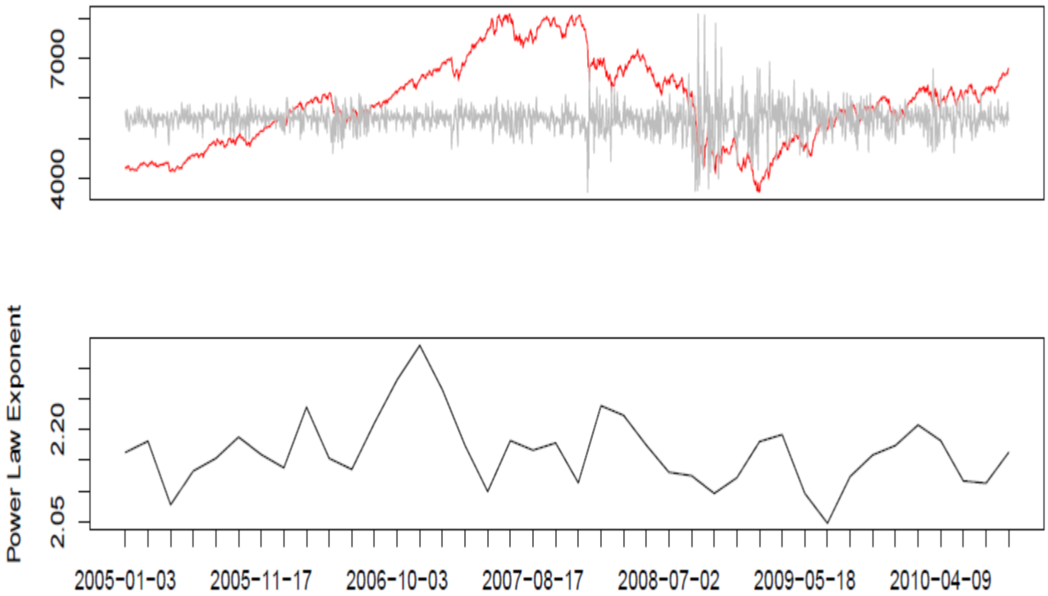

Figure 4-18: Time evolution of the tail index of the power law edge degree distribution of the filtered DAX network. The distribution exhibits fat tails indicating scale-free networks.

During the financial crisis, there are large structural disruptions in the MST topology as was deduced from the T1. We simply sum the T1 distances of all nodes between two consecutive time observation points which results in the following figure:

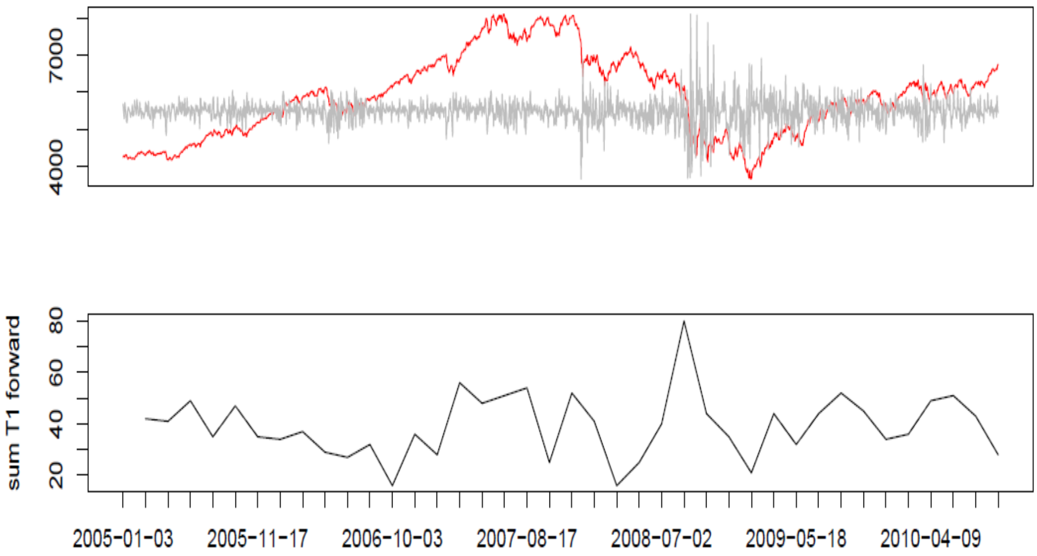

Figure 4-19: Structural breaks of the filtered DAX network as observed by a measure based on T1 distance.

Aste et al. (2010) comment that with respect to the dynamic T1 edge rewiring there is a need for frequent restructuring to make maximum use of diversification benefits.

Some disruptions can even be observed even before the crisis being an early warning indicator. The disruptions even starting before the crisis can also be observed in the temporal change of average centrality and average. For example, the average betweenness as a measure of centrality seems to have risen since 2007 which may be a trend to more integrated markets and globalization:

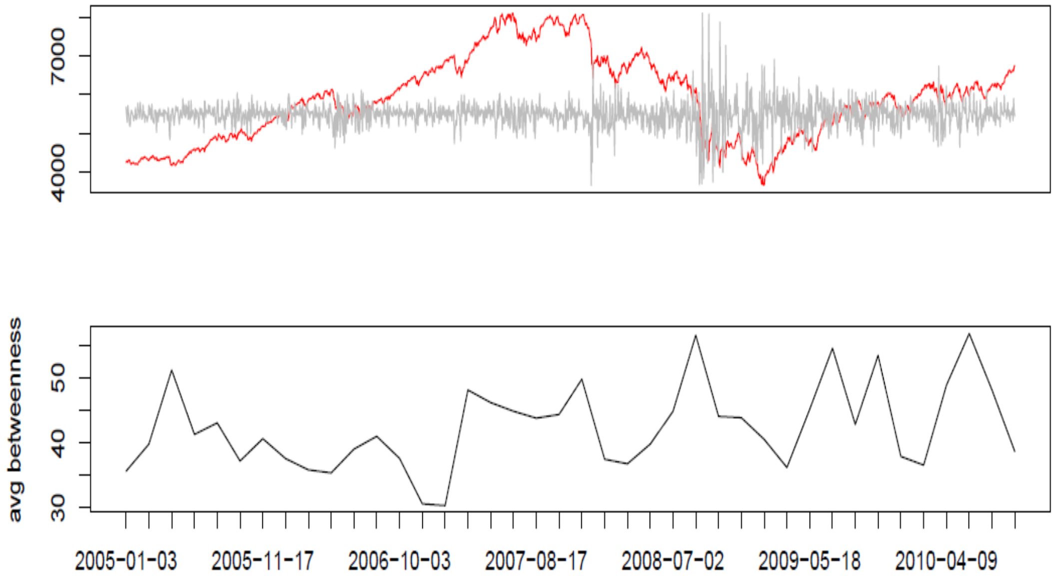

Figure 4-20: Average betweenness of the DAX stocks in time.

The same applies to the peripherality expressed by eccentricity:

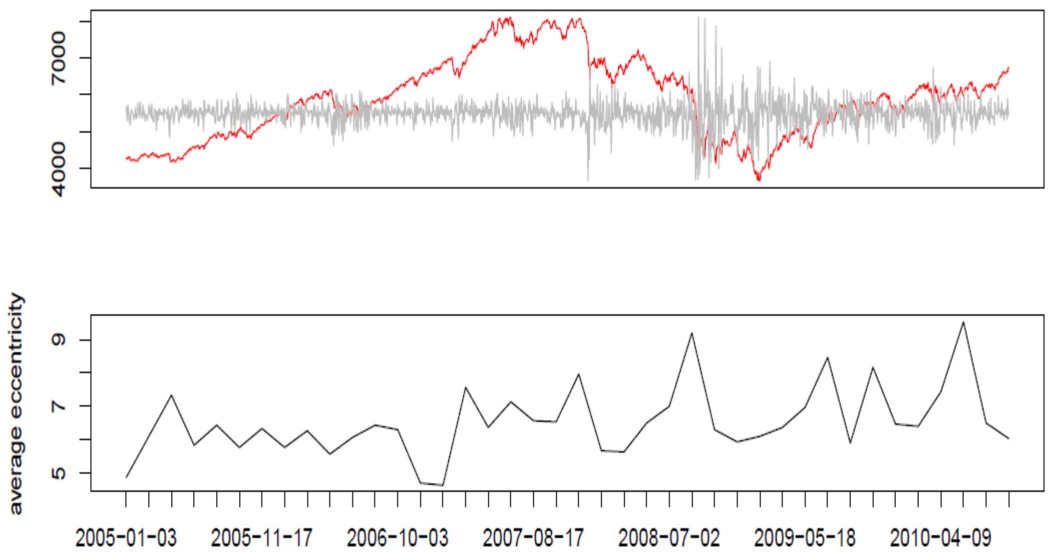

Figure 4-21: Average eccentricity of the DAX stocks in time.

It is interesting that since 2007 the market seems to have moved to extremes in two ways and has not yet recovered: it is more centralized on the one hand and also more eccentric on the other. Also, it is clear that the two banks of the

DAX "Deutsche Bank" and "Commerzbank" have lost their former central position in the German economy for some time after the major crash in October 2008. Throughout the beginning of the crash in 2008, these banks are very central. The following graph shows the relative betweenness of the two banks in comparison to the other DAX members:

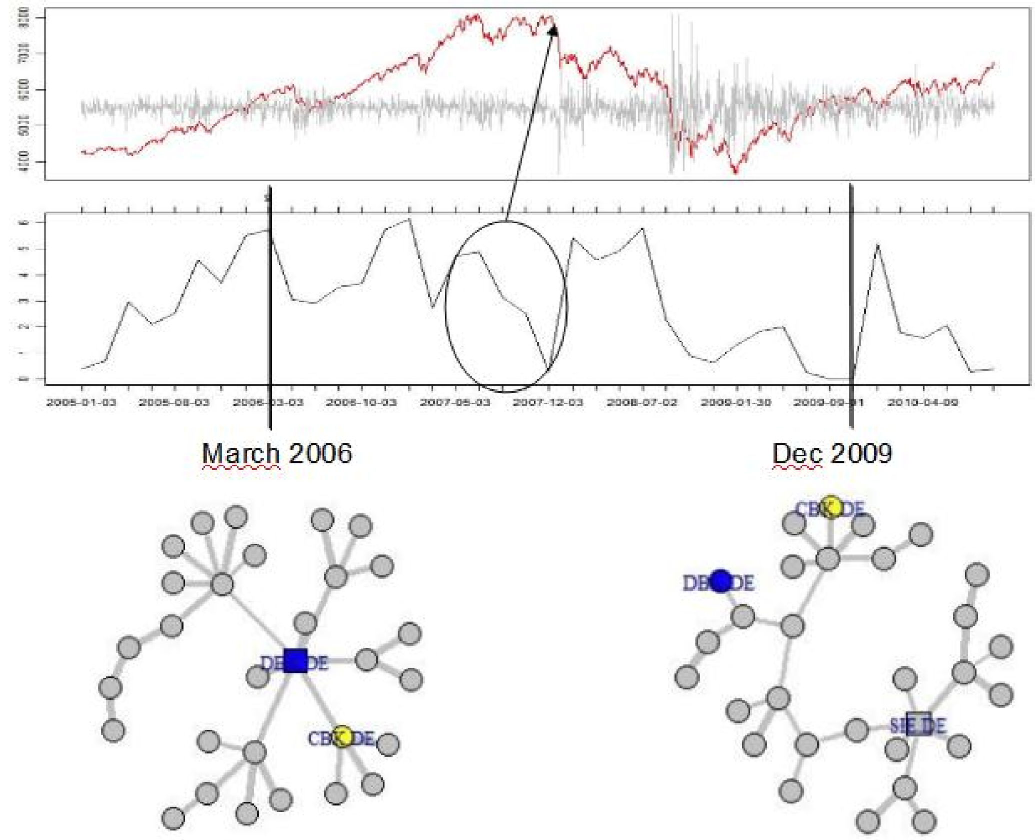

Figure 4-22: The middle chart is the timeline of the relative average betweenness of the German banks Deutsche (blue) and Commerzbank (yellow) in comparison to the other DAX stocks. The red curve shows the DAX index. Snapshots of the networks are taken in March 2006 and December 2009. It can be observed that the banks have become decentralized.

In the years 2005 to 2007 especially Deutsche Bank together with Allianz had been one of the most central nodes in the German economy. Since the crisis, the two banks "Deutsche Bank" and "Commerzbank" seem to be rather eccentric for some time which can be observed relative to the other DAX members:

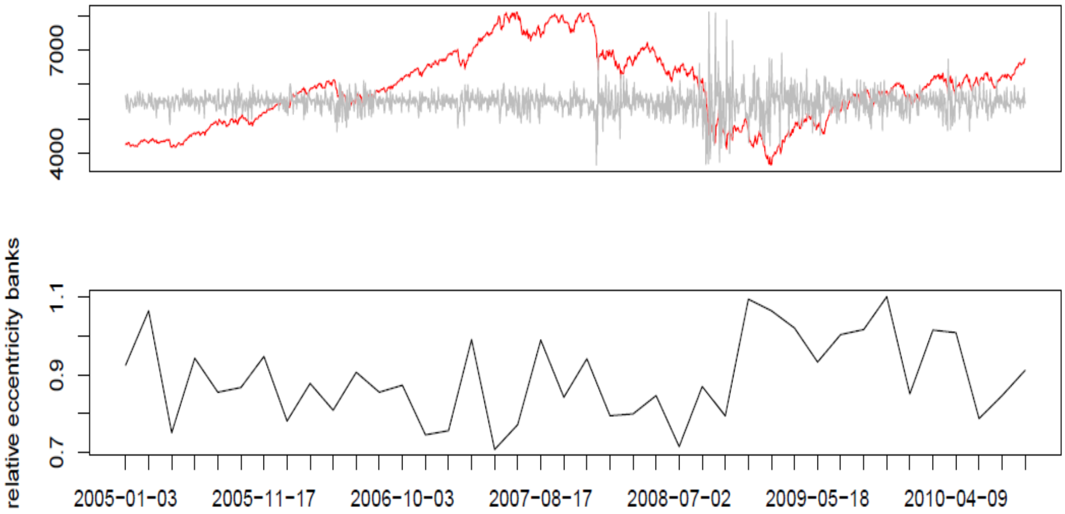

Figure 4-23: Relative eccentricity of the banking sector in time.

There might be some economical interpretation of this banking phenomenon as Aste et al. (2010) explain: "This might be a consequence of the increasingly speculative nature of the financial activity that […] has relatively reduced its original role of service and support to other firms. An interpretation could be that after the financial crisis banks started fighting with market eruptions, the debt crisis and legal proceedings and are thus decoupled from the rest of the economy.

Especially in times of crises banks are very network central. Also, the early drop of centrality before the financial crisis can be seen as an early warning indicator before the crisis.

Finally, other companies and sectors than banks could be analysed by this dynamic analysis of the DAX as support for economical studies of the German stock market structure and the German economy.

Each MST of the different time observation point can be plotted by means of a layout. This is the layout at observation point August 2006 with the banks highlighted by larger nodes:

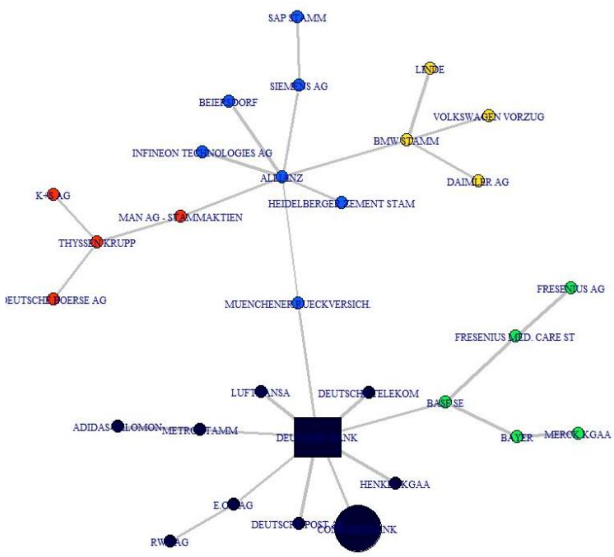

Figure 4-24: DAX MST in August 2006.

The squared shape of Deutsche Bank shows that its node exhibits the highest number of edge degrees in the DAX. The two banks are in the centre of the DAX network. The same layout algorithm used for the observation point January 2009 can be seen here:

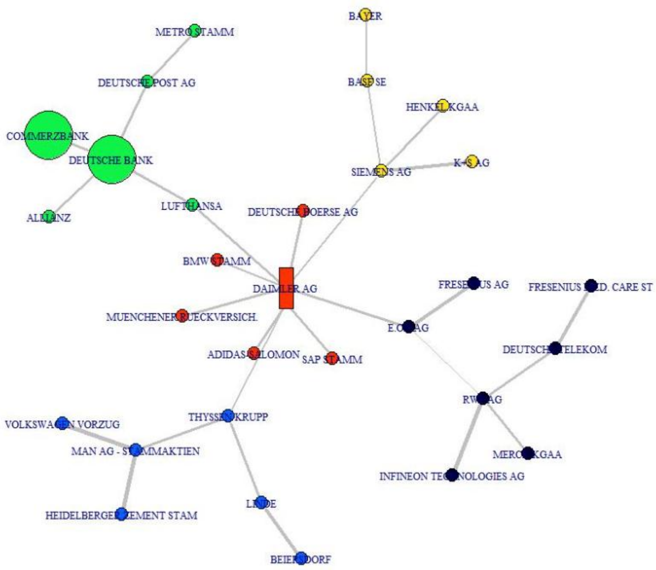

Figure 4-25 DAX MST in January 2009.

Now Daimler is at the centre of the network whereas the two banks are even further away than a stock like Lufthansa.

The colours in the network layouts correspond to the community detection algorithm of Newman and Girvan (2004). It is interesting to see that some of the colour codes correspond to classical industry classification schemes – and some exactly don't. This is the ideal case to show the potential of explorative methods and can be interpreted in the following way: as outlined at the outset, explorative cluster or community detection is just derived from the data without using any economical classification approach. There are some obvious relations like

- Deutsche Bank, Commerzbank and Allianz being in one community (financials),
- Daimler and BMW being in one community (automotive),
- EON und RWE being in one community (energy),

and numerous other examples for correspondence to industry sectors. This is successfully replicated by the data-driven approach so it produces economically reasonable results. Having replicated some of the economically explainable relations it is not necessary to ideally replicate traditional industry classifications by explorative methods as these traditional classifications exhibit some drawbacks as outlined above. Rather it is the deviation from the norm of the explorative methods that are of major interest as these are useful hidden pieces of information.

It this section it was shown that dynamic MST analysis is able reveal complex market activities and major market dynamics including delivering a detailed picture of the anatomy of market eruptions and crashes. The presented methods and analyses will now be consolidated in the final section about our new approach in risk management and portfolio optimization.

4.3.3 Portfolio Optimization Based on Asset Networks

In our approach we plan to apply MST analysis and its graph based measures and community detection algorithms and combine it with further information concerning risk and reward measures in order to create an intelligence amplification system for decision makers in risk management and portfolio optimization.

As an example, we will start with an MST in January 2005 with the modules colour coded and the assets with large variance (as some other risk measure) visualized with large node size:

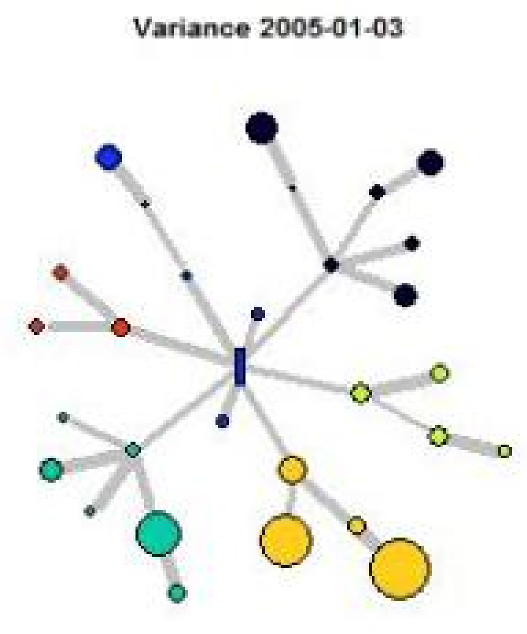

Figure 4-26: DAX MST with community colour codes and node size according to variance.

We would expect a portfolio optimization program to avoid large risk nodes as well as to spread the weights across the different modules equally (module diversification). At the same time, we would expect that weights are positioned at the outer branch nodes avoiding a too centralized weighting as too much weight at the centre nodes would reduce diversification benefit.

In approach we test the waterfall approach based on the single linkage clustering:

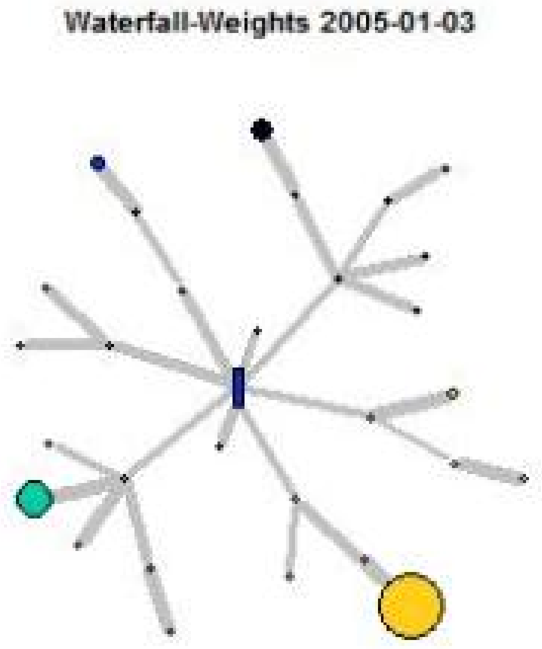

Figure 4-27: DAX MST with community colour codes and node size according to the cluster-based waterfall approach.

The waterfall approach puts strong weight on decentralized nodes and also spreads across different modules. It suggests extreme and eccentric weighting schemes as expected. Also, it has no information about the variance so a large fraction of the weighting volume is put on a large variance node (yellow) which might explain some market exaggeration of the waterfall approach. The following weighting scheme might be another good starting point:

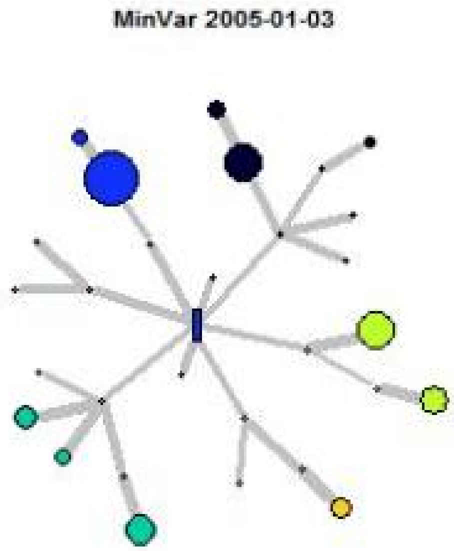

Figure 4-28: DAX MST with community colour codes and node size according to the Minimum-Variance optimization approach.

It is the Minimum Variance portfolio. It can be observed the weights are positioned at the outskirts of the graph, and that weights are spread across the different modules, and that large variance nodes are avoided. This scheme already fulfils most of the requirements of a low risk portfolio. A similar weighting scheme results form the Mean-Variance-(Markowitz)-Approach:

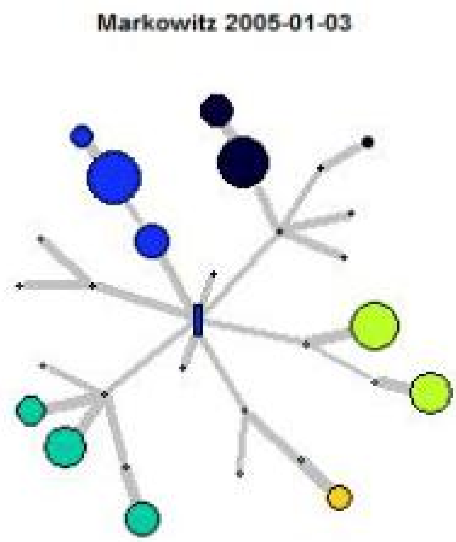

Figure 4-29: DAX MST with community colour codes and node size according to the mean-variance (Markowitz) approach.

However, both the Minimum Variance approach as well as the Markowitz approach exhibit crude allocation errors as can be seen from the following examples:

- The outer yellow node has some weight although it has the highest variance,
- the red module as source of diversification is neglected,
- high weights can be found near the centre node, and
- assets directly connected and thus highly correlated exhibit large weights (especially in the blue and dark blue modules).

These allocation errors might be explained by the outlined “error maximization” properties due to parameter estimation problems of the approaches. Also, variance as a central optimization parameter in these approaches might not be a proper choice of a risk measure as variance estimates rather means “return potential” (in both directions up and down) and also it is no coherent financial risk measure like Expected Tail Loss.

For these reasons we suggest a decision support system for portfolio optimization and risk management with the following visual appearance for suggesting optimal portfolio weights:

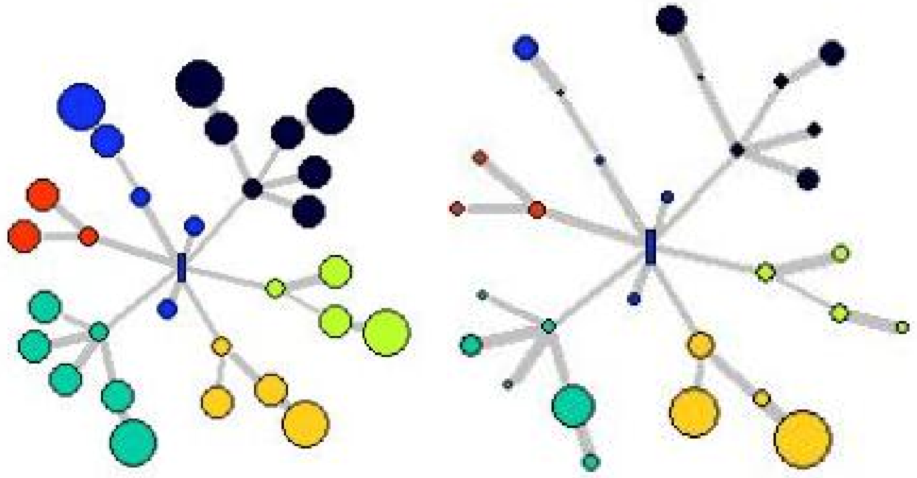

Figure 4-30: DAX MST with community colour codes and node size according to eccentricity (left) and example of a DAX MST based on some just-adjusted performance weighting scheme (right).

The upper network in the upper illustration shows different modules with eccentric or decentralized nodes being large. The purpose is to distribute the weights evenly across the modules and push them into the most eccentric/decentralized nodes. This weighting scheme is combined with the lower network of the illustration which exemplarily highlights assets with a high potential for future risk-adjusted performance based on quantitative or qualitative models. In summary, the large nodes of both network illustrations in combination should receive high weights. This can be accomplished with the support of a simple ranking scheme or rules based engines. Finally, the weights could be adjusted for constraints like institutional investment constraints, or subjective adjustments by the portfolio manager, or adjustments by other automatic model-based approaches. The respective MST constructions can be validated and tested for significance with the methods presented. The result is a highly diversified portfolio with large weights put on high performance assets, meeting investment constraints.

Alternatively, this approach can be used as a warning tool for other portfolio construction systems in that potentially dangerous allocation schemes are identified in the way we showed the errors of the Mean Variance or Markowitz approaches. Finally, the dynamic MST analysis can be used as an early warning tool for major market disruptions and this information could also be used in the asset allocation process.

5 New Approaches to Credit Portfolio Risk Management

Banks and other financial institutions are required to manage their risk concentrations, especially when risk exposures are on a low diversification level. Concentrations of credit risk are usually measured by first defining sectors like industry or geographical regions and then identifying the amount of risk per sector. Other standard approaches determine the risk distribution across names. According to Deutsche Bundesbank (2006), risk concentration in credit portfolios arise from an uneven distribution of credit exposures to individual borrowers (addresses/ names), and as sector concentration in terms of industrial and service sectors, or in geographic regions and countries. Another category of concentration risk is derived from business-linked and intertwined borrowers. The resulting risk of contagion in a credit event of one of these borrowers, however, has only received attention in recent research and is probably hard to model due to complexity and lack of information.

International risk management standards have already been supplemented by the risk perspective concentrations and regional supervisory authorities have already concretised the importance of concentration risk in their latest amendments. This can for example be seen in the "Minimum Requirements of Risk Management" (MaRisk) for German banks.[8] Similar minimum risk management requirements with respect to risk concentration were enacted for insurance companies and asset management firms (see for example MaRisk (VA) und MaRisk (Inv)[9]).

According to MaRisk, risk concentrations have to be paid attention for in the business and risk strategy of the institution. The reason for this can be found

[8] Bundesanstalt für Finanzdienstleistungsaufsicht [Hrsg.] (2010): Rundschreiben 11/2010 (BA) - Mindestanforderungen an das Risikomanagement – MaRisk.
[9] To be found as “Rundschreiben” on http://www.bafin.de.

in many business models of banks and other financial institutions that permanently build up risk concentrations, for example due to special expertise in certain financial instruments, products, and regions. Risk concentrations should also be embedded in every part of the risk management cycle (identification, assessment, management and monitoring of risks) and should be included in risk reporting, stress testing, limitations of credit risk, and in institution-wide scenario analysis. Finally, financial regulation requires to prove that an institution's capital endowments are sufficient to bear risk concentrations and to bear stress tests being based on risk concentration results. So from the regulatory as well as from the internal risk management perspective of a bank it is fundamental to adequately measure risk concentrations.

A prerequisite for the measurement of sector concentration risk is an appropriate sector definition. The definition of sectors should ideally be association with respective risk factors. In simple terms, a sector classification is ideal if asset correlations within a sector are high and low between various sectors.[10]

Practical approaches to the measurement of sector concentration start with definitions of sectors like industry, regions, products, etc. These sector definitions were not primarily intended for the purpose of risk measurement and do not necessarily fulfil a major criterion: the adequate grouping of borrowers into specific sectors, whose credit risk depends on the same risk factor. When measuring the risk of all products associated with credit risk in a large commercial bank it is not clear which sector should be chosen when measuring risk concentrations. For example, when distinguishing sector risk concentration by industry or region, the two kinds of sector concentration risk distinguished differ from a theoretical point of view: The exposure concentration in industries is a typical corporate credit risk, while country risk plays a role in the credit risk of governmental borrowers like public finance business as well as in private borrowing and retail business. Hence, in concentration measurement with respect to the whole bank's credit portfolio or the banking book,

[10] In standard approaches like the structural model by Merton (1974), the term pair wise asset correlation is the correlation of value change of two companies.

differentiating just by industry or region is less precise as there are very different products like retail, asset backed securities, loans, bonds, public finance, commercial real estate, even credit derivatives and many others. Another example is the ambiguous definition of industry sectors as several industries can be strongly connected or a single borrower operates in several industries. This effect is quite similar when partitioning the credit portfolio for example according to country as several countries could be politically interconnected. So the definition of sectors like industry or region requires some kind of causal model which is subject to economical assumptions and thus model risk (also see explanations in the chapters above).

As a result, standard measurement approaches to sector risk concentration according to industry or region are rather incapable of finding the underlying risk factors in the bank's whole credit portfolio. It is indicated to develop a holistic risk measurement approach whereas the sector definitions are model free and result from a data-driven approach in order to discover the underlying risk concentrations. It is thus useful to explore the hidden connections among the obligor names to identify credit risk concentrations. We present several approaches based on Credit Clusters and Credit Networks to find an adequate formation of sectors of the credit portfolio and to model the global mechanics as well as the microstructure of risk concentrations.

As a basis, we use a standard credit portfolio model incorporating both credit exposure size as well as diversification effects. When combining the cluster and network approach with the credit portfolio model it is possible to identify those loans which exhibit little contribution to general portfolio diversification and to formulate concrete risk management initiatives. Since the implementation of the Basel II Capital Accord, banks using the Advanced Internal Rating Based Approach (IRBA) often run an Asymptotic Single Risk Factor (ASRF) model based on internal ratings and the structural approach (Merton 1974) in order to quantify their regulatory and capital requirements.

Some commercial banks employ sophisticated internal portfolio models in order to quantify their economic capital more precisely than just using the regulatory capital computations. These models are often based on several correlated systematic factors, incorporate some group/conglomerate structure and measure risk coherently based on Expected Shortfall (ES).

The approach assumes some process of the firm value V_i of company i and when its value drops below some barrier D_i there is a default event. That is the point where all equity is used up.

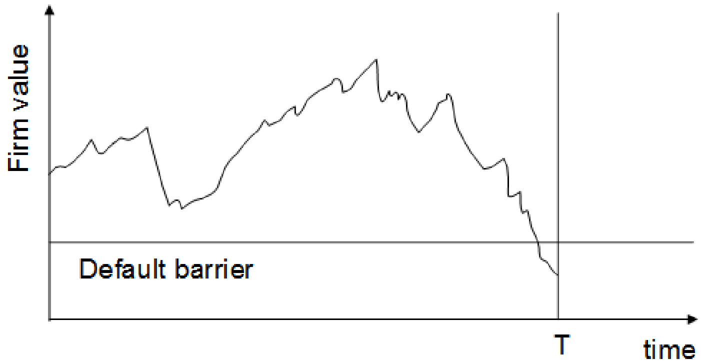

Figure 5-1: The firm value has dropped below the default barrier in time T.

The change of the firm value A_i is assumed to have standard normal distribution $A_i \approx N_{0,1}$ and it is calibrated against the one-year default probability PD_i so that the respective barrier is $D_i = \{A_i \leq c_i\}$ with $c_i = N^{-1}(PD_i)$. In an extended version there are K correlated systematic factors X with $X \approx N_{0,\Sigma}$. Company i has exposure weights to these factors: $w_i = (w_i^1, ..., w_i^K)$. The idiosyncratic risk ε_i of company i are distributed $\varepsilon_i \approx iid\ N_{0,1}$ and the correlation of the firm value change with systematic factors is $R_i^2 = w_i^T \sum w_i$ so that

$$A_i = w_i X + \sqrt{1 - R_i^2}\,\varepsilon_i \text{ with } \begin{array}{ll} Cov(X, \varepsilon_i) = 0 & \forall\ i \\ Cov(\varepsilon_i, \varepsilon_j) = 0 & \forall\ i \neq j \end{array}.$$

The expected loss (EL) of a loan portfolio is the product of PD, EAD (exposure at default) and LGD (loss given default): $EL = PD * EAD * LGD$. The Credit Value-at-Risk (CVaR) is the maximal loss at a certain confidence level α . Roughly described, in the one-factor model of Basel II the unexpected loss UL is defined as $CVaR - EL$ and it has to be smaller than 8% of the risk-weighted assets (RWA):

$$UL = RWA * 8\% = EAD * RW(PD, LGD) * 8\%,$$

where RW is a risk weight function with PD and LGD as inputs. There are capital requirements to cover the UL. The risk weight is constructed as a very bad outcome of the systematic factor (change of the macro economy in the magnitude of the 99,9% quantile) so the following formular holds

$$RW(PD, LGD) = 12{,}5 \cdot LGD \cdot \left(N\left(\frac{N^{-1}(PD) + \sqrt{R^2} \cdot N^{-1}(0{,}999)}{\sqrt{1-R^2}} \right) - PD \right).$$

5.1 From Concentrations to Risk Concentrations

Standard approaches to measuring risk concentrations combine a coefficient measuring distributional concentration with some sort of risk measure in order to express risk concentration as some unequal degree of risk volume distribution.[11] The new approach is to consider the standard risk concentration measures in combination with data-driven sector definitions instead of the generally known industry or country classification.

Prominent representatives of sector concentration measures are the Herfindahl Hirschman Index, the Gini coefficient and Concentration Ratio. If the input of the concentration measure is related to risk then the risk concentration can be simply measured based on the concentration measures.

The Lorenz curve is a graphical representation of the cumulative distribution function of the empirical probability distribution of risk across segments. The

[11] See for example Deutsche Bundesbank (2006).

Gini coefficient is the area between the line of perfect equality and the observed Lorenz curve L, as a percentage of the area between the line of perfect equality and the line of perfect inequality:

$$G = 1 - 2\int_0^1 L(X)dX \ .$$

The higher the coefficient, the more unequal the distribution is. The standardized Gini coefficient is defined between 0 and 1.

The input to the volume concentration measures could be credit exposure, EL, RWA, Economic Capital (EC) and CVaR as they can be summed up per sector. An example is a commercial bank measuring the portfolio risk concentration in terms of CVaR: the Gini coefficient is 0 when total equality is reached and a value of 1 for maximal inequality. This allows computing the amount of misallocated risk budget which may be defined as the distance from an equally distributed portfolio.

The measures presented above just give a general idea of the risk concentration of a portfolio. It would be favourable instead to identify the risk concentration drivers on micro level in order to address just a few obligors who are responsible for most of the risk concentration. After their identification it would be possible to introduce some risk mitigation technique or even sell those assets in order to manage risk concentrations. Most prominent candidates are of course loans with very high exposure or high CVaR but in practical applications it is not possible to easily mitigate or even sell such large loan entities. Also, there may be a large quantity of smaller loans which are highly correlated and thus jointly “behave” like a single large synthetic loan. It is desirable to find small groups of loans that are responsible for large CVaR marginal contributions also called ‘jumps” of the portfolio CVaR as these are the drivers of risk concentration.

In the following section we will analyse a data set of a loan portfolio from a commercial bank with the help of networks and clusters in order to measure general portfolio risk concentration and in order to identify the risk concentration drivers.

5.2 Practical Application: Detection of Credit Risk Concentrations

The test data set is completely artificial but exhibits some real life characteristics. It includes the asset correlation matrix and the CVaR contributions of 1000 loans. The loans originate from diverse credit products, countries, industry sectors and other economic classifications in order to analyse a very heterogeneous portfolio. We introduce two different economic classifications like the one listed before and compare it to a data-driven classification. There are 29 sectors in classification scheme 1 and 30 sectors in classification scheme 2. The CVaR model is based on a few tens of systematic factors and the asset correlation matrix is estimated on the basis of data time windows of several years. We construct an MST from the asset correlation distance which results in the following distribution of edge weights in comparison to the distribution of asset correlations:

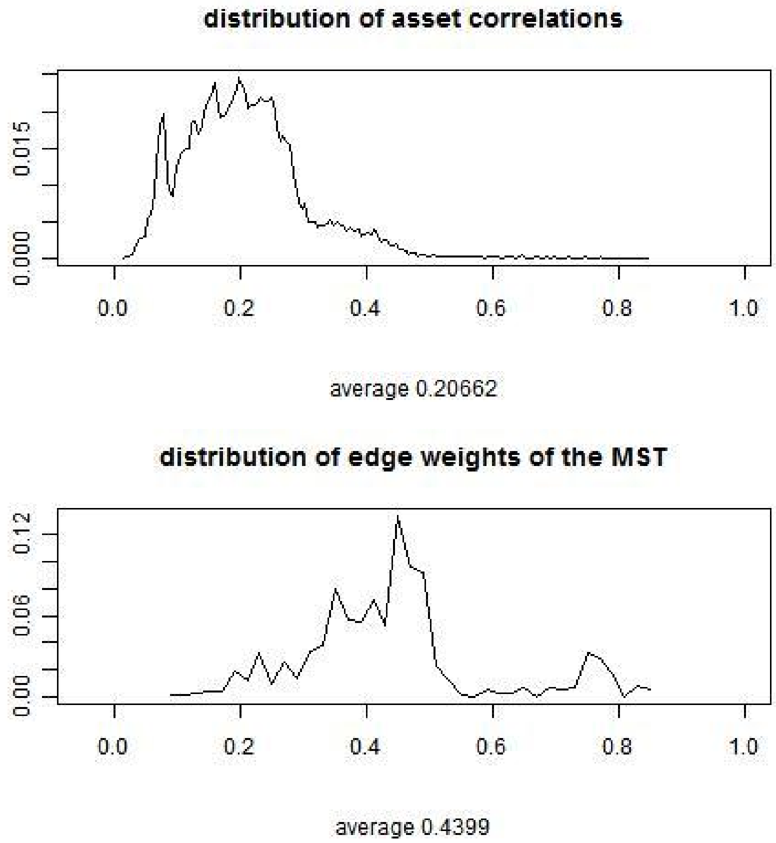

Figure 5-2: Distribution of the asset correlations averaging to 20 % and distribution of the MST edge weights averaging to 44%.

Again, the network in our analysis is laid out with the Fruchterman/Reingold algorithm. In the following picture, the network is colour coded according to two different classification schemes and the node size corresponds to CVaR amount per loan:

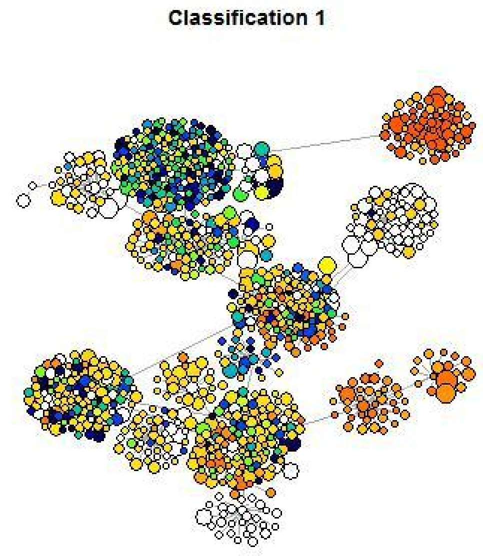

Figure 5-3: Network visualization with a force-based network layout and colour codes according to the sectors of classification 1.

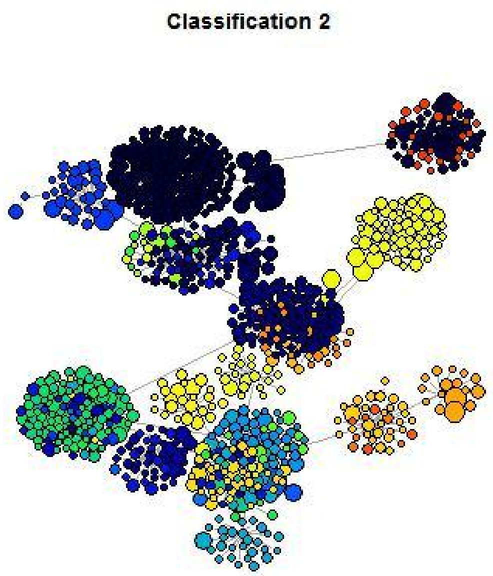

Figure 5-4: Network visualization with a force-based network layout and colour codes according to the sectors of classification 2.

From the colour codes it can be seen that classification 2 matches the network structure better than classification 1 as dense regions of the network mostly consist of a single colour. This is quite informative as it shows that dense regions of the network are economically meaningful assuming that the classification 2 is economically meaningful. Furthermore, assuming that the data-driven structure may even be a better classification for a heterogeneous portfolio than classification 1 or 2, the errors of the latter can be observed in network regions with mixed colours. The following example shows the colour codes resulting from the community detection algorithm of Newman and Girvan (2004). The number of detected communities is 19.

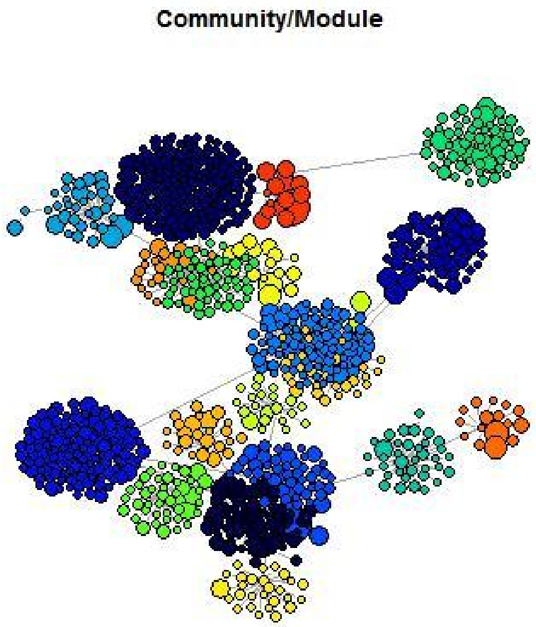

Figure 5-5: Network visualization with a force-based network layout and colour codes according to the network communities.

It can be clearly seen that the data-driven community detection matches the network structure very well with only a few attribution errors. These insights are underpinned by measuring the modularity of the three classifications:

Classification Scheme	Modularity
Classification 1	0.05301498
Classification 2	0.6470625
Community/Module	0.8903278

The modularity measure expresses what has already been observed visually: classification 2 is better than classification 1, and the community structure is better than the two standard classifications. The following graph shows how CVaR is distributed according to the two classifications and the community structure:

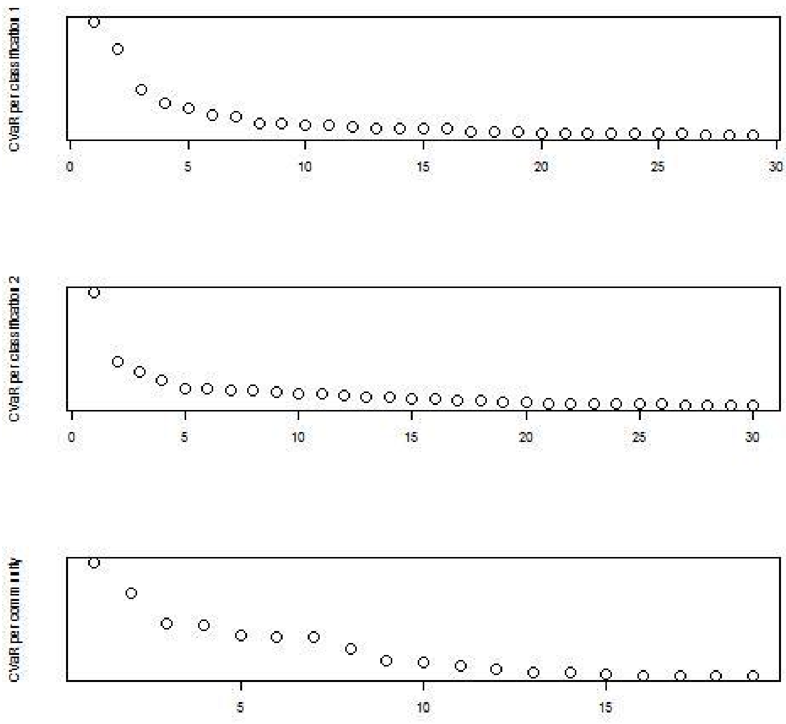

Figure 5-6: Distribution of Credit Value-at-Risk amount for the sectors of classification scheme 1, the sectors of classification scheme 2 and for the communities.

With the explored community structure it is now possible to measure general portfolio risk concentration more adequately. The following graph compares the Gini coefficient of the portfolio CVaR distribution according to the two classification schemes and the community structure:

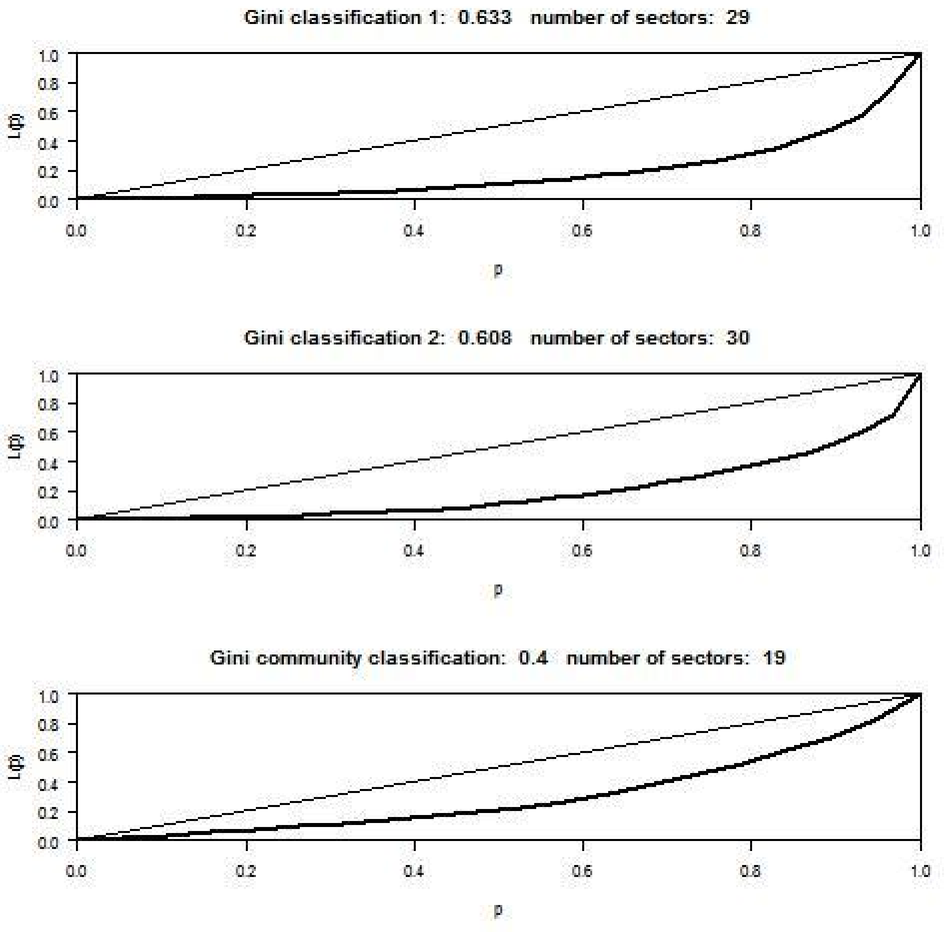

Figure 5-7: Lorenz curve and Gini coefficient for the 29 sectors of classification scheme 1, for the 30 sectors of classification scheme 2 and for the 19 graph-based communities.

It can be observed that the data-driven risk concentration measure is much lower (Gini: 0.4) than the risk concentration measurements based on classification 1 and 2 (Gini 0.633 and 0.608). An explanation could be that the bank's portfolio management has of course not been solely based on reducing sector risk based on classification 1 or 2 but on more sophisticated risk management approaches which is better captured by the data driven risk concentration measurement approach. Also, it is interesting that there are only 19 distinct sector in the data-driven approach but 29 or 30 for classification 1 or to, respectively. This could mean that some classes are actually so interconnected that they collapse to one cluster or module.

Summarizing, we showed that in heterogeneous loan portfolios it is not adequate to report sector risk concentration according to sectors like country, industry or product. Rather, the sectors should be discovered by data-driven methods.

Finally, it is possible to report the substructure of each module by hierarchical clustering as in the following example:

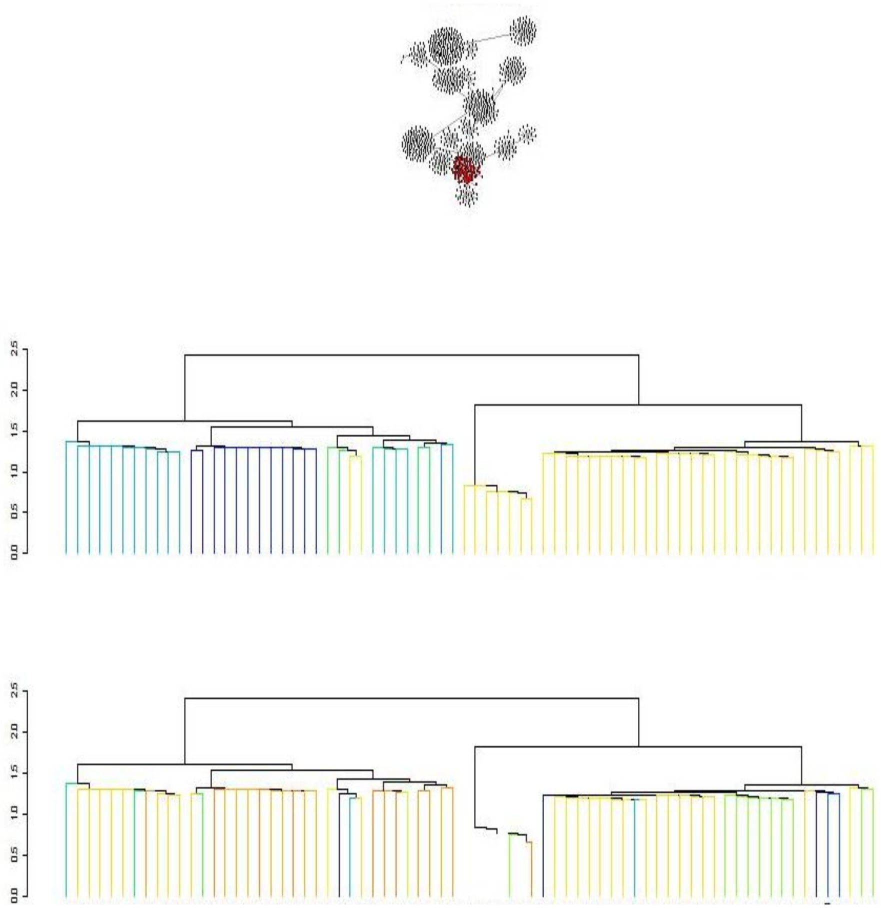

Figure 5-8: Identifying a certain module (colour red in the network) and drawing dendrograms of this community, once colour coded by classification scheme 1 and once by classification scheme 2. This shows the dominance of few sectors, countries and products within each community.

The first hierarchical tree is coloured according to classification scheme 1 and the second according to classification scheme 2. Once again, it can be observed that the clusters are mostly arranged in line with the economical classifications according to single colours in most clusters.

5.3 The Effect of Eliminating Risk Concentration Drivers

The network topology and the community detection offer concrete hints which loans are hardly contributing to portfolio diversification effects and rather resemble risk concentration drivers. For example, very central nodes with a large number of edge degrees are predestined to exhibit changes in company value in synchronicity with many other loans. This applies to the whole network as well as to each community.

It seems straightforward to use centrality measures like edge degree and edge betweenness in order to find very special nodes in the network as well as in each community. We filter the central degrees in the network and in each community by different measures and gather a list of 22 with 14 distinct entries. These are highlighted in the following network graphics, whereas the layout is the same as before:

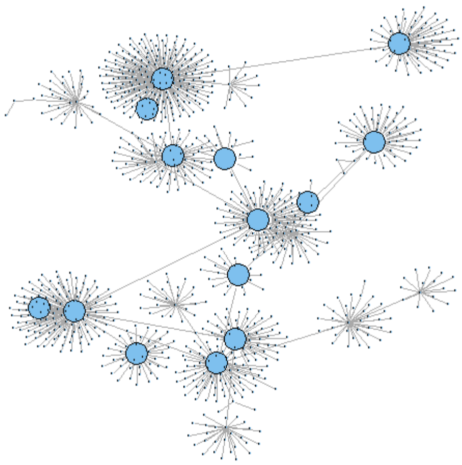

Figure 5-9: Nodes with highest centrality are colour coded in blue and may be candidates for risk concentration driving forces.

The CVaR sum of the 14 loans is the theoretical optimum that could be saved if these loans were eliminated (for example sold) from the portfolio. In reality, the CVaR reduction will be less as eliminating loans will reduce portfolio diversification effects. Taking this information into account, a test can be designed of how successful the network detection of risk concentration is: the nearer the CVaR reduction is to the theoretical optimum the better the method. This is empirically tested against loans of similar size which are arranged in a decentralized and isolated way, thus contributing to diversification. As expected, the elimination of the concentration drivers (central nodes) from the portfolio resulted in the higher portfolio CVaR reduction than eliminating the diversification drivers (decentralized nodes). The presented method can naturally be reversed in order to find “diversification helpers” in the portfolio. Exposures of those helpers could be increased marginally without compromising general portfolio diversification.

5.4 Scenario Based and Dynamic Analysis of Risk Concentrations

As we showed in the approach for the equity portfolio optimization, risk concentrations can be analysed dynamically in time. For example, it is of interest which edges are maintained in the MSTs and which change over time, or if risk concentration drivers are always the same or change over time. Also, modularity and the corresponding explored sectors may stay the same or evolve in time. It can be stated, that a dynamic analysis of the networks can be a valuable contribution to risk management.

In financial stress testing applications and scenario analysis it is common practice to apply historic risk parameters to the current portfolio situation. Accordingly, it is possible to use historic asset correlations (e.g. estimated in times of crises and recession) as current credit portfolio parameters in order to construct historical stress test scenarios. Examples like these show the potential of cluster and network approaches to even more sophisticated risk management applications.

Both dynamic analyses of the commercial banking portfolio as well as stress testing are currently investigated. Since the latest MaRisk amendment it is required to consider risk concentrations in stress testing and scenario analysis. It would be an interesting approach to use dynamic network analysis in this integrated view.

An example is the analysis of the Credit Value-at-Risk of the portfolio in time (yellow line in the figure below), the average correlation (orange), the average centrality of the MST and the number of modules/communities.

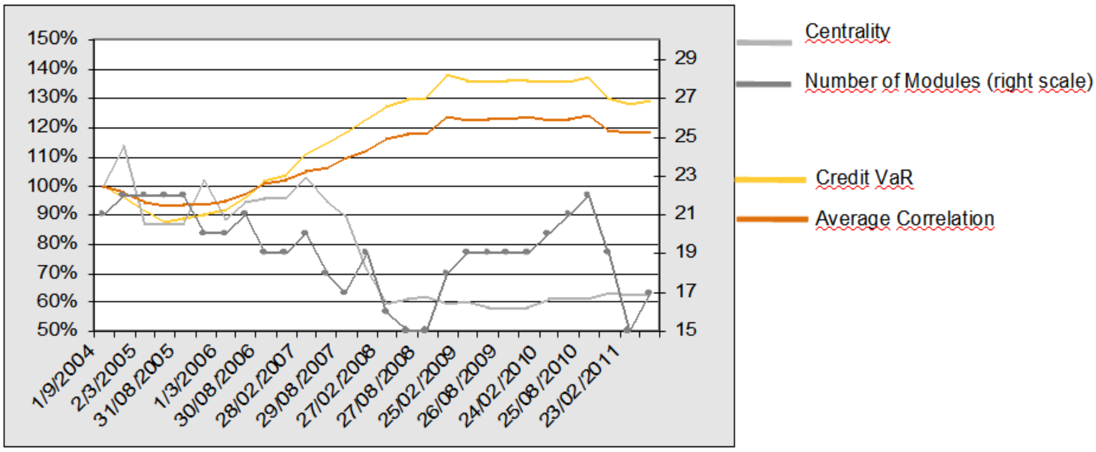

Figure 5-10: The number of communities shrinks in times of stress and crisis.

It can be seen that the risk measure and the correlation rise during the financial crisis. Also, there are fewer modules and there is lower centrality. The network statistics have answers to the following questions:

1. Which clusters merge during the crisis? What are the loan characteristics of those super clusters?
2. Which clusters remain isolated?
3. Which clusters are born?

There is a saying in finance that in times of crisis correlation tend to 1. As this "rule" is too simple, the dynamic community analysis in contrast gives a detailed picture which loans really group and which groups remain isolated.

6 Further Potential of Clustering and Networks in Finance

Besides using Asset Clustering and Asset Networks for the optimization of equity portfolios or for the management of credit portfolio risk there can be imagined several other applications of these technologies to related problems in finance. Here is a list of possible applications:

- data-driven classification of assets
- (multi-) asset allocation and portfolio construction
- analysis of the diversification potential of markets and investment universes
- integrated approaches of qualitative and quantitative information in portfolio optimization
- analysis of market dynamics and integration into limit allocation and early warning systems of financial institutions
- data-driven nested factor modelling
- cost reduced replication of (diversification) indices
- data-driven attribution of risk budgets
- classification and analysis of different investment styles and trading strategies
- enhancement of fundamental economic analyses of markets and investment universes
- analysis of exogenous and endogenous market shocks and anatomical studies of market crashes

The potential of visualization techniques of clusters and networks at different time observation points has already been shown. However, adapting static layout procedures to dynamic tasks is not a trivial problem and there have to be used metrics and statistics to assess and identify change and evolution across networks. The main concern in working with dynamic graphs is to maintain a stable view of the layout. For the most part, algorithmic work on network layouts has been devoted to static graphs. The design of algorithms to analyse and generate visualisations of dynamic networks poses both a technical and perceptual problem. We have experimented with a modified version of the Kamada/Kawai spring embedder layout algorithm (see Kamada/Kawai 1989) which is implemented in the open-source project SoNIA (Social Network Image Animator) described in Bender-deMoll/McFarland (2006). The results are “movies” of the dynamic asset tree evolving in time. Our impression is that these visualization technologies have enormous potential for in-depth dynamical analysis of evolving investment universes and financial portfolios.

Other topics for future research of networks for risk management and portfolio optimization are the following:

- using different filtering techniques like planar maximally filtered graphs as described in Tumminello et al. (2005) and Di Matteo et al. (2010)
- using more sophisticated community structure detection algorithms like those described in the benchmark analysis in Lancichinetti/Fortunato (2009)
- using different distance measures incorporating non-linearities like for example GARCH-distance, mutual information, etc.
- test much more applications in market analysis, risk management and portfolio optimization
- check usability in crisis analysis and early warning
- use high-frequency data
- try other distance measures based on extremes and econometrics
- dynamic application (structural breaks, regime switches, etc.)
- dynamic visualization
- big data and big visualization
- application in scenario analysis and stress testing

7 Bibliography

Arnott R. D. (1980): “*Cluster Analysis and Stock Price Comovement*”, Financial Analysts Journal, vol.37, no. 6, 56-62

Aste T., Shaw W., Di Matteo T. (2010): “*Correlation structure and dynamics in volatile markets*”, New Journal of Physics 12, 085009 1-21

Bender-deMoll S., McFarland D.A. (2006): “*The Art and Science of Dynamic Network Visualization*”, Journal of Social Structure. Volume 7, Number 2

Bouchaud J.-P. (2008): “*Economics Needs a Scientific Revolution*”, Nature 455, 1181

Carlsson, G., Mémoli, F. (2010): “*Characterization, Stability and Convergence of Hierarchical Clustering methods*”, Journal of Machine Learning Research, Vol. 11, pp. 1425-1470

Choueifaty Y., Coignard Y. (2008): “*Toward Maximum Diversication*”, Journal of Portfolio Management 35, 1: 40–51

DeMiguel, V., Garlappi, L., Uppal, R. (2009): “Optimal Versus Naive Diversification: How inefficient is the 1/N Portfolio Strategy?”, The Review of Financial Studies, 22(5), 1915-1953

Deutsche Bundesbank [Hrsg.] (2006): “*Konzentrationsrisiken in Kreditportfolios*“, Monatsbericht Juni 2006, Frankfurt a. M.

Di Matteo T., Pozzi F., Aste T. (2010): "*The use of dynamical networks to detect the hierarchical organization of financial market sectors*", Eur. Phys. J. B 73,

Dose C., Cincotti S. (2005): “*Clustering of financial time series with application to index and enhanced index tracking portfolio*”, Physica A, 355:145-151

Fama, E. (1963): “*Mandelbrot and the stable Paretian hypothesis*”, Journal of Business

Farmer J.D., Geanakoplos J. (2009): “*The Virtues and Vices of Equilibrium and the Future of Financial Economics*”, Complexity 14:11-38

Farrell J. Jr. (1974): “*Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings*”, Journal of Business, vol. 47, no. 2 (April), 186-207

Fruchterman T.M.J., Reingold E. M. (1991): "*Graph Drawing by Force-Directed Placement*", Software – Practice & Experience (Wiley) 21 (11)

Gower J.D. (1966): "*Some distance properties of latent root and vector methods used in multivariate analysis*", Biometrika, 53, 325-338

Heimo T., Saramäki J., Onnela J.-P-, Kaski K. (2007) "*Spectral and network methods in the analysis of correlation matrices of stock returns*", Physica A 383, 147

Jain A.K., Dubes R. C. (1988): "Algorithms for Clustering Data", Prentice Hall

Jain, A. K. (2010): "*Data Clustering: 50 Years Beyond K-Means*", Pattern Recognition Letters, Vol. 31, No. 8, 651-666

Jain, A.K., Murty, M.N, Flynn, P.J. (1999): "*Data Clustering: A Review*", in: Advances in Image Understanding: A Festschrift for Azriel Rosenfeld, IEEE Computer Society

Kalpakis K., Gada D., Puttagunta V. (2001): "*Distance measures for effective clustering of ARIMA time-series*", Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, November 29–December 2, pp. 273–280

Kamada T., Kawai S. (1989): "*An algorithm for drawing general undirected graphs*", Information Processing Letters (Elsevier) 31 (1): 7–15

Keogh E., Kasetty S. (2002): "*On the need for time series data mining benchmarks: A survey and empirical demonstration*", in D. Hand, D. Keim, and R. Ng, editors, Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02), pages 102-111. ACM Press

Keogh E., Lonardi S., Ratanamahatana C. (2004): "*Towards parameter-free data mining*", in W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel, editors, Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), 206-215., ACM Press

Khandani A.E., Lo A.W. (2007): "*What Happened To The Quants In August 2007*", Journal of Investment Management, 5:5-54

King B. (1966): "*Market and Industry Factors in Stock Price Behavior*", Journal of Business, Vol. 39

Kleinberg J. (2002): "*An impossibility theorem for clustering*", In: NIPS 15. pp. 463–470

Kritzman M., Page S., Turkington, D. (2010): "*In Defense of Optimization: The Fallacy of 1/N*", Financial Analysts Journal, 66(2)

Laloux L., Pierre C., Bouchaud J.-P., Potters M. (1999): "*Noise Dressing of Fincancial Correlation Matrices*", Physical Review Letters

Lancichinetti A., Fortunato S. (2009): "*Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities*", Phys. Rev. E 80, 016118

Lee, W. (2011): "*Risk-Based Asset Allocation: A New Answer To An Old Question?*", Journal of Portfolio Management

Lisi F., Corazza M. (2008): "*Clustering Financial Data for Mutual Fund Management*", Springer. Pages 157-164

Maillard S., Roncalli T., Teletche, J. (2010), "*The Properties of Equally Weighted Risk Contribution*", Journal of Portfolio Management, vol. 36, n°4

Mandelbrot B. (1963) "*The variation of certain speculative prices*", Journal of Business

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

Mantegna R. N., Stanley H. E. (2000): "*An Introduction to Econophysics: Correlations and Complexity in Finance*", Cambridge University Press, Cambridge UK, ISBN 0 521 62008

Markowitz H. M. (1959): "*Portfolio Selection: Efficient Diversification of Investments*", Wiley

Martellini L. (2008): "*Toward the Design of Better Equity Benchmarks Rehabilitating the Tangency Portfolio from Modern Portfolio Theory*", Journal of Portfolio Management

Merton R.C. (1974): "*On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*", Journal of Finance, 29, 449-70

Merton R.C. (1980): "*On Estimating the Expected Return on the Market: An Exploratory Investigation*", Journal of Financial Economics 8, 323-61

Murtagh F. (2004): "*On ultrametricity, data coding, and computation*", Journal of Classification, 21, 167-184

Nanopoulos A., Alcock R., Manolopoulos Y. (2001) “*Feature-based classification of time-series data*”, in N. Mastorakis and S. D. Nikolopoulos, editors, Information processing and technology, pages 49-61. Nova Science

Newman M., Girvan M. (2004): “*Finding and evaluating community structure in networks*”, Physical Review E 69, 026113

Onnela J.-P, Chakraborti, A., Kaski, K., Kertesz J., Kanto, A. (2003): “*Dynamics of market correlations: Taxonomy and portfolio analysis*”, Phys. Rev. E 68, 056110

Onnela J.-P., Chakraborti A., Kaski K., Kertesz J. (2002): “*Dynamic asset trees and portfolio analysis*”, European Physical Journal B 30, 285-288

Onnela J.-P., Chakraborti A., Kaski, Kertesz J., Kanto A. (2003) “*Asset trees and asset graphs in financial markets*”, Physica Scripta T 106, 48

Onnela J.-P., Kaski K, Kertész J. (2004) “*Clustering and information in correlation based financial networks*”, European Physical Journal B 30, 353–362

Papenbrock J., Rachev S.T., Höchstötter M., Fabozzi, F. J. (2009): “*Price Calibration and Hedging of Correlation Dependent Credit Derivatives*”, Applied Financial Economics

Porter M. A., Onnela J.-P., Mucha P. J. (2009) “*Communities in Networks*”, Notices of the American Mathematical Society 56, 1082

Rachev S.T. Mittnik S. (2000): “Stable Paretian Models in Finance”, Series in Financial Economics and Quantitative Analysis. John Wiley \& Sons, Inc., New York, London

Rachev S.T., Mittnik S., Fabozzi F.J., Focardi S., Jasic,T. (2007): “*Financial Econometrics – From Basics To Advanced Modeling Techniques*”, Wiley Finance

Rachev S.T., Stoyanov S.V., Fabozzi F.J. (2008): “*Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization*”, Wiley

Schweitzer F., Fagiolo G., Sornette D., Vega-Redondo F., Vespignani A., White D. R. (2009): "*Economic Networks: The New Challenges*", Science, 24 July, Vol. 325, No. 5939, pp. 422-425

Simon H. A. (1962): “*The Architecture of Complexity*”, Proceedings of the American Philosophical Society, Vol. 106, No. 6. (Dec. 12), pp. 467-482

Song D.-M., Tumminello M., Zhou, W.-X., Mantegna R.N. (2011): "*Evolution of worldwide stock markets, correlation structure and correlation based graphs*", Physical Review E 84 (2), 026108

Tan P.-N., Steinbach M., Kumar, V. (2005): „*Introduction to data mining*", Addison Wesley

Tola, V., Lillo, F., Gallegati, M., Mantegna, R.N. (2008): "*Cluster analysis for portfolio optimization*", Journal of Economic Dynamics and Control, 32(1), 235-258

Tumminello M., Aste T., Di Matteo T., & Mantegna, R.N. (2005): "*A tool for filtering information in complex systems*", Proceedings of the National Academy of Sciences of the United States of America

Tumminello M., Di Matteo T., Aste T., Mantegna R.N. (2007): "*Correlation based networks of equity returns sampled at different time horizons*", European Physical Journal B 55, 209-217

Tumminello M., Lillo F., Mantegna R.N. (2010): "*Correlation, hierarchies, and networks in financial markets*", Journal of Economic Behavior and Organization, Volume 75, Issue 1, July, 40-58

Ultsch A. (2007): "*Analysis and practical results of U*C clustering*", Proceedings 30th annual conference of the german classification society (GfKl 2006), Berlin, Germany

Wang X., Smith K., Hyndman R. (2006): "*Characteristic-Based Clustering for Time Series Data*" Journal on Data Mining and Knowledge Discovery, vol. 13, no. 3, pp. 335-364

Ward J. H. (1963): "*Hierarchical Grouping to Optimize an Objective Function*", Journal of the American Statistical Association, 58(301)

Zadeh R. B., Ben-David, S. (2009): „*A Uniqueness Theorem for Clustering*", in Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, AUAI Press

Zhang J., Maringer D. (2010): "*A Clustering Application in Portfolio Management*", Springer Science+Business Media, Chap. 27, 309-321