

FINANCIAL NETWORKS AND THEIR APPLICATIONS TO THE STOCK MARKET

Edward Mandere

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

May 2009

Committee:

Dr Haowen Xi, Advisor

Dr Hassan Rajaei

Dr Bruno Ullrich

Dr Ronald Lancaster

ABSTRACT

Dr Haowen Xi, Advisor

Complex networks exist in many different fields of study. Recently financial networks have garnered a lot of interest mainly as a way of visualizing the relationships between financial entities. This makes them useful in assessing current market dynamics and in predicting future market conditions. The most studied network in econophysics is the correlation network of stock price returns. Stock price return correlations are used to determine a metric distance between stocks. A minimum spanning tree algorithm (Prim's or Kruskal's) is then applied to find closest stocks to create a tree. These networks can then be used to identify and cluster stocks into economic sectors based on their proximity. Some characteristics of these networks such as topology, connectedness and size will be studied in this thesis. The use of these correlation networks in modern portfolio optimization problem will be considered. Using correlation networks to predict large and coordinated price movements will also be discussed.

TABLE OF CONTENTS

	Page
CHAPTER 1. INTRODUCTION	1
1.1 What are networks	1
1.2 Background and Motivation	2
1.3 Studying the hierarchy of the Stock Market	2
1.4 Optimizing Asset Portfolios	3
1.5 Revealing dynamics of stock market during extreme market events	4
1.6 Thesis Organization	4
CHAPTER 2. MINIMUM SPANNING TREE NETWORKS	6
2.1 Creating correlation network using the Minimum Spanning Tree (MST)	6
2.2 Basic Graph Theory Concepts	7
2.3 Minimum Spanning Tree	8
2.4 Kruskal's Minimum Spanning Tree Algorithm	8
2.5 Generated Trees and Index effect	9
2.6 Time Scale effects	10
2.7 Correlation Network Topology	11
2.8 Conclusion	13
CHAPTER 3. METRIC DISTANCE IN MINIMUM SPANNING TREE	19
3.1 How do we measure how close stocks are?	19
3.2 Derivation of metric measurement $d_{ij} = \sqrt{2(1 - \rho_{ij})}$	19
3.3 Conclusion	21
CHAPTER 4. USING NETWORKS TO OPTIMIZE PORTFOLIOS	22
4.1 Modern Portfolio Theory	22

4.2	Using Random Matrix Theory to clean correlation matrices	25
4.3	Using clustering to clean correlation matrices	28
4.4	Ultrametric distance	29
4.5	Single Linkage Cluster Analysis	29
4.6	Average Linkage Cluster Analysis	30
4.7	Observations and conclusions	30
CHAPTER 5. CORRELATION NETWORKS AND STOCK MARKET DYNAMICS		36
5.1	Stock Market Criticality	36
5.2	Correlation Network Size and Market Movements	36
5.3	Dow Jones Index and the correlation network of DJIA 30 Stocks	38
5.4	Conclusion	39
CHAPTER 6. CONCLUSION		46
6.1	summary	46
6.2	Visualization of the hierarchy and interconnectedness of the stock market . .	46
6.3	Portfolio Optimization	46
6.4	Sharp market movements	47
REFERENCES		48

LIST OF FIGURES

Figure		Page
1.1	Map of the Internet	1
1.2	Comparison of correlation network and sector	3
2.1	Complete Graph	7
2.2	Minimum Spanning Tree of 400 of SP500 stocks ($t = 1, T = 50$) showing financial stocks circled	10
2.3	Minimum Spanning Tree of 400 of SP500 stocks ($t = 1, T = 800$) showing financial stocks circled	11
2.4	Dow Jones Graph	14
2.5	MST of Dow Jones 30 + Index	14
2.6	Correlation network of SP500 stocks in March-May 2008	15
2.7	Degree distribution of correlation network for SP500 stocks in March-May 2008	16
2.8	Distribution of Coefficients of correlation matrix vs distribution in MST . . .	17
2.9	Growth of distance with network size	18
4.1	Graph of Risk Return for optimal and non-optimal portfolios	23
4.2	Graph of Risk Return for optimal portfolio at time $t=1$ and time $t=190$. . .	24
4.3	Graph of Risk Return for optimal portfolios with RMT cleaned matrix in time t and time $t+T$ compared to unaltered correlation matrix	28
4.4	Minimum Spanning Tree for four stocks the circles represent the clusters formed	32
4.5	Dendrogram formed using the metric distances between stocks using ALCA .	33
4.6	Risk-return curves generated from ALCA cleaned matrix at on day 1 and day 190	34
4.7	Measure of how much predicted risk differed from observed risk over a period of 190 days	35

5.1	Average size of network during crash of 1987	37
5.2	Hedge fund strategy correlation strengths (thick black lines $\geq 50\%$) calculated from 1994-2007 with and without August 2007	39
5.3	DJIA from (9/7/1996-8/27/2008)	40
5.4	Measure of Tree Length over time using $T=25$ (9/7/1996-8/27/2008)	41
5.5	Measure of Tree Length over time using $T=75$ (9/7/1996-8/27/2008)	42
5.6	Measure of Tree Length over time using $T=300$ (9/7/1996-8/27/2008)	43
5.7	Measure of Tree Length over time using $T=600$ (9/7/1996-8/27/2008)	44
5.8	Measure of Tree Length over time using $T=900$ (9/7/1996-8/27/2008)	45

CHAPTER 1

INTRODUCTION

1.1 What are networks

Networks help us understand the relationships between entities. They can be found in many diverse disciplines. Popular examples include the communication networks like the Internet or ecological networks like the food web and social networks found in kinships and genealogy.

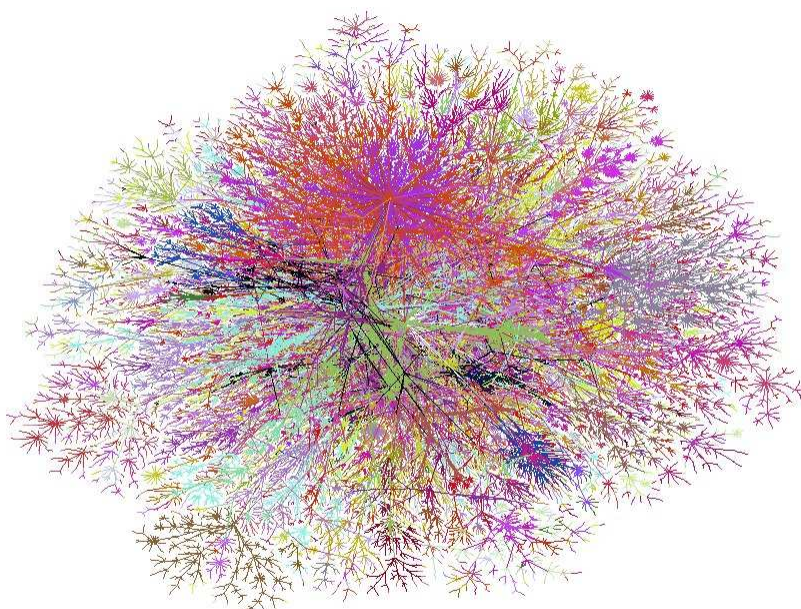


Figure 1.1: A visualization on the Internet

In recent years there has been a growing interest in financial networks. These networks not only help visualize the relationship between different financial entities (Stocks, Companies, Hedge Funds among others) they may also be used to predict future market conditions. I will ex-

plore these financial networks, their properties as well as how to in cooperate this properties in decision making.

1.2 Background and Motivation

Different types of financial networks can be considered. Networks based on company ownership have been studied in the Italian and US stock market. This study showed how company ownership was a power law distribution with small number of people controlling the bulk of companies.

Another network studied involves board membership on companies listed on the NYSE. if a person sits on the board of two companies then there is a connection between the companies. Here too board membership was a power law distribution with a small number of board members sitting on most boardrooms. This leads to a concentration of decision making in the hands of a few people.

The other more common type of financial network is based on stock price correlations. This study of this network will form the study of this thesis. There are three main uses of this network that have emerged in recent years

1.3 Studying the hierarchy of the Stock Market

The initial work on financial networks showed how without prior knowledge of industry sectors you could generate a hierarchical structure that grouped stocks in the S&P 500 very closely to their industries based on their stock price correlations [1]. This initial work by Mantegna has been replicated in various stock markets around the world such the Korean Stock Market. Foreign Exchange (FX) market [6].

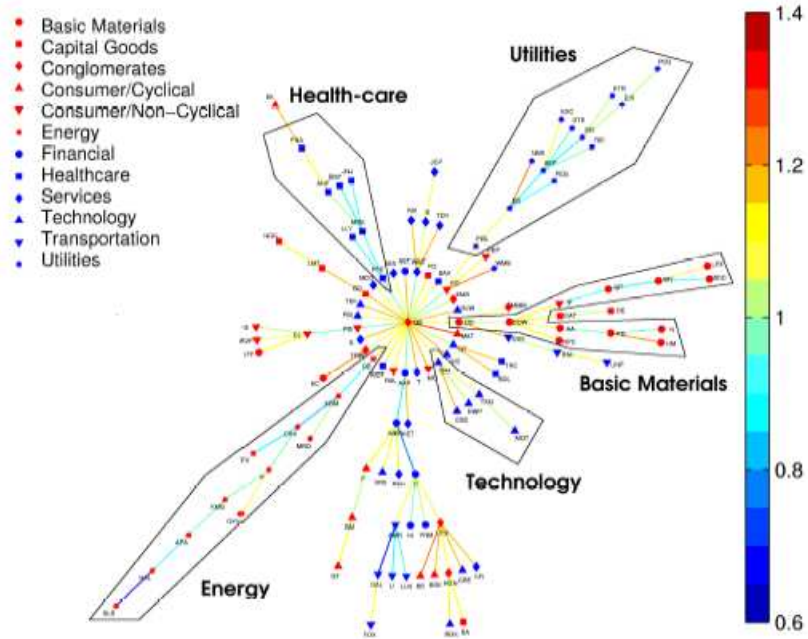


Figure 1.2: Stock Market Sectors compared to networks

1.4 Optimizing Asset Portfolios

There has also been work on using correlation networks to optimize portfolio[2]. This work done by Vincenzo Tola and Fabrizio Lillo involves replacing the correlation matrix used in Markowitz solution to portfolio optimization with a new matrix that uses the ultrametric distance between stocks. This was then compared to previous attempts to clean the correlation matrix using Random Matrix Theory and it showed better results in predicting the future risk/return function of the portfolio[2].

1.5 Revealing dynamics of stock market during extreme market events

Correlation networks have shown some promise in predicting sudden market movements. One example has been the average distance between stocks in these networks[3]. This was first observed when the average length of correlation networks was plotted against time during the period 1985-1990. It showed a significant change before and after the crash of 1987.

Work by Onnela has also hinted at the possibility that such networks may be used to predict rare catastrophic market events such as the 1987 crash [3]. More recently a paper from MIT's Andrew Lo and Khandaniy showed how the events leading to the large losses incurred by hedge funds in August 2007 could be observed through the correlation network of different hedge fund strategies[4].

1.6 Thesis Organization

The focus of this thesis is how we can use the information from financial networks in decision making. Do these networks truly have a predictive power? Are they more or less sensitive to noise that plagues financial data? To answer these questions we will first need understand the methodology used to create the networks. We will then test certain properties of the networks to see how they hold up against real stock market data.

Chapter 2 will introduce the algorithm used to create the network. In order to do so we will introduce some concepts of graph theory such as edges, vertices and minimum spanning trees.

Chapter 3 will look at the metric distance used to measure how close or how far stocks are. We will use the standard metric found in most literature and then try other metrics to see what difference it makes.

Chapter 4 deals with how the correlation networks have been used to clean correlation matrices and therefore improve risk analysis. It will be compared to another common technique known as Random Matrix Theory. Comparisons show that matrices cleaned using network techniques give slightly better characterisation of risk.

Chapter 5 looks at how correlation networks can be used to possibly predict sharp market movements based on the size of the correlation network.

CHAPTER 2

MINIMUM SPANNING TREE NETWORKS

2.1 Creating correlation network using the Minimum Spanning Tree (MST)

To create the correlation network between financial entities (in this case stocks) we need to first select the variable of interest. In the stock market case the stock price appreciation or depreciation is the variable of interest. This is also known as the return on the stock. If we let P_t be the price at time t and P_{t-1} be price at previous interval the the return R_t is defined as.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (2.1)$$

In finance the logarithmic return is used since it makes working with returns over continuous time easier. This is especially true in the case of compound returns over time. Therefore we will work with $\ln(R_t)$ instead of R_t .

It is important to note the time scale t used in measuring the return. Usually returns are measured as daily, weekly, quarterly and yearly. Changing the time scale changes the topology and metric size of the network. We will explore this properties (topology and metric size) once we have described the Minimum Spanning Tree.

After selecting the time scale for return, we need to select a second time scale T to calculate the correlation. This time scale T consists of all the returns over a certain historical period. For example you could use the returns for the prior hundred days to create the correlation. The choice of this time scale leads to considerable changes in the topology and metric size of the minimum spanning tree as well.

There is also a more fundamental question regarding how far back one should look for correlations. For economist most strong advantageous correlations are arbitrated away

quickly thus the Efficient Market Hypothesis[5]. This is part of what also what makes financial data noisy. Past correlations may be unreliable in predicting future correlations.

Once we have determined the variable to correlate ($\ln(R_t)$) the time scale of the return and the time scale of their historic window. The formula to calculate correlations is [6].

$$\rho_{ij}(\Delta t) = \frac{\langle p_i p_j \rangle - \langle p_i \rangle \langle p_j \rangle}{\sqrt{(\langle p_i^2 \rangle - \langle p_i \rangle^2)(\langle p_j^2 \rangle - \langle p_j \rangle^2)}}. \quad (2.2)$$

By correlating every stock with every other stock we end up with a matrix known as the correlation matrix. This matrix plays an important role in modern finance. It is especially prevalent in risk analysis and asset portfolio management. The coefficients present in the correlation matrix are then used to generate the Minimum Spanning tree. In order to understand what a Minimum Spanning Tree is we will need to introduce some concepts from graph theory.

2.2 Basic Graph Theory Concepts

Graph Theory is used when trying to define objects and their relationships. A graph $G = \{V, E\}$ consists of a set of vertices V and edges E . The vertices represent the objects and the edges represent the relationships between the objects. A graph is considered complete when every vertex has an edge to every other vertex

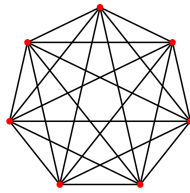


Figure 2.1: A Complete Graph

In some graphs the edges of the graphs can be assigned values or weights that represent the "distance" between vertices.

2.3 Minimum Spanning Tree

A minimum spanning tree is a graph that connects all the vertices in a graph without forming any cycles[7]. The size of the tree is also minimized. More formally if we define a graph $G = \{V, E\}$ and $(u, v) \in E$ and $w(u, v)$ is the weight associated with edge (u, v) then a tree T is a subset of E and $w(T)$ is minimized. $w(T)$ is defined as

$$w(T) = \sum_{(u,v) \in T} w(u, v) \quad (2.3)$$

In order to create a minimum spanning tree from stock data the correlation matrix ρ_{ij} is first generated. Using this matrix a complete graph is created with a metric distance d_{ij} is assigned to the edge connecting stock i and j where d_{ij} can be defined as

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \quad (2.4)$$

Equation 2.4 will be derived in the next chapter.

2.4 Kruskal's Minimum Spanning Tree Algorithm

An algorithm like Kruskal's's algorithm can then be used to extract the Minimum spanning tree from the complete graph. The algorithm works by create disjoint sets of vertices. The edges are then sorted by their weight and the smallest weight is picked each time. If the edge picked connects vertices in the same disjoint set then a cycle is formed so the algorithm avoids edges with vertices in the same disjoint set. The algorithm continues picking edges using the previous algorithm until all vertices are connected [7].

MST-KRUSKAL(G, w)

$A \Leftarrow \emptyset$

for each vertex $v \in V[G]$ **do**

MAKE-SET (v)

```

end for

sort the edges of  $E$  into nondecreasing order by weight  $w$ 

for each edge  $(u, v) \in E$ , taken in nondecreasing order by weight do

    if FIND-SET( $u$ )  $\neq$  FIND-SET ( $v$ ) then

         $A \leftarrow A \cup \{(u, v)\}$ 

    end if

    Union( $u, v$ )

end for

return  $A$ 

```

2.5 Generated Trees and Index effect

Previous trees generated by Mantegna showed that the resulting tree shows stocks groupings that closely relate to their sectors. This sector identification is the root of this technique. Clustering stocks with some regularity demonstrates that there are valid relationships being revealed in the network. Intuitively the closest relationships should exist between companies in the same sector or companies in closely related sectors such as Industrial and Energy. Using 400 Stocks from the SP500 with $t = 1$ and $T = 50$ the sector identification was still evident as demonstrated by the clustering of Financials (Fig 2.2).

However when the the time scale was lengthed to $T = 800$ large number of stock clustered around the market index and less hierachy appeared (Fig 2.5). When a minimum spanning tree was created using Dow Jones 30 stocks plus the DJIA Index as shown in figure 2.5 almost all nodes clustered around the index.

2.6 Time Scale effects

Previous work on this subject has not explored the question of time scale, usually staying with an arbitrary time period T . The assumption is that smaller historical time windows are dominated by noisy and therefore should not be used. However the kind of change in hierarchy that appears in small time scales still appears meaningful suggesting that there is still information to be gained. The domination of market index in longer time scale suggest that in the long term the average market movements dominates most of the price movements on the market.

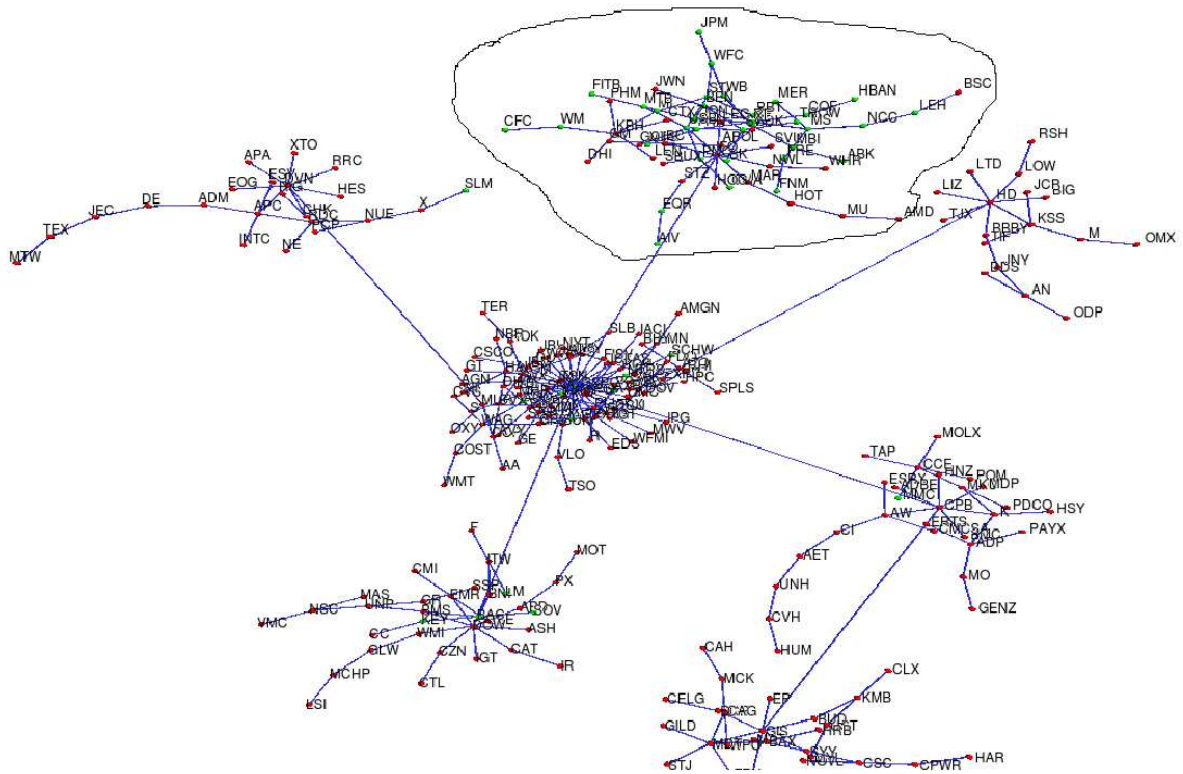


Figure 2.2: Minimum Spanning Tree of 400 of SP500 stocks ($t = 1, T = 50$) showing financial stocks circled

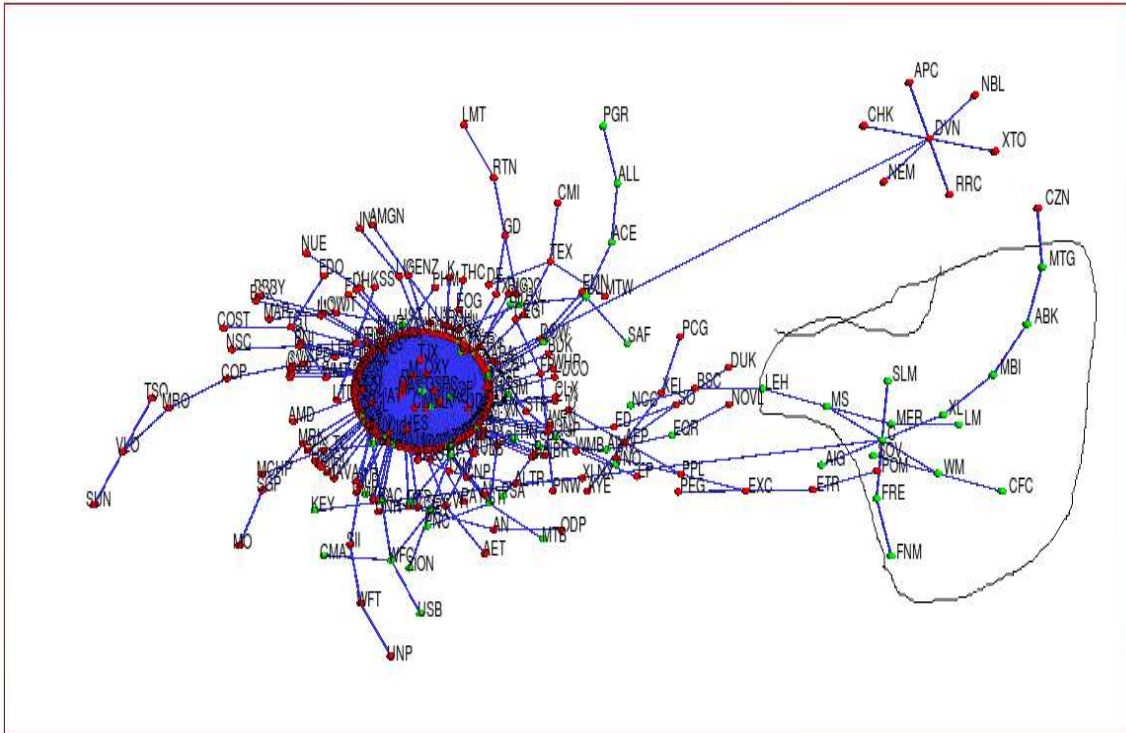


Figure 2.3: Minimum Spanning Tree of 400 of SP500 stocks ($t = 1, T = 800$) showing financial stocks circled

2.7 Correlation Network Topology

To further explore the correlation network a number of important concepts from complex networks shall be introduced.

1. Degree This is a measure of how many connections(edges) a node (vertex) has. It is defined as k . The degree distribution $P(k)$ can be used to classify different networks based on the distribution of their connectness[10].
2. Clustering coefficient This is a measure of the tendency of neighbors of given vertex to be connected. If k is the degree of a node then the cluster coefficient C_i for node i is defined as

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (2.5)$$

Where E_i are the number of neighbours of i . The average cluster coefficient C for the whole network can be defined as

$$C = \frac{\sum_{i=1}^N C_i}{N} \quad (2.6)$$

However for spanning trees by definition cannot have cycles. Therefore $C = 0$ for the generated correlation network.

3. Average Path length The average path $\langle d \rangle$ length is the measure of how many hops it should take, on average, to go from any node to any other node.

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{u,v \in E} d_{uv} \quad (2.7)$$

To study these properties in correlation networks, a network of 400 stocks from SP500 was used. When the degree distribution was plotted it was found to obey a power law distribution. A power law distribution is indicative of what is known as a scale free network. Scale free networks are abundant in various systems and fields. Two common examples are the Internet and the airline flight network. In the Internet the nodes are websites and the edges are their links. In the case of the airlines the nodes are the airports and the edges are connecting flights to the different airports. If the degree distribution is a power law it means a small number of nodes contain a large number of edges or connections [10]. In the case of the Internet this means a web traffic is concentrated to a small number of websites or in the case of the airlines there are certain airports that most airline would transit through e.g. Chicago.

In the case of the correlation network we discovered that the market index would have the largest links. Other nodes with large edges are K(Coca Cola) and C (Citigroup). Varying the time scale also changed the degree distribution especially with the addition of the SP500 Index. In longer time scale the central nodes have a much higher concentration of edges.

The average path length for a the minimum spanning tree was shown to grow as $O(N)$

$$\log(\log(N)) < O(N) < \log(N) \quad (2.8)$$

A scale free network grows at the rate of $\log(\log(N))$ and a random network grows at $\log(N)$. This shows that the minimum spanning tree may lie somewhere in between in terms of its characteristics. This could also be attributed to noise present in the data.

2.8 Conclusion

This chapter has shown that the minimum spanning trees created from stock price returns can be useful in understanding the hierarchy of the stock market. The trees created have a correspondence to market sectors such as the Financials. The characteristics of the degree distribution were also discussed. The next chapter will look into how networks have been used to clean correlation matrices which is important in portfolio optimization.

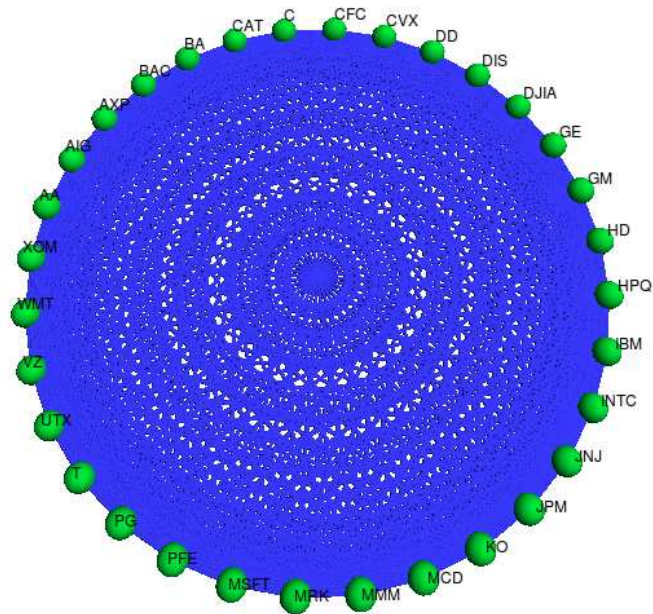


Figure 2.4: Complete Graph of Dow Jones 30 stocks + Index

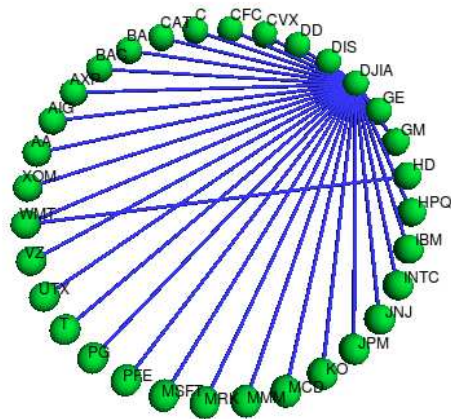


Figure 2.5: MST from Dow Jones 30 stocks + Index

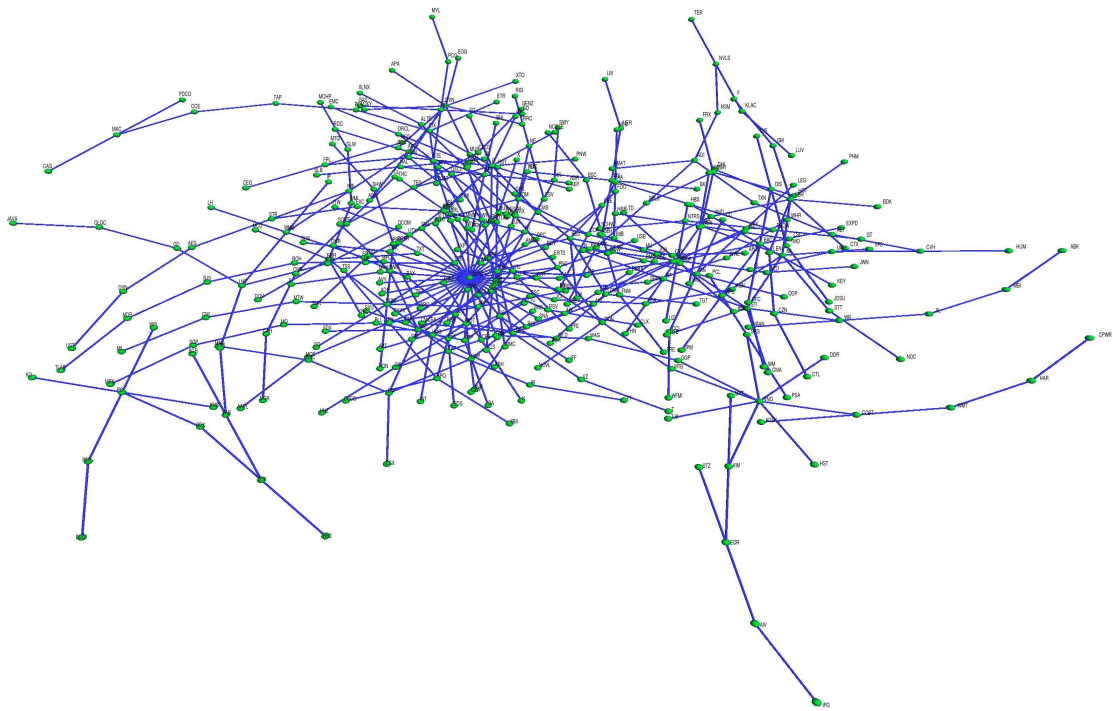


Figure 2.6: Correlation network of SP500 stocks in March-May 2008

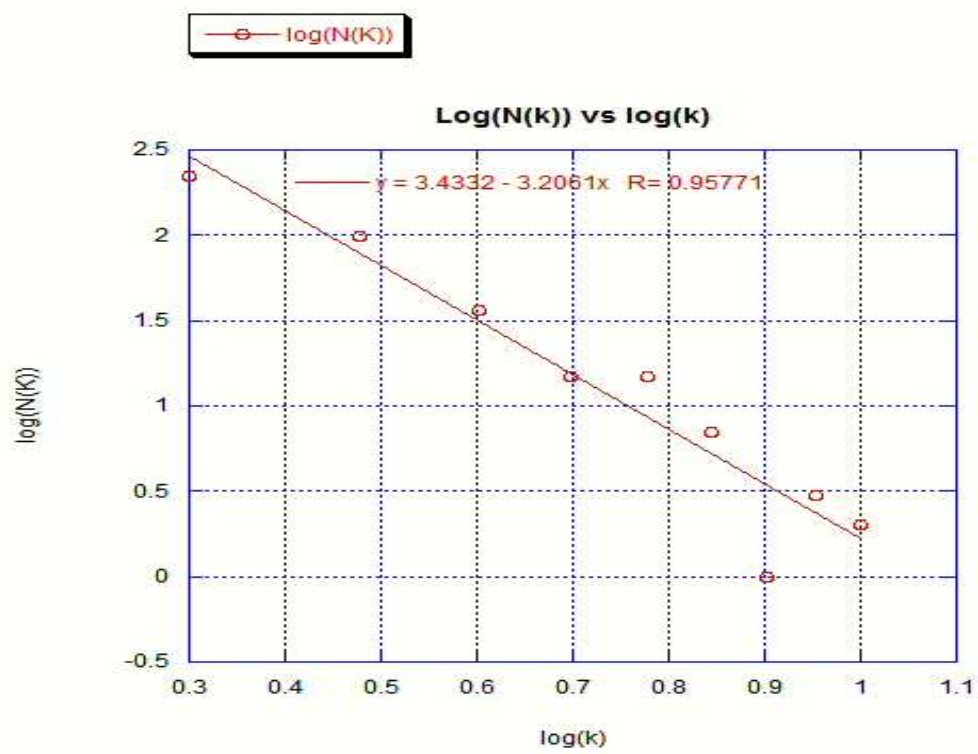


Figure 2.7: Degree distribution of correlation network for SP500 stocks in March-May 2008

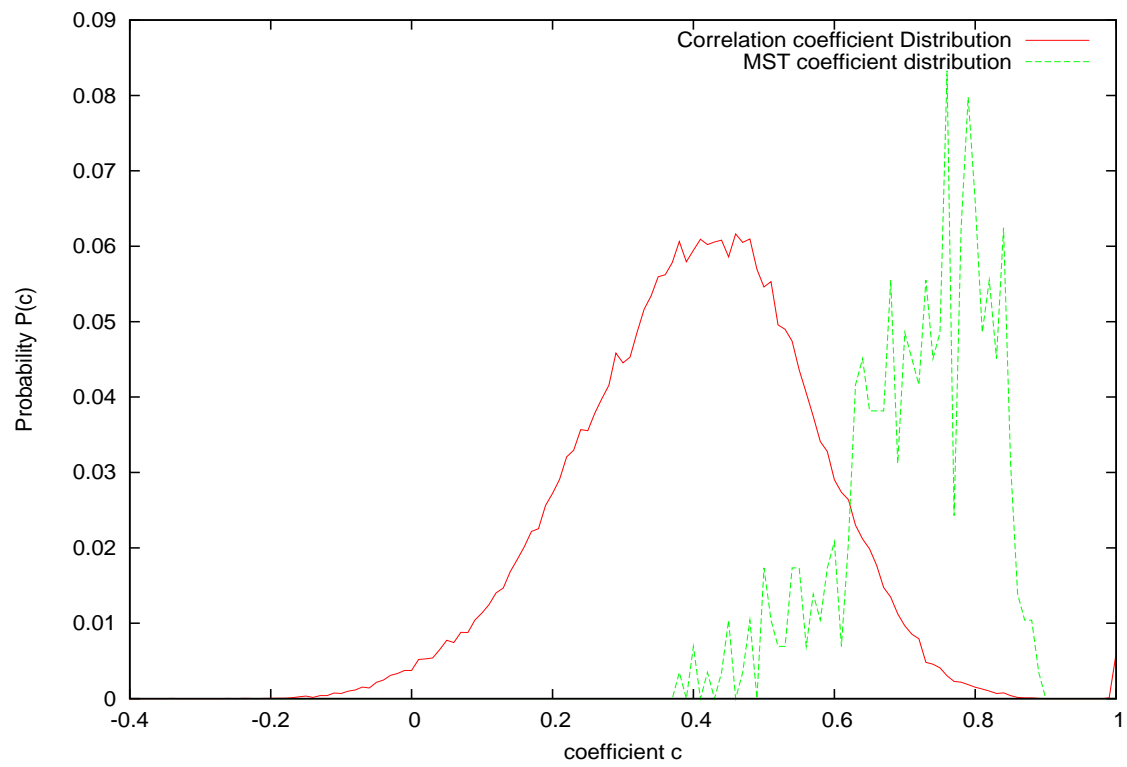


Figure 2.8: Distribution of Coefficients of correlation matrix vs distribution in MST

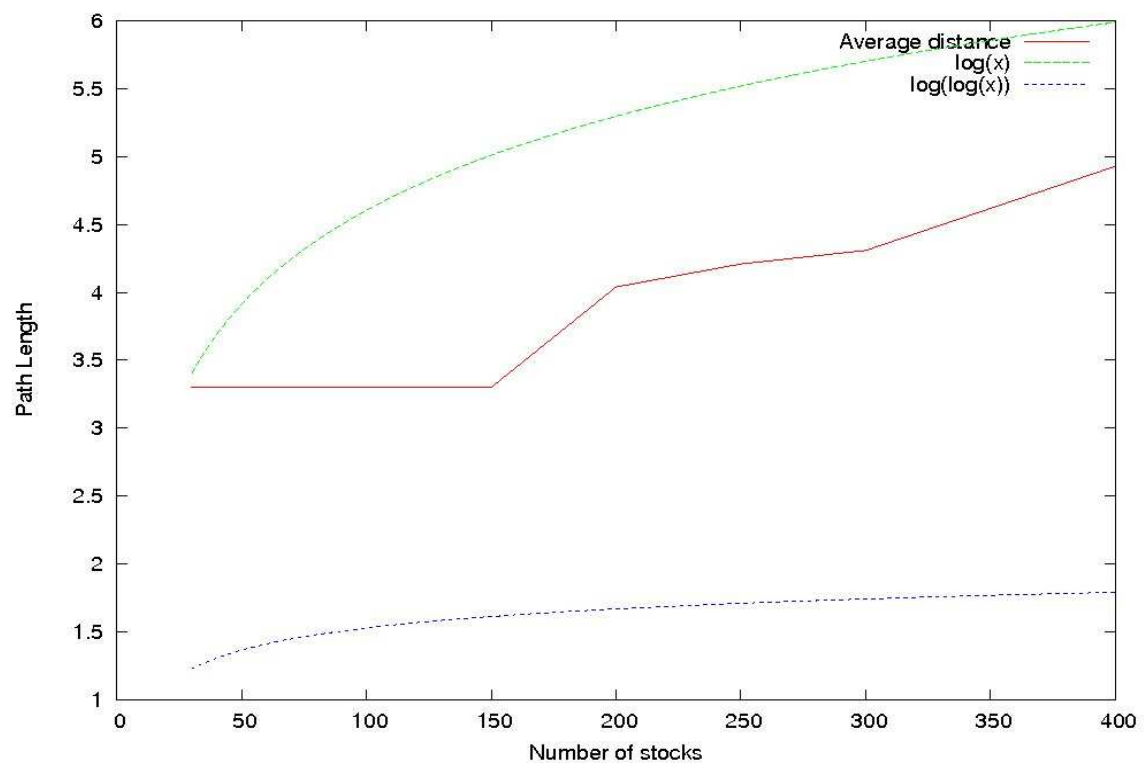


Figure 2.9: Growth of distance with network size

CHAPTER 3

METRIC DISTANCE IN MINIMUM SPANNING TREE

3.1 How do we measure how close stocks are?

In the previous chapter the notion of distance d_{ij} was introduced. This measurement was used to generate the minimum spanning tree. The distance represents how closely stocks move together based on their correlation. This chapter will explore the derivation of this distance and how to interpret it. Another way of measuring distance will be introduced and a comparison of the two will be done.

3.2 Derivation of metric measurement $d_{ij} = \sqrt{2(1 - \rho_{ij})}$

Let the i^{th} stock return with time scale Δt .

$$x_i(t; \Delta t) = \frac{S_i(t)}{S_i(t - \Delta t)} - 1 \quad (3.1)$$

In most literature on correlation networks Δt is the daily return. If we collect all the returns over a period of time T then we can generate a stock return vector for each stock defined as [6]

$$\vec{x}_i = \begin{pmatrix} x_i(t) - \mu_i \\ x_i(t-1) - \mu_i \\ \cdot \\ \cdot \\ x_i(t-T+1) - \mu_i \\ x_i(t-T) - \mu_i \end{pmatrix}$$

The normalized stochastic return vector for i^{th} stock is defined as

$$\vec{x}_i = \begin{pmatrix} (x_i(t) - \mu_i)/\sigma_i \\ (x_i(t-1) - \mu_i)/\sigma_i \\ \cdot \\ \cdot \\ (x_i(t-T+1) - \mu_i)/\sigma_i \\ (x_i(t-T) - \mu_i)/\sigma_i \end{pmatrix}$$

The metric distance between the i^{th} stock and j^{th} stock was defined by Mantega using the relationship

$$d_{ij} = \sqrt{\sum_{k=0}^T \left(\frac{x_{ik}}{|\vec{x}_i|} - \frac{x_{jk}}{|\vec{x}_j|} \right)^2} \quad (3.2)$$

$$d_{ij} = \sqrt{\sum_{k=0}^T \left(\frac{x_{ik}^2}{|\vec{x}_i|^2} + \frac{x_{jk}^2}{|\vec{x}_j|^2} - \frac{2x_{ik}x_{jk}}{|\vec{x}_i||\vec{x}_j|} \right)^2} \quad (3.3)$$

Since $\sum_{k=0}^T x_{ik}^2 = |\vec{x}_i|^2$ and $\sum_{k=0}^T x_{jk}^2 = |\vec{x}_j|^2$ then d_{ij} can be simplified to

$$d_{ij} = \sqrt{2 \left(1 - \sum_{k=0}^T \frac{x_{ik}x_{jk}}{|\vec{x}_i||\vec{x}_j|} \right)} \quad (3.4)$$

Using the correlation matrix $\rho_{ij} = \sum_{k=0}^T \frac{x_{ik}x_{jk}}{|\vec{x}_i||\vec{x}_j|}$ we can define the distance as

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \quad (3.5)$$

In this thesis a non-normalized distance will also be used to see which distance (normalized or non-normalized). If the distance is non-normalized then

$$d_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2 + 2\sigma_i\sigma_j\rho_{ij}} \quad (3.6)$$

An advantage of the normalized distance is that all distances lies within the range

$$0 \leq d_{ij} \leq 2 \quad (3.7)$$

Since the correlation coefficients lie within

$$-1 \leq \rho_{ij} \leq 1 \quad (3.8)$$

The non-normalized distance is less bounded and lies within the range

$$\sigma_i - \sigma_j \leq d_{ij} \leq |\sigma_i + \sigma_j| \quad (3.9)$$

The normalized distance is therefore more advantageous because

1. The range of values lie within 0 and 1
2. Since the derivation of both is so similar it has little impact on the network topology
3. If we are given the distance measure d_{ij} is easier to deduce the ρ_{ij} .

3.3 Conclusion

This chapter showed how the metric distance can be derived from stock price returns. The normalized d_{ij} can be used to derive ρ_{ij} . This will improve important in the next section as correlation networks are used to reconstruct the correlation matrices.

CHAPTER 4

USING NETWORKS TO OPTIMIZE PORTFOLIOS

4.1 Modern Portfolio Theory

The traditional model for portfolio optimization comes from the Modern Portfolio Theory (MPT) proposed by Harry Markowitz . In this model the return of a portfolio characterised by the mean (average) of the return of a historical period. The risk is defined as the standard deviation of these sampled returns. If we define r_p as the return and σ_p then we define the risk of a portfolio as

$$\sigma_p^2 = \sum_{i=0}^N \sum_{j=0}^N p_i p_j \sigma_{ij} \quad (4.1)$$

using the constraint

$$\sum_{i=0}^N p_i = 1 \quad (4.2)$$

The portfolio consists of N assets each assigned a weight p_i . The quantity σ_{ij} forms the covariance matrix that defines how each assets mean return is correlated the other. The goal of MPT is to deliver a portfolio that minimizes the risk σ_p for a given return r_p . This results is a quadratic equation to deliver an optimal portfolio of weights p_i which meets the constraint in eq 4.1. The solution to this problem was defined by Markowitz as.

$$\mathbf{p}^* = \lambda \Sigma^{-1} \mathbf{1}^T + \gamma \Sigma^{-1} \mathbf{m} \quad (4.3)$$

Where Σ is the covariance matrix, $\mathbf{1}^T = (1, \dots, 1)$ and \mathbf{m} is the vector containing the mean returns of each asset. The other parameters can be defined using r_p which is the expected return as

$$\begin{aligned} \lambda &= \frac{C - r_p}{\Delta} & \gamma &= \frac{r_p A - B}{\Delta} \\ A &= \mathbf{1}^T \Sigma^{-1} \mathbf{1} & B &= \mathbf{1}^T \Sigma^{-1} \mathbf{m} \\ C &= \mathbf{m}^T \Sigma^{-1} \mathbf{m} & \Delta &= AC - B^2 \end{aligned}$$

If we graph the risk-return graph of these optimal solution we get what is known as the Markowitz bullet. In Figure 4.1 of a three stock portfolio. The optimised portfolio forms the parabola around the random points (unoptimized portfolios).

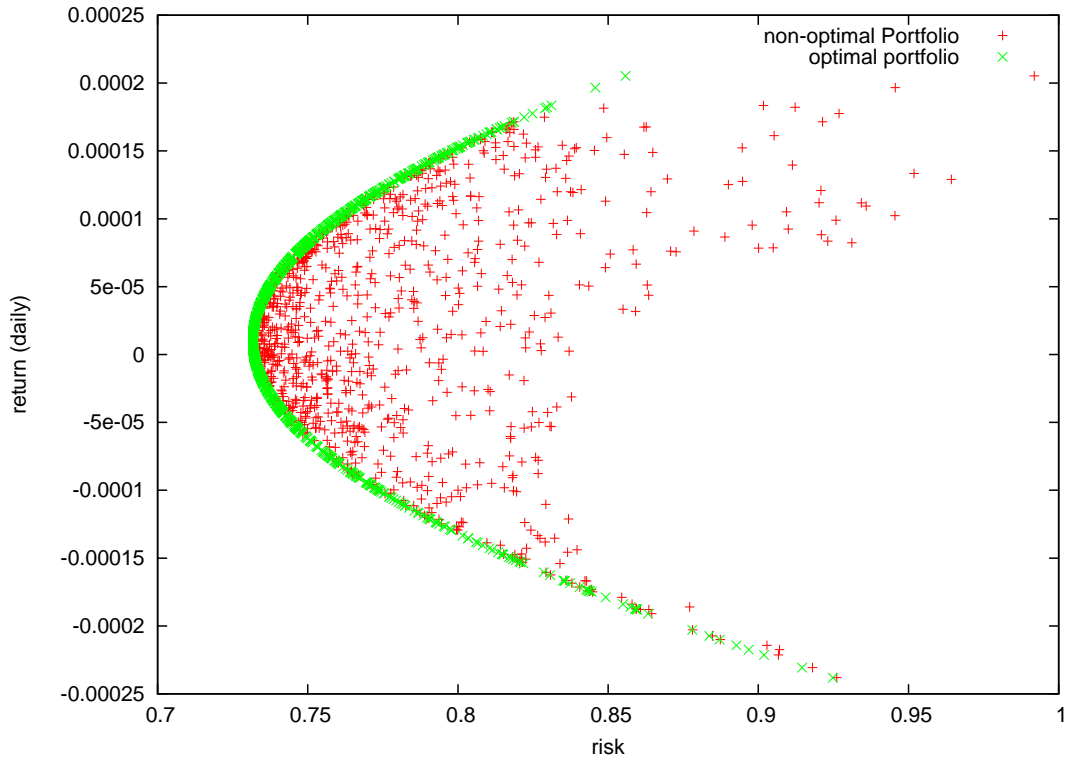


Figure 4.1: Graph of Risk-Return for optimal and non-optimal portfolios

One important assumption made in this model is that the returns are normally distributed and therefore the first and second moments of the distribution suffice (mean and standard deviation). Recent work has shown that this assumption of a normally distributed return for most assets does not hold up to real financial data. Most returns instead show a return that is a fat tail which means MPT would underestimate the occurrence of extreme moves in returns of a portfolio. Another assumption made by the MPT is that the correlation matrix accurately represents the correlation between different assets. However recent studies have shown that the matrix is actually dominated by noise and only a small amount of information is contained. Different methods have therefore been proposed to "clean" these

correlation matrix. Two common techniques have been proposed, Random Matrix Theory and clustering algorithms. In order to compare how efficient which technique is we will compare the predicted risk of an optimal portfolio with the realized risk that is observed after a specified period. So we define

$$R(t) = \frac{|\sigma_{p,t+1} - \sigma_{p,t}|}{\sigma_{p,t}} \quad (4.4)$$

The risk-return curve over the period will also reveal how "clean" or predictive the correlation matrix is

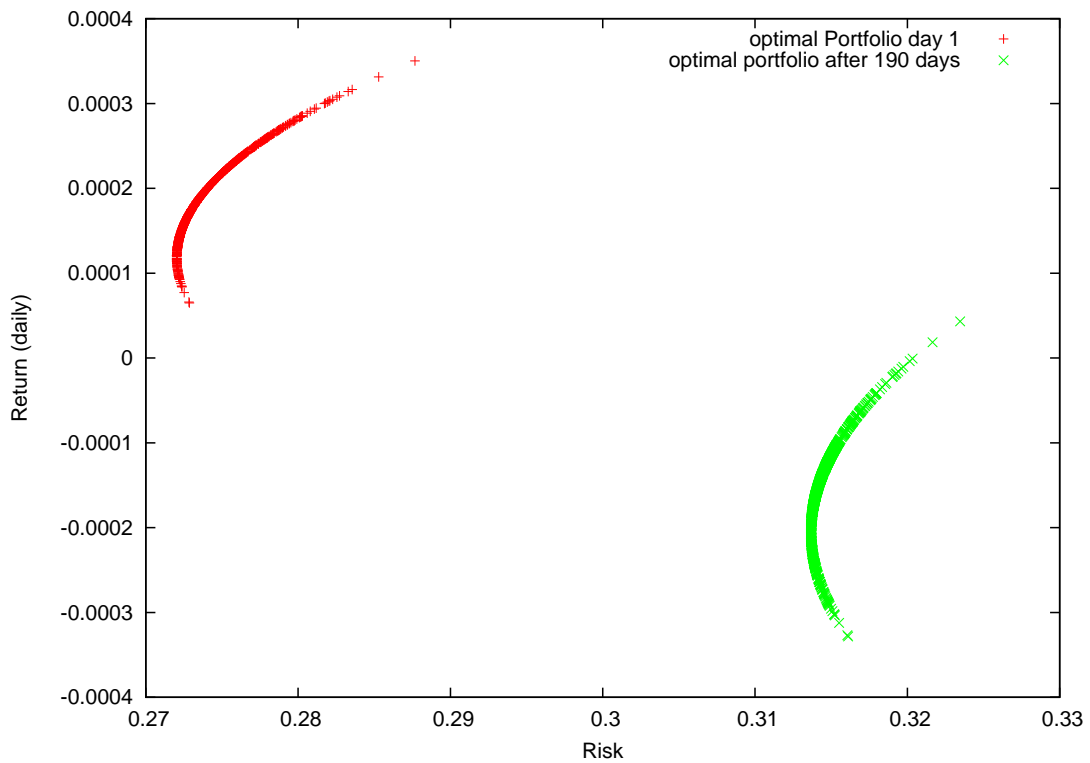


Figure 4.2: Graph of Risk Return for optimal portfolio at time $t=1$ and time $t=190$

4.2 Using Random Matrix Theory to clean correlation matrices

The first is the Random Matrix Theory which is an import from nuclear physics. It involves decomposing the correlation matrix into eigenvalues and eigenvectors. The basic idea is that most of the information is contained in the larger eigenvectors. Therefore using a determined limit we can eliminate eigenvectors that can be explained as noise. We then reconstruct the correlation matrix using only the largest eigenvectors.

More formally the procedure involves decomposing the matrix C with $N \times N$ dimensions into $\lambda_1 \dots \lambda_n$ eigenvalues and their respective eigenvectors $\vec{v}_1 \dots \vec{v}_N$. The eigenvalues and corresponding eigenvectors are sorted such that

$$\lambda_k < \lambda_{k+1} \quad (4.5)$$

If Λ is defined as $\Lambda = \text{diag}(\lambda_i)$ where the Λ is an $N \times N$ matrix with all zeros except for the diagonals which are set to corresponding eigenvalues.

$$\Lambda_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \lambda_i & \text{if } i = j \end{cases}$$

The eigenvectors are also used to construct a matrix $W = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_N)$. These two matrices can then be used to reconstruct the original matrix C using the equation

$$C = W \Lambda W \quad (4.6)$$

This is where Random Matrix Theory comes into play. According to the theory most of the information is contained in the largest eigenvectors. Eigenvalues and corresponding eigenvectors that lie within a range $]\lambda_{min}, \lambda_{max}[$ are considered noise and the range is calculated as.

$$\lambda_{min}^{max} = \sigma^2 \left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \right) \quad (4.7)$$

where $Q = \frac{T}{N}$, T is the interval, N is the number of variables in our case stocks

$$\sigma^2 = 1 - \frac{\lambda_1}{N} \quad (4.8)$$

and λ_1 is the smallest eigenvalue.

In order to 'clean' the matrix the eigenvectors with corresponding eigenvalues within the range $]\lambda_{min}, \lambda_{max}[$ are set to zero and a new Λ^* matrix is used to get the new clean matrix C^* using the equation

$$C^* = W\Lambda^*W \quad (4.9)$$

For a correlation matrix the diagonals of the cleaned matrix are set to one to keep it meaningful. To illustrate this procedure we will use a matrix

$$C = \begin{pmatrix} 1 & 0.20658 & 0.332996 & 0.30024 \\ 0.20658 & 1 & 0.377469 & 0.410961 \\ 0.332996 & 0.377469 & 1 & 0.504519 \\ 0.30024 & 0.410961 & 0.504519 & 1 \end{pmatrix}$$

Then the matrix eigenvalues are

$$\lambda_i = \begin{pmatrix} 0.49157 & 0.616292 & 0.807498 & 2.08464 \end{pmatrix}$$

and the Λ matrix is

$$\Lambda = \begin{pmatrix} -0.0605881 & 0.343165 & -0.842199 & 0.411421 \\ 0.10458 & 0.705514 & 0.513184 & 0.477446 \\ 0.679824 & -0.48672 & 0.0205718 & 0.548197 \\ -0.723348 & -0.384176 & 0.164072 & 0.54978 \end{pmatrix}$$

The time window (in days) used to create this correlation matrix was $T = 800$ and since there are only four stocks $Q = \frac{T}{N} = 200$ and using equation 4.6 and 4.7 we have $\lambda_{min} = 0.849885$ and $\lambda_{max} = 1.12824$. After eliminating eigenvalues between this range then

Λ^* becomes

$$\Lambda^* = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.08464 \end{pmatrix}$$

and the new correlation matrix C^* using the equation 4.8 is

$$C^* = \begin{pmatrix} 0.352862 & 0.409489 & 0.47017 & 0.471527 \\ 0.409489 & 0.475203 & 0.545622 & 0.547197 \\ 0.47017 & 0.545622 & 0.626477 & 0.628285 \\ 0.471527 & 0.547197 & 0.628285 & 0.630098 \end{pmatrix}$$

Setting the diagonals to one gives the new matrix

$$C^* = \begin{pmatrix} 1 & 0.409489 & 0.47017 & 0.471527 \\ 0.409489 & 1 & 0.545622 & 0.547197 \\ 0.47017 & 0.545622 & 1 & 0.628285 \\ 0.471527 & 0.547197 & 0.628285 & 1 \end{pmatrix}$$

This new matrix C^* can be used in the portfolio optimization in the place of Σ to generate \mathbf{p}^* the optimal portfolio. The risk associated with this portfolio σ_t^* is then compared with the observed risk after a period of time σ_{t+T}^* . The risk is calculated as

$$\sigma_t^* = \sum_{i=0}^N \sum_{j=0}^N p_i^* p_j^* \sigma_{ij,t} \quad (4.10)$$

and

$$\sigma_{t+T}^* = \sum_{i=0}^N \sum_{j=0}^N p_i^* p_j^* \sigma_{ij,t+T} \quad (4.11)$$

The 30 stock portfolio would look like figure 4.3. It is important to note how the optimised curve becomes more risky and has less returns after time interval

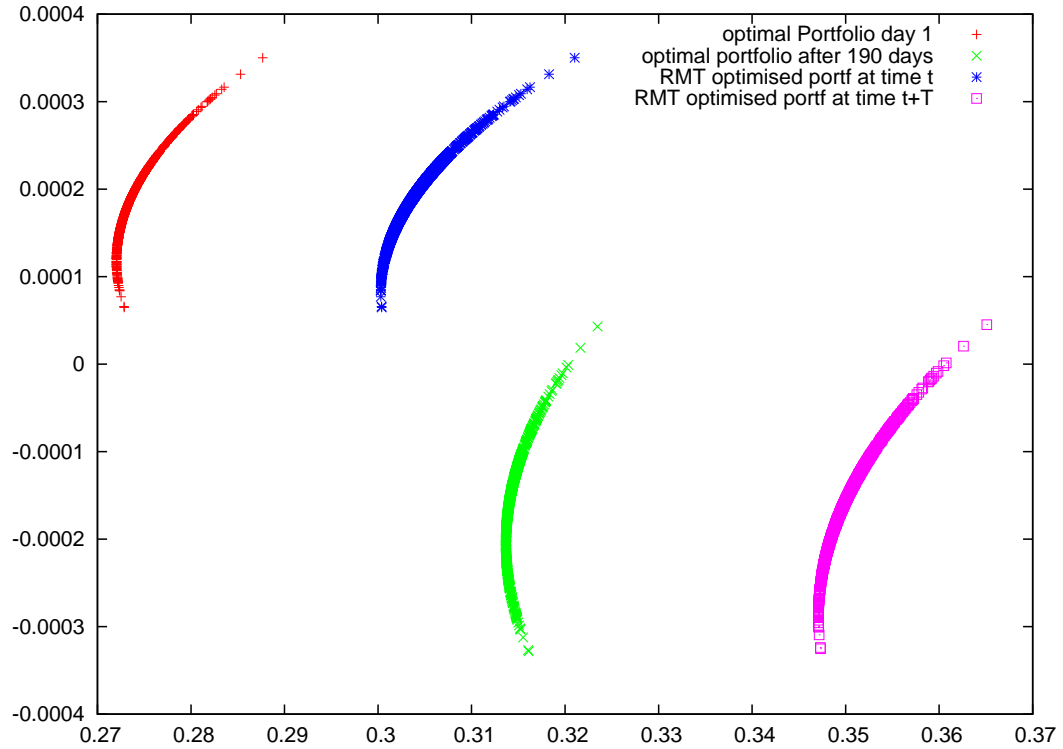


Figure 4.3: Graph of Risk-Return for optimal portfolios with RMT cleaned matrix in time t and time $t+T$ compared to unaltered correlation matrix

4.3 Using clustering to clean correlation matrices

Another technique used to clean correlation matrices are clustering algorithms. Generally clustering techniques involve grouping assets in our case stocks based on a predefined criteria. By grouping stocks into clusters one reduces the dimensionality of the problem from $N(N-1)/2$ to $N-1$. This can be clearly seen from a minimum spanning tree where for N vertices there are $N-1$ edges. Reducing dimensionality creates a matrix that is less sensitive to noise and should ideally give us a more robust measure of risk.

Two types of clustering algorithms will be studied. They are the Single Linkage Cluster Analysis (SLCA) and Average Linkage Cluster Analysis (ALCA). The two techniques are very similar in that they not only create clusters but reveal a hierarchy of the stock market

that should closely resemble different sectors. They differ in how they measure the distance between a stock and a cluster. The distance being the previously defined metric distance

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

4.4 Ultrametric distance

Before dealing with the details of the clustering algorithms an important concept known as ultrametricity should be introduced. If three objects x, y, z exist in a metric space then a distance $d(x, y)$ can be defined that meets the following requirements

$$\begin{aligned} d(x, y) &> 0 & \text{for} & & x \neq y \\ d(x, y) &= 0 & \text{for} & & x = y \\ d(x, y) &= d(y, x) & & & \forall x, y \\ d(x, y) &\leq d(x, z) + d(y, z) & (\forall x, y, z) & \text{triangle inequality} \end{aligned}$$

For an ultrametric space the triangle inequality is more strict and is defined as

$$d(x, y) \leq \max(d(x, z), d(y, z)) (\forall x, y, z) \quad (4.12)$$

Ultrametric distances are important to clustering since they redefine the distance between two objects as the distance between their closest ancestors. If we generate a metric distance matrix D with $N \times N$ dimensions there will be $N(N - 1)/2$ unique distances. However an ultrametric distance matrix can be defined by only $N - 1$ distances

4.5 Single Linkage Cluster Analysis

In order to generate an ultrametric matrix using a single linkage cluster analysis (SLCA) one uses the following procedure.

1. Begin with N objects with distance matrix D as a distance matrix

2. Find the closest objects $x, y \in N$ s.t $d(x, y) = \min(d(i, j)) \forall i, j \in$ and $i \neq j$
3. Merge the objects into a cluster $u = (x, y)$ and recalculate the distance matrix D using the new object u and removing the x, y columns row and columns and adding a u column. The distance for any $z \in N$ to u are calculated as $d(z, u) = \min(d(z, x), d(z, y))$.
4. Repeat step one with $N - 1$ clusters (remaining objects plus new cluster) and new distance matrix. If $N = 1$ then stop

The ultrametric tree or dendrogram generated by the SLCA directly corresponds to the tree produced by the Minimum Spanning Tree algorithm.

4.6 Average Linkage Cluster Analysis

The Average Linkage Cluster analysis (ALCA) is very similar to the SLCA in the construction except when it comes to calculating the distance between the new cluster $u = (x, y)$ and z in step 3. To calculate the distance on uses the equation

$$d(z, u) = \frac{1}{2}(d(z, x) + d(z, y)) \quad (4.13)$$

The risk-return curves generated by the SLCA and ALCA were similar for the Dow 31 stocks plus index used. They both maintained the same risk profile even after 190 days

Using equation 4.4 we can show that R , the measure of how risk predicted varies with actual result is lowest for portfolios generated with cleaned matrices from ALCA.

4.7 Observations and conclusions

The correlation matrix cleaning done using clustering method yielded more dramatic results than the RMT cleaning process. This may go to show that the cluster method is better suited for cleaning that RMT. However RMT has been used in larger correlation matrices for

example with 1000 stock portfolio's and has produces better risk profiles[8]. This is because the statistical assumptions underlying random matrix theory such eigenvalue distribution are best observed with large $N \times N$ matrices. The removal of a large bulk of eigenvectors using RMT has also been shown to eliminate a lot of vital information from the correlation matrix [9]. This therefore suggests that clustering techniques may be more suited for correlation cleaning especially for smaller portfolios

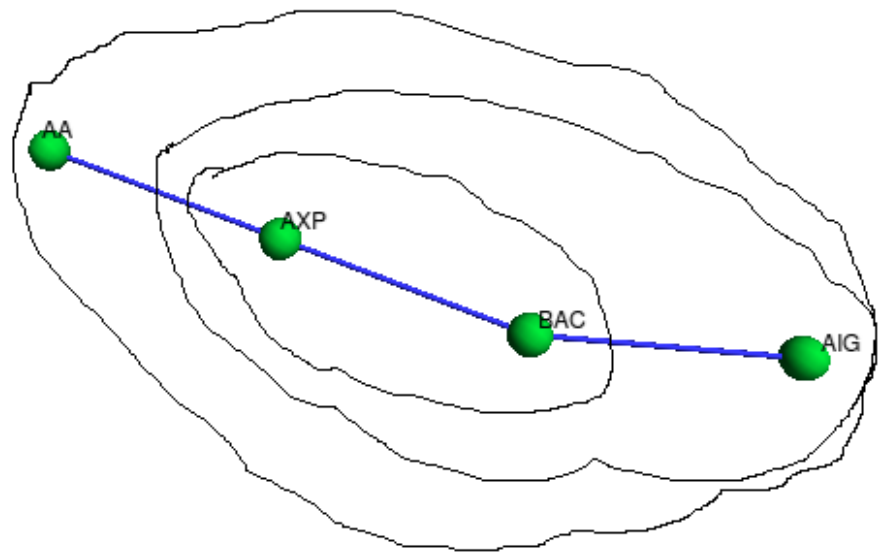


Figure 4.4: Minimum Spanning Tree for four stocks the circles represent the clusters formed

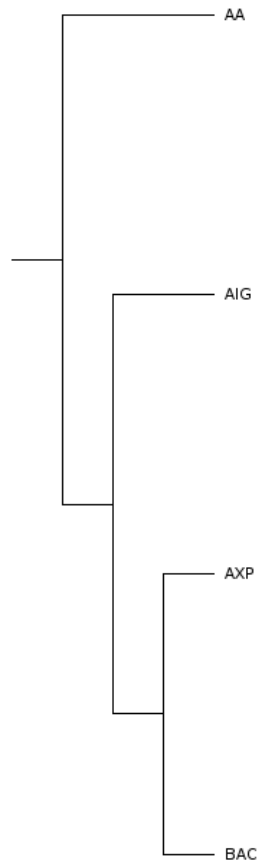


Figure 4.5: Dendrogram formed using the metric distances between stocks using ALCA

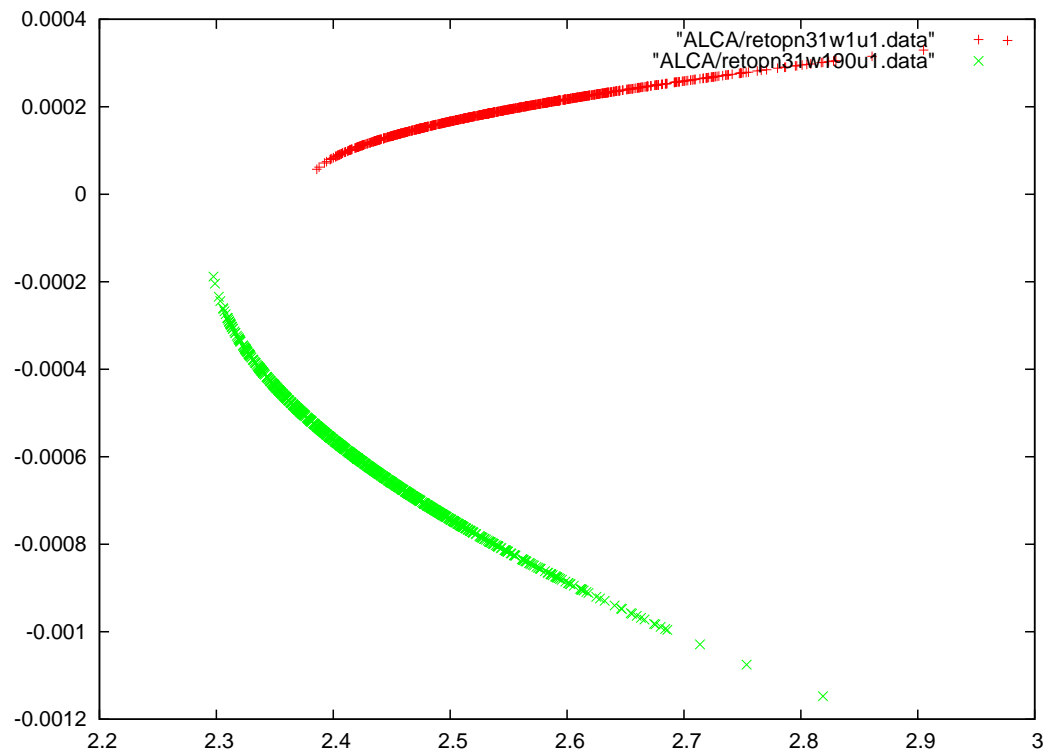


Figure 4.6: Risk-return curves generated from ALCA cleaned matrix at on day 1 and day 190

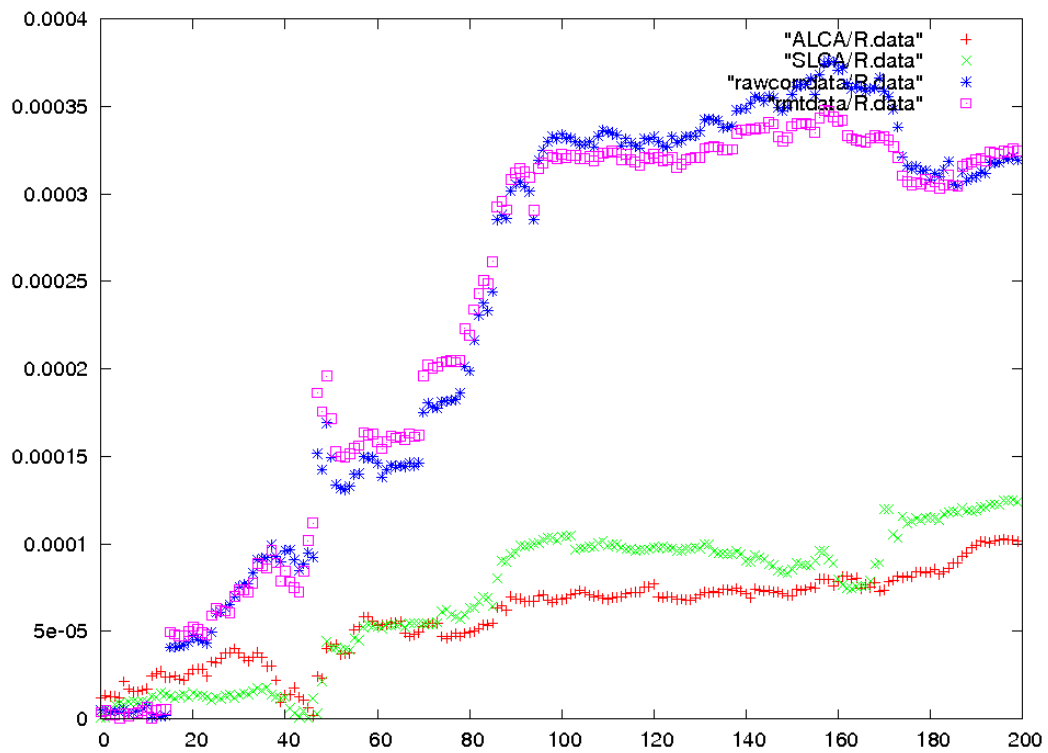


Figure 4.7: Measure of how much predicted risk differed from observed risk over a period of 190 days

CHAPTER 5

CORRELATION NETWORKS AND STOCK MARKET DYNAMICS

5.1 Stock Market Criticality

Critical behaviour is observed in systems that are undergoing a phase transition from one state to another for example transition from solid to liquid or ferromagnetic demagnetization at critical Curie temperatures. The stock market has been shown to exhibit critical behaviour more precisely self organized criticality where the critical behaviour emerges from the interactions with elements within the system as opposed to heat based transitions where an external effect is added.

Self organized criticality is characterised by a system undergoing sharp movements (critical points) beyond the expected Gaussian expectations. In the stock market these are characterised by large crashes and rallies such as Black Monday (October 19, 1987) as well as the August 2007.

5.2 Correlation Network Size and Market Movements

Researchers like Onnela have proposed using correlation networks to predict large market movements. By plotting the total distance of the tree $w(T) = \sum_{(u,v) \in T} w(u,v)$ over time and looking for sharp movements in the graph Onnela noticed that during 1987-1988 the average total distance dropped at the same time the stock market experienced a crash and then returned to previous levels.

A similar observation was made by Andrew Lo when he studied the huge losses that occurred to simulated hedge fund strategies in August 2007. In the previous month (July) the

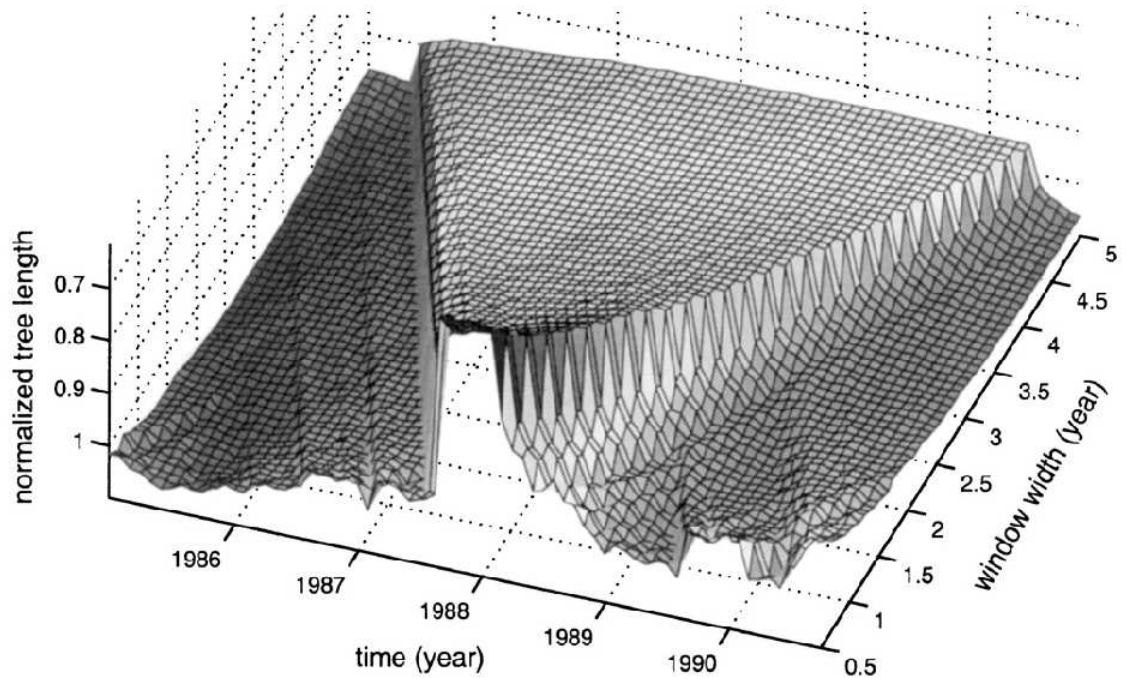


Figure 5.1: Crash of 1987

two hedge fund owned by Bear Stearns lost all their value. Within days hedge funds invested in non-mortgage related assets also took large losses. Interestingly most of these hedge funds used quantitative strategies that traded a large variety of assets based on "statistical arbitrage" and should therefore be immune to market risk (beta).

Lo measured the correlation matrices of these assets and generated correlation networks before and after August 2007. Thick edges in the graph represent correlations $> 50\%$. If the month of August was removed from the calculation then the strength of the correlation between strategies reduced significantly.

From his analysis Lo concludes that hedge fund losses in August of 2007 were due to

1. The losses to quant funds during the second week of August 2007 were initiated by the temporary price impact resulting from a large and rapid unwinding of one or more quantitative equity market-neutral portfolios. The speed and magnitude of the price impact suggests that the unwind was likely the result of a sudden liquidation of a multi-strategy fund or proprietary-trading desk, perhaps in response to margin calls from a deteriorating credit portfolio, a decision to cut risk in light of current market conditions, or a discrete change in business lines
2. The price impact of the unwind on August 78 caused a number of other types of equity funds long/short, 130/30, and long-only to cut their risk exposures or de-leverage, exacerbating the losses of many of these funds on August 8th and 9th

In other words a single hedge fund tried sell the assets it sold and this triggered other hedge funds to do the same causing losses. Since the hedge funds were heavily leveraged (borrowed money to invest) this forced hedge funds to sell due to margin requirements (a hedge fund can only lose a certain amount, margin, otherwise the creditors will demand their money. This leads to an event similar to a run on the bank where all depositors want their money at once which could lead to the collapse of a financial institution.

5.3 Dow Jones Index and the correlation network of DJIA 30 Stocks

To measure how the correlation distance changed over time the historical data for the DJIA 30 (9/7/1996-8/27/2008) was downloaded from finance.yahoo.com. The daily return $t = 1$ was used and the historical window T was varied from monthly (25 days) to quarterly (75 days) and yr (300) and 2yr (600) and 3yr (900). The results showed that with larger windows (1 year, 2 year and 3 year) there were less fluctuations but the increase and reduction in the size of the networks was large. This sharp lengthening and contractions

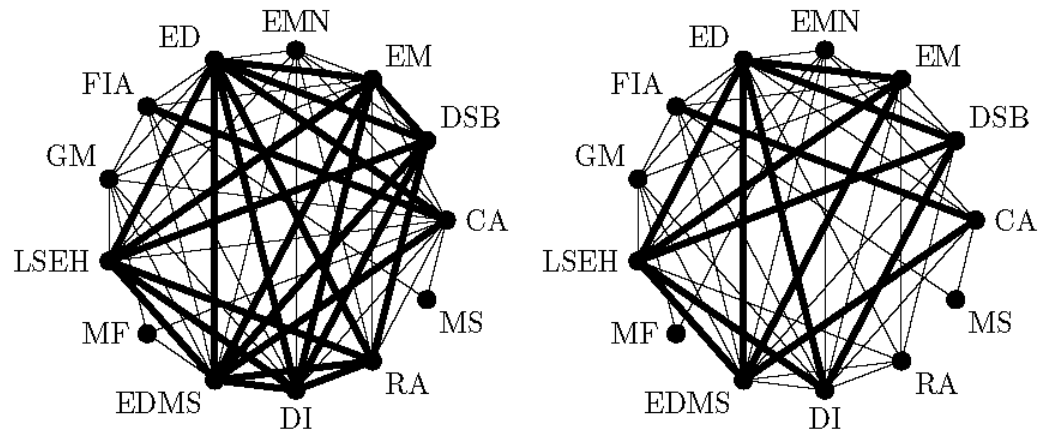


Figure 5.2: Hedge fund strategy correlation strengths (thick black lines $\geq 50\%$) calculated with and without August 2007

suggest changes in stock correlation happen over short time periods. The contraction of the correlation network length also preceded by sharp movements in the Index.

5.4 Conclusion

The size of the correlation network could prove useful in predicting sharp market movements. A smaller size correlation network means that the market as a whole is moving sharply in one direction.

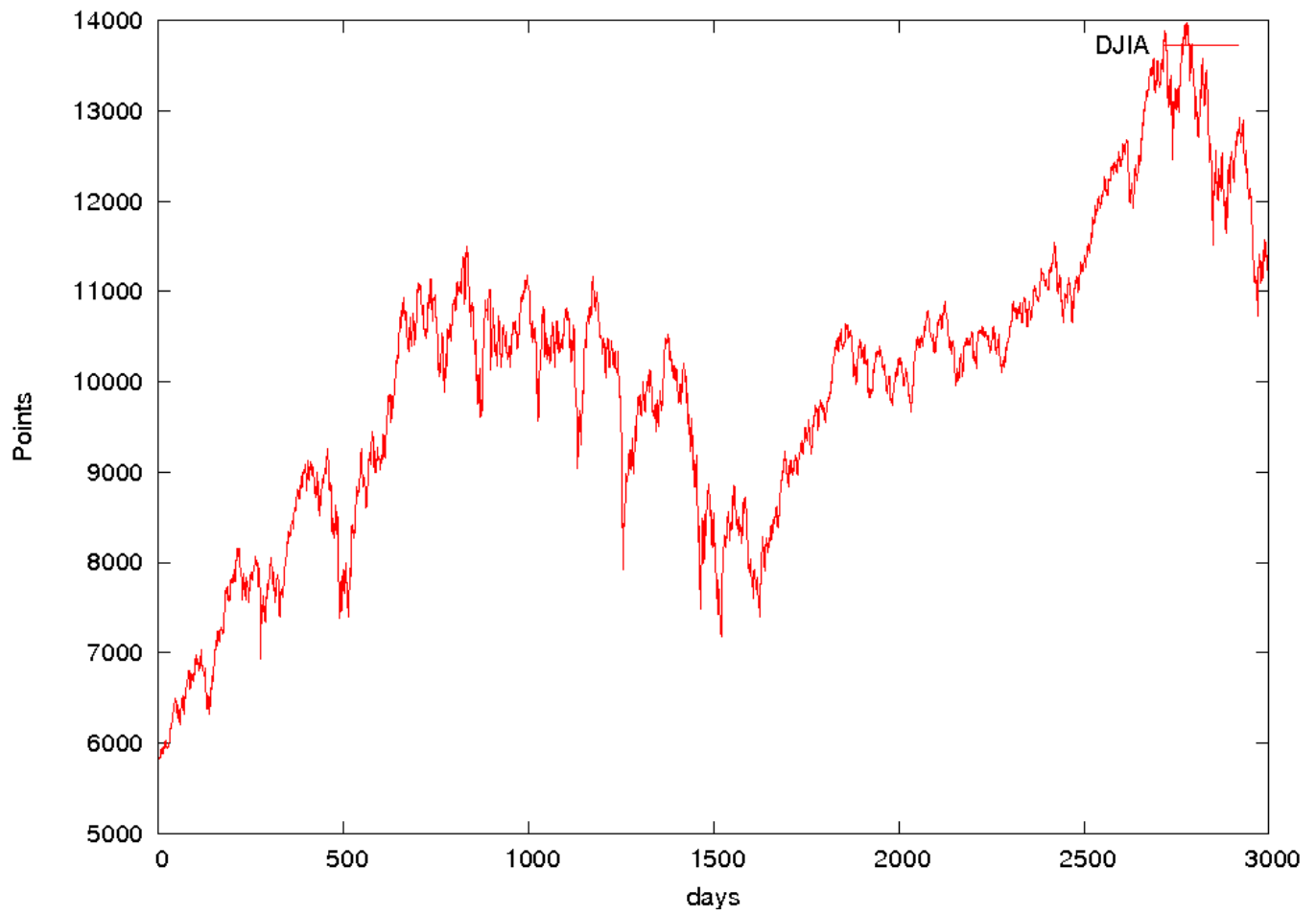


Figure 5.3: DJIA from (9/7/1996-8/27/2008)

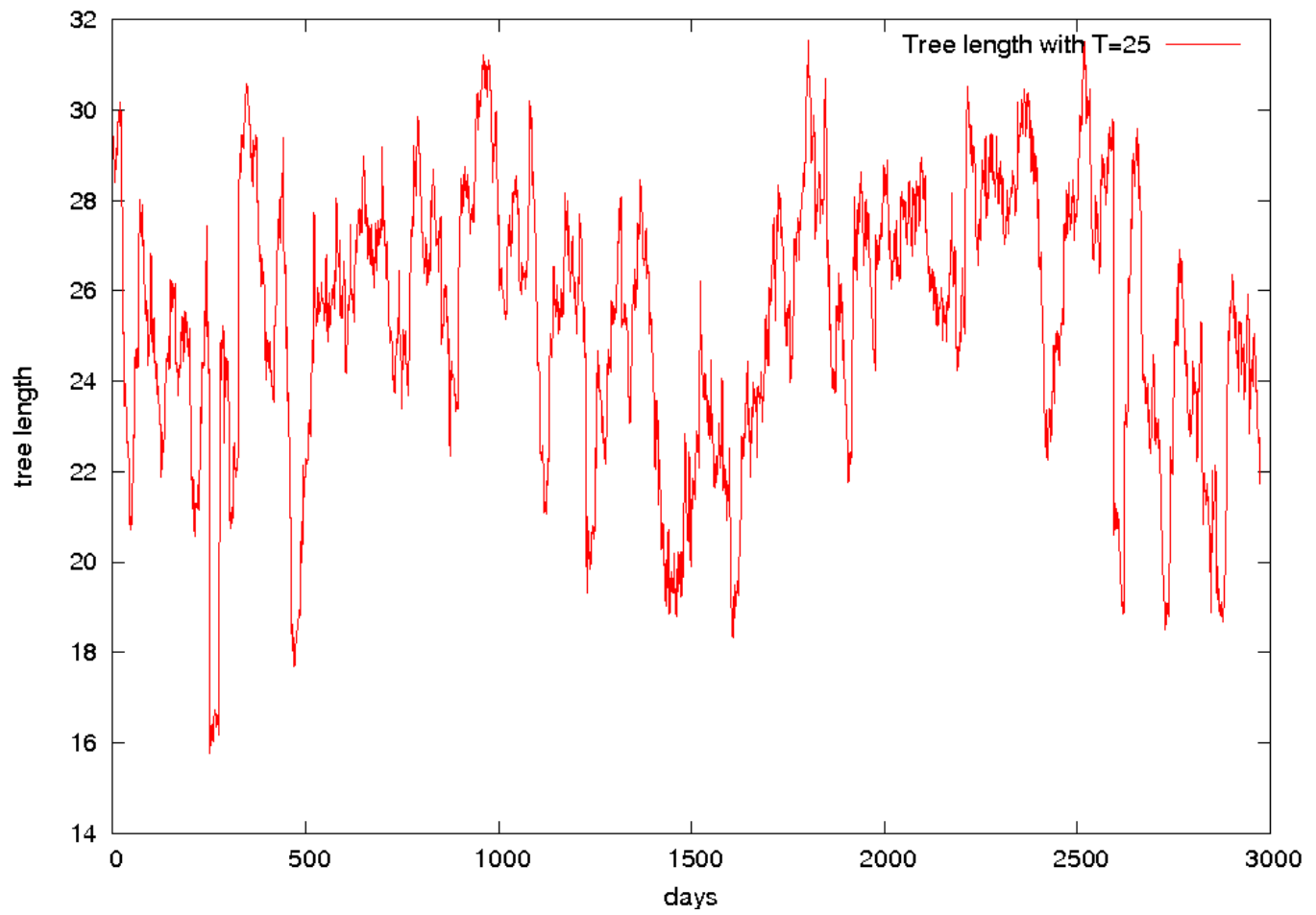


Figure 5.4: Measure of Tree Length over time using $T=25$

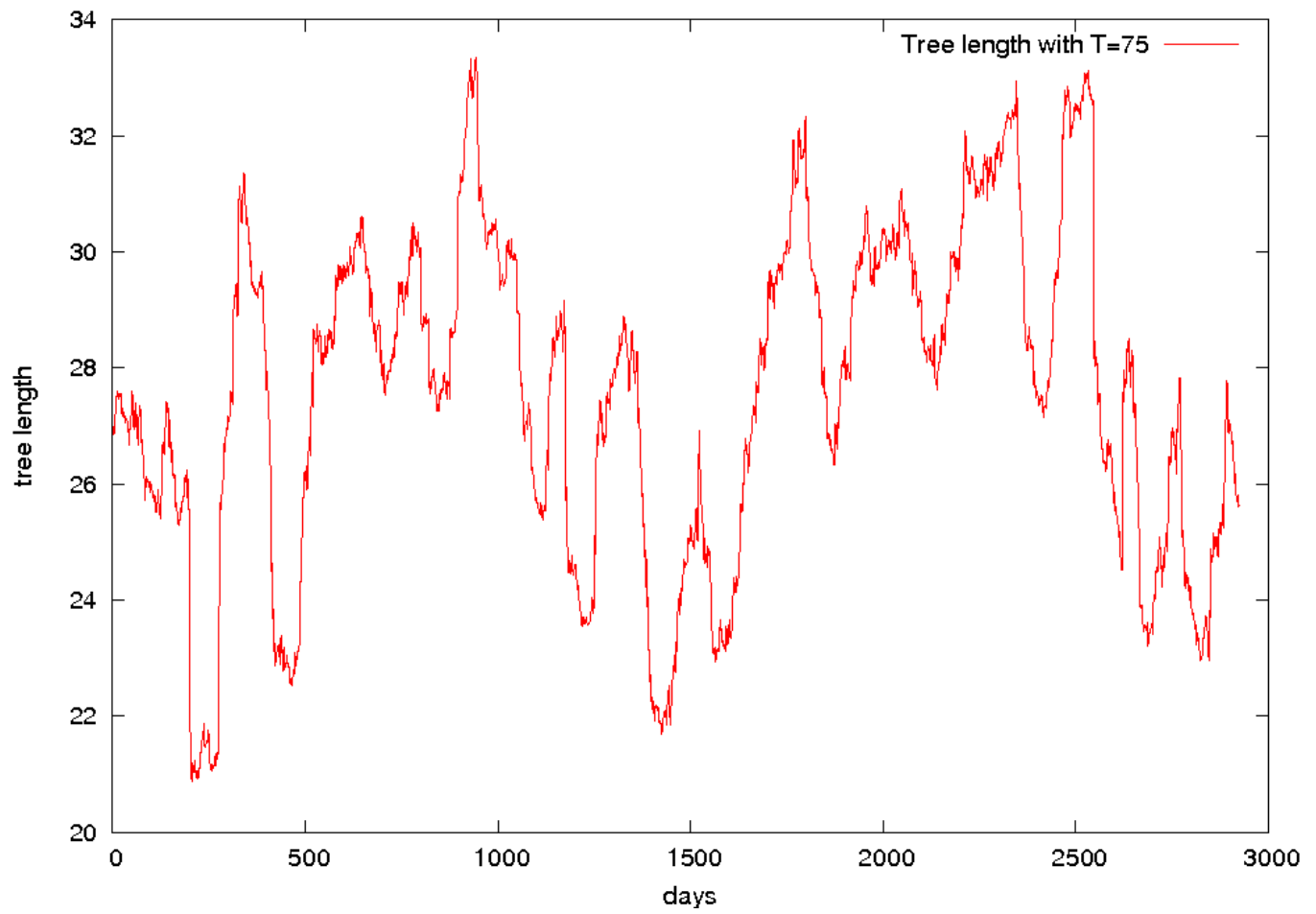


Figure 5.5: Measure of Tree Length over time using $T=75$

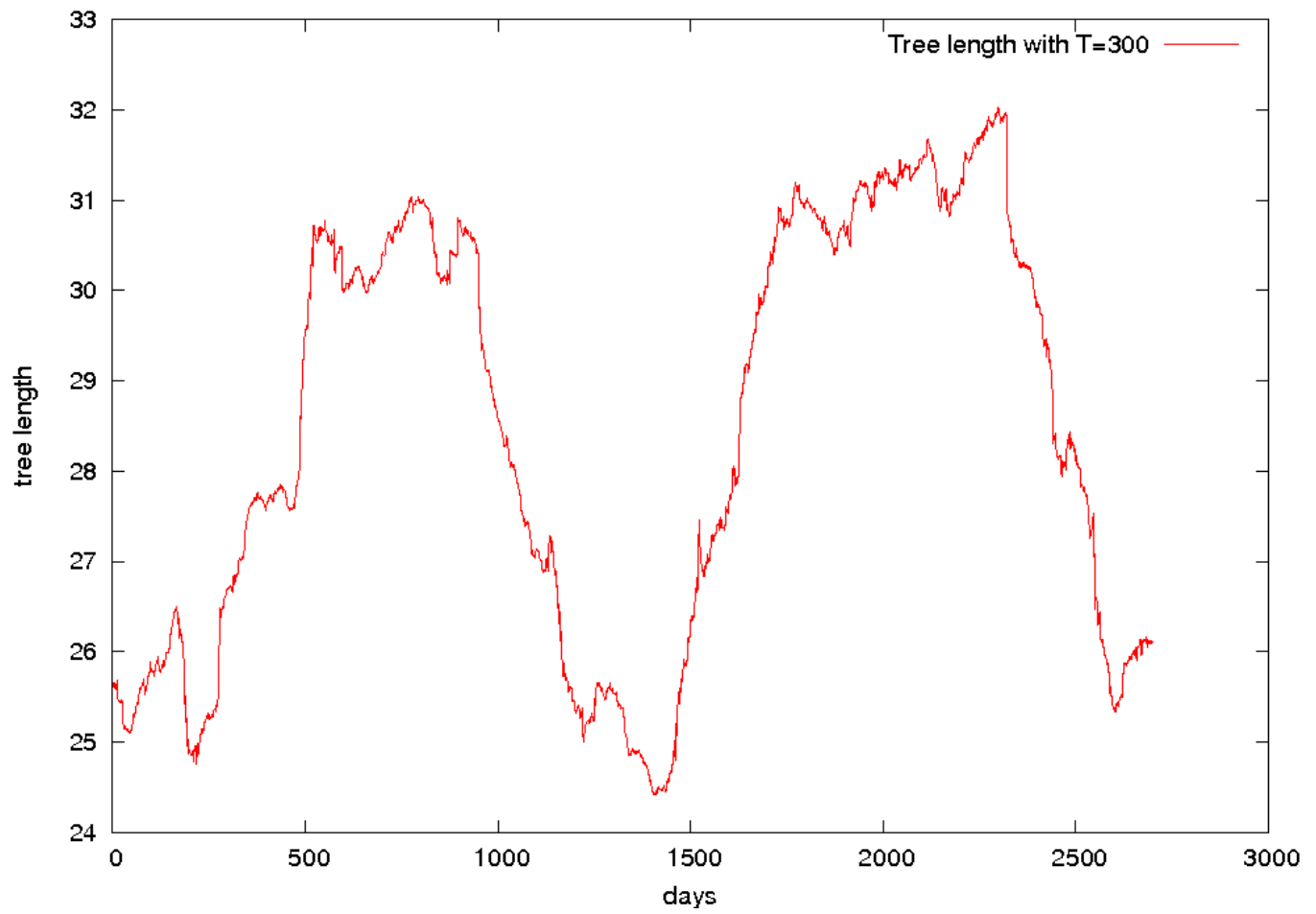


Figure 5.6: Measure of Tree Length over time using $T=300$

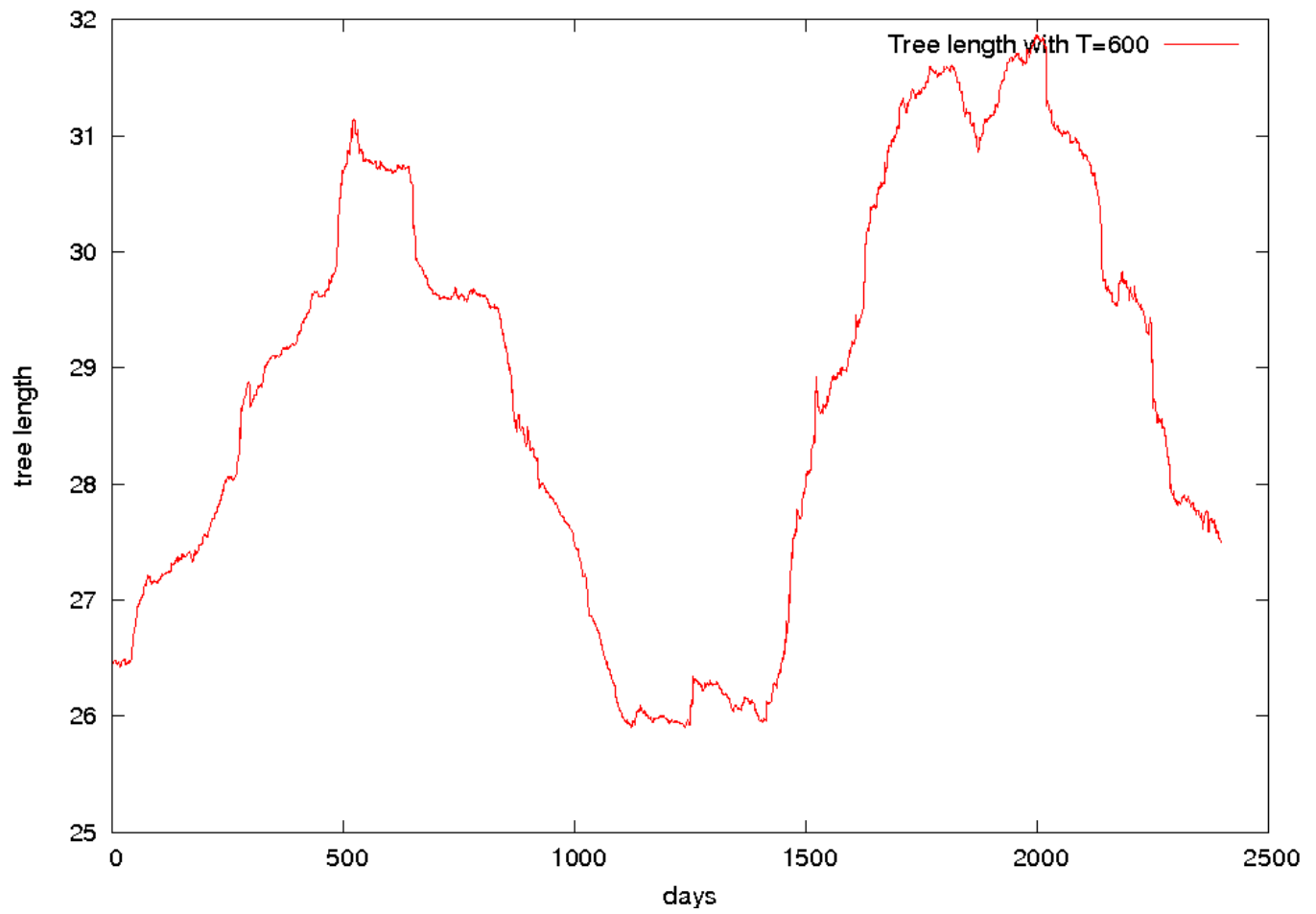


Figure 5.7: Measure of Tree Length over time using $T=600$

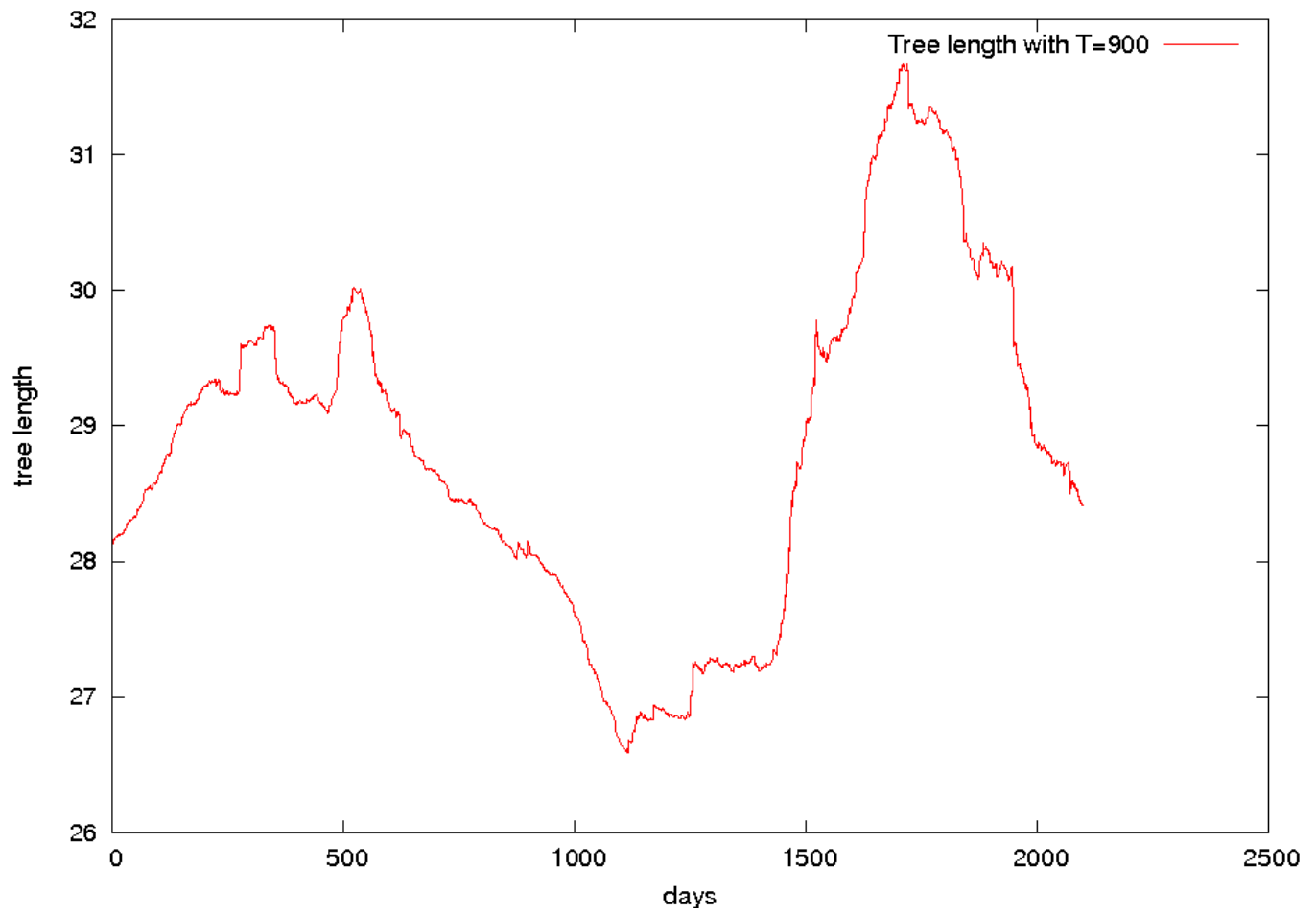


Figure 5.8: Measure of Tree Length over time using $T=900$

CHAPTER 6

CONCLUSION

6.1 Summary

This thesis has shown how correlation networks generated from Minimum Spanning Trees can be applied to different problems in finance

6.2 Visualization of the hierarchy and interconnectedness of the stock market

The correlation networks generated exposed the stock market hierarchy that closely resembled different sectors of the stock market. That is, stocks within the same sector were more likely to be closer together. This was demonstrated by the financial sector stocks

The correlation network can be used to show which stocks or entities play a key role in the movements of stock prices. A clear example of this is the way in which the Dow Index forms the center of the network of the 30+1 DJIA stocks. The SP 500 index GSPC also forms the center of the network of 400+ stocks from the SP 500.

Another property that the networks illuminate is the way in which a few stocks have a large number of connections represented by the power law degree distribution. Stocks with higher degrees (more edges) may be considered market movers especially the large capital stocks.

6.3 Portfolio Optimization

The weighted adjacency matrix of the correlation network can be used to clean the correlation matrix. The adjacency matrix is used to cluster stocks together and since this

creates a matrix with less coefficients, it is considered cleaner than the raw correlation matrix. Portfolios created using these cleaner matrices have a risk return curve that does not diverge as much as portfolios created by other matrices.

The clustering cleaning technique was compared to the raw correlation matrix and the Random Matrix theory technique (RMT). For smaller stocks the clustering technique outperformed the RMT. However the technique was less useful for large portfolios.

6.4 Sharp market movements

The size of the correlation network may also be useful in identifying an oncoming sharp market movement. This is because a smaller network indicates a strong correlation across the network which may mean the stocks will all move together up or down. Data from Onnela was used to show how this size reduction coincided with the 1987 crash. The same technique was later applied to the August 2007 crash .

In all we have seen how we can use the correlation network have been applied to these problems and how understanding the interconnectedness of stocks affects various financial decisions.

REFERENCES

- [1] Rosario Mantegna. Correlation based hierarchical clustering in financial time series. *Preprint*, 2005.
- [2] Vincenzo Tola. Cluster analysis for portfolio optimization. *arXiv:physics/0507006 v1*, 2005.
- [3] A. Chakrabortia J.P. Onnela. Dynamic asset trees and black monday. *Physica A*, 2003.
- [4] Andrew W. Lo Amir E. Khandaniy. What happened to the quants in august 2007? *Preprint*, 2007.
- [5] JOSEPH L. MCCAULEY. What economists should learn from econophysics. *Dynamics of Complex Interconnected Systems: Networks and Bioprocesses*, 2006.
- [6] Mark McDonald. An investigation into the dynamics of correlation networks in the foreign exchange market. Master's thesis, University of Oxford, 2005.
- [7] Thomas H. Cormen. *Introduction to Algorithms*, chapter 23. MIT PRESS, 2002.
- [8] Bernd Rosenow. Random matrix theory for financial markets. *Phys Rev E*, 2001.
- [9] Stacy Williams Omer Suleman, Mark McDonald. Implications of correlation cleaning for risk management. *Preprint*, 2008.
- [10] Pedro G. Lind. The network approach: basic concepts and algorithms. *arXiv:0711.3972v1*, 2007.