INVESTIGATIONS OF ISTANBUL STOCK EXCHANGE NATIONAL 100 INDEX (BIST–100) BY USING DATA MINING AND FINANCIAL NETWORK TECHNIQUES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES OF
KARABUK UNIVERSITY

BY

YUSUF YARGI BAYDİLLİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE IN
DEPARTMENT OF
COMPUTER ENGINEERING

May 2014

I certify that in my opinion the thesis submitted by Yusuf Yargı BAYDİLLİ titled "INVESTIGATIONS OF ISTANBUL STOCK EXCHANGE NATIONAL 100 INDEX (BIST–100) BY USING DATA MINING AND FINANCIAL NETWORK TECHNIQUES" is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Şafak BAYIR ....
Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a master thesis. May 8, 2014

<u>Examining Committee Members (Institutions)</u>	<u>Signature</u>
Chairman: Prof. Dr. Fatih V. ÇELEBİ (YBU)	....
Member : Assist. Prof. Dr. Şafak BAYIR (KBU)	....
Member : Assist. Prof. Dr. Salih GÖRGÜNOĞLU (KBU)	....

...../...../2014

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Graduate School of Natural and Applied Sciences, Karabük University.

Prof. Dr. Mustafa BOZ ....
Head of Graduate School of Natural and Applied Sciences

“I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well.”

Yusuf Yargı BAYDİLLİ

ABSTRACT

M. Sc. Thesis

INVESTIGATIONS OF ISTANBUL STOCK EXCHANGE NATIONAL 100 INDEX (BIST–100) BY USING DATA MINING AND FINANCIAL NETWORK TECHNIQUES

Yusuf Yargı BAYDİLLİ

Karabük University
Graduate School of Natural and Applied Sciences
The Department of Computer Engineering

Thesis Advisor:
Assist. Prof. Dr. Şafak BAYIR
May 2014, 131 pages

Stock market is a complex system. Stocks in this system are in various relationships with stocks in their sector and as well as the other sectors. One of the methods which, that is used to analyze this relationship, is stock correlation network technique. In this type of analysis, a network including all stocks is created and this network is used to examine market movements, both computing minimum spanning tree (MST) by using various algorithms and/or hierarchical structural (HT) techniques.

In this thesis, shares in Istanbul Stock Exchange National 100 Index (BIST–100), which gives direction to Istanbul Stock Exchange, were investigated. In first part of thesis, two years of data (2011–2013) of the index was evaluated to observe usage of financial network technique in analysis of stock market dynamics in particular. In second part, ten years of data (2003–2013) of BIST–100, local and global factors, also, were analyzed to investigate factors affecting whole trade market, results of specific crisis periods into stock correlation network and possibility of prediction of these movements by using various fields of studies such as; data mining, topology, statistics and econophysics disciplines.

Briefly, the study of correlation based financial networks is a successful method, which provides great information from the correlation coefficient matrix. Besides, the MST and HT are useful in analyzing stock markets and defining factors affecting groups of stocks. In addition, it was showed that stock correlation network concept is skillful for portfolio optimization, risk management and crisis analysis. These results proved that time series of stocks prices carry valuable economic information and provide an insight into market behavior.

Key Words : Finance, stock market, econophysics, financial network, stock correlation network, data mining, topology, graph theory, minimum spanning tree, hierarchical tree.

Science Code : 902.2.042

ÖZET

Yüksek Lisans Tezi

İSTANBUL BORSASI ULUSAL 100 ENDEKSİNİN (BİST–100) VERİ MADENCİLİĞİ VE FİNANSAL AĞ TEKNİKLERİYLE İNCELENMESİ

Yusuf Yargı BAYDİLLİ

Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:
Yrd. Doç. Dr. Şafak BAYIR
Mayıs 2014, 131 sayfa

Borsa karışık bir sistemdir. Bu sistem içerisinde bulunan hisse senetlerinin, hem kendi sektörü içindeki hisse senetleriyle, hem de diğer sektörlerdekilerle çeşitli ilişkileri mevcuttur. Bu ilişkilerin analizinde kullanılan yöntemlerden bir tanesi de korelasyon ağı analizidir. Bu analiz tipinde tüm hisse senetlerini içeren bir ağ oluşturulur ve çeşitli algoritmalarla en kısa yol ağacı hesaplanarak ve/veya hiyerarşik sınıflandırma teknikleriyle, piyasa hareketleri incelenir.

Bu tezde, İstanbul Borsasına yön veren İstanbul Borsası Ulusal 100 Endeksi'nde (BİST–100) bulunan hisseler incelenmiştir. Tezin ilk bölümünde, endeksin 2 yıllık (2011–2013) verileri detaylıca değerlendirilerek, finansal ağ tekniklerinin mevcut borsa dinamikleri analizinde kullanımı incelenmiştir. İkinci bölümde ise; endeksin, ayrıca, yerel ve küresel etkenlerin 10 yıllık (2003–2013) verileri kullanılarak tüm

marketi etkileyen faktörler, kriz dönemlerinin sonuçlarının borsa korelasyon ağında oluşturdukları hareketler ve bu hareketlerin tahmin edilebilme olasılığı veri madenciliği, topolololoji, istatistik ve ekonofizik gibi bilim dallarından yararlanılarak analiz edilmiştir.

Sonuç olarak, korelasyon merkezli finansal ağ teknikleri, korelasyon matrisinden faydalanan ve bilgi sağlayan başarılı bir yöntemdir. Ayrıca, MST ve HT borsa analizinde ve hisse gruplarını etkileyen faktörleri tanımlamakta faydalıdır. Diğer yandan, hisse korelasyon ağı konseptinin portfolyo optimizasyonu, risk yönetimi ve kriz analizinde oldukça yetenekli olduğu gösterilmiştir. Bu sonuçlar, hisse senetlerinin fiyatlarının zaman serilerinin değerli ekonomik bilgi taşıdığını ve borsa davranışlarını anlamamızı sağladığını kanıtlamıştır.

Anahtar Kelimeler : Finans, borsa, ekonofizik, finansal ağ, korelasyon ağı, veri madenciliği, topoloji, graf teori, en kısa yol ağacı, hiyerarşik sınıflandırma ağacı.

Bilim Kodu : 902.2.042

ACKNOWLEDGMENT

First of all, I would like to give thanks to my advisor, Assist. Prof. Dr. Şafak BAYIR, for his great interest and assistance in preparation of this thesis.

I wish to express my gratitude to Prof. Dr. Resul ERYİĞİT for provided me to choose this topic by studying this subject in my graduation project.

I would like to thank to BIST (Istanbul Stock Exchange) for their data support.

Moreover, I would like to thanks to my former director Prof. Dr. Nizamettin KAHRAMAN, my former associate directors Assoc. Prof. Dr. Yavuz SUN and Assist. Prof. Dr. Metin ZEYVELİ, secretary-general Ümran ÇELİK and all staff for their association. In addition, I should also thank to my roommate Res. Asst. Mustafa GÖÇENOĞLU for his good friendship.

Besides, I wish to give my grateful to my director Prof. Dr. Mustafa BOZ and my associate directors Assist. Prof. Dr. Hüseyin DEMİREL and Assist. Prof. Dr. A. Mustafa ERER for their good cooperation.

Finally, I have to thank to my family and my spouse Esra BAYDİLLİ for their endless confidence, support, faith and love.

Last but not the least, I thanks to ALLAH who makes all happen.

CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

Page

INDEX OF SYMBOLS & ABBREVIATIONS

SYMBOLS

C : correlation matrix
d : distance
D : distance matrix
E : edge
G : graph
k : degree
L : length
N : nodes
O : big-O notation
$P(k)$: power-law distribution
$r(t)$: return in time
t : time-period
T : tree
V : vertex
w : weight
ρ : correlation coefficient
σ : standard deviation
μ : mean
γ : power-law exponent
σ^2 : variance
: number of

ABBREVIATIONS

ALCA	: Average Linkage Cluster Analysis
BIST	: Bourse Istanbul
BIST–100	: Bourse Istanbul 100 Index
cov	: covariance
GDP	: Gross Domestic Product
HT	: Hierarchical Tree
ISE	: Istanbul Stock Exchange
log	: logarithmic
MPT	: Modern Portfolio Theory
MST	: Minimum Spanning Tree
ROI	: Rate of Interest
SLCA	: Single Linkage Cluster Analysis
var	: variance
XU030	: National 30 Index
XU050	: National 50 Index
XU100	: National 100 Index
XUHIZ	: service index
XUMAL	: financial index
XUSIN	: industrial index
XUTEK	: technology index

PART 1

INTRODUCTION

Stock market (or bourse) is a highly organized market where stocks and shares are bought and sold. The stocks in market exist in several sectors. All stocks in same sectors and all sectors in market act in a relationship among them. This relationship can be in related or far from each other. Analyzing these relationships helps to investigate current market dynamics, predict market movements and determine major stocks giving direction to sectors and portfolio.

Stock market prediction is an act of trying to determine the future condition of company stock or other financial instruments traded on a financial exchange. There are some methodologies to observe conditions of stock market. One of them is financial network techniques.

In this technique, data in range is captured and processed to create a financial network by using some statistic functions and data mining techniques. After that, a minimum spanning tree (MST) and/or hierarchical tree (HT) are computed by using some algorithms and topological operations. These trees are used to analyze stock market.

In this thesis, Istanbul Stock Exchange National 100 Index was studied. This index consists of exactly 100 companies, which are chosen according to some circumstances determined before, such as; capability of sectoral representation and their trading volume. In addition, this index is a general indicator of all market. By analyzing this index, it was aimed to have an idea about whole market.

In Part 2, literature review was introduced. In the last decade, studies and their benefits to investigate stock markets were demonstrated.

In Part 3, theoretical background of financial network system was investigated. Before starting to create a financial network, the following questions were taken into consideration: which knowledge should be known, what kind of calculations should be done and what are the steps for initializing a network according to topology.

In Part 4, which stock market and index were used, it's reasons, scale of chosen data, results of calculations, local and global agents effecting market, the observation and view of stock market, also, its dynamics from a financial network were introduced.

In Part 5, the outcome of this study was put forward. Acquired information and observation of results were introduced. It was discussed that creating a financial network how provides benefits to study finance dynamics. Moreover, some ideas in order to develop this study were presented in Recommendations chapter.

PART 2

LITERATURE REVIEW

In the last decade, financial networks have attracted more attention from the research community. The efficient market paradigm states that stock returns of financial price time series are unpredictable. Within this paradigm, time evolution of stock returns is well described by random process. Several empirical analyses of real market data have proven that returns of time series are approximately described by non-redundant time series. The absence of redundancy is not complete in real markets and the presence of residual redundancy has been detected. A minimized degree of redundancy is required to avoid the presence of arbitrage opportunities. There are many studies about this topic, which was firstly introduced by Edward Mantegna.

In his work, the motivation of the study was twofold. The first motivation concerned the search for the kind of topological arrangement, which was present among the stocks of a portfolio traded in a financial market. The second motivation was the search of empirical evidence about the existence and nature of common economic factors, which drove the time evolution of stock prices. The observable, which was used to detect the topological arrangement of the stocks, was the synchronous correlation coefficient of the daily difference of logarithm of closure price of stocks.

He created HT and MST by using Dow Jones 30 and S&P (Standard & Poor's) 500 Index in a time from *July 1989* to *October 1995*. The reason of choosing these indices was that they mainly describe the performance of the New York Stock Exchange. With this study, he showed that the MST and the associated sub-dominant ultra-metric HT, which were obtained starting from the distance matrix and selected a topological space for the stocks of a portfolio traded in a financial market, are able to give an economic meaningful taxonomy.

According to him, this topology is useful for theoretical description of financial markets and search of economic common factors affecting specific groups of stocks. The topology and hierarchical structure associated to them could be obtained by using information in the time series of stock prices only. This result showed that time series of stock prices are carrying valuable (and detectable) economic information [1].

In 2000, Giovanni Bonanno et al. investigated two sets of data – (i) the non-synchronous time evolution of *N=24* daily stock market indices computed in local currencies during the time-period from *January 1988* to *December 1996*, and (ii) the closure value of the 51 Morgan Stanley Capital International (MSCI) country indices daily computed in local currencies or in US Dollars in the time-period from *January 1996* to *December 1999*. The stock indices, which were used in their research, belonged to stock markets distributed all over the world in five continents.

They presented a study showed that meaningful information could be extracted by a set of stock indices in time series. In their study, different levels of interdependence and complexity of data were elucidated by considering multiple applications of the same methodology on modified sets of the investigated time series. They were able to show that it is possible to extract a group of taxonomies, which directly reflected geographical and economic links between several countries all over the years. This was obtained by using the almost non-redundant time series of several stock indices of financial markets locate all over the world only. In conclusion, they showed that sets of stock index time series locate all over the world could be used to extract economic information about the links between different economies provided, that the effects of the non-synchronous nature of the time series and the different currencies used to compute the indices, were properly taken into account [2].

In 2001, Giovanni Bonanno et al. created a financial network for highly capitalized *N=100* stocks traded in US equity market during the period *January 1995–December 1998*. Their investigations showed that the degree and nature of cross-correlation between pairs of stock returns vary as a function of the time horizon, which was used to compute them. Specifically, intraday time series of returns showed a lowest degree of correlation that manifested itself in the partial disruption of the hierarchical structure observed for longer (e.g. daily) time horizons. The disruption of the correlation structure was more pronounced for intra-sector correlation than for inter-sector correlation inducing a modification in the correlation based clustering of a stock portfolio, whose results become time horizon dependent. In addition, they showed that at shorter time horizons, a few stocks (in the present case a single stock) become a hierarchical reference of the large majority of stocks. The determination of a hierarchical structure for each time horizon allowed them to follow the mechanism of cluster formation as a function of the time horizon. They interpreted fine details of the cluster formation as evidence that the process of rational price formation take a finite amount of time to occur [3].

In 2001 again, Nicolas Vandewalle et al. analyzed the cross-correlations of daily fluctuations for $N=6358$ US stock prices during the year 1999. By using $\frac{1}{2}[n\ (n-1)]$ correlation coefficients, the MST was built. They put into evidence the emergence of some order out of the apparent disorder of the stock markets. Indeed, they found signatures of non-trivial correlations in the topology of stock markets. They did the same analysis for sub-sets of financial data and they obtained the same behavior (a power-law for $f\ (n)$ and a divergence of σ). For shifted periods, the main results did not change: the structure evolved slowly and locally. Indeed, month after month, a node kept the majority of its neighbors. The non-randomness of the stock market topology was thus a robust property [4].

In 2003, Giovanni Bonanno et al. presented a topological characterization of the minimal spanning tree, which could be obtained by considering the price return correlations of stocks traded in a financial market. They compared the minimal spanning tree obtained from a large group of stocks traded at the New York Stock Exchange during *12-year* trading period with the one obtained from surrogated data simulated by using simple market models. They found that the empirical tree had features of a complex network that could not be reproduced, even as a first approximation by a random market model and by the one-factor model.

These results showed that the topology of the MST for the real and for the considered artificial markets is different for node with both high and low degree. They defined the importance of a node as its degree (or its in-degree component) from their analysis emerged, that the real market has a hierarchical distribution of importance of the nodes, whereas the considered models are not able to catch such a hierarchical complexity. Specifically, in the random model, the fluctuations selected randomly few nodes and assigned them small values of degree. Thus, the MST of the random model was essentially non-hierarchical. On the other hand, the MST of the one factor model showed a simple one-center hierarchy. The topology of stock return correlation based MST showed large-scale correlation properties and characteristic of complex networks in the native as well as in an oriented form. Such properties could not be reproduced at all, even as a first approximation, by simple models as a random model or the widespread one-factor model [5].

In 2003, Salvatore Micciché et al. investigated the statistical properties of cross-correlation among volatility and price return time series for the *N=93* most capitalized stocks traded in US (United States) equity markets during *12-year* time-period. The parallel investigation of MSTs, which were obtained from price return and volatility time series of a set of stocks, allowed them to conclude that the stability of the degree of MSTs is lower for volatility time series than for price return time series. In their study, for price return time series, the stability of stock degree dynamics increased, when the time window *t* used to compute, MSTs were increased. A similar but much weaker trend was also observed in MSTs obtained starting from volatility time series.

In addition, relevant economic information was stored in the degree of MSTs obtained from price return and volatility time series. The dynamics of stock degree were statistically more stable for price return than for volatility MSTs and it had a slow dynamics characterized by a time scale of the order of *3* calendar years for price return MSTs and longer than *6* calendar years for volatility MSTs [6].

In 2004, Giovanni Bonanno et al. reviewed the recent approach of correlation-based networks of financial equities. They investigated portfolio of stocks at different time horizons, financial indices, volatility time series and they showed that meaningful economic information could be extracted from noise dressed correlation matrices. They showed that the method could be used to falsify widespread market models by directly comparing the topological properties of networks of real and artificial markets. They built MST of $N=100$ highly capitalized stocks traded in the US equity markets. Lastly, they showed that the study of correlation based financial networks is a fruitful method, which is able to filter out economic information from the correlation coefficient matrix of a set of financial time series. The topology of the detected network could be used to validate or falsify simple, although widespread, market models [7].

In 2005, Hokky Situngkir and Yohanes Surya analyzed the evolving price fluctuations by using ultra-metric distance of minimally spanning financial tree of stocks traded in Jakarta Stock Exchange (Indonesia) *2000–2004*. Ultra-metrication was derived from transformation of correlation coefficients into the distances among stocks. Their analysis evaluated the performance of ups and downs of stock prices and discovered the evolution towards the financial and economic stabilization in Indonesia. This was partly recognized by mapping the HTs upon the realization of liquid and illiquid stocks [8].

In 2006, Ricardo Coelho et al. showed how the asset tree evolved over time and described the dynamics of its normalized length, mean occupation layer, and single- and multiple-step linkage survival rates. Over the period studied, *1997–2006*, the tree showed a tendency to become more compact. This implied that global equity markets were increasingly interrelated. The use of the MST provided a way to extract a manageable amount of information from a large correlation matrix of global stock returned to reveal patterns of links between different markets. It provided an insight into market behavior, which is not as easily obtained from the correlation matrix as a whole. Applied dynamically, the analysis let them to observe consistencies as well as evolutions in patterns of market interactions over time. As would be expected, there was a strong tendency for markets to organize by geographical location, although other, related factors such as economic ties, might also play important roles. Developed European countries, with France and Germany at their center, had consistently constituted the most tightly linked markets within the MST. There had also been a limited tendency of the CEE (Central and Eastern Europe) accession countries to link more closely with the more developed EU (European Union) countries [9].

In 2007, again, Antonios Garas and Panos Argyrakis studied the year-after-year properties of three different portfolios traded in the Athens Stock Exchange (ASE) for the time-period *1987–2004*. They used the MST technique and the random matrix theory (RMT) which makes it possible to examine at the same time the temporal evolution of the portfolios and of the market as a whole. The first four moments of the distribution of correlations and the normalized tree lengths of the MST showed a similar behavior for all three portfolios. However, by studying topological properties of the MST, such as the node *degree k*, they were able to identify changes to the MST associated to each portfolio that were due to a crisis in the market, like the one that happened during the period *1999–2001*. They also showed that, while the effect of the market to the information content of the correlation matrix for all three portfolios was almost the same, the market was affected differently by different economic sectors at different time-periods.

Their analysis showed that the mean value of the correlation coefficient matrix and the associated normalized tree length $L(t)$ were almost identical each year for all portfolios considered which leaded them to suggest that in the case of the Greek stock market all stocks were (on the average) affected the same way by the price movements of other stocks. An interesting observation in their study was that the behavior of the market was completely different during a crisis period. During such crisis, they found that the correlations between stocks become stronger and the MST shrank. They found that for every year the contribution of the market was the same to each of the investigated portfolios. This had implications to the way one selected a trading portfolio, since it meant that this portfolio would be affected by changes in the market the same way as any other portfolio.

They also found that during a crisis period, the effect of the market become dominant to the correlation matrix. This is a verification of that stocks are largely correlated during such periods, and the market moves as a whole. They found that for the early years of the Greek stock market, there were specific sectors, which seemed to play more significant role than others did, while for the later years the contribution of all the economic sectors to the *"market"* became homogenous. According to them, this could be a signal that such a market is becoming more mature since the almost equal contribution from all economic sectors to the eigenvector that corresponded to the largest eigenvalue of the correlation matrix is verified for the case of mature markets, such as the NYSE (New York Stock Exchange) and LSE (London Stock Exchange) [10].

In 2010, Chi K. Tse et al. constructed complex networks to study correlations between the closing prices for all US (United States) stocks that were traded from *July 1, 2005* to *August 30, 2007*. The nodes were the stocks, and the connections were determined by cross-correlations of the variations of the stock prices, price returns and trading volumes within a chosen period. A winner-take-all approach was used to determine if two nodes were connected by an edge. They reported that all networks based on connecting stocks of highly correlated stock prices, price returns and trading volumes, displayed a scale-free degree distribution. The results from this work clearly suggested that the variation of stock prices is strongly influenced by a relatively small number of stocks.

They proposed a new approach for selecting stocks for inclusion in stock indices and compared it with existing indices. From the composition of the highly connected stocks, it could be concluded that the market was heavily dominated by stocks in the financial sector. It had been found that the networks formed by connecting stocks of highly correlated closing prices, price returns and trading volumes were all scale-free. This result suggested that a relatively small number of stocks are exerting much of the influence over the majority of stocks. New indices, which reflected on the performance of the majority stock, might be defined based on a relatively small number of highly connected stocks. It had been found that the market was actually heavily influenced by stocks in the financial sector [11].

In 2010, Michele Tumminello et al. discussed how to define and obtain HTs, correlation based trees and networks from a correlation matrix. Their example considered the correlation matrix of *N=10* daily stock returns traded at the New York Stock Exchange during the time-period from *January 2001* to *December 2003*. The hierarchical clustering and other procedures performed on the correlation matrix to detect statistically reliable aspects of it are seen as filtering procedures of the correlation matrix. They also discussed a method to associate a hierarchically nested factor model to a HT obtained from a correlation matrix. The information retained in filtering procedures and its stability with respect to statistical fluctuations was quantified by using the Kullback-Leibler distance.

They also showed that the information present in correlation based tree and graphs provide additional clues about the interrelations among stocks of different economic sectors and sub-sectors. The information obtained from what they call the *"filtering procedure"* of the correlation matrix was subjected to statistical uncertainty. For this reason, they discussed a bootstrap methodology able to quantify the statistical robustness of both the HTs and correlation based trees or graphs. The amount of information and the statistical stability of filtering procedures of the correlation matrix were quantified by using the Kullback-Leibler distance. They reported and also discussed analytical results both for Gaussian and for Student's t-distributed multivariate time series [12].

In 2011, Ersin Kantar et al. examined the hierarchical structures of Turkey's foreign trade by using real prices of their commodity export and import moved together over time. They obtained the topological properties among the countries based on Turkey's foreign trade during the *1996–2010* periods by using the concept of hierarchical structure methods MST and HT. These periods were divided into two sub-periods, such as *1996–2002* and *2003–2010*, in order to test various time-windows and observe the temporal evolution. They performed the bootstrap techniques to investigate a value of the statistical reliability to the links of the MSTs and HTs. They also used a clustering linkage procedure in order to observe the cluster structure much better. From the structural topologies of these trees, they identified different clusters of countries according to their geographical location and economic ties. Their results showed that the DE (Germany), UK (United Kingdom), FR (France), IT (Italy) and RU (Russia) were more important within the network, due to a tighter connection with other countries. They also found that these countries played a significant role for Turkey's foreign trade and had important implications for the design of portfolio and investment strategies [13].

In 2011, Ersin Kantar et al. presented a study, within the scope of econophysics, of the hierarchical structure of $N=98$ stocks among the largest international companies including *18* among the largest Turkish companies, namely Banks, Automobile, Software-hardware, Telecommunication Services, Energy and the Oil-Gas sectors, viewed as a network of interacting companies. They analyzed the daily time series data of the Boerse-Frankfurt and Istanbul Stock Exchange. They examined the topological properties among the companies over the period *2006–2010* by using the concept of hierarchical structure methods MST and the HT. They also used average linkage clustering analysis (ALCA) in order to better observe the cluster structure. From these studies, they found that the interactions among the Banks/Energy sectors and the other sectors were reduced after the global economic crisis; hence, the effects of the Banks and Energy sectors on the correlations of all companies were decreased. Telecommunication Services were also greatly affected by the crisis. They also observed that the Automobile and Banks sectors, including Turkish companies as well as some companies from the USA, Japan and Germany were strongly correlated with each other in all periods [14].

In 2011, Leonidas Sandoval Junior employed some techniques in order to filter random noise from the information, provided by MSTs obtained from the correlation matrices. The first technique established a threshold above which connections were considered affected by noise, based on the study of random networks with the same probability density distribution of the original data. The second technique was to judge the strength of a connection by its survival rate, which is the amount of time a connection between two stock market indices endures. The idea was that true connections would survive for longer periods, and that random connections would not. This information was then combined with the information obtained from the first technique in order to create a smaller network, where most of the connections were either strong or enduring in time.

In his research, the represented networks were obtained from the correlations between international stock exchange indices prior to and during years of well known international financial crises, namely the Black Monday (1986 and 1987), the Asian Financial Crisis (1997), the Russian Crisis (1998), the crisis after September, 11 (2001) and the Subprime Mortgage Crisis (2008).

In the approach of this article, he followed the path of limiting the connections to those, which were strong meaning that they were related with distances below a threshold, which was obtained by simulations with randomized data, and to those, that were not strong, but that were very enduring in time. This made it possible to obtain diagrams that showed sensible relations between the many indices that were studied. The graphs showed that the connections between stock market indices were strongly influenced by the geographic positions of the stock exchanges, as there were well-defined clusters related with continents, reinforcing results previously obtained by regression methods. They also showed that cultural relations or international treaties has a strong effect in stock market indices correlations, as one could see by the strong connections between Israel and South Africa with Europe, and between Greece and Cyprus, as examples. The centrality of France had been detected and the closeness between Russia and Turkey was detected in.

In article, the other factor to be observed was the nature of the centrality measures in a network. They were built in such a way that if many nodes move together, then, one or more of them is better representatives of the collective movement. In addition, this work contributed to a better understanding of the MST representation, when dealing with weakly interacting indices and offered a way to prune some of them away from the network. The process showed that weakly interacting nodes rarely survive for long, so that, in practice, it made little difference if one pruned them away based only on a distance threshold obtained by randomizing the original data. It also showed how the world stock markets evolved in the past decades, mainly in the proximity of crises, and how stock markets become more integrated with time [15].

Moreover, in 2012, Ersin Kantar et al. analyzed the topology of *N=50* important Turkish companies for the period *2006–2010* by using the concept of hierarchical methods MST and HT. They investigated the statistical reliability of links between companies in the MST by using the bootstrap technique. They also used the average linkage cluster analysis (ALCA) technique to observe the cluster structures. They used these techniques to obtain analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods. They found that Turkish companies were not very affected by the global financial crisis [16].

As can be seen in this part, many studies were done by using financial networks techniques that handle cross-correlation or bootstrap to analyze stock markets. They showed that these techniques could be used to investigate stocks, sectors, sub-sectors, moreover, its connections with the other stock markets. On the other hand, this technique can be used to investigate crisis and their results to whole market and the other markets related with this markets.

To the best of our knowledge, there are not any financial network study provides a general view for Bourse Istanbul by using BIST–100 index, which is a general indicator of whole market. Therefore, in this thesis, Bourse Istanbul's dynamics was investigated with discussing most efficient techniques to analyze a stock market.

PART 3

THEORETICAL BACKGROUND

Papers introduced previous chapter showed that the results of these researches are very useful for understanding whole stock market and dynamics. Moreover, they provide information about predicting market movements. To earn these benefits, much knowledge must be known and series of calculations must be done. In this part, this knowledge and calculations will be represented and the way should be followed to construct a financial network for a stock market would be explained.

3.1. NETWORKS AND GRAPH THEORY

3.1.1. Basic Definition

In information technology, a network is a series of points or nodes interconnected by communication paths. Networks can interconnect with other networks and contain sub-networks, as can be seen in Figure 3.1 [17].

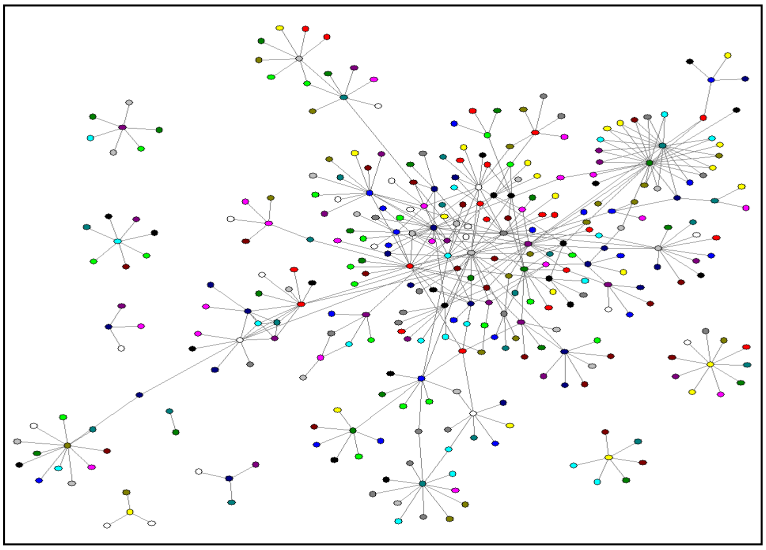


Figure 3.1. Visualization of a network [18].

A graph is a symbolic representation of a network and its connectivity. It implies an abstraction of the reality, so, it can be simplified as a set of linked nodes. The following elements are fundamental at understanding graph theory:

- A graph G is a set of vertex (nodes) v connected by edges (links) e. Thus $G = (v, e)$.
- <u>Vertex (Node)</u>: A node v is a terminal point or an intersection point of a graph.
- <u>Edge (Link)</u>: An edge e is a link between two nodes. The link (i, j) is initial extremity of i and terminal extremity of j.
- <u>Buckle (Loop or Self edge)</u>: A link that makes a node correspond to itself is a buckle [19].

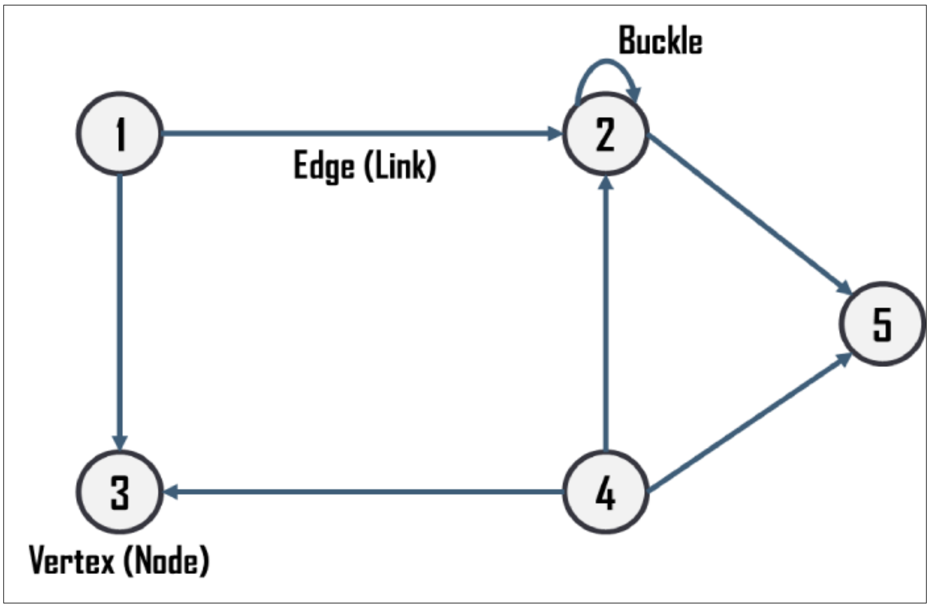


Figure 3.2. Basic graph representation [20].

This simple graph in Figure 3.2 has the following definition: $G = (v, e)$

$v = (1, 2, 3, 4, 5)$

$e = (1, 2), (1, 3), (2, 2), (2, 5), (4, 2), (4, 3), (4, 5)$

- <u>Sub-Graph</u>: A sub-graph is a sub-set of a graph G where p is the number of sub-graphs. For instance $G' = (v', e')$ can be a distinct sub-graph of G.

- Planar Graph: A graph where all the intersections of two edges are a vertex. Since this graph is located within a plane, its topology is two-dimensional.
- Non-Planar Graph: A graph where there are no vertices at the intersection of at least two edges.
- Simple Graph: A graph that includes only one type of link between its nodes.
- Multi-Graph: A graph that includes several types of links between its nodes [19].

3.1.2. Links and Their Structures

Graph theory must thus offer the possibility of representing movements as linkages that can be considered over several aspects:

- Connection: A set of two nodes as every node is linked to the other. Considers if a movement between two nodes is possible, whatever its direction. Knowing connections makes it possible to find if it is possible to reach a node from another node within a graph.
- Path: A sequence of links that are traveled in the same direction. For a path to exist between two nodes, it must be possible to travel an uninterrupted sequence of links.
- Chain: A sequence of links having a connection in common with the other. Direction does not matter.
- Length of a Link, Connection or Path: Refers to the label associated with a link, a connection or a path. The length of a path is the number of links (or connections) in this path.
- Cycle: Refers to a chain where the initial and terminal node is the same and that does not use the same link more than once is a cycle.
- Circuit: A path where the initial and terminal node corresponds. It is a cycle where all the links are traveled in the same direction.
- Clique: A clique is a maximal complete sub-graph where all vertices are connected.

- Cluster: Also called community, it refers to a group of nodes having denser relations with each other than with the rest of the network. A wide range of methods are used to reveal clusters in a network, notably they are based on modularity measures (intra-versus inter-cluster variance) [19].
- Node Degree: The degree of a node is the number of connections it has in the network.
- Node Strength: The strength of a node is the sum of the correlations of the node with all other nodes to which it is connected [15].

3.1.3. Basic Structural Properties

The organization of nodes and links in a graph convey a structure that can be described and labeled. The basic structural properties of a graph are:

- Symmetry and Asymmetry: A graph is symmetrical if each pair of nodes linked in one direction is also linked in the other. By convention, a line without an arrow represents a link where it is possible to move in both directions.
- Assortativity and Disassortativity: Assortative networks are those characterized by relations among similar nodes, while disassortative networks are found when structurally different nodes are often connected. Transport (or technological) networks are often disassortative when they are non-planar, due to the higher probability for the network to be centralized into a few large hubs.
- Completeness: A graph is complete if two nodes are linked in at least one direction. A complete graph has no sub-graph and all its nodes are inter-connected.
- Connectivity: A complete graph is described as connected if for all its distinct pairs of nodes there is a linking chain. Direction does not have importance for a graph to be connected, but may be a factor for the *level* of connectivity. If $p>1$ the graph is not connected, because it has more than one sub-graph (or component). There are various levels of connectivity depends on the degree at which each pair of nodes is connected.
- Complementarily: Two sub-graphs are complementary if their union results in a complete graph. Multimodal transportation networks are complementary as

each sub-graph (modal network) benefits from the connectivity of other sub-graphs.

- Root: A node r where every other node is the extremity of a path coming from r is a root. Direction has an importance. A root is generally the starting point of a distribution system.
- Trees: A connected graph without a cycle is a tree. A tree has the same number of links than nodes plus one $(e = v-1)$. If a link is removed, the graph ceases to be connected. If a new link between two nodes is provided, a cycle is created. A branch of root r is a tree where no links are connecting any node more than once.
- Articulation Node: In a connected graph, a node is an articulation node if the sub-graph obtained by removing this node is no longer connected. It therefore contains more than one sub-graph $(p > 1)$. It is also called a bridge node.
- Isthmus: In a connected graph, an isthmus is a link that is creating, when removed, two sub-graphs having at least one connection. Most central links in a complex network are often isthmuses, which removal by reiteration helps revealing dense communities (clusters) [19].

3.1.4. Advanced Properties

While a small network can be visualized directly by its graph (N, g), larger networks can be more difficult to envision and describe. Therefore, a set of summary statistics or quantitative performance measures can be defined to describe and compare networks (focus on undirected graphs):

- Diameter and Average Path Length: Let $l\ (i, j)$ denote the length of the shortest path (or geodesic) between node i and j (or the distance between i and j). The *diameter* of a network is the largest distance between any two nodes in the network. The *average path length* is the average distance between any two nodes in the network.
- Clustering Coefficient: It measures the extent to which nodes are friends with one another.

- Centrality: A micro measure that captures the importance of a node's position in the network. Different measures of centrality:
  - Degree Centrality: For node *i*,
  - Closeness Centrality: Tracks how close a given node is to any other node: for node *i*, one such measure is,
  - Betweenness Centrality: Captures how well situated a node is in terms of paths that it lies on,
- Degree Distributions: The degree distribution, *P (d)*, of a network is a description of relative frequencies of nodes that have different degrees *d*. For a *random graph model*: *P (d)* is a probability distribution.
  - Two types of degree distributions:
    - $P(d) \leq ce^{-\alpha d}$, for some $\alpha > 0$ and $c > 0$: The tail of the distribution *falls off faster than an exponential*, i.e., large degrees are unlikely.
    - $P(d) \leq cd^{-\gamma}$, for some $\gamma > 0$ and $c > 0$. *Power-law distribution*: The tail of the distribution is fat, i.e., there tend to be many more nodes with very large degrees.
      - Appear in a wide variety of settings including networks describing incomes, city populations, WWW, and the Internet.
      - Also known as a *scale-free distribution*: a distribution that is unchanged (within a multiplicative factor) under a rescaling of the variable.
      - Appear linear on a log-log plot [21].

3.2. CLUSTERING

Clustering can be considered the most important *unsupervised learning* problem; so, as every other problem of this kind, it deals with finding a *structure* in a collection of unlabeled data. A loose definition of clustering could be *"the process of organizing objects into groups whose members are similar in some way"*. A *cluster* is therefore, a collection of objects that are *"similar"* between them and are *"dissimilar"* to the objects belonging to other clusters. This can be seen with a simple graphical example in Figure 3.3:

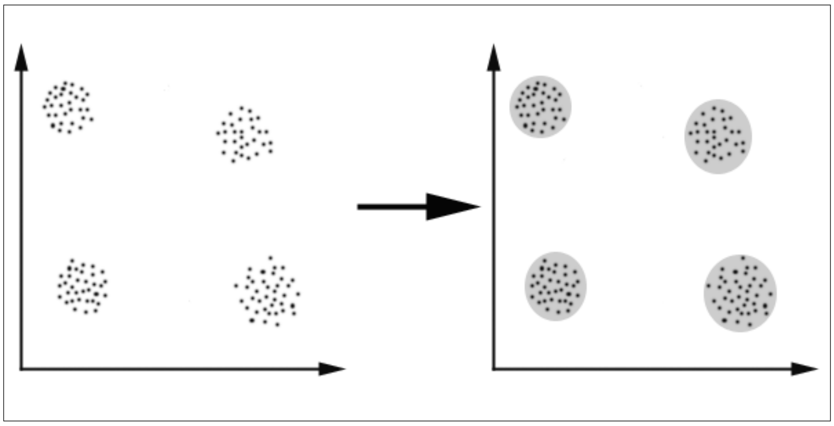


Figure 3.3. Graphical example of clustering.

In this case as can be easily identified, the 4 clusters into which the data can be divided; the similarity criterion is *distance:* two or more objects belong to the same cluster if they are *"close"* according to a given distance (in this case geometrical distance). This is called *distance-based clustering*. Another kind of clustering is *conceptual clustering:* two or more objects belong to the same cluster if this one defines a concept *common* to all that objects. In other words, objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures.

So, the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data, but what constitutes provide a good clustering? It can be shown that there is no absolute *"best"* criterion, which would be independent of the final aim of the clustering. Consequently, it is the user, which must supply this criterion, in such a way that the result of the clustering will suit their needs. For instance, one could be interested in finding representatives for homogeneous groups *(data reduction)*, in finding *"natural clusters"* and describe their unknown properties *("natural" data types)*, in finding useful and suitable groupings *("useful" data classes)* or in finding unusual data objects *(outlier detection)*. The main requirements that a clustering algorithm should satisfy are:

- scalability;
- dealing with different types of attributes;
- discovering clusters with arbitrary shape;

- minimal requirements for domain knowledge to determine input parameters;
- ability to deal with noise and outliers;
- insensitivity to order of input records;
- high dimensionality;
- interpretability and usability.

3.2.1. Clustering Algorithms

Clustering algorithms may be classified as listed below:

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering
- Probabilistic Clustering

In the first case, data are grouped in an *exclusive* way, so that if a certain data belongs to a definite cluster then it could not be included in another cluster. On the contrary, the second type, the *overlapping* clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value.

Instead, a *hierarchical* clustering algorithm is based on the union between the two nearest clusters. The beginning condition is realized by setting every datum as a cluster. After a few iterations, it reaches the final clusters wanted. Finally, the last kind of clustering uses a completely *probabilistic* approach. There are four of the most used clustering algorithms:

- K-means
- Fuzzy C-means
- Hierarchical Clustering
- Mixture of Gaussians

Each of these algorithms belongs to one of the clustering types listed above, so that K-means is an *exclusive clustering* algorithm, Fuzzy C-means is an *overlapping clustering* algorithm, Hierarchical Clustering is obvious and lastly, Mixture of Gaussian is a *probabilistic clustering* algorithm [22].

3.2.2. Hierarchical Clustering

A hierarchical clustering is a sequence of partitions in which each partition is nestled into the next partition in the sequence. An agglomerative algorithm for hierarchical clustering starts with the disjoint clustering, which places each of the *n* objects in an individual cluster. The clustering algorithm being employed dictates how the proximity matrix should be interpreted to merge two or more of these trivial clusters, thus, nesting the trivial clustering into a second portion.

The process is repeated to form a sequence of nested clustering in which the number of clusters decreases as the sequence progresses until a single cluster containing all *n* objects, called the conjoint clustering, remains. A divisive algorithm performs the task in the reverse order. Three specific hierarchical clustering methods are called the single-link, the complete-link and the average-link methods [23].

3.2.2.1. Distance Measure

An important component of a clustering algorithm is the distance measure between data points. Some commonly used metrics for hierarchical clustering are [22]:

- Euclidean distance
- squared Euclidean distance
- Manhattan distance
- maximum distance
- Mahalanobis distance
- cosine similarity

Euclidean distance formula is:

$$\|a-b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \tag{3.1}$$

3.2.2.2. Single-Link, Complete-Link & Average-Link Clustering

Hierarchical clustering treats each data point as a singleton cluster, and then successively merges clusters until all points have been merged into a single remaining cluster. A hierarchical clustering is often represented as a dendrogram. A dendrogram is a tree-structured graph used in heat maps to visualize the result of a hierarchical clustering calculation. The result of a clustering is presented either as the distance or the similarity between the clustered rows or columns depending on the selected distance measure [24]. An example of dendrogram can be seen in Figure 3.4.

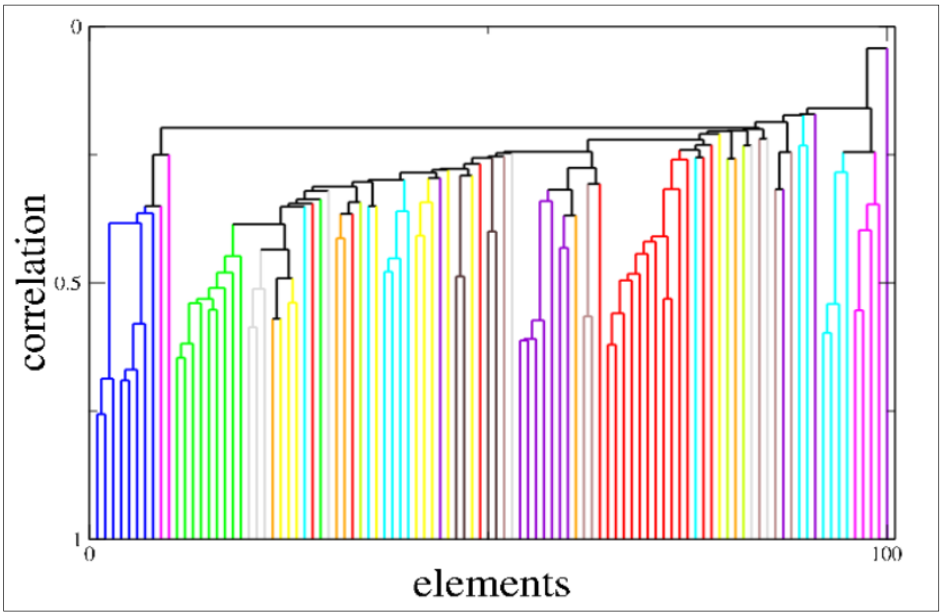


Figure 3.4. A dendrogram [25].

In complete-link (or complete linkage) hierarchical clustering, merge in each step the two clusters whose merger has the smallest diameter (or: the two clusters with the smallest *maximum* pair wise distance). Complete-link clustering can also be described using the concept of clique. Let *d (n)* is the diameter of the cluster created in step *n* of complete-link clustering. Define graph *G (n)* as the graph that links all data points with a distance of at most *d (n)*. Then the clusters after step *n* are the cliques of *G (n)*. This motivates the term complete-link clustering.

In single-link (or single linkage) hierarchical clustering, merge in each step the two clusters whose two closest members have the smallest distance (or: the two clusters with the smallest *minimum* pair wise distance). Single-link clustering can also be described in graph theoretical terms. If *d (n)* is the distance of the two clusters merged in step *n*, and *G (n)* is the graph that links all data points with a distance of at most *d (n)*, then the clusters after step *n* are the connected components of *G (n)*. A single-link clustering also closely corresponds to a weighted graph's minimum spanning tree.

Average-link (or group average) clustering is a compromise between the sensitivity of complete-link clustering to outliers and the tendency of single-link clustering to form long chains that do not correspond to the intuitive notion of clusters as compact, spherical objects. Average-link clustering merges in each step the pair of clusters with the highest cohesion. If the data points are represented as normalized vectors in a Euclidean space, it can be defined the cohesion *G* of a cluster *C* as the average dot product [26].

3.3. CORRELATION BASED NETWORKS

This approach provides a new system level analysis of the activity and topology of directed networks. The approach extracts causal topological relations between the networks nodes (when the network structure is analyzed), and provides an important step towards inference of causal activity relations between the network nodes (when analyzing the network activity). In the case of network activity, the analysis is based on partial correlations, which are becoming ever more widely used to investigate complex systems. In simple words, the partial (or residual) correlation is a measure of the effect (or contribution) of a given node, say *j*, on the correlations between another pair of nodes, say *i* and *k*.

Using this concept, the dependency of one node on another node, is calculated for the entire network. This results in a directed weighted adjacency matrix, of a fully connected network. Once the adjacency matrix has been constructed, different algorithms can be used to construct the network, such as a threshold network, Minimal Spanning Tree (MST), Planar Maximally Filtered Graph (PMFG), and others. The partial correlation based Dependency Networks is a revolutionary new class of correlation based networks, which is capable of uncovering hidden relationships between the nodes of the network [27].

3.3.1. Minimum Spanning Tree (MST)

Suppose there are given a connected, undirected, weighted graph. This is a graph $G = (V, E)$ together with a function $w: E \rightarrow IR$ that assigns a weight $w\ (e)$ to each edge e. For this lecture, assume that the weights are real numbers. The task is to find the *minimum spanning tree* of G, that is, the spanning tree T minimizing the function:

$$w\ (T)\ = \sum_{e\ \in\ T} w\ (e) \tag{3.2}$$

To keep things simple, assume that all the edge weights are distinct: $w\ (e) \neq w\ (e')$ for any pair of edges e and $e\ (0)$. Distinct weights guarantee that the minimum spanning tree of the graph is unique. Without this condition, there may be several different minimum spanning trees. For example, if all the edges have weight 1, then every spanning tree is a minimum spanning tree with weight V - 1. A minimum spanning tree example can be seen in Figure 3.5.

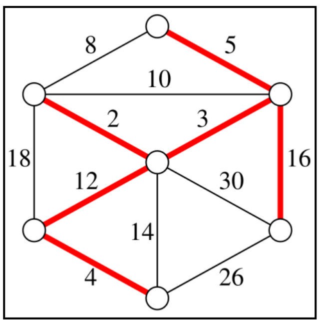


Figure 3.5. A weighted graph and its minimum spanning tree.

If there is, an algorithm that assumes the edge weights are unique, one can still use it on graphs where multiple edges have the same weight, as long as there have a consistent method for breaking ties. One way to break ties consistently is to use the following algorithm in place of a simple comparison. SHORTEREDGE takes as input four integers *i, j, k, l* and decides which of the two edges *(i, j)* and *(k, l)* has *'smaller'* weight.

SHORTEREDGE *(i, j, k, l)*:

- *if w (i, j) < w (k, l) return (i, j)*
- *if w (i, j) > w (k, l) return (k, l)*
- *if min (i, j) < min (k, l) return (i, j)*
- *if min (i, j) > min (k ,l) return (k, l)*
- *if max (i, j) < max (k, l) return (i, j)*
- *⟨⟨if max (i, j) < max (k, l) ⟩⟩ return (k, l)*

3.3.1.1. The Only Minimum Spanning Tree Algorithm

There are several different methods for computing minimum spanning trees, but really, they are all instances of the following generic algorithm. The generic minimum spanning tree algorithm maintains an acyclic sub-graph *F* of the input graph *G*, which one will call an *intermediate spanning forest*. *F* is a sub-graph of the minimum spanning tree of *G*, and every component of *F* is a minimum spanning tree of its vertices. Initially, *F* consists of *n* one-node trees. The generic algorithm merges trees together by adding certain edges between them. When the algorithm halts, *F* consists of a single *n-node* tree, which must be the minimum spanning tree. Obviously, have to be careful about which edges added to the evolving forest, since not every edge is in the minimum spanning tree.

The intermediate spanning forest *F* induces two special types of edges. An edge is *useless*, if it is not an edge of *F*, but both its endpoints are in the same component of *F*. For each component of *F*, they associated a safe edge – the minimum – weight edge with exactly one endpoint in that component. Different components might or might not have different safe edges. Some edges are neither safe nor useless called these edges undecided. All minimum spanning tree algorithms are based on two simple observations.

Lemma: *The minimum spanning tree contains every safe edge and no useless edges.*
Proof: Let T be the minimum spanning tree. Suppose F has a *'bad'* component whose safe edge $e = (u, v)$ is not in T. Since T is connected, it contains a unique path from u to v, and at least one edge e' on this path has exactly one endpoint in the bad component. Removing e' from the minimum spanning tree and adding e gives a new spanning tree. Since e is the bad component's safe edge, there is $w(e') > w(e)$, so the new spanning tree has smaller total weight than T. Nevertheless, this is impossible – T is the *minimum* spanning tree. Therefore, T must contain every safe edge. Adding any useless edge to F would introduce a cycle as can be seen in Figure 3.6.

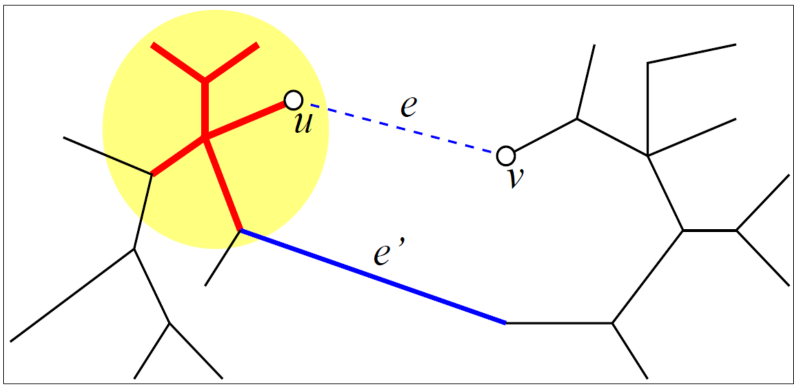


Figure 3.6. The 'bad' component of F.

So generic minimum spanning tree algorithm repeatedly adds one or more safe edges to the evolving forest F. Whenever add new edges to F, some undecided edges become safe, and others become useless. To specify a particular algorithm, must decide which safe edges to add, and how to identify new safe and new useless edges, each iteration of our generic template.

3.3.1.2. Borûvka's Algorithm

The earliest known algorithm for finding a minimum spanning tree was given by Otakar Borûvka back in 1926. In a Borûvka step, every *super-vertex* selects its smallest adjacent edge. These edges are added to the MST, avoiding cycles. Then the new super-vertices, i.e., the connected components, are calculated by contracting the graph on the edges just added to the MST. This process is repeated until only one super-vertex is left. In other words, there are $n - 1$ edges contracted. The union of these edges gives rise to a minimum spanning tree.

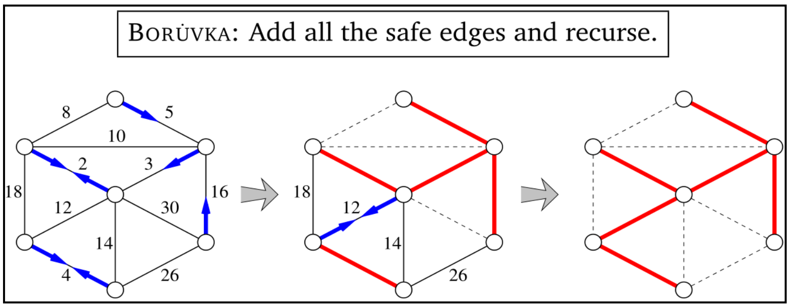


Figure 3.7. Borûvka's algorithm.

Borûvka's algorithm runs on the example graph in Figure 3.7. Thick edges are in *F* arrows point along each component's safe edge. Dashed edges are useless.

Borûvka's algorithm:

make a list L of n trees, each a single vertex

while (L has more than one tree)

for each T in L, find the smallest edge connecting T to G–T

add all those edges to the MST

(causing pairs of trees in L to merge)

Notice that each Borûvka step reduces the number of vertices by a factor of at least two. Therefore the while loop will be executed at most $O(\log n)$ times. In each step, all the contraction can be done in $O(m)$ time. In total, the Borûvka algorithm has a running time of $O(m \log n)$.

3.3.1.3. Prim's Algorithm

Prim's algorithm was conceived by computer scientist Robert Prim in 1957. It starts from an arbitrary vertex, and builds upon a single partial minimum spanning tree, at each step adding an edge connecting the vertex nearest to but not already in the current partial minimum spanning tree. It grows until the tree spans all the vertices in the input graph. This strategy is greedy in the sense that at each step the partial spanning tree is augmented with an edge that is the smallest among all possible neighboring edges. In Jarník's algorithm, the forest *F* contains only one nontrivial component *T*; all the other components are isolated vertices. Initially, *T* consists of an arbitrary vertex of the graph. The algorithm repeats the following step until *T* spans the whole graph:

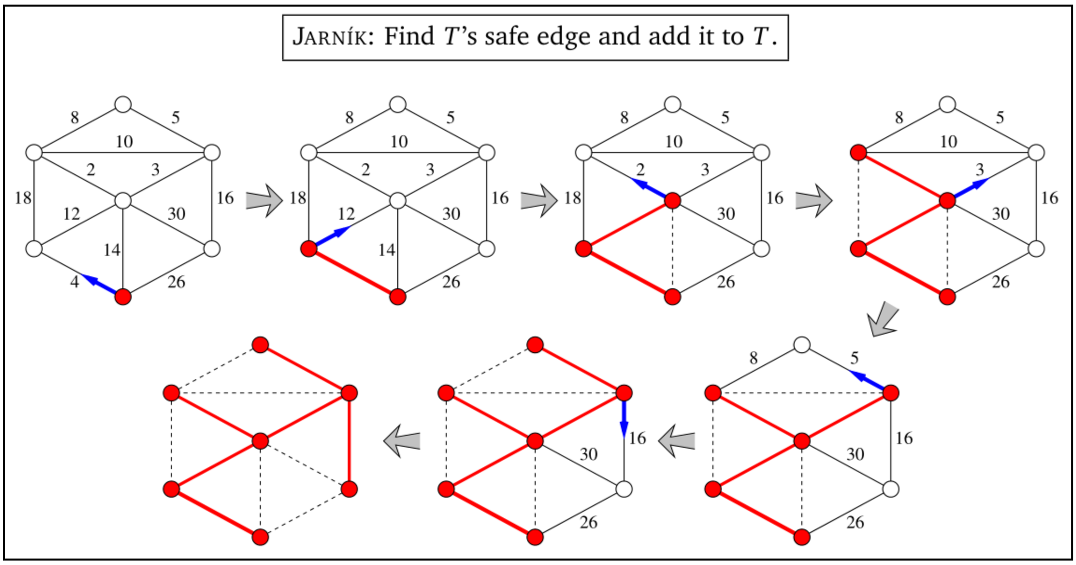


Figure 3.8. Jarník's algorithm.

Jarník's algorithm runs on the example graph in Figure 3.8, starting with the bottom vertex. At each stage, thick edges are in *T*, an arrow points along *T*'s safe edge, and dashed edges are useless.

<u>Prim's algorithm</u>:

let T be a single vertex x

while (T has fewer than n vertices)

{

find the smallest edge connecting T to G–T

add it to T

}

Prim's algorithm appears to spend most of its time finding the smallest edge to grow. A straightforward method finds the smallest edge by searching the adjacency lists of the vertices in *V*; then each iteration costs *O (m)* time, yielding a total running time of *O (mn)*. By using binary heaps, this can be improved to *O (m log n)*. By using Fibonacci heaps, Prim's algorithm runs in *O (m + n log n)* time.

3.3.1.4. Kruskal's Algorithm

Kruskal's algorithm was given by Joseph Kruskal in 1956. It creates a forest where each vertex in the graph is initially a separate tree. It then sorts all the edges in the graph. For each edge *(u, v)* in sorted order, the following is done. If vertices *u* and *v* belong to two different trees, then *(u, v)* is added to the forest, combining two trees into a single tree. It proceeds until all the edges have been processed.

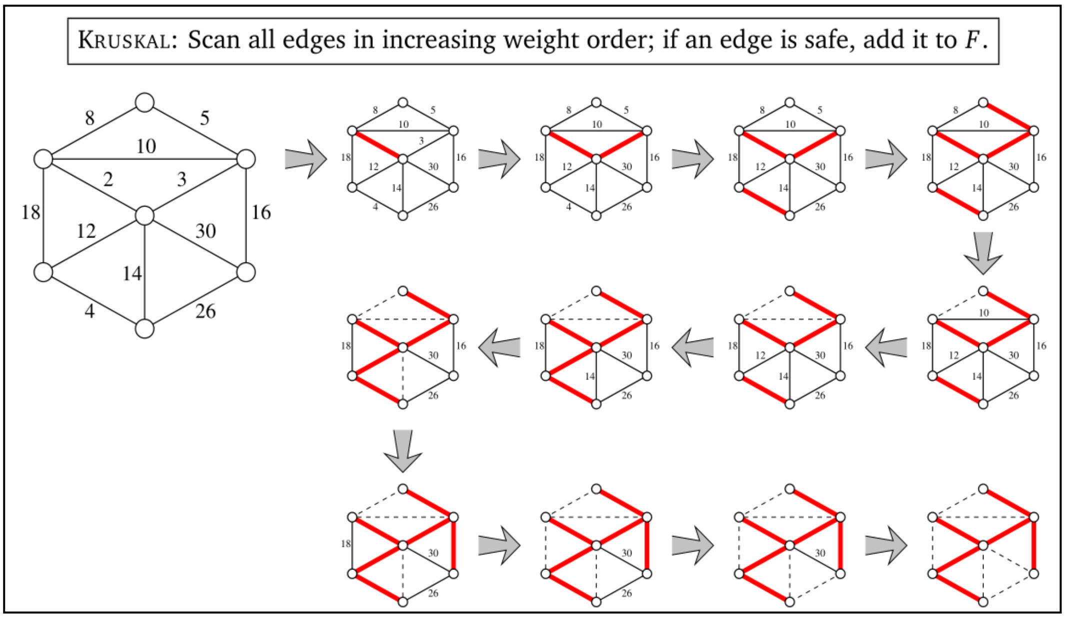


Figure 3.9. Kruskal's algorithm.

Kruskal's algorithm runs on the example graph in Figure 3.9. Thick edges are in *F* and dashed edges are useless. Since examining the edges in order from lightest to heaviest, any edge examine is safe if and only if its endpoints are in different components of the forest *F*. To prove this, suppose the edge *e* joins two components *A* and *B* but is not safe. Then there would be a lighter edge *e′* with exactly one endpoint in A. However, this is impossible, because (inductively) any previously examined edge has both endpoints in the same component of F.

Kruskal's algorithm:

sort the edges of G in increasing order by length
keep a sub-graph S of G, initially empty
for each edge e in sorted order
  if the endpoints of e are disconnected in S
  add e to S
return S

Sorting the edges in non-decreasing order takes $O(m \log m)$ time. The total running time of determining if the edge joins two distinct trees in the forest is $O(m\, \alpha(m, n))$ time, where α is the functional inverse of Ackermann's function defined. Therefore the asymptotic running time of Kruskal's algorithm is $O(m \log m)$, which is the same as $O(m \log n)$ since $\log m = \Theta(\log n)$ by observing that $m = O(n2)$ and $m = \Omega(n)$ [28–30].

3.4. STOCK CORRELATION NETWORK

Stock correlation network is a type of financial network based on stock price correlation used for observing, analyzing and predicting the stock market dynamics. The basic approach for building the stock correlation network involves two steps:

- The first step aims at finding the correlation between each pair of stock considering their corresponding time series.
- The second step applies a criterion to connect the stocks based on their correlation.

The popular method for connecting two correlated stocks is the *minimum spanning tree* method. The other methods are, planar maximally filtered graph, and winner-take-all method. In all three methods, the procedure for finding correlation between stocks remains the same.

3.4.1. Networks of Financial Time Series

To build a stock correlation network, firstly, these steps would follow:

- Entities generating financial time series (stocks, indices, hedge funds or currencies) are represented by nodes.
- Weighted edges between nodes represent the correlation between the time series generated by these entities.
- This gives a fully connected network with $\frac{1}{2}[n(n-1)]$ edges where n is the number of nodes [31].

3.4.1.1. Stocks

In accounting, there are two common uses of the term *stock*. One meaning of stock refers to the goods on hand which is to be sold to customers. In that situation, stock means inventory. The term *stock* is also used to mean the ownership shares of a corporation. For example, an owner of a corporation will have a stock certificate, which provides evidence of his or her ownership of a corporation's common stock or preferred stock. The owner of the corporation's common or preferred stock is known as a stockholder [32].

3.4.1.2. Liquid and Illiquid Stocks

Liquid Stocks

An asset that can be converted into cash quickly and with minimal impact to the price received. Liquid assets are generally regarded in the same light as cash because their prices are relatively stable, when they are sold on the open market. For an asset to be liquid, it needs an established market with enough participants to absorb the selling without materially influencing the price of the asset. There also needs to be a relative ease in the transfer of ownership and the movement of the asset. Liquid assets include most stocks, money market instruments and government bonds. The foreign exchange market is deemed to be the most liquid market in the world because trillions of Dollars exchange hands each day, making it impossible for any one individual to influence the exchange rate [33].

Illiquid Stocks

It is the state of a security or other asset that cannot easily be sold or exchanged for cash without a substantial loss in value. Illiquid assets also cannot be sold quickly because of a lack of ready and willing investors or speculators to purchase the asset. The lack of ready buyers also leads to larger discrepancies between the asking price (from the seller) and the bidding price (from a buyer) than would be found in an orderly market with daily trading activity. Illiquid securities carry higher risks than liquid ones; this becomes especially true during times of market turmoil, when the ratio of buyers to sellers may be thrown out of balance. During these times, holders of illiquid securities may find themselves unable to unload them at all, or unable to do so without losing a lot of money [34].

3.4.1.3. Sectors

Sectors are a distinct sub-set of a market, society, industry, or economy, whose components share similar characteristics. Stocks are often grouped into different sectors depending upon the company's business. For example, Standard & Poor's breaks the market into 11 sectors. Two of these sectors, utilities and consumer staples, are said to be defensive sectors, while the rest tend to be more cyclical in nature. The other nine sectors are transportation, technology, health care, financial, energy, consumer cyclical, basic materials, capital goods, and communications services. Other groups break up the market into different sector categorizations, and sometimes break them down further into sub-sectors [35].

3.4.1.4. Indices

It is a statistical indicator providing a representation of the value of the securities, which constitute it. Indices often serve as barometers for a given market or industry and benchmarks against which financial or economic performance is measured [36].

3.4.1.5. Desired Time Series Data

The time series data can be daily closing prices, daily trading volumes, daily opening prices, and daily price returns.

- <u>Daily closing prices</u>: The final price at which a security is traded on a given trading day. The closing price represents the most up-to-date valuation of a security until trading commences again on the next trading day [37].
- <u>Daily trading volumes</u>: The average amount of individual securities traded in a day or over a specified amount of time. Trading activity relates to the liquidity of a security; therefore, when average daily trading volume is high, the stock can be easily traded and has high liquidity. As a result, average daily trading volume can have an effect on the price of the security. If trading volume isn't very high, the security will tend to be less expensive because people are not as willing to buy it [38].
- <u>Daily opening prices</u>: The price at which a security first trades upon the opening of an exchange on a given trading day. A security's opening price is an important marker for that day's trading activity, especially for those interested in measuring short-term results, such as day traders. Additionally, securities, which experience very large intra-day gains and losses, will have those swings measured relative to their opening price for the day [39].

Daily Price Returns

A performance measure used to evaluate the efficiency of an investment or to compare the efficiency of a number of different investments. To calculate ROI (Rate of Interest), the benefit (return) of an investment is divided by the cost of the investment; the result is expressed as a percentage or a ratio [40]. The initial value of an investment, V_i, does not always have a clearly defined monetary value, but for purposes of measuring ROI, *the expected value must be clearly stated* along with the rationale for this initial value. Similarly, the final value of an investment, V_f, also does not always have a clearly defined monetary value, but for purposes of measuring ROI, *the final value must be clearly stated* along with the rationale for this final value. There are two type of calculation:

- Arithmetic return is:

$$r_{arith} = \frac{V_f - V_i}{V_i} \tag{3.3}$$

r_{arith} sometimes refers to as yield.

- Let $V_i(\tau)$ be the stock-price of a company $i = i(1, \ldots, N)$ at time τ. Then, the return of the stock-price at a time interval defined as;

$$r = \frac{ln\left(\frac{V_f}{V_i}\right)}{\tau} \quad \text{or} \quad r_i(\tau) = ln\, V_i(\tau) - ln\, V_i\, (\tau - \Delta t) \tag{3.4}$$

meaning the geometrical change of $V_i(\tau)$ during the interval Δt [41].

3.4.2. Correlation

The Pearson Product-Moment Correlation Coefficient *(r)*, or correlation coefficient for short is a measure of the degree of linear relationship between two variables, usually labeled *X* and *Y*. While in regression, the emphasis is on predicting one variable from the other, in correlation, the emphasis is on the degree to which a linear model may describe the relationship between two variables. In regression, the interest is directional, one variable is predicted and the other is the predictor; in correlation, the interest is non-directional, the relationship is the critical aspect.

3.4.2.1. Correlation Coefficient

The correlation coefficient may take on any value between plus and minus one.

$$-1.00 \ \leq \rho \leq +1.00 \tag{3.5}$$

The sign of the correlation coefficient *(+, -)* defines the direction of the relationship, either positive or negative. A positive correlation coefficient means that as the value of one variable increases, the value of the other variable increases and as one decreases the other decreases. A negative correlation coefficient indicates that as one variable increases, the other decreases, and vice-versa.

Taking the absolute value of the correlation coefficient measures the strength of the relationship. A correlation coefficient of $\rho = 0.50$ indicates a stronger degree of linear relationship than one of $\rho = 0.40$. Likewise a correlation coefficient of $\rho = -0.50$ shows a greater degree of relationship than one of $\rho = 0.40$. Thus a correlation coefficient of zero $(\rho = 0.0)$ indicates the absence of a linear relationship and correlation coefficients of $\rho = +1.0$ and $\rho = -1.0$ indicate a perfect linear relationship [42].

Correlation coefficient can be calculated as;

$$\rho_{X,Y} = \frac{cov\ (X,Y)}{\sigma_X \sigma_Y} \tag{3.6}$$

where *cov* means covariance and σ is standard deviation.

Covariance provides a measure of the strength of the correlation between two or more sets of random varieties. The covariance for two random varieties X and Y, each with sample size N, is defined by the expectation value;

$$cov\ (X,Y) = \langle (X - \mu_X)(Y - \mu_Y) \rangle = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N} \tag{3.7}$$

where μ_X and μ_Y are the respective means [43].

The standard deviation (σ) of a probability distribution is defined as the square root of the variance (σ^2),

$$\sigma = \sqrt{\langle X^2\rangle\langle X\rangle^2} \tag{3.8}$$

where $\langle X\rangle$ is the mean and the square root of the sample variance of a set of N values is the sample standard deviation [44].

$$S_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3.9}$$

If these two equations are combined, a correlation coefficient formula for two series can be obtained as;

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}} \tag{3.10}$$

The Pearson correlation can be expressed in terms of uncentered moments. Since $\mu_X = E(X)$, $\sigma_x^2 = E[X - E(X)^2] = E(X^2) - E^2(X)$ and likewise for Y, and since;

$$E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = E(XY) - E(X)E(Y) \tag{3.11}$$

The correlation can also be written as;

$$\rho_{X,Y} = \frac{\langle XY\rangle - \langle X\rangle\langle Y\rangle}{\sqrt{(\langle X^2\rangle - \langle X\rangle^2)(\langle Y^2\rangle - \langle Y\rangle^2)}} \tag{3.12}$$

3.4.2.2. Distance

The weights of the links between nodes are based on the correlation between them. The most intuitive measure of distance is the Euclidian distance between the time series;

$$d_{ij} = 1 - \rho_{ij} \tag{3.13}$$

Since;

$-1 \leq \rho_{ij} \leq 1$, have $0 \leq d_{ij} \leq 2$ (3.14)

This is a non-linear transformation of the correlation which gives a metric distance between nodes [31].

3.4.3. Statistical Moments

The *mean* and the *variance* provide information on the location and variability (spread, dispersion) of a set of numbers, and by doing so; provide some information on the appearance of the distribution of the numbers. The mean and variance are the first two *statistical moments*, and the third and fourth moments provide information on the shape of the distribution. For comparison;

$$First\ moment = \sum_{i=1}^{n} (x_i - \bar{X})^1 \tag{3.15}$$

is by definition is equal to zero. One might think of the *mean* as being that value of *x* that makes the above statement true, and consequently indicates where the individual numbers generally lie. The second moment is recognized as the numerator of the *variance*;

$$Second\ moment = \sum_{i=1}^{n} (x_i - \bar{X})^2 \tag{3.16}$$

this gives information on the spread or scale of the distribution of numbers. The third moment $\sum_{i=1}^{n}(x_i - \bar{X})^3$ is used to define the *skewness* of a distribution;

$$Skewness = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^3}{ns^3} \tag{3.17}$$

Skewness is a measure of the symmetry of the shape of a distribution. If a distribution is symmetric, the skewness will be zero. If there is a long tail in the positive direction, skewness will be positive, while if there is a long tail in the negative direction, skewness will be negative. The fourth moment $\sum_{i=1}^{n}(x_i - \bar{X})^4$ is used to define the *kurtosis* of a distribution;

$$Kurtosis = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^4}{ns^4} \tag{3.18}$$

Kurtosis is a measure of the flatness or peakedness of a distribution. Flat-looking distributions are referred to as *"platykurtic"*, while peaked distributions are referred to as *"leptokurtic"* [45].

3.4.4. Normal (Gaussian) Distribution

The normal distribution is the most widely known and used of all distributions. Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems. Distribution model and its properties can be seen in Figure 3.10.

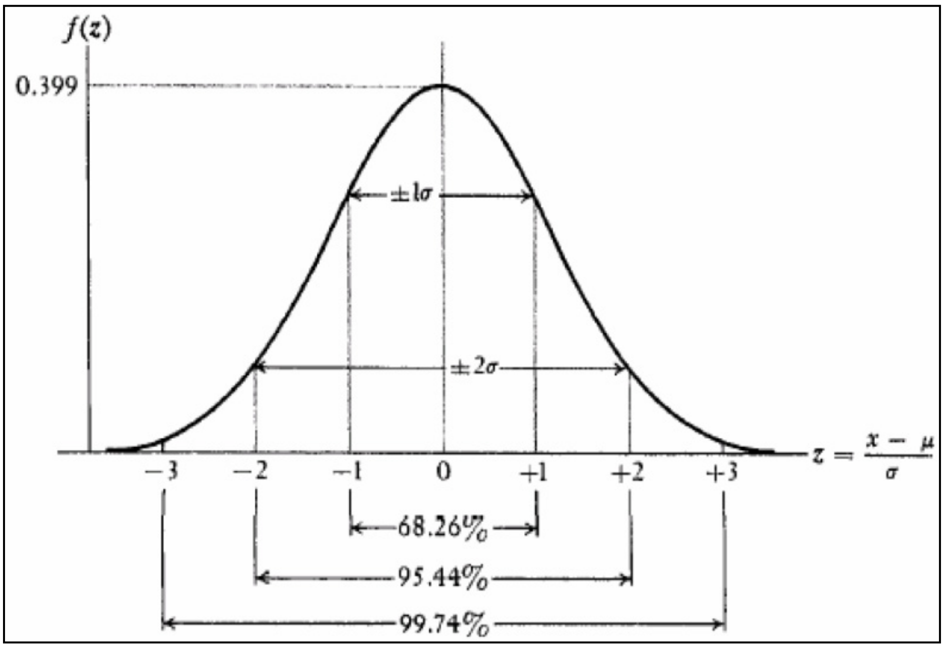


Figure 3.10. Distribution model.

Characteristics of the Normal Distribution:

- Symmetric and bell shaped,
- Continuous for all values of *X* between -∞ and ∞ so that each conceivable interval of real numbers has a probability other than zero.
- $-\infty \leq X \leq \infty$
- Two parameters, μ and σ. Note that the normal distribution is actually a family of distributions, since μ and σ determine the shape of the distribution.
- The rule for a normal density function is;

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{3.19}$$

- The notation $N(\mu, \sigma^2)$ means normally distributed with mean μ and variance σ^2. If $X \sim N(\mu, \sigma^2)$ mean that *X* is distributed $N(\mu, \sigma^2)$.
- About *2/3* of all cases fall within one standard deviation of the mean that is $P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826$.
- About *95%* of cases lie within *2* standard deviations of the mean, that is $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9544$.

Benefit of Normal Distribution:

- Many things actually are normally distributed, or very close to it. For example, height and intelligence are approximately normally distributed; measurement errors also often have a normal distribution.
- The normal distribution is easy to work with mathematically. In many practical cases, the methods developed using normal theory work quite well even, when the distribution is not normal.
- There is a very strong connection between the size of a sample *N* and the extent to which a sampling distribution approaches the normal form. Many sampling distributions based on large *N* can be approximated by the normal distribution even though the population distribution itself is definitely not normal [46].

PART 4

METHODOLOGY

In brief, it has been shown that many natural and social systems display unexpected statistical properties of links connecting different elements of the system and cannot therefore be described in terms of random graphs. The fact that financial markets behave as a complex system with huge amounts of available data has resulted in bringing in new approaches developed during the past decades such as, network structures and characterizations, which help towards our understanding of the dynamics of economic systems. The process of clustering a set of economic entities can improve economic forecasting and modeling of composed financial entities, for example, stock portfolios.

The study of correlations in the parameters of a system is great importance for a variety of physical phenomena. The calculation or the explicit measurement of a correlation function can provide a wealth of information, for example, about the structure of the system. Similarly the study of correlations between different stocks, based on analyzing the properties of their correlation matrix, regarding the fluctuations in their price, is considerable importance for understanding the behavior of financial markets [3,7,10].

Besides, the correlation matrix of the time series of a multivariate complex system can be used to extract information about aspects of hierarchical organization of such a system. The clustering procedure is done by using the correlation between pairs of elements as a similarity measure and by applying a clustering algorithm to the correlation matrix. Because of the clustering procedure, a hierarchical tree of the elements of the system is obtained. The correlation based clustering procedure allows also associating a correlation based network with the correlation matrix [12].

First of all, to construct a stock correlation network, a stock market, number of nodes, data type and date period should be chosen. Then, these quantities should be greatly investigated according to their specific attributions.

In this part of thesis, chosen and investigated stock market would be introduced, its dynamics, stocks and their sectors would be demonstrated, calculations, their results and analyses done would be represented. Finally, the stock correlation network would be constructed and results of financial network techniques would used to examine this stock market would be represented.

As said before, to construct a stock correlation network, a scale of data should be chosen. This data shows the change in market in scale and the effects for future movements. In next, the chosen market and its indices would be discussed.

4.1. ISTANBUL STOCK EXCHANGE

Istanbul Stock Exchange (ISE) or Bourse Istanbul (BIST) began its operation in 1986 and has been the only stock exchange in Turkey. It has demonstrated a considerable growth since its establishment in 1986. The total market capitalization of the firms traded increased from *US$ 938 million* at the end of 1986, to *US$ 30.8 billion* at the end of 1996 and *US$ 202.8 billion* at the end of 2012. Another noticeable growth was observed in the trading value, which sharply increased from only *US$ 13 million* in 1986, to over *US$ 51 billion* in 1995 and *US$ 1.5 trillion* at the end of 2012. The listing requirements for the securities presenting partnership are regulated by both the ISE and the Capital Market Board. To get the listing of a security at exchange, the following conditions are required: the number of shareholders must be above *100*; at least *15%* of the paid-in capital must have been publicly offered; at least *3* years must have elapsed since the incorporation date. The exchange administration normally determines and approves a financial structure, which must be at a level to enable the company to carry out its activities. The firm is also required to show a profit in the previous *2* consecutive years [47].

4.1.1. Bourse Istanbul Indices

ISE price indices are computed and published throughout the trading session while the return indices are calculated and published at the close of the session only.

- ISE National–All Shares Index is composed of all National Market companies except investment trusts.

- ISE National–30 is composed of National Market companies except investment trusts and will be used for trading in the Derivatives Market. The constituent *30* companies are selected based on pre-determined criteria directed for the companies to be included in the indices.

- ISE National–50 is composed of National Market companies except investment trusts. The constituent *50* companies are selected based on pre-determined criteria directed for the companies to be included in the indices. ISE National–50 Index contains the ISE National–30 Index companies.

- ISE National–100, which has been calculated since the inception of the ISE, is composed of National Market companies except investment trusts. The constituents *100* companies of the index are selected based on pre-determined criteria directed for the companies to be included in the indices. ISE National–100 Index contains the ISE National–50 and ISE National–30 Index companies.

- Sector and sub-sector indices are composed of National Market companies excluding investment trusts.

In this thesis ISE National–100 (BIST–100) index was used due to it is used as a main indicator of the National Market. The existing stock market indices and sub-sectors of companies in the Istanbul Stock Exchange are shown in Table 4.1.

Table 4.1. ISE indices and sub-sectors [48].

CODE	INDICES and SUB-SECTORS
XU030	ISE National–30
XU050	ISE National–50
XU100	ISE National–100
XUTUM	ISE National–All Shares
XUSIN	ISE National–Industrials
XGIDA	Food, Beverage
XKAGT	Wood, Paper, Printing
XKMYA	Chemical, Petroleum, Plastic
XMADN	Mining
XMANA	Basic Metal
XMESY	Metal Products, Machinery
XTAST	Non-metal Mineral Products
XTEKS	Textile, Leather
XUHIZ	ISE National–Services
XELKT	Electricity
XILTM	Telecommunications
XINSA	Construction
XSPOR	Sport
XTCRT	Wholesale and Retail Trade
XTRZM	Tourism
XULAS	Transportation
XUMAL	ISE National–Financials
XBANK	Banks
XFINK	Leasing, Factoring
XGMYO	Real Estate Investment Trusts
XHOLD	Holding and Investment
XSGRT	Insurance
XUTEK	ISE National–Technology
XBLSM	Information Technology
XSVNM	Defense

4.1.2. BIST–100 Companies

The composition of the ISE National–30, ISE National–50 and ISE National–100 indices are reviewed and adjusted 4 times on a quarterly basis for the periods *January–March*, *April–June*, *July–September* and *October–December*.

In next Table 4.2, the companies and sectors, which existed in BIST–100 for used *2-year* data scale (2011–2013), specified at the *October–December* period of *2012* can be seen. Data was taken from ISE [49].

Table 4.2. BIST–100 stocks in October–December 2012.

CODE	COMPANY	INDEX	SECTOR	CODE	COMPANY	INDEX	SECTOR	CODE	COMPANY	INDEX	SECTOR
AEFES	ANADOLU EFES	XUSIN	XGIDA	FENER	FENERBAHÇE FUTBOL	XUHIZ	XSPOR	NETAS	NETAŞ TELEKOM.	XUTEK	XBLSM
AFYON	AFYON ÇİMENTO	XUSIN	XTAST	FROTO	FORD OTOSAN	XUSIN	XMESY	NTHOL	NET HOLDİNG	XUMAL	XHOLD
AKBNK	AKBANK	XUMAL	XBANK	GARAN	GARANTİ BANKASI	XUMAL	XBANK	NTTUR	NET TURİZM	XUHIZ	XTRZM
AKENR	AK ENERJİ	XUHIZ	XELKT	GLYHO	GLOBAL YAT. HOLD.	XUMAL	XHOLD	OTKAR	OTOKAR	XUSIN	XMESY
AKSA	AKSA	XUSIN	XKMYA	GOLDS	GOLDAS KUYUMCULUK	XUMAL	XHOLD	PETKM	PETKİM	XUSIN	XKMYA
AKSEN	AKSA ENERJİ	XUHIZ	XELKT	GOLTS	GÖLTAŞ ÇİMENTO	XUSIN	XTAST	PRKME	PARK ELEK. MADEN.	XUSIN	XMADN
ALARK	ALARKO HOLDİNG	XUMAL	XHOLD	GOODY	GOOD-YEAR	XUSIN	XKMYA	RHEAG	RHEA GİRİŞİM	XUMAL	XHOLD
ANHYT	ANADOLU HAYAT EMEK.	XUMAL	XSGRT	GOZDE	GÖZDE GİRİŞİM	XUMAL	XHOLD	SAFGY	SAF GMYO	XUMAL	XGMYO
ANSGR	ANADOLU SİGORTA	XUMAL	XSGRT	GSDHO	GSD HOLDİNG	XUMAL	XHOLD	SAHOL	SABANCI HOLDİNG	XUMAL	XHOLD
ARCLK	ARÇELİK	XUSIN	XMESY	GSRAY	GALATASARAY SPORTİF	XUHIZ	XSPOR	SASA	SASA POLYESTER	XUSIN	XKMYA
ASELS	ASELSAN	XUTEK	XSVNM	GUBRF	GÜBRE FABRİK.	XUSIN	XKMYA	SISE	ŞİŞE CAM	XUMAL	XHOLD
ASYAB	ASYA KATILIM BANKASI	XUMAL	XBANK	HALKB	T. HALK BANKASI	XUMAL	XBANK	SKBNK	ŞEKERBANK	XUMAL	XBANK
AYGAZ	AYGAZ	XUSIN	XKMYA	HURGZ	HÜRRİYET GZT.	XUSIN	XKAGT	SNGYO	SİNPAŞ GMYO	XUMAL	XGMYO
BAGFS	BAGFAŞ	XUSIN	XKMYA	IHEVA	İHLAS EV ALETLERİ	XUSIN	XMESY	TAVHL	TAV HAVALİMANLARI	XUMAL	XHOLD
BANVT	BANVİT	XUSIN	XGIDA	IHLAS	İHLAS HOLDİNG	XUMAL	XHOLD	TCELL	TURKCELL	XUHIZ	XILTM
BIMAS	BİM MAĞAZALAR	XUHIZ	XTCRT	IPEKE	İPEK DOĞAL ENERJİ	XUSIN	XMADN	THYAO	TÜRK HAVA YOLLARI	XUHIZ	XULAS
BIZIM	BİZİM MAĞAZALARI	XUHIZ	XTCRT	ISCTR	İŞ BANKASI (C)	XUMAL	XBANK	TIRE	MONDİ TİRE KUTSAN	XUSIN	XKAGT
BJKAS	BEŞİKTAŞ FUTBOL YAT.	XUHIZ	XSPOR	ISFIN	İŞ FİN.KİR.	XUMAL	XFINK	TKFEN	TEKFEN HOLDİNG	XUMAL	XHOLD
BOYNR	BOYNER MAĞAZACILIK	XUHIZ	XTCRT	ISGYO	İŞ GMYO	XUMAL	XGMYO	TOASO	TOFAŞ OTO. FAB.	XUSIN	XMESY
BRISA	BRİSA	XUSIN	XKMYA	ISYHO	IŞIKLAR YAT. HOLD.	XUMAL	XHOLD	TRCAS	TURCAS PETROL	XUSIN	XKMYA
BRSAN	BORUSAN MANNESMANN	XUSIN	XMANA	ITTFH	İTTİFAK HOLDİNG	XUMAL	XHOLD	TRGYO	TORUNLAR GMYO	XUMAL	XGMYO
CCOLA	COCA COLA İÇECEK	XUSIN	XGIDA	IZMDC	İZMİR DEMİR ÇELİK	XUSIN	XMANA	TRKCM	TRAKYA CAM	XUSIN	XTAST
CEMAS	CEMAŞ DÖKÜM	XUSIN	XMANA	KARSN	KARSAN OTOMOTİV	XUSIN	XMESY	TSKB	T.S.K.B.	XUMAL	XBANK
DEVA	DEVA HOLDİNG	XUSIN	XKMYA	KARTN	KARTONSAN	XUSIN	XKAGT	TSPOR	TRABZONSPOR SPORTİF	XUHIZ	XSPOR
DOAS	DOĞUŞ OTOMOTİV	XUHIZ	XTCRT	KCHOL	KOÇ HOLDİNG	XUMAL	XHOLD	TTKOM	TÜRK TELEKOM	XUHIZ	XILTM
DOHOL	DOĞAN HOLDİNG	XUMAL	XHOLD	KLGYO	KİLER GMYO	XUMAL	XGMYO	TTRAK	TÜRK TRAKTÖR	XUSIN	XMESY
DYHOL	DOĞAN YAYIN HOL.	XUMAL	XHOLD	KONYA	KONYA ÇİMENTO	XUSIN	XTAST	TUPRS	TÜPRAŞ	XUSIN	XKMYA
ECILC	ECZACIBAŞI İLAÇ	XUMAL	XHOLD	KOZAA	KOZA MADENCİLİK	XUSIN	XMADN	ULKER	ÜLKER BİSKÜVİ	XUSIN	XGIDA
ECZYT	ECZACIBAŞI YATIRIM	XUMAL	XHOLD	KOZAL	KOZA ALTIN	XUSIN	XMADN	VAKBN	VAKIFLAR BANKASI	XUMAL	XBANK
EGEEN	EGE ENDÜSTRİ	XUSIN	XMESY	KRDMD	KARDEMİR (D)	XUSIN	XMANA	VESTL	VESTEL	XUSIN	XMESY
EGGUB	EGE GÜBRE	XUSIN	XKMYA	METRO	METRO HOLDİNG	XUMAL	XHOLD	YKBNK	YAPI VE KREDİ BANK.	XUMAL	XBANK
EKGYO	EMLAK KONUT GMYO	XUMAL	XGMYO	MGROS	MİGROS TİCARET	XUHIZ	XTCRT	ZOREN	ZORLU ENERJİ	XUHIZ	XELKT
ENKAI	ENKA İNŞAAT	XUHIZ	XINSA	MNDRS	MENDERES TEKSTİL	XUSIN	XTEKS				
EREGL	EREĞLİ DEMİR CELİK	XUSIN	XMANA	MUTLU	MUTLU AKÜ	XUSIN	XMESY				

4.1.3. Data Scale

Between 2011 and 2013, *N=100* stocks bargained during *L=506* days. In this thesis, daily opening prices of stocks were used to calculate their logarithmic returns. Thence, after calculation, each stock had *505* values. The main advantage of logarithmic return is that the continuously compounded return is symmetric, while the arithmetic return is not: positive and negative percent arithmetic returns are not equal. Also, because of different stocks have different opening price scales, logarithmic return is more stable. Figure 4.1 shows opening prices of two *XTAST (Non-metal Mineral Products)* stocks for *506* days.

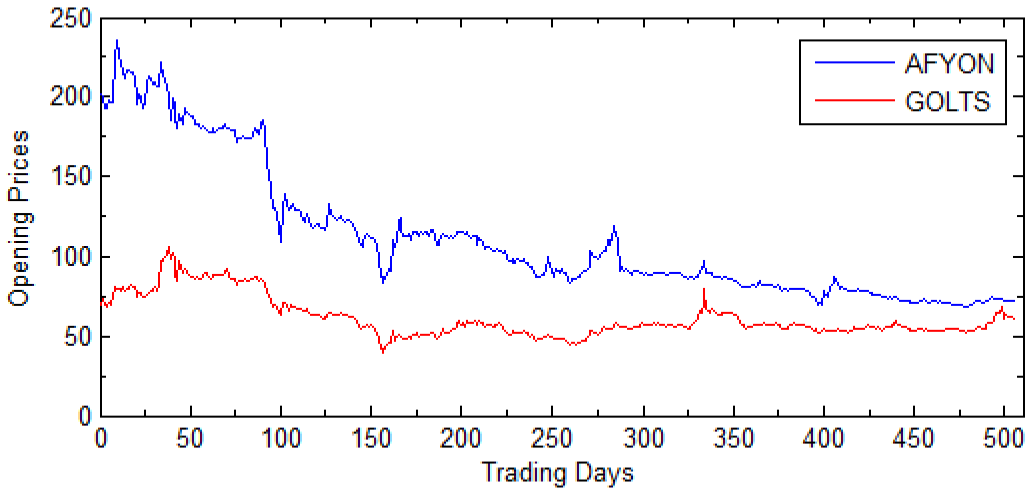


Figure 4.1. Opening prices of two cement companies.

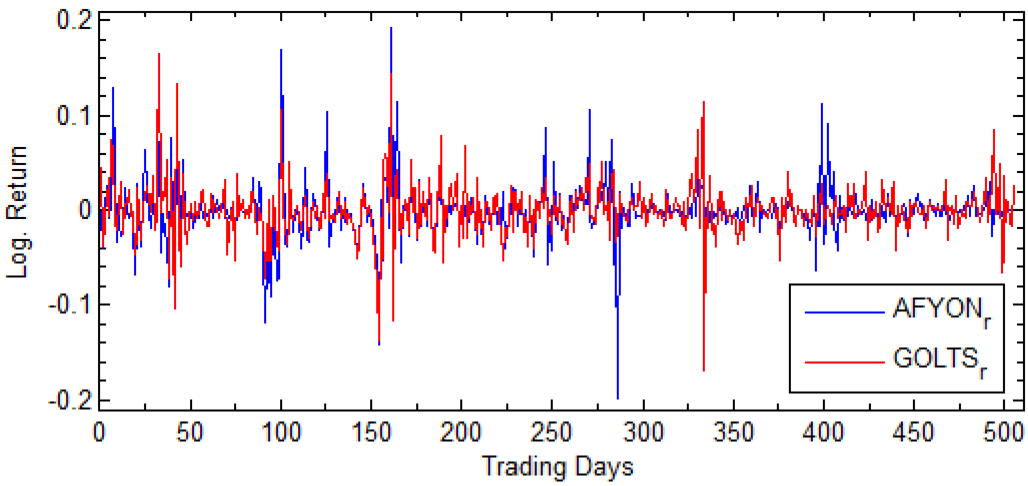


Figure 4.2. Logarithmic return scale of two stocks.

As can be seen in Figure 4.1 *AFYON* starts from *200s* level unlike *GOLTS*. So, for a realistic correlation coefficient, logarithmic return was used. Figure 4.2 is a demonstration of logarithmic return of data that swing in a scale and its correlation is more significant.

4.2. BASIC CALCULATIONS

As specified before, many calculations are done to construct a network, which shows relationships among stocks. In this section, basic calculations and their results would be demonstrated.

4.2.1. Calculation of Correlation Coefficients

To compute correlation coefficient between two stocks, logarithmic return of prices was used in this thesis as mentioned before. Equation 3.10 was used due to the values of these stocks are time series. So;

- As an example, *AFYON* and *GOLTS* were chosen for computation. Firstly, covariance was calculated as there would be *N=505* values for two series. Also, mean values of these two series must be computed. Mean values was found as $\bar{x} = 111.3483$ and $\bar{y} = 61.4962$. If, these values are placed on Equation 3.7;

$$\sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N} = \sum_{i=1}^{505} \frac{(x_i - 111.35)(y_i - 61.50)}{505} = 5.5060x10^{-4} \tag{4.1}$$

  This is covariance between two series. After that, standard deviation of these series must be calculated. Standard deviation of first series is $\sigma_x = 0.0296$ and the second $\sigma_y = 0.0292$. Now, correlation coefficient of these two stocks can be calculated;

$$\rho_{X,Y} = \frac{cov\,(X,Y)}{\sigma_X \sigma_Y} = \frac{5.5060x10^{-4}}{0.0296 \times 0.0292} = 0.6375 \tag{4.2}$$

- According to this value, there is a strong relationship between these two cement stocks can be said.

Now, a correlation matrix can be created by applying this calculation to all time series. This matrix consists of *100* rows and *100* columns due to there are *100* stocks. This *100 x 100* matrix that applied coloring method can be seen in Figure 4.3.

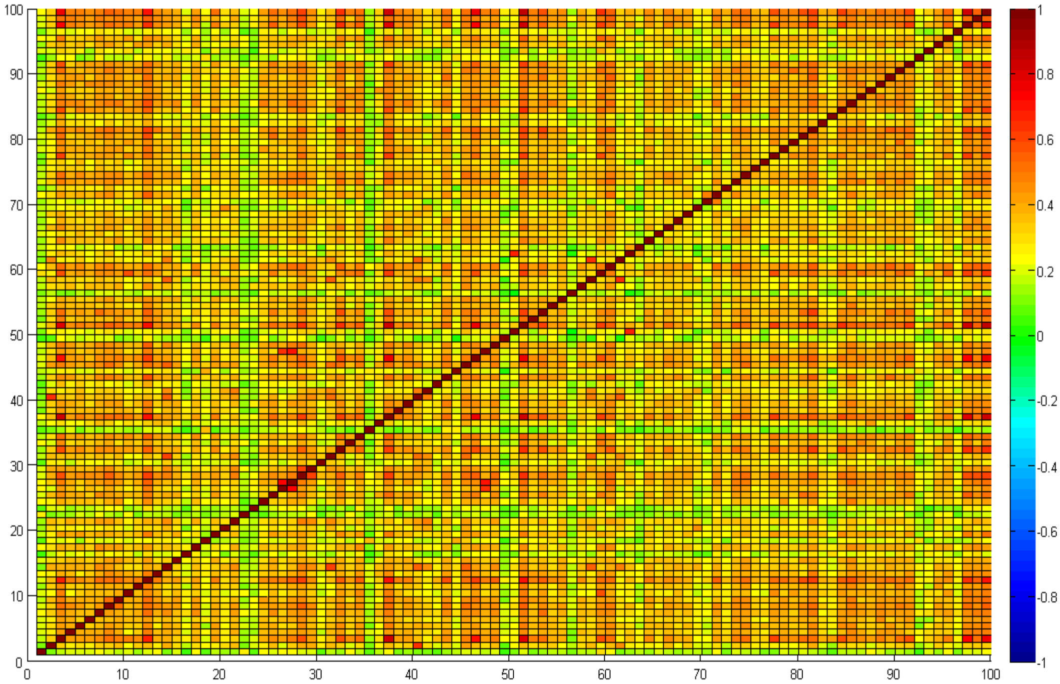


Figure 4.3. Correlation coefficient matrix of ISE National 100 Index.

Diagonal of matrix shows correlation coefficient between same stocks that equal *1*. As can be seen, matrix generally consists of green, yellow, orange and near of red points. So, correlations among stocks mostly changes between *0* and *1*. Some stocks act random (near zero), some ones low (green-yellow), medium (orange) and high correlated (red) according to correlation coefficient values.

4.2.2. Calculation of Metric Distances

In 2000, R. N. Mantegna and H. Stanley mentioned a metric distance that provides the relative distance between the stocks of a given portfolio. A method of determining a distance between stocks i and j evolving in time in a synchronous fashion is the following;

$$\check{S}_i \equiv \frac{S_i - \langle S_i \rangle}{\sqrt{\langle S_i^2 \rangle - \langle S_i \rangle^2}} \tag{4.3}$$

The Euclidean distance d_{ij} between vectors $\check{S}_i$ and $\check{S}_j$ is obtainable from the Pythagorean relation;

$$d_{ij}^2 = \left\| \check{S}_i - \check{S}_j \right\|^2 = \sum^{n} \left(\check{S}_{ik} - \check{S}_{jk} \right)^2 \tag{4.4}$$

The vector $\check{S}_i$ has unit length because, from Equation 4.3,

$$\sum_{k=1}^{n} \check{S}_{ik}^2 = 1 \tag{4.5}$$

Hence, Equation 4.4 can be rewritten as;

$$d_{ij}^2 = \left(\sum_{k=1}^{n} \check{S}_{ik}^2 + \check{S}_{jk}^2 + \check{S}_{ik}\check{S}_{jk} \right) = 2 - 2\sum_{k=1}^{n} \check{S}_{ik}\check{S}_{jk} \tag{4.6}$$

The sum on the right side of Equation 4.6 coincides with ρ_{ij}. Hence, it leads to;

$$d_{ij} = \sqrt{2\left(1 - \rho_{ij}\right)} \tag{4.7}$$

[50].

As an example, *AFYON* and *GOLTS* were used again to compute metric distance between these stocks. Correlation coefficient of these stocks calculated above was *0.6375*. So, the metric distance was calculated as seen below;

$$d_{ij} = \sqrt{2(1-\rho_{ij})} = \sqrt{2(1-0.6375)} = 0.7250 \tag{4.8}$$

So, values in correlation matrix can be used to calculate metric distances.

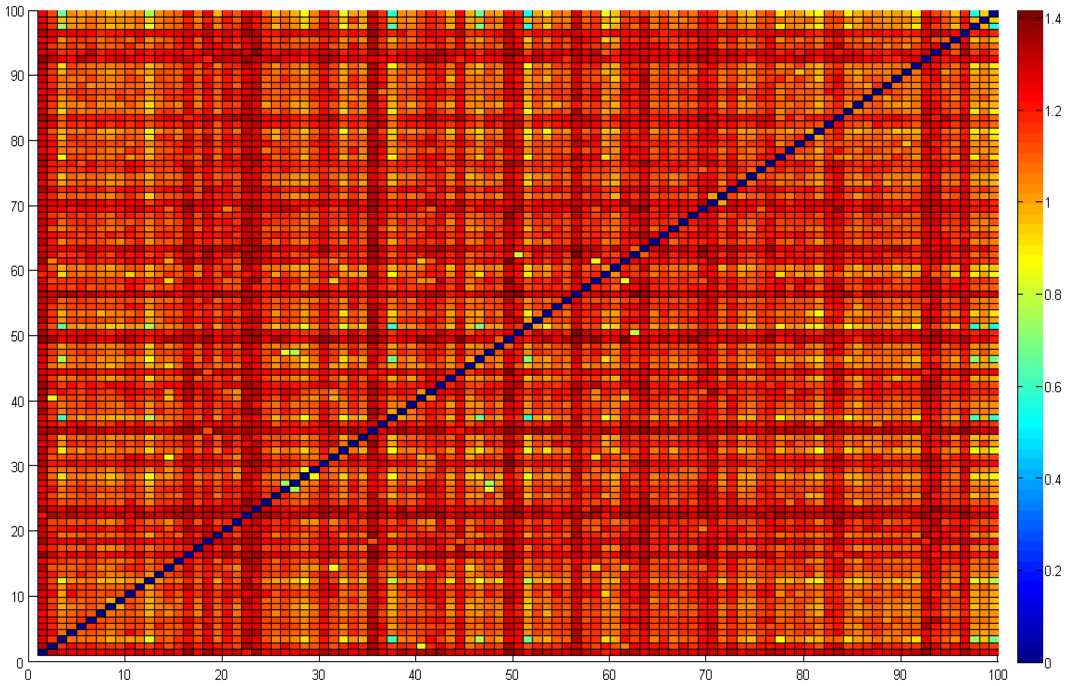


Figure 4.4. Distance matrix.

New matrix representation in Figure 4.4 shows that distances among stocks lie between *0* and *1.4*. Blue diagonal is the calculations between same stocks that have *0* values owing to correlation coefficients between same stocks are *1*. In addition, connections between stocks decrease from blue to red points.

4.3. STATISTICAL CALCULATIONS

First, daily values of BIST–100 Index were analyzed. This compound index will help to figure out trend of whole market during *2011–2013*. Here are used daily opening values of BIST–100 Index and plot a graph in Figure 4.5. Between 2011 and 2013, BIST–100 index showed an exponential behavior. Until *Feb-2012*, there was a decrease, but increase after it. Its cubic spline proved this movement of the index. Mean value *(μ)* of index was *62230* and standard deviation *(σ)* was *6155*.

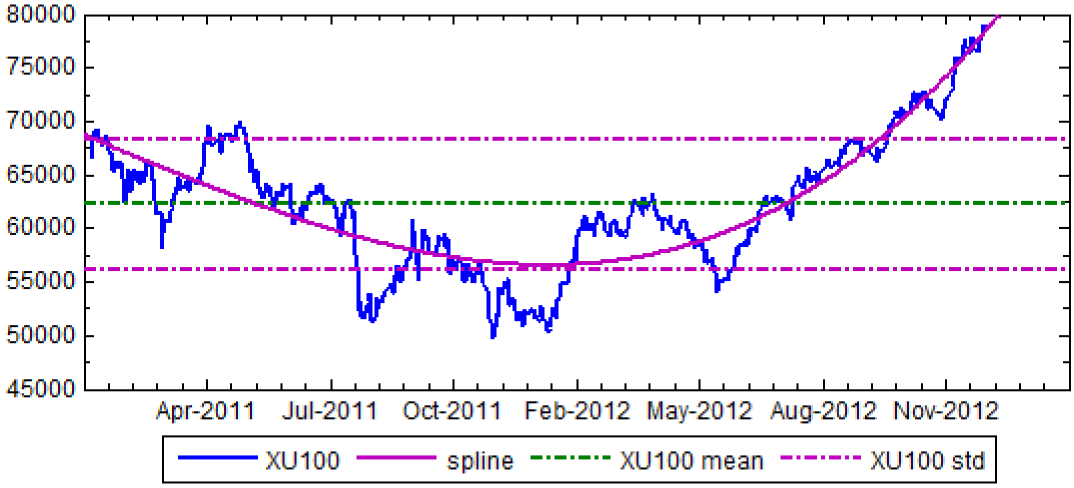


Figure 4.5. BIST–100 opening values.

4.3.1. Distribution of Correlation Coefficients

To figure out evolution of market in more detail, probability distribution of correlation coefficients was plotted in Figure 4.6 and to capture how the correlation coefficients evolve through time, *six-month time windows* (t=126 trading days) were created and calculated the statistical moments of each. Finally, these data combined and distributions of *2-years* were analyzed in Figure 4.7.

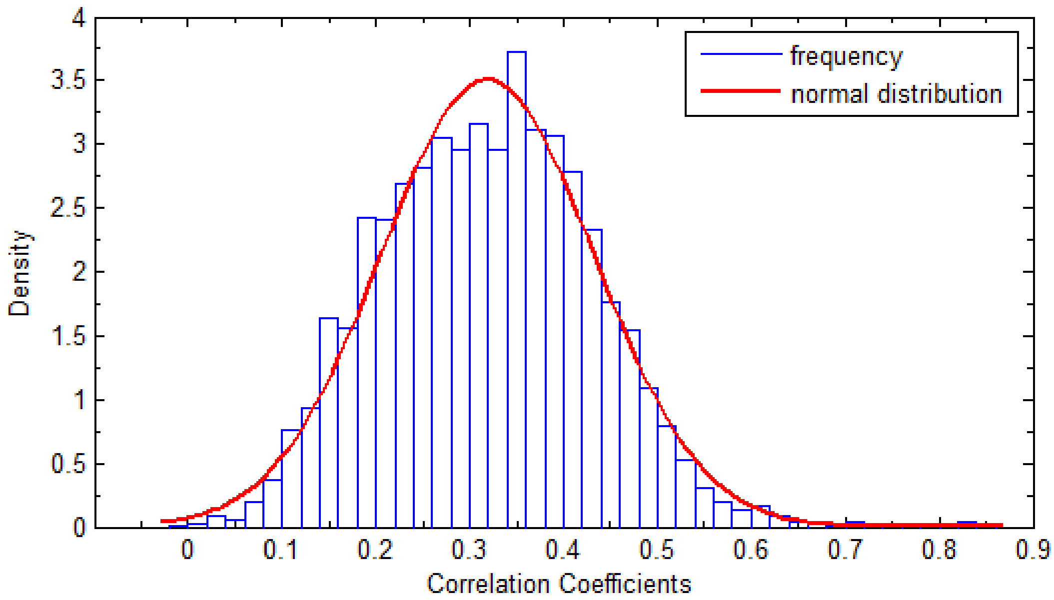
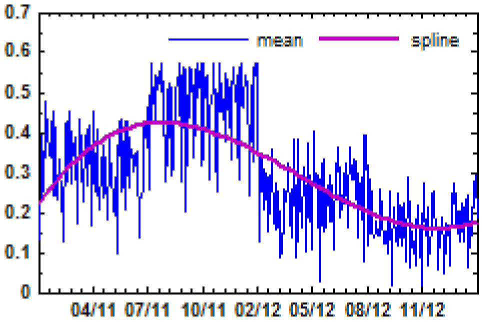
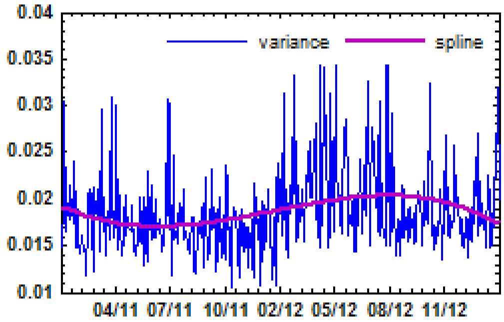


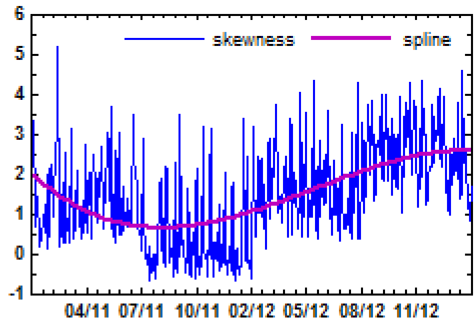
Figure 4.6. Distribution of correlation coefficients.



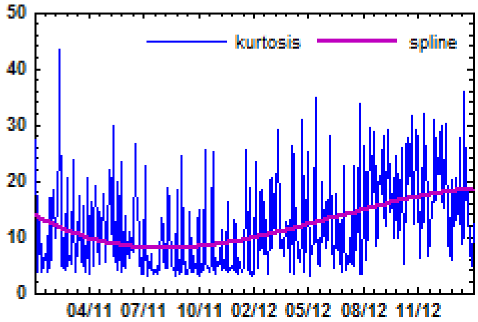
a) Mean of correlation coefficients.



b) Variance of correlation coefficients.



c) Skewness of correlation coefficients.



d) Kurtosis of correlation coefficients.

Figure 4.7. Statistical moments of correlation coefficients.

Correlation coefficients of BIST–100 lie generally between *0.1* and *0.6*. Its mean value (μ) is *0.32* and standard deviation (σ) is *0.11*. Correlation coefficients of BIST–100 stocks changes in $0.1<\rho_{ij}<0.6$ and all of them are larger than *-0.1*. This distribution similar to *"normal"* and indicates *"up and down in the same direction"* means prices of most stocks in BIST–100 fluctuate towards the same direction [51].

When statistical moments of coefficients in six-month period were computed, as can be seen there is clearly an increase in *mean of correlation coefficients* between *May 2011* and *February 2012*. In addition, the *skewness* decreases toward zero, implying that the distribution of the correlations becomes more Gaussian. According to our knowledge, the high values of mean correlation coefficient is a sign of crisis period [8–10,51–53].

The reason of decreasing until *Feb-2012* can be shown crisis in EU Markets, US credit rating downgrade, some important decisions Turkish Central Bank made and sales of stocks by foreign investors because of panic [54].

4.4. STOCK CORRELATION NETWORK OF BIST–100

From a set of *n* time series, the correlation coefficient between any pair of variables can be extracted. If the different time series with the nodes of the network are identified, each pair of nodes can be thought to be connected with a weight related to the correlation coefficient. Therefore, the network would be completely connected.

The correlation coefficient cannot be used as a distance for networks because it does not fulfill the three axioms that define a Euclidean metric. So, the distance matrix *D* is used to determine the weights of edges connecting the *N* stocks of the portfolio. With computing a metric distance, the *minimum spanning tree* and a *hierarchical tree* associated with each correlation coefficient matrix can be obtained. Put differently, *geometrical* (MST) and *taxonomic* (HT) aspects of the correlation present between the stock pairs of a set of stocks can be sorted out using the information contained in the correlation coefficient matrix.

4.4.1. Constructing MST

MST approach is derived from graph theory and has been used as a simple way to investigate the correlations of stocks in a stock market. One of the advantages that MST analysis has over traditional finance perspectives and it provides a parsimonious representation of the network of all possible inter-connectedness. With *N* equity indices, the number of possible nodal connections is large, $\frac{1}{2}[n\ (n-1)]$. Also, useful information in terms of the centrality or otherwise of individual equity markets (nodes) in the overall system are provided by MST.

Besides, this kind of tree is particularly useful to represent complex networks, filtering the information about the correlations between all nodes and presenting it in a planar graph. It is repulsive due to providing an arrangement of stocks, which selects the most relevant connections of each point of the network. Moreover, the minimal spanning tree gives, in a direct way, the sub-dominant ultra-metric hierarchical organization of the points (stocks) of the investigated portfolio. The minimal spanning tree associated with the distance matrix *D* is great tool for an economic point of view. In particular, by assuming this kind of topology, it is able to isolate groups of stocks that make sense by starting from the information carried by the time series of prices only. The MST captures the cooperative behaviors behind the stock market, too. Because of its simplicity and advantageous, they have been widely used to represent some important financial structures such as, namely the structures of stock exchanges, of currency exchange rates, of world trade, commodities, of GDPs (Gross Domestic Products), corporations and, of the world financial markets.

In this thesis, the variable under investigation is the daily price return $r_i(t)$ of asset i on day t. The $N \times N$ correlation matrix is extracted for given a portfolio composed of N assets traded simultaneously in a time-period of T trading days. Each correlation coefficient $\rho_{i,j}$ can be associated to a metric distance $d_{i,j} = \sqrt{2\,(1 - \rho_{i,j})}$ between asset i and j. Then, the distance matrix can be used to determine the MST connecting all the assets. The method of constructing the MST linking N objects is known in multivariate analysis as the nearest neighbor single linkage cluster algorithm [1,3–5, 7,9,15].

Firstly, six-month of data scale were used to construct MSTs as can be seen in Figure Appendix A.1. After that, these networks were used to investigate market dynamics in their period. To construct MSTs, *Prim's* algorithm was used due to it is faster than *Kruskal's* [28–30]. The networks that constructed to compute MSTs are multi, non-planar, directed, weighted, connected, symmetric, complete and isthmus. They have sub-graphs, paths, loops, chains, cycles, circuits, cliques and trees. However, MSTs are simple, planar, undirected, weighted, connected, and asymmetric graphs that have paths, clusters but loops, cycles and circuits in. They have assortative, articulation nodes and root, also.

Normalized tree length of MSTs can be calculated owing to they were divided into periods *(t=126 days)*. More specific information can be obtained by decreasing t window, likewise, statistical moments. Distances among stocks in MSTs, in other term, normalized tree length of MSTs can be shown in Figure 4.8.

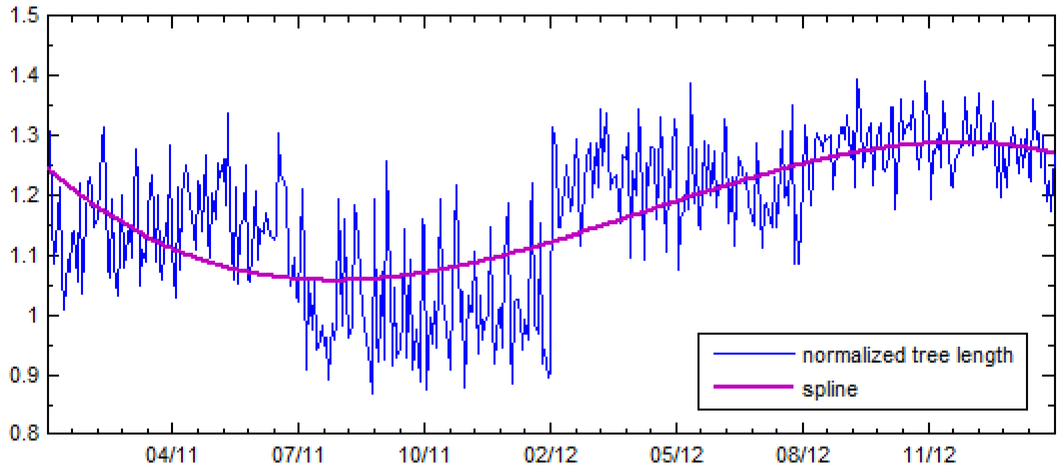


Figure 4.8. Normalized tree length of MSTs.

As well as statistical moments, during the crisis period *(second six-month period)* the tree shrank considerably which means that the distances between the stocks were rapidly decreasing [10].

Before analyzing MST networks, it should be said that grouping of stocks was not perfect at the branch level, it is defined that a smaller subset whose nodes are more homogeneous according to their sector classifications. The term *cluster* is described as a subset of a branch. A *complete cluster* contains all the companies of the studied set belonging to the corresponding business sector, so that none is left outside of the cluster. In practice, however, clusters are mostly incomplete, containing most, but not all, of the companies of the given business sector, and the rest are to be found somewhere else in the tree. While they are analyzed, also it was seen that, all trees had banks cluster in center of tree. High-degree nodes of banks clusters changed but this fact did not. After the crisis period hub nodes of clusters became more meaningful, also [55].

While the four MSTs were analyzed, it was seen that, they behaved differently from each other. The reason of that, nodes (stocks) had variable opening values that were decreasing before slightly crisis period, greatly during and increasing slightly after crisis period, greatly long after. This reason affected their correlation coefficient as well as distances; so, distances among stocks became closer and the two trees, which were formed before crisis period, acted like had great random clusters. Stocks moved especially according to their return of rates values and ignored their sectors. After the crisis period, distances increased and nodes of the other two trees joined to clusters of same kind of stocks according to their time series, sectors, and sub-sectors. Furthermore, they joined to their main holding which would be discussed after.

After investigating four periods of BIST–100 from MST approach, a MST was plotted to see general view of BIST–100 by using all data between *2011* and *2013*. This plot helped to analyze main structure of BIST–100 and examined behaviors of stocks according to their sectors. This tree can be seen in Figure 4.9 (Larger view in Figure Appendix A.2).

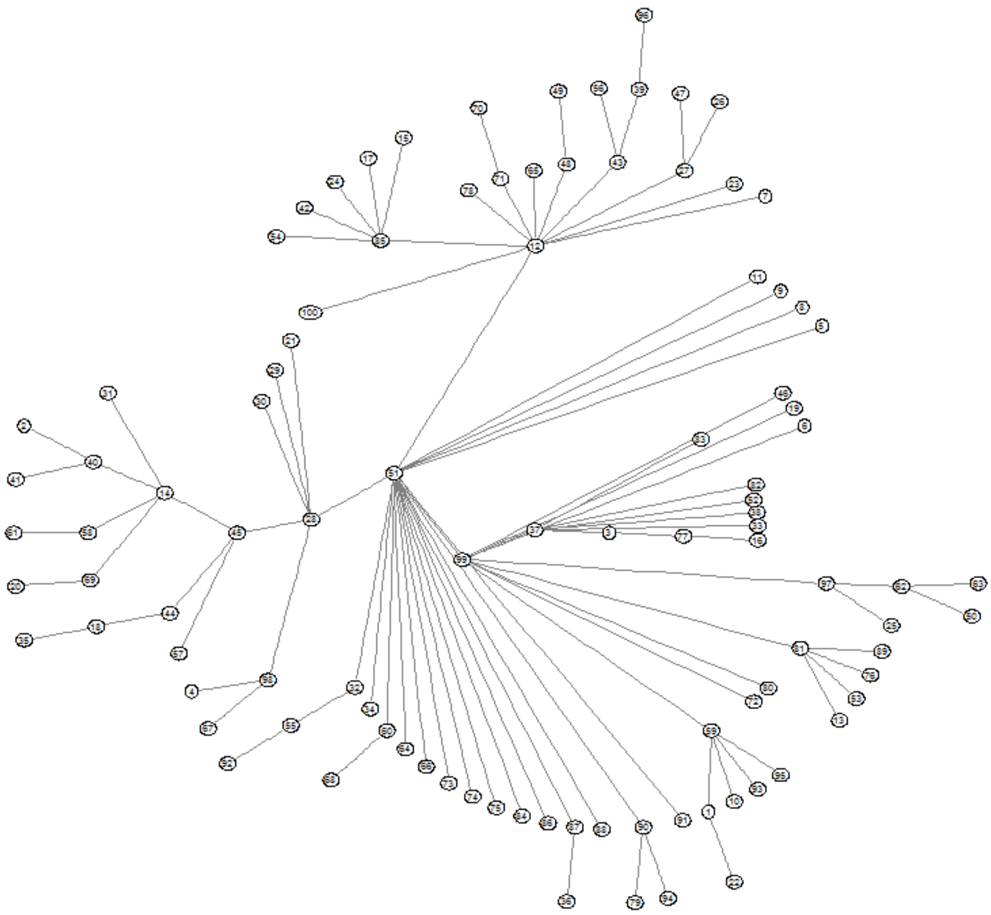


Figure 4.9. MST of BIST–100 in 2011–2013.

Before analyzing MST in details, properties of MST network were computed and investigated. These properties provide to understand that what conditions and quantities MST have and how stock correlation network processes.

4.4.1.1. Properties of MST

To have more specific information about BIST–100 network, some properties were computed and trying to determine characteristics of the network. These properties are; Network Properties: # of nodes, # of edges, average clustering coefficient, diameter, average degree and average path length. Node and Edge Properties: degrees, degree distribution *k*, power-law fitting, node betweenness, edge betweenness and closeness. Computed network properties can be seen in Table 4.3.

Table 4.3. Properties of MST network.

Properties	**Value**
# of nodes	100
# of edges	99
Average clustering coefficient	0.1667
Diameter	9.0449
Average degree	1.0980
Average path length	3.9512

Investigated network has *100* nodes due to tree had *100* stocks and *99* edges. The clustering coefficient of a node is the number of triangles (3-loops) that pass through this node, relative to the maximum number of 3-loops that could pass through the node. Sum of these coefficients are computed and divided total number of nodes and average cluster coefficient is obtained. *Average clustering coefficient* of the network is *1/6=0.1667*. This small value showed that mostly nodes of the network passes *1* or *2* triangle through it, due to it is known that clustering coefficient lays between *0* and *1*, and these nodes showed meaningful clusters.

Moreover, *diameter* is the maximum shortest path length between any two nodes. The value calculated from network showed that the distances between nodes were small and they were connected highly correlated. *Average degree* value pointed that this network had some nodes, which had high node degrees, but most of degrees of nodes were approximately *2*. Finally, by computing *average path length* value, it was shown that nodes of these network so close to each other’s.

Table 4.4. Node properties of MST network.

Node	Deg	In	Out	NBetw	Closeness	Node	Deg	In	Out	NBetw	Closeness
1	2	2	0	0.0354	0.0024	**51**	21	13	8	2.9013	0.0047
2	1	1	0	0.0000	0.0017	**52**	1	0	1	0.0000	0.0026
3	2	2	0	0.2408	0.0029	**53**	1	1	0	0.0000	0.0025
4	1	1	0	0.0000	0.0022	**54**	1	1	0	0.0000	0.0023
5	1	1	0	0.0000	0.0032	**55**	2	1	1	0.0374	0.0025
6	1	1	0	0.0000	0.0030	**56**	1	0	1	0.0000	0.0022
7	1	1	0	0.0000	0.0028	**57**	1	0	1	0.0000	0.0023
8	1	1	0	0.0000	0.0033	**58**	2	1	1	0.0517	0.0020
9	1	1	0	0.0000	0.0032	**59**	5	3	2	0.2474	0.0032
10	1	1	0	0.0000	0.0025	**60**	2	1	1	0.0357	0.0033
11	1	1	0	0.0000	0.0032	**61**	1	0	1	0.0000	0.0017
12	11	10	1	1.2053	0.0039	**62**	3	2	1	0.1383	0.0026
13	1	1	0	0.0000	0.0024	**63**	1	0	1	0.0000	0.0020
14	5	5	0	0.4590	0.0024	**64**	1	0	1	0.0000	0.0032
15	1	1	0	0.0000	0.0023	**65**	1	0	1	0.0000	0.0028
16	1	1	0	0.0000	0.0019	**66**	1	0	1	0.0000	0.0032
17	1	1	0	0.0000	0.0023	**67**	1	1	0	0.0000	0.0022
18	2	2	0	0.0360	0.0018	**68**	1	0	1	0.0000	0.0025
19	1	1	0	0.0000	0.0030	**69**	2	0	2	0.0388	0.0019
20	1	1	0	0.0000	0.0016	**70**	1	1	0	0.0000	0.0022
21	1	1	0	0.0000	0.0027	**71**	2	0	2	0.0444	0.0029
22	1	0	1	0.0000	0.0018	**72**	1	1	0	0.0000	0.0029
23	1	0	1	0.0000	0.0027	**73**	1	0	1	0.0000	0.0033
24	1	1	0	0.0000	0.0023	**74**	1	0	1	0.0000	0.0032
25	1	1	0	0.0000	0.0026	**75**	1	0	1	0.0000	0.0031
26	1	1	0	0.0000	0.0024	**76**	1	1	0	0.0000	0.0024
27	3	1	2	0.1124	0.0029	**77**	2	0	2	0.1070	0.0024
28	6	5	1	1.2436	0.0038	**78**	1	0	1	0.0000	0.0028
29	1	0	1	0.0000	0.0029	**79**	1	1	0	0.0000	0.0025
30	1	0	1	0.0000	0.0027	**80**	1	1	0	0.0000	0.0031
31	1	0	1	0.0000	0.0020	**81**	5	2	3	0.1891	0.0032
32	2	2	0	0.1024	0.0034	**82**	1	0	1	0.0000	0.0026
33	1	1	0	0.0000	0.0026	**83**	1	0	1	0.0000	0.0026
34	1	1	0	0.0000	0.0033	**84**	1	0	1	0.0000	0.0033
35	1	0	1	0.0000	0.0015	**85**	6	0	6	0.2208	0.0030
36	1	1	0	0.0000	0.0024	**86**	1	0	1	0.0000	0.0032
37	7	5	2	0.7445	0.0035	**87**	2	0	2	0.0357	0.0032
38	1	0	1	0.0000	0.0026	**88**	1	0	1	0.0000	0.0032
39	2	2	0	0.0424	0.0023	**89**	1	0	1	0.0000	0.0025
40	3	1	2	0.0935	0.0020	**90**	3	1	2	0.0724	0.0033
41	1	0	1	0.0000	0.0017	**91**	1	0	1	0.0000	0.0033
42	1	1	0	0.0000	0.0023	**92**	1	0	1	0.0000	0.0020
43	3	1	2	0.1513	0.0029	**93**	1	0	1	0.0000	0.0024
44	2	1	1	0.0903	0.0023	**94**	1	0	1	0.0000	0.0025
45	4	1	3	0.8341	0.0030	**95**	1	0	1	0.0000	0.0025
46	1	1	0	0.0000	0.0033	**96**	1	0	1	0.0000	0.0018
47	1	0	1	0.0000	0.0024	**97**	3	1	2	0.3304	0.0035
48	2	1	1	0.0423	0.0029	**98**	3	0	3	0.0814	0.0028
49	1	0	1	0.0000	0.0022	**99**	10	0	10	2.2625	0.0043
50	1	1	0	0.0000	0.0021	**100**	1	0	1	0.0000	0.0029

When node properties were investigated, it was obtained that some nodes had high-degree compared to others, such as *12, 14, 28, 37, 51, 59, 81, 85,* and *99*. These nodes, especially, *51* acted like a hub node and gave direction to clusters and namely to network. *In-degrees* are incoming edges to nodes and *out-degrees* are outgoing. As can be seen in table, some of high-degree nodes have high in-degrees, some of them not. This property showed importance and role of nodes connecting other nodes and clusters in network. Then, *node betweenness* and *closeness centrality* measures helped to analyze role of nodes in network. Nodes with a high betweenness centrality were interesting, because, they control information flow in a network. It was clearly seen in Table 4.4 that high-degree nodes had high-value of betweenness and closeness. It is said that high valued nodes are more important than others are.

On the other hand, *closeness centrality* is a value shows how fast information spreads from the node to other nodes in the network. As well as node betweenness centrality, high-degree nodes had higher values. This means nodes that had higher closeness measure were more capable than others at connecting nodes were and had more important role in clusters than other nodes in network. To investigate edge centrality of network, a 3-D graph that seen in Figure 4.10 was plotted.

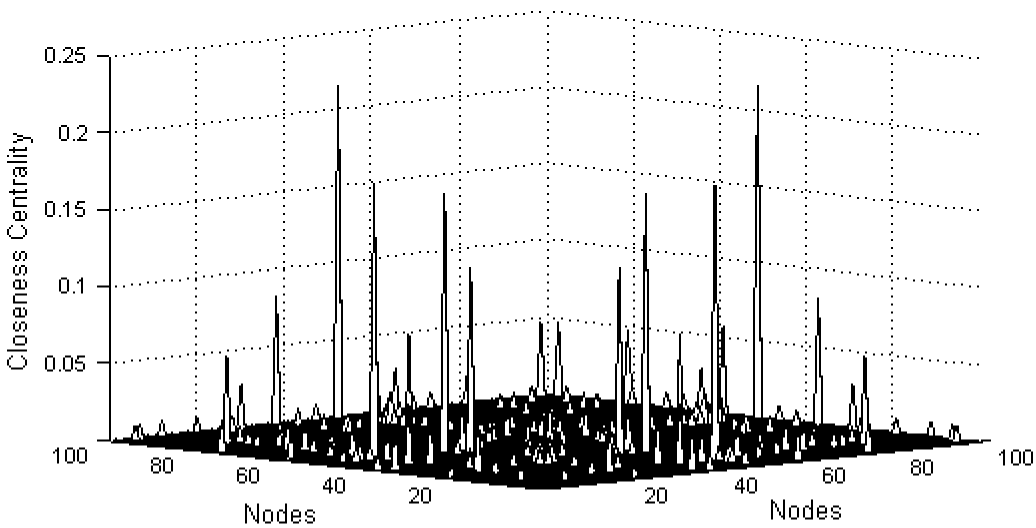


Figure 4.10. Edge betweenness of MST.

As can be seen in graph, *edge betweenness* graph was symmetric due to two nodes connected via same edge. Betweenness is a measure of the number of times an edge occurs on a geodesic. In addition, it is the total amount of flow it carries counting flow between all pairs of nodes using this edge. Shortly, edge betweenness is a way to show important edges. In MST, most high-betweenness valued edge is *51-99* or *99-51* edge according to graph. Deeply, it was seen that edges which connecting hub nodes had higher values than other edges.

In order to get more information from the MST, the *distribution of the degree k*, which is the number of links that are connected to each node of the network, was examined. After that, the distribution for all data MST was plotted with the function in Equation 4.9. The log-log plot of the topological distribution is shown in Figure 4.11.

$$P(k) \propto k^{-\gamma} \tag{4.9}$$

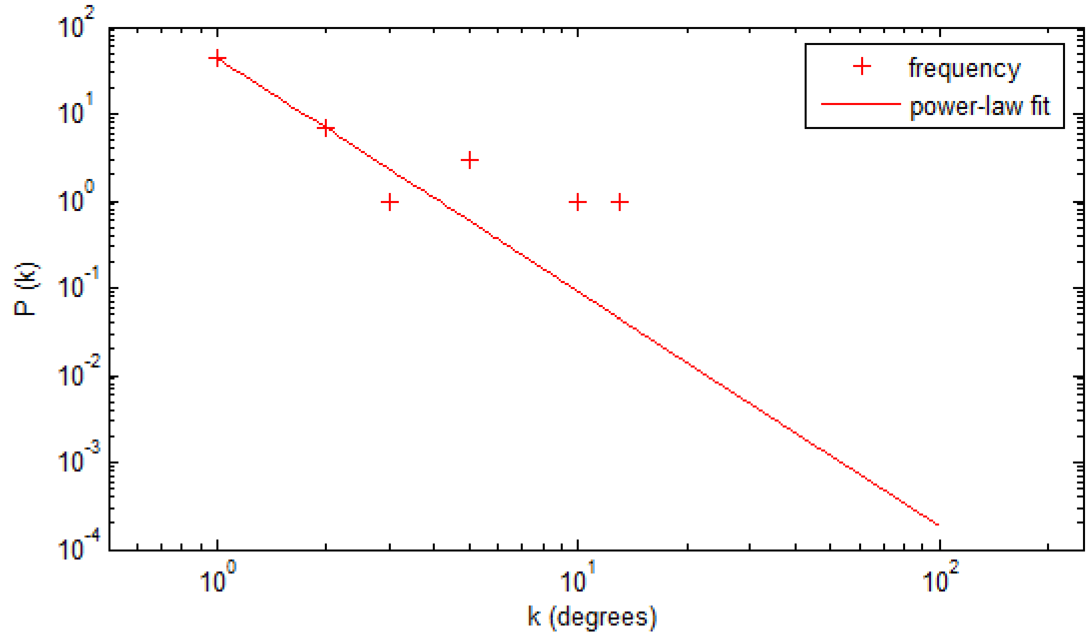


Figure 4.11. Degree distribution-k of MST.

The network having power-law vertex degree distribution is a <u>scale-free network</u>. In such a network, most of the nodes have small-degree and a few nodes have high-degree, called the latter as *"hub"* nodes. Especially, in the stock correlation network, stock *i* having a high-degree means that it is correlated with many other stocks in the sense of price fluctuation. So, it can be analyzed the most of the stocks that can most accurately reflect market behaviors by the degree of stocks.

The market seems to be self-organized in a coordination invariant structure, similarly to self-organized critical structures. The higher degree of a stock is the more stocks it is correlated with in the sense of price fluctuation. From the point of price fluctuation influence, the stocks that have high-degree generally have a *"higher status"* and *"stronger market influence power"*. Therefore, the scale-free property of the stock correlation network demonstrates that most of the stocks are at the same level, at the same time, a small quantity of stocks have a "higher status" and "stronger market influence power". The latter play a very important role in the price fluctuation correlations of the whole network. Besides, this exponent implies that the *second moment* of the distribution would diverge in the infinite market limit, or in other words, the second moment of the distribution is always dominated by the rare but extremely high-connected vertices. The degree distribution for the MST of the real data was used to construct in this thesis showed a power-law behavior with exponent 2.68 ± 0.1 with *% 95* confidence bounds. A power-law result with a similar exponent value was observed in previous studies, also, such as, Vandewalle et al., Bonanno et al., Onnela et al., Garas and Argyrakis, Huang et al., Tse et al. and L.S. Junior [4,5,10,11,15,51,55].

When power-law fitting procedure was applied on other periods MSTs, in crisis period, it was seen that scale-free behavior of networks was affected with proofs that power-law exponent value is 2.11 ± 0.1. A similar situation was observed in Onnela et al.'s research that investigated Black Monday (1987) period [55].

4.4.1.2. Probing MST

Now, *sectors* and *sub-sectors* can be investigated so that specific information of nodes and edges of MST were known. First, MST was colored according to their general sectors in Figure 4.12 (Larger view in Figure Appendix A.3). As can be seen in graph, most of nodes (stocks) of MST were formed by industrial stocks and spread to all clusters. Besides, stocks of service sectors were generally existed at side parts of clusters. Two technology stocks were in unessential position. However, it was clearly said that financial stocks were central nodes of network. Thus, it can be said that financial nodes give direction to MST, so, BIST–100.

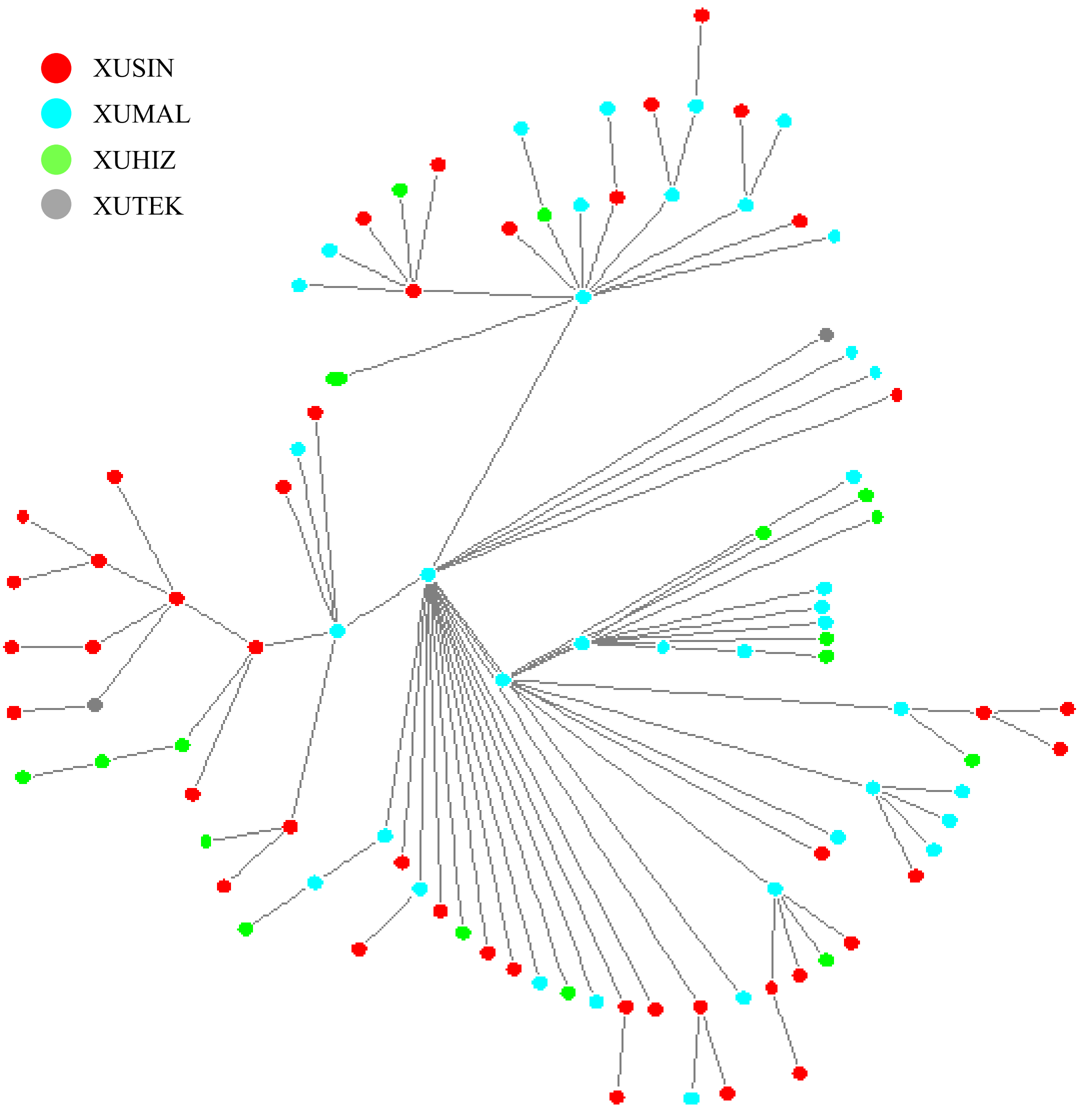


Figure 4.12. Sector view of BIST–100 from MST.

As can be seen in graph, generally, stocks in same sector were in the same clusters or connected to each other in direct relationship. This behavior pointed that economical sectors of stocks are important factor to group stocks. Thereafter, these grouping of stocks would be investigated, as four big clusters to have more detail information and to study their states in MST.

After investigating main indices and sectors of stocks, now, *sub-sectors* of stocks could be investigated. The MST includes sub-sector information of stocks can be seen in Figure 4.13 (Larger view in Figure Appendix A.4).

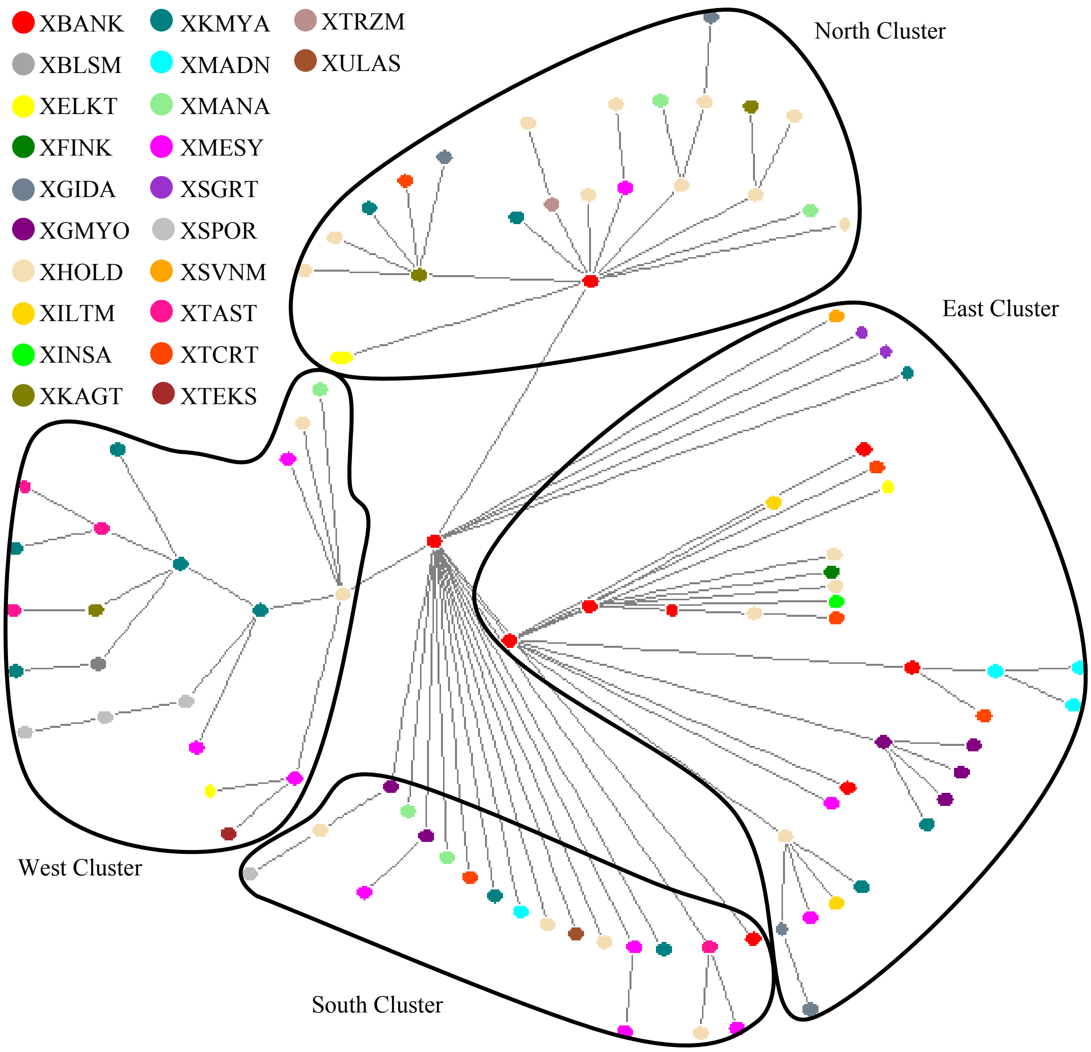


Figure 4.13. Sub-sectoral view and clusters of BIST–100 in MST.

<u>Central Cluster</u>: Central cluster of MST was formed of financial stocks. It has three high-degree (hub) nodes *ISCTR, ASYAB* and *YKBNK*. It is clearly seen in graph that node *51 (ISCTR)* is root of MST and reference stock for other clusters. It is the central node of *XBANK (Banks)* cluster, too. Besides, Node *99 (YKBNK)* is in a considerable condition according to its degree and it has strong relationships with other Bank nodes. Banks cluster consists of *nine* nodes and it is a complete cluster. When weights of edges that connect nodes were investigated, it was seen that distances are short, means: Banks are in strong contact.

<u>North Cluster</u>: This cluster holds stocks of three main sectors, financial, service and industrial. Nodes in these cluster spread via second high-degree node *12 (ASYAB)*.

When studied this cluster deeper, it was seen that most of stocks are in *XHOLD (Holdings)* sub-sector. Otherwise, some unexpected situations were sighted in this cluster. Some company stocks connected to their main company, such as, *70 (NTHOL)* and *71 (NTTUR)*. These company stocks are sub-assets of *NET Holding*. Besides, they have direct connection to *ASYAB* and this means these stocks moved in a strong relationship. As well as, *NET Holding*, *48 (IHEVA)* and *49 (IHLAS)* are sub-assets of *IHLAS Holding* and connected to *ASYAB*. Similar condition was observed for *DOGAN Holding*, also. *26 (DOHOL)* and *47 (HURGZ)* were connected to *27 (DYHOL)*, which connected to *12 (ASYAB)*, that was *"hub"* of North Cluster. This case showed that these stocks ignored their sub-sectoral act and set up a relation according to their corporative state.

<u>West Cluster</u>: This cluster had stocks of all sectors: financial, industrial, service and technology. On the other term, this cluster can be named as <u>Industrial Cluster</u> because of that most stocks in cluster are industrial stocks. Node *28 (ECILC)* is reference stock for this cluster. As well as <u>North Cluster</u>, *28 (ECILC)* and *29 (ECZYT)*, which are stocks of *ECZACIBASI Holding*, had direct connection. In addition, most of *XKMYA (Chemical, Petroleum and Plastic)* stocks and *XTAST (Non-metal Mineral Products)* stocks were in this cluster. Except *92 (TSPOR)*, all *XSPOR (Sports)* stocks existed in this cluster and connected directly. This means *18 (BJKAS)*, *35 (FENER)* and *44 (GSRAY)* moved together and were affected from each other, briefly.

<u>East Cluster</u>: As well as <u>West Cluster</u>, this cluster had all sectors, either. Some of nodes in this cluster connected to root node of MST. Interesting thing is *8 (ANHYT)* and *9 (ANSGR)*, which are in *XSGRT (Insurance)* sub-sector, had direct connection to root node *51 (ISCTR)*, which is main holding asset of them. The other part of cluster spread via another financial sector node *99 (YKBNK)* and this means East Cluster was directed by financial nodes like <u>West Cluster</u> and <u>North Cluster</u>. Another attractive observation in this cluster is attaching of *SABANCI Holding*'s sub-assets, *3 (AKBNK)* and *77 (SAHOL)*. Only two beverage company in *XGIDA (Food, Beverage)* in BIST–100, *1 (AEFES)* and *22 (CCOLA)* existed in this cluster and had direct connection. Besides, most of *XGMYO (Real Estate Investment Trusts)* stocks,

53 (ISGYO), *76 (SAFGY)*, *81 (SNGYO)*, *89 (TRGYO)* were in this cluster and formed a sub-cluster, but *13 (AYGAZ)*. Furthermore, three of four *XMADN (Mining)* stocks existed in here and formed a complete cluster. Interestingly, all stocks in this sub-cluster, *50 (IPEKE)*, *62 (KOZAA)* and *63 (KOZAL)* are sub-assets of *KOZA-IPEK Holding*. Moreover, *10 (ARCLK)*, *13 (AYGAZ)*, *59 (KCHOL)*, *72 (OTKAR)*, *95 (TUPRS)* and *99 (YKBNK)* are sub-assets of *KOC Holding* and existed in this cluster. These observations showed that, corporative condition was important as well as sub-sectoral classification.

<u>South Cluster</u>: Nodes in this cluster had direct relation and connection with the root node of MST. This cluster had all type of sectors excluding *XUTEK (Technology)*, too. Remarkable facts in this cluster are that some *XMESY (Metal Products, Machinery)* stocks, two auto-manufacturing stocks, which are sub-assets of *KOC Holding*, *36 (FROTO)* and *87 (TOASO)* and two glass-product companies *79 (SISE)* and *90 (TRKCM)* connected by one same edge. Although, *SISE* was in *XHOLD (Holdings)* sub-sector and *TRKCM* was *XTAST (Non-metal Mineral Products)*, they ignored their sub-sectoral behavior and product manufacturing defined states of them in cluster.

Therefore, in MST, all clusters spread via financial nodes and generally, their sub-sectors decided their grouping. However, it was clearly seen that, in some conditions stocks ignored their sectoral and/or sub-sectoral action and moved according to their main holdings. Product manufacturing affected their situation in clusters, too. It can be said that only sectoral conditions do not form clusters, the other conditions are important for stocks for grouping them. Finally, to investigate BIST–100 according to *average daily trading volume* and *liquidity*, stocks were grouped by their indices. It is known that *XU030*, *XU050* and *XU100* are defined by these quantities and it is a very effective way for grouping stocks [48]. View of BIST–100 from indices is seen in Figure 4.14 (Larger view in Figure Appendix A.5).

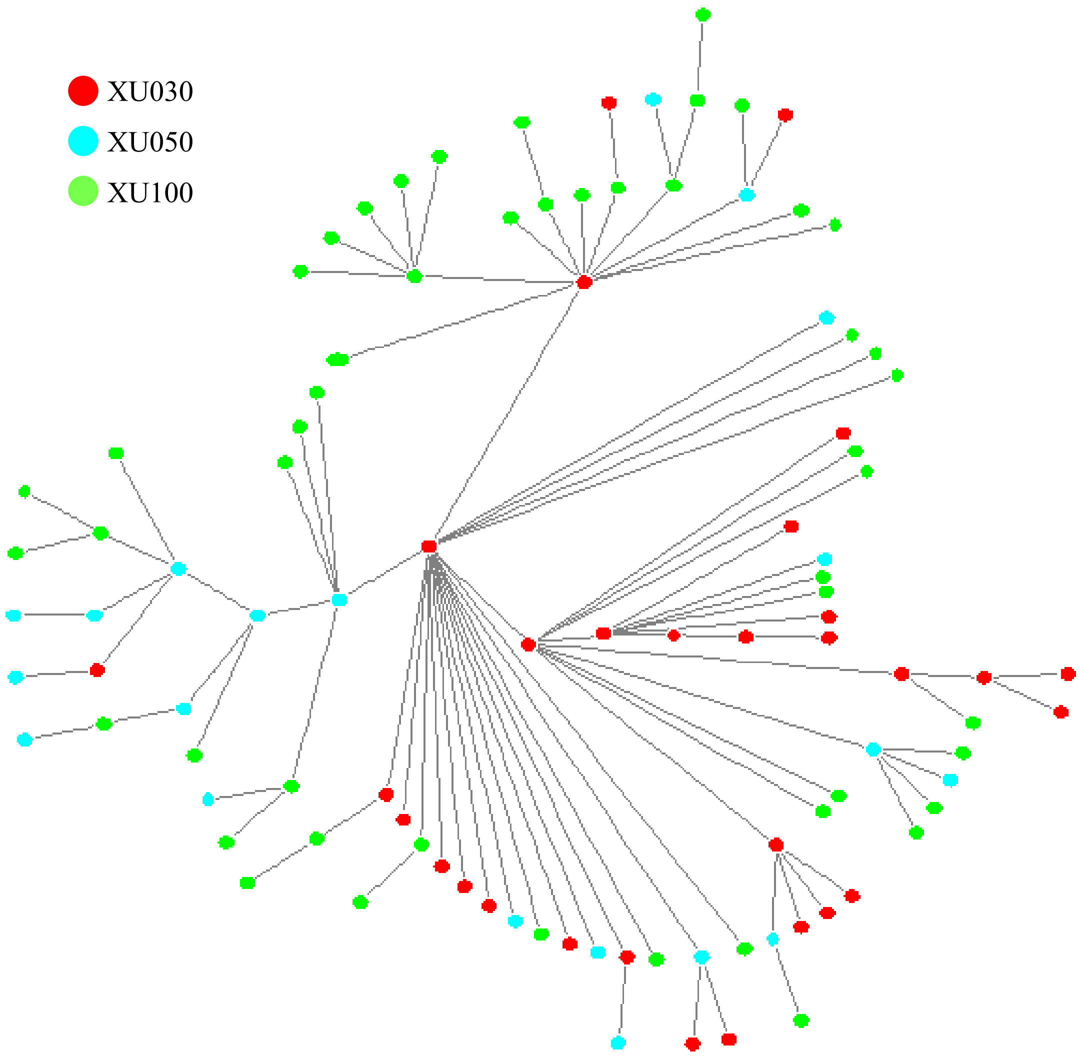


Figure 4.14. Classification of stocks by indices.

It was seen that central nodes consisted of *XU030* index stocks in MST; these nodes are liquid and have higher trading volumes than others have. When it was thought that central nodes were financial stocks, especially Banks, it can be said that these stocks are in very important place at BIST–100. As expected, the other *"hub"* nodes, which spread nodes to clusters, in other words, connects to root node, were situated in *XU050* index. On the other hand, *XU100* stocks were existed in side parts of MST. As a result, it was obvious that the most liquid stocks are laid in the center of the tree and become the reference for many other stocks that are relatively more illiquid [8].

4.4.2. Constructing Hierarchical Tree

The hierarchical tree characterizing a complex system can also be used to extract a factor model with independent factors acting on different elements in a nested way. Hereafter, it was discussed that a simple example illustrating two filtering procedures of a correlation matrix performed with methods of hierarchical clustering. By using the correlation between elements as the similarity measure and by applying a given hierarchical clustering procedure, hierarchical tree is obtained. The information present in the hierarchical tree is completely equivalent to the information stored in the filtered matrix and, when the correlation is non-negative for each pair of elements, this matrix is positive definite.

The correlation-based clustering procedure introduced is based on the computation of the *sub-dominant ultra-metric distance* associated with a metric distance that one may obtain from the correlation coefficient. The sub-dominant ultra-metric distance can be used to obtain a hierarchical tree and a MST. The sub-dominant ultrametric distance between i and j objects, i.e. the element $d_{ij} < of\ the\ D < (\Delta t)$ matrix, is the maximum value of the metric distance $d_{k,l}$ detected by moving in single steps from i to j through the path connecting i and j in the MST. The method of constructing a MST linking a set of n objects is direct and it is known in multivariate analysis as the *nearest neighbor single linkage cluster analysis* as said before.

Mantegna et al. constructed a correlation matrix for *N=100* stocks for *T=748* days and they considered two filtered correlation matrices, obtained by applying the *ALCA* and the *SLCA* to the empirical correlation matrix respectively. After that, they compared these clustering techniques according to stability and storing information skills. As can be seen in result of this comparison in Figure 4.15, *SLCA* more stable, though, correlation matrix filtered by *average linkage cluster analysis* has more information. Because of MST is a presentation of *SLCA*, in forward of this part, HT (Hierarchical Tree) of BIST–100 was created by using *ALCA*.

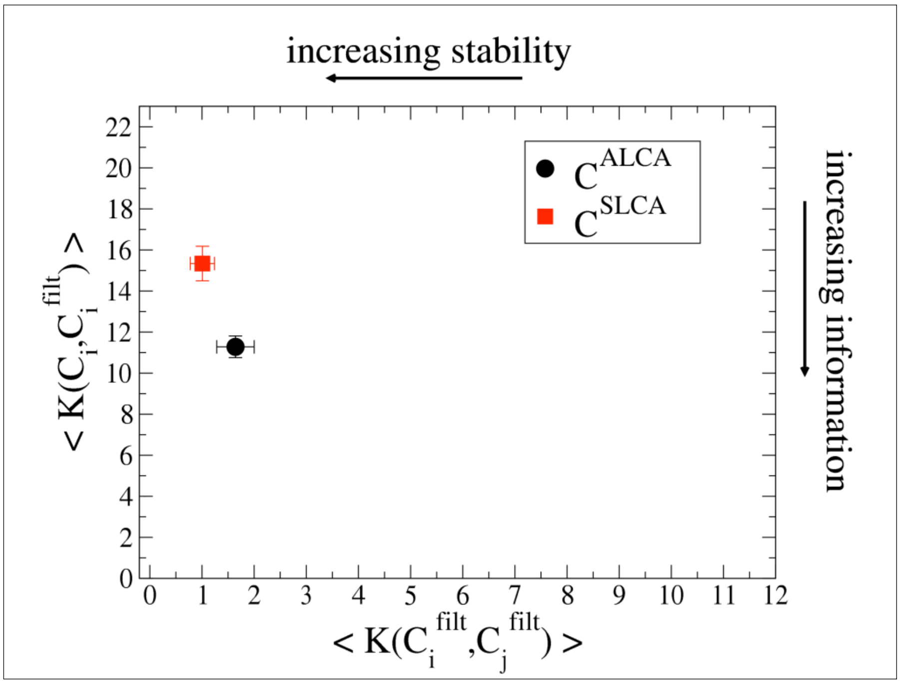


Figure 4.15. Comparison of filtered correlation matrices.

The starting point of both the procedures is the empirical correlation matrix C. The following procedure performs the $ALCA$ giving as an output a hierarchical tree and a filtered correlation matrix $C^{<}_{ALCA}$:

- *(i)* Set $B = C$.
- *(ii)* Select the maximum correlation b_{hk} in the correlation matrix B. Note that h and k can be simple elements *(i.e. clusters of one element each)* or clusters *(sets of elements)*. $\forall i \in h$ and $\forall j \in k$ one sets the elements $\rho^{<}_{ij}$ of the matrix $C^{<}_{ALCA}$ as $\rho^{<}_{ij} = \rho^{<}_{ji} = b_{hk}$.
- *(iii)* Merge cluster h and cluster k into a single cluster, say q. The merging operation identifies a node in the rooted tree connecting clusters h and k at the correlation b_{hk}.
- *(iv)* Redefine the matrix B:

$$\begin{cases} b_{qj} = \frac{n_h b_{hj} + n_k b_{kj}}{n_h + n_k} \, if \ j \notin h \, and \, j \notin k \\ b_{ij} = b_{ij} \qquad otherwise, \end{cases}$$

  where n_h and n_k are the number of elements belonging respectively to the cluster hand to the cluster k before the merging operation. Note that if the dimension of B is $m\,x\,m$ then the dimension of the redefined B is $(m-1)\,x\,(m-1)$ because of the merging of clusters h and k into the cluster q.

- *(v)* If the dimension of B is larger than 1 then go to step *(ii)*, else stop. By replacing point *(iv)* of the above algorithm with the following item
- Redefine the matrix B:

$$\begin{cases} b_{qj} = Max\left[b_{hj}, b_{kj}\right] if\ j\ \notin h\ and\ j\ \notin k \\ \qquad b_{ij} = b_{ij} \qquad\qquad otherwise, \end{cases}$$

one obtains an algorithm performing the *SLCA* and the associated filtered correlation matrix $C^{<}_{SLCA}$ [3,6,10,12–14,16,25].

4.4.2.1. Probing HT

Constructed HT by following ALCA procedure seen in Figure 4.16 (Larger view in Figure Appendix B.1) was investigated and edges were colored if least two stocks, which were in same sector, existed next to each other. *XBANK (red)* stocks formed a cluster and they had shortest distance among them. This means Bank nodes generally moved together; also, their cluster was the biggest cluster that has most of one kind stock. *XHOLD (cyan)* stocks spread over HT as well as MST and two *XSGRT (teal)* stocks, *8 (ANHYT)* and *9 (ANSGR)* formed a cluster. Unlike MST, two *XKMYA (olive)* stocks, which were in <u>South Cluster</u> in MST, formed a section. *73 (PETKM)* and *88 (TRCAS)*, which are sub-assets of *PETKIM-PETROKIMYA Holding*, specified a connection that was not noticed in MST. ALCA HT helped to see them. Also, unlike MST, *32 (EKGYO)*, *53 (ISGYO)* and *81 (SNGYO)* generated a *XGMYO (maroon)* cluster. Besides, *XTAST (purple)* stocks; *2 (AFYON)* and *40 (GOLTS)*, another *XKMYA (olive)* and manure stocks; *14 (BAGFS)* and *31 (EGGUB)*, *XGIDA (magenta)* stocks; *1 (AEFES)* and *22 (CCOLA)* generated clusters. Finally, *XHOLD* formed cluster whose stocks spread in HT were *39 (GOLDS)* and *43 (GSDHO)*; *59 (KCHOL)* and *77 (SAHOL)*; *26 (DOHOL)* and *27 (DYHOL)*; *54 (ISYHO)* and *65 (METRO)*; *49 (IHLAS)* and *70 (NTHOL)*.

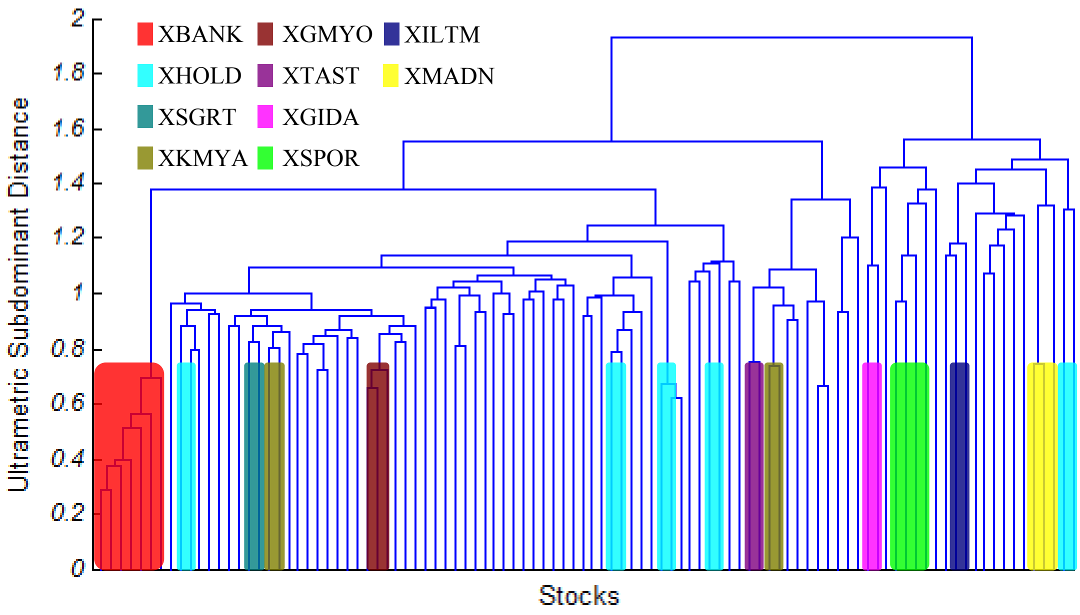


Figure 4.16. Hierarchical tree of BIST–100 stocks.

In contrast MST, two *XILTM (navy) – Telecommunications* – stocks *83 (TCELL)* and *93 (TTKOM)* created a cluster. These stocks were in different clusters in MST. Moreover, it was seen that all-four *XSPORT (lime)* stocks took place in same cluster. This means ALCA HT analysis is skillful as well as MST to investigate sub-sectors. This kind of analysis captured some important points that MST was not. Parallel to MST, ALCA HT pointed three *XMADN* (yellow) stocks, which are sub-assets of *KOZA-IPEK Holding*, *50 (IPEKE)*, *62 (KOZAA)* and *63 (KOZAL)*. To give more example for this situation, *8 (ANHYT)* and *9 (ANSGR)* that are sub-assets of *51 (ISCTR)* were next to each other. Besides, *DOGAN Holding* sub-assets *26 (DOHOL)*, *27 (DYHOL)* and *47 (HURGZ)* formed a cluster. It can be said that HT also inspects sub-asset behavior in BIST–100. However, MST has more ability to investigate cooperative states.

When HT seen in Figure 4.17 (Larger view in Figure Appendix B.2) was studied according to sectoral states, it was seen that stocks were moving with their sectoral information. In addition, it was seen that *XUMAL* stocks had the shortest distances among them. This showed that financial stocks had high-relation in own sector. Besides, *XUMAL* stocks direct the BIST–100 index due to these stocks had the shortest distances in each cluster formed in HT. Moreover, while *XUSIN* stocks were investigating, it was seen that generally, these stocks moved together and they had second shortest distances after financial ones. Although, *XUHIZ* stocks moved together, it was seen that they had longest distances between them and the other sector stocks in clusters. On the other hand, *XUTEK* stocks act insignificant in HT.

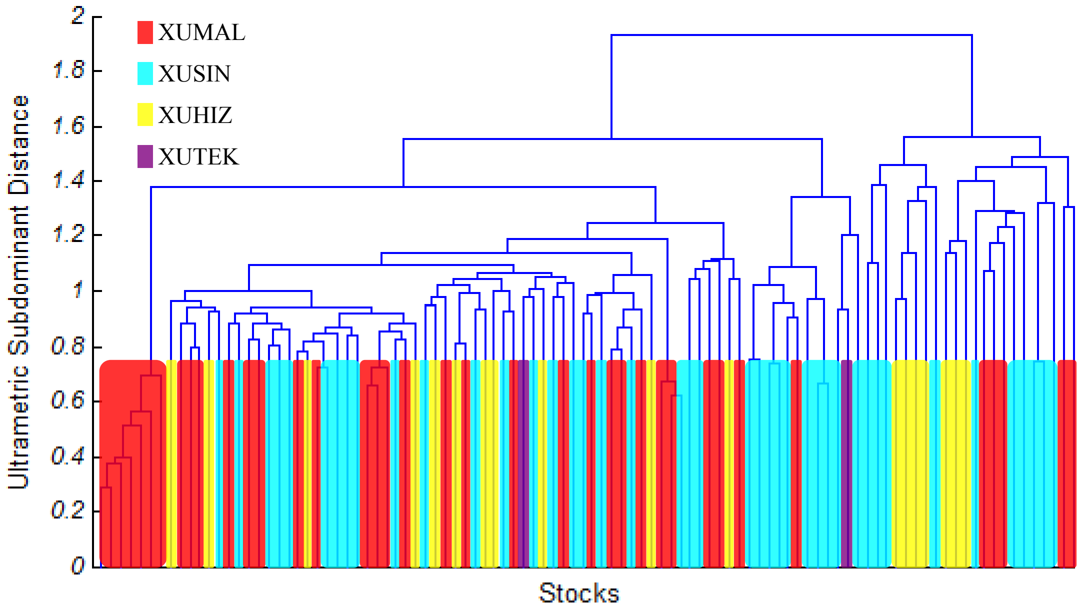


Figure 4.17. Sectoral view of BIST–100 stocks from HT.

In summary, financial stocks had shortest *ultra-metric sub-dominant distances*. This showed that this type of stocks were in affiliation and gave direction to BIST–100. In addition, industrial stocks had short distances and strong relation. Despite, service stocks moved together, they had longest distances compare to the other sectors. So, this showed that service stocks had strong relationships among them, but, weak with the other sector stocks.

In Figure 4.18 (Larger view in Figure Appendix B.3), to probe liquidness of stocks, ALCA HT was divided into four clusters according to branching of HT to investigate stocks in detail by taking advantage of their indices. It was seen that same indices had strong relations but some exceptions. As first cluster were probed, it was clearly seen that all stocks in this cluster consisted of *XU030 (most liquid)* assets and they were in central point of BIST–100 because of having shortest distances. When second cluster were investigated, it was seen that liquid stocks split from illiquid ones and moved together. Third cluster of HT was formed of mostly illiquid stocks moved related. This proved that same kind of stocks had strong connection according to index or liquidness as well as sectors. Thus, fourth cluster kept stocks that had longest distances. Although, liquid and illiquid stocks existed in this cluster, they were separated and were grouped according to their indices, so, liquidness.

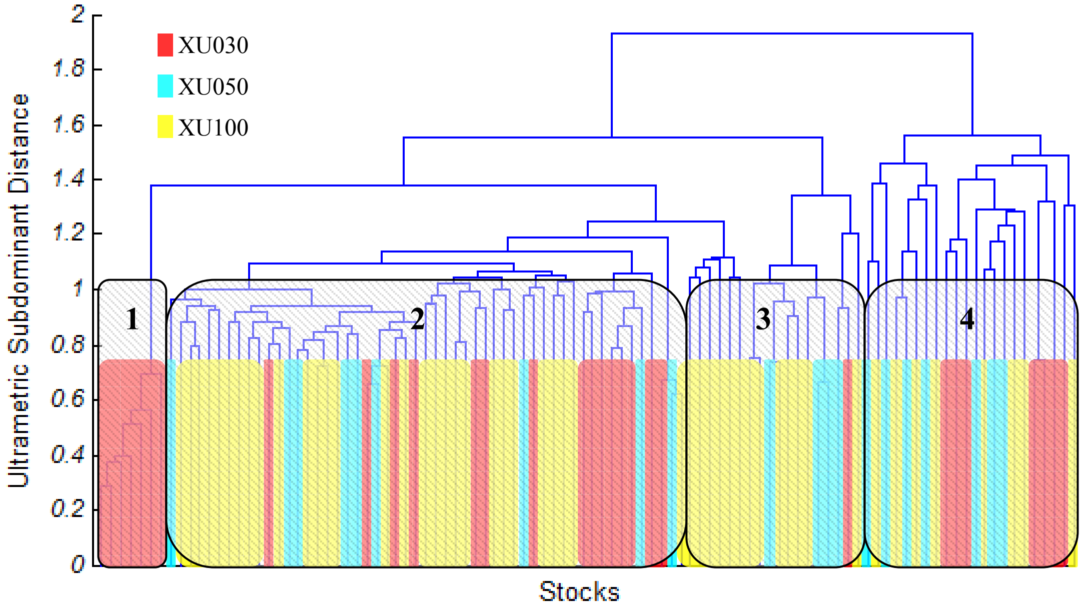


Figure 4.18. Index view of BIST–100 from HT.

4.4.3. Portfolio Optimization

After completing MST and HT analysis, it was indented to notice effectiveness of *"Stock Correlation Network"* for risk management of market by paralleling with *"Modern Portfolio Theory"*; also, with taking advantage of this kind of comparing, it would be provided to see accuracy of MST and HT for using as a investment guide or not.

4.4.3.1. Modern Portfolio Theory

Modern Portfolio Theory (MPT), a hypothesis put forth by Harry Markowitz, is a Nobel Prized (1990) investment theory based on the idea that risk-averse investors can construct portfolios to optimize or maximize expected return based on a given level of market risk, emphasizing that risk is an inherent part of higher reward. It is one of the most important, influential economic theories dealing with finance, and investment suggests that it is possible to construct an *"efficient frontier"* of optimal portfolios, offering *the maximum possible expected return* for a given level of risk. It suggests that it is not enough to look at the expected risk and return of one particular stock. By investing in more than one stock, an investor can reap the benefits of diversification, particularly a reduction in the riskiness of the portfolio [56]. Figure 4.19 is *mean-variance-efficient frontier graphs* for six-month periods of BIST–100.

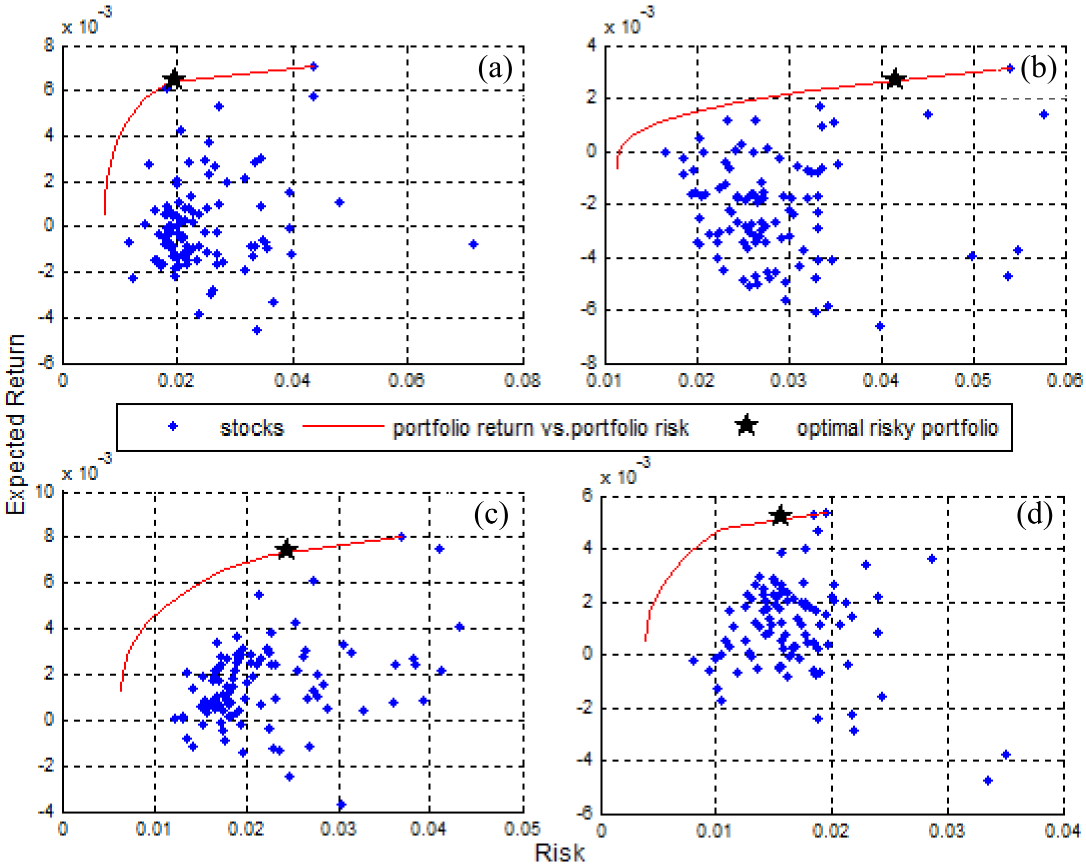


Figure 4.19. Efficient frontiers of six-month periods of BIST–100; a) January–June 2011, b) July–December 2011, c) January–June 2012 and d) July–December 2012.

As the figure investigated, it was clearly seen that *mean-variance-efficient frontier* graphs adapted with *statistical moments* graph seen in Figure 4.7 and *normalized tree length* graph seen in Figure 4.8. While these graphs were investigating, it was observed that second and third period of data scale had been affected negatively because of crisis. Likewise, MPT showed that risk of these periods were greater than the other periods. Hence, MST concept is a *good-guide for risk management*.

4.4.3.2. Probing MPT

Following, MPT was investigated by plotting *mean-variance-efficient frontier* graph for all data scale. Plotted graph showed that BIST–100 stocks existed mostly in same risk rate area, except some stocks. Well, which stocks moved out of common risk area and which stocks were more riskless. Yet more important, these acts and states can be guessed from MST or/and HT? Figure 4.20 was helped to answer these questions.

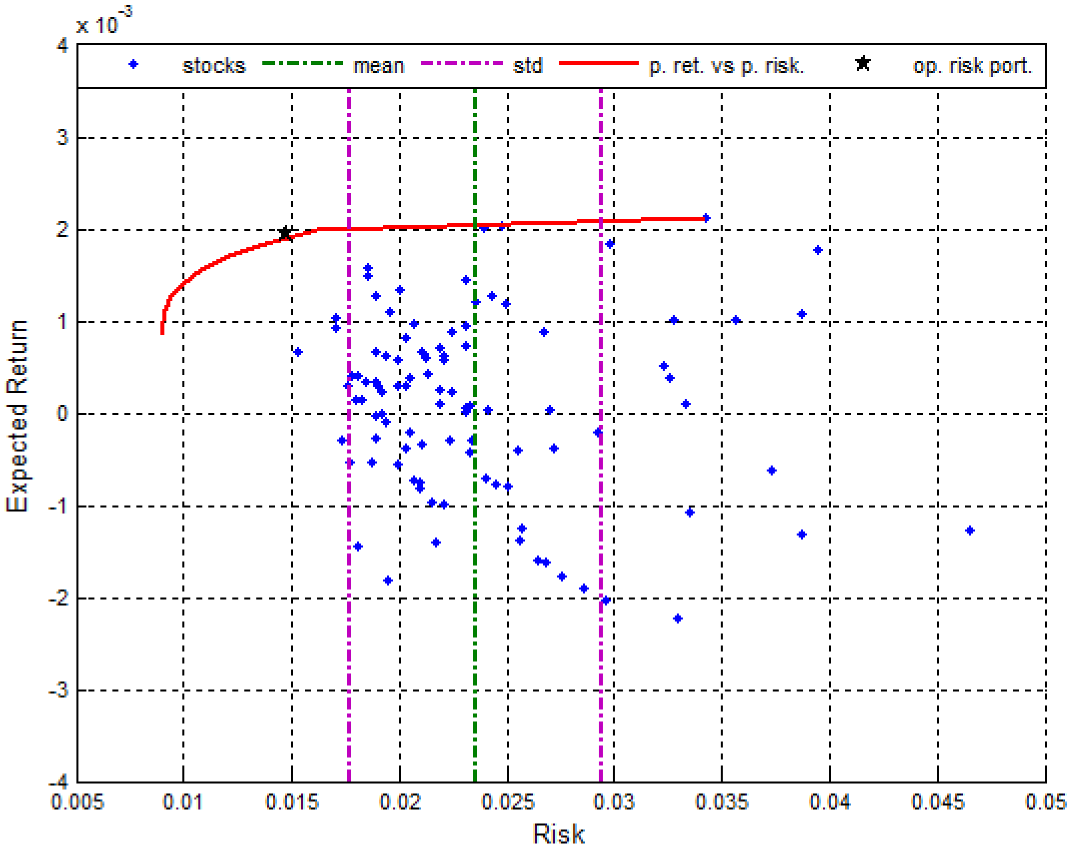


Figure 4.20. Efficient frontier graph of BIST–100.

Firstly, to examine risks of stocks in detail, stocks, which are out of range, must be determined. Standard deviation ($\sigma = 0.0059$) and mean ($\mu = 0.0235$) of risks were computed to observe extraordinary stocks and their volatilities, which were spotted by using plot, were given in Table 4.5 (Full-list can be seen in Table Appendix C.1).

In first place, when stocks, which are more riskless, were investigated, it was seen that these stocks were mostly financial stocks and took place closely to *"hub"* stocks. More generally, *ALARK*, *NTHOL*, *BIMAS*, *ECILC* and *ISGYO* had direct connection with bank cluster. Besides, it was seen that *ECILC* had direct connection hub stock *ISCTR*.

Table 4.5. Out-range stocks.

Stock	Sector	Index	Risk	Volatility
ALARK	XUMAL	XU100	0.0153	24.20%
NTHOL	XUMAL	XU100	0.0170	26.96%
BIMAS	XUHIZ	XU030	0.0171	27.10%
ECILC	XUMAL	XU050	0.0174	27.53%
ISGYO	XUMAL	XU100	0.0176	27.85%
AFYON	XUSIN	XU100	0.0296	46.90%
IPEKE	XUSIN	XU030	0.0298	47.24%
SAFGY	XUMAL	XU050	0.0323	51.23%
KONYA	XUSIN	XU050	0.0326	51.69%
GOZDE	XUMAL	XU100	0.0328	51.96%
GSRAY	XUHIZ	XU050	0.0329	52.20%
KARTN	XUSIN	XU050	0.0333	52.84%
BJKAS	XUHIZ	XU100	0.0335	53.11%
GOODY	XUSIN	XU100	0.0342	54.25%
BRISA	XUSIN	XU050	0.0357	56.54%
FENER	XUHIZ	XU050	0.0373	59.14%
TSPOR	XUHIZ	XU100	0.0387	61.33%
NETAS	XUTEK	XU030	0.0387	61.42%
EGEEN	XUSIN	XU100	0.0394	62.49%
CEMAS	XUSIN	XU100	0.0465	73.77%

When more risky stocks were investigated, it was seen that these stocks existed far away from hub stocks, especially *ISCTR*. On the other hand, interestingly, most of these stocks existed in West Cluster, which consisted of mostly *XKMYA* and *XTAST*, so, *XUSIN* stocks. In addition, these stocks had behaved meaningless according to their sectoral and/or sub-sectoral states and had more variety, as can be seen previous parts. Furthermore, it was seen that all sport stocks, which were *XUHIZ* stocks in this cluster, were riskier. Although, *TSPOR* was in South Cluster, it was risky, too. It was only sport stock in South Cluster and it behaved differently from the other stocks in own sub-sector. On the other hand, *CEMAS* and *GOZDE*, which were in North Cluster, had more risk rates than others have. When these stocks were probed, it was seen that these stocks had close relation with hub stocks. However, while ALCA was investigated, it was seen that *ultra-metric sub-dominant distances* of these stocks are much and close to *XSPOR* stocks, which were more risky.

Last two risky stocks stated in East Cluster. One of them, *SAFGY* had longer edge-weight than the other *XGMYO* stocks in own sub-cluster. Otherwise, it was observed that it was far away from the other sub-cluster stocks in ALCA and connected with *GOZDE*, which was another risky stock. The other risky stock, *IPEKE*, was situated in a *XMADN* sub-cluster, which connected directly to hubs and all nodes consisted of sub-assets of same holding. However, this stock had long sub-dominant distance in ALCA. Farther, all stocks in this cluster formed a cluster, which had long sub-dominant distances in ALCA, too. To prove effectiveness of ALCA for risk management, it was taken look at risk rates of the other stocks in this cluster and seen that its risk rates were more than mean value.

With this last part, it was showed that MST and HT could be used as a *portfolio risk management tool*, briefly. For MST concept, it can be said that stocks, which move away from central cluster or hub stocks, are riskier than the others are. Moreover, stocks, which form insignificant clusters according to their sectoral and sub-sectoral information, are more risky than the other stocks that act in respect of these cases. Anyhow, some stocks were too risky, they were pretty close to hubs and existed center of MST, though. At this point HT concept helped to separate these stocks. It was seen that, risky stocks had long ultrametric sub-dominant distances and took place of right side of ALCA HT. Besides, some stocks had risk in despite of taking right places according to both *closeness* and *sub-sectoral* information. At this point, HT again was used to observe situation of clusters in ALCA to classify risks of these stocks and the others in this sub-cluster. Thus, it was seen that risky stocks and the other stocks in its cluster risky, too. Therefore, it was proven that MSTs and HTs are well enough for studying portfolio optimization.

4.5. BIST–100 CASE STUDY

In previous section, it was seen that stock correlation network technique could be used to analyze stock movements in market, briefly. *Clustering of stocks*, *portfolio optimization* and *risk management* can be investigated by using financial time series in specific data scale, only. In this section of thesis, ten years of data *(2003–2013)* of BIST–100 Index, local and global factors were analyzed to investigate effects of other economical factors and crisis periods into stock market and possibility of prediction of these movements by *stock correlation network technique*.

4.5.1. Crisis: Local and Global Economical Factors

Movements in the stock market can be quite changeable and sometimes movements in share prices can be influenced from economic factors. There are certain factors, which affect share prices and the stock markets in general. These factors are:

- Economic growth: Higher economic growth will help firms to be more profitable, because, there will be more demand for goods and services. This will help to boost company dividends and therefore share prices.
- Interest rates: Lower interest rates can make shares more attractive. Lower rates help to boost economic growth making firms more profitable. Moreover, lower rates make shares more attractive than saving money in a bank or holding bonds.
- Stability, Confidence and Expectations: Stock markets dislike shocks threaten economic stability and future growth. Thus, they will tend to undermine commodities prices and exchange rates. They will also dislike political instability, which makes it difficult to follow up strong economic policies. This situation may lead to change some important economical parameters, namely, stock market.
- Related markets: Often investors rather than investing in stock market, they could buy government bonds or commodities. If investors feel bonds are overpriced and likely to fall, then the stock market can benefit as people move into shares. Also, the other world markets have influence on investing market.

For example, if a strong stock market rally happens in the US, with the Dow and the NASDAQ registering impressive gains, we are likely to see a large influx of foreign money into the US, as international investors rush in to join the party. This influx of money would be very positive for the US Dollar, because in order to participate in the equity market rally, foreign investors would have to sell their own domestic currency and purchase US Dollars. The opposite also holds true; if the stock market in the US is doing poorly, foreign investors will most likely rush to sell their US equity holdings and then reconvert the US Dollars into their domestic currency that would have a substantially negative impact. This logic can be applied to all the other currencies and equity markets around the world [57,58].

In next section, some economical parameters were investigated to observe economical situation of Turkish Economy. These are:

- Gross Domestic Product (GDP): The monetary value of all the finished goods and services produced within a country's borders in a specific time-period, though GDP is usually calculated on an annual basis. GDP is commonly used as an indicator of the economic health of a country, as well as to gauge a country's standard of living. The GDP numbers are reported in two forms: *current currency* and *constant currency*.
  - Current currency GDP is calculated by using today's currency and makes comparisons between time-periods difficult because of the effects of inflation.
  - Constant currency GDP solves this problem by converting the current information into some standard era currency, such as 1997 Dollars. This process factors out the effects of inflation and allows easy comparisons between periods [59,60].
- Growth Rates of Gross Domestic Product: A measure of economic growth from one period to another in percentage terms. In practice, it is a measure of the rate of change that a nation's gross domestic product goes through from one year to another. The economic growth rate provides insight into the general direction and magnitude of growth for the overall economy. Growth rates are

reported in two forms as well as GDP numbers: *current currency* and *constant currency* [61].

- External Debt Stock: The portion of a country's debt that was borrowed from foreign lenders including commercial banks, governments or international financial institutions. A debt crisis can occur if a country with a weak economy is not able to repay external debt due to the inability to produce and sell goods and make a profitable return. The International Monetary Fund (IMF) is one of the agencies that keep track of the country's external debt. This debt can be defined in two forms: *gross debt stocks* and *net debt stock* [62].
- Current Account: The difference between a nation's savings and its investment. The current account is an important indicator about an economy's health. It is defined as the sum of the balance of trade (goods and services exports less imports), net income from abroad and net current transfers. A positive current account balance indicates that the nation is a net lender to the rest of the world, while a negative current account balance indicates that it is a net borrower from the rest of the world. The current account and the capital account are the two main components of a nation's balance of payments [63].
- Balance of Trade (BOT): The difference between a country's imports and its exports. Balance of trade is the largest component of a country's balance of payments. Debit items include imports, foreign aid, domestic spending abroad and domestic investments abroad. Credit items include exports, foreign spending in the domestic economy and foreign investments in the domestic economy. A country has a trade deficit if it imports more than it exports; the opposite scenario is a trade surplus. Also referred to as *"trade balance"* or *"international trade balance"* [64].

In Appendix D, Turkish Economy was investigated and specified critical points by using plots. Data was captured from Turkish Statistical Institute, Republic of Turkey Prime Ministry Undersecretaries of Treasury, Republic of Turkey Ministry of Economy and Central Bank of the Republic of Turkey [65–68].

When plots were investigated, it was seen that *current account* decreased significantly from *3rd quarter of 2008* to *3rd quarter of 2011*. This behavior can be summarized as reduction was observed in *2009-2012*. Although, *gross and net external debt* of Turkey increasing from *2007* to *2013*, actual raise was seen after *2009*. *Interest rates* showed raise in *2008*, however, after that, continued to drop until *2011*. After this time, rates increased for both *late liquidity* and *overnight*. *Gross domestic product in current and constant prices* increased from *2009* to *2013*, briefly. In addition, *growth rates growth rates by constant and current prices* felt in *2008–2009* and *20011–2012*. Although, *gross domestic product by purchasers in constant and current prices* increased, it was seen that *growth rates* for this parameter decreased. Finally, *export and import* values in *2007–2013* investigated and seen that in *2008* and *2009* both export and import values dropped significantly. After this time, a cliff was formed in export and import.

In next, global parameters were investigated to observe effects of them into Turkish Economy and Turkish Stock Market. These parameters are global stock markets; Australia: *S&P/ASX 200 (Standard & Poor's/Australian Stock Exchange 200 Index)*, Brazil: *Bovespa (Índice Bolsa de Valores de São Paulo)*, Canada: *S&P/TSX Composite (Standard & Poor's/Toronto Stock Exchange Composite Index)*, China: *Shanghai-Composite (Shanghai Stock Exchange Composite Index)*, Eurozone: *Eurostoxx–50 (Eurex Exchange 50 Index)*, *France: CAC–40 (NYSE Euronext Paris Cotation Assistée en Continu 40 Index)*, Germany: *DAX (Deutsche Börse Deutscher Aktien IndeX)*, Hong-Kong: *Hang-Seng (Hong Kong Stock Exchange Hang Seng Index)*, India: *BSE Sensex (Standard & Poor's Bombay Stock Exchange Sensitive Index)*, Indonesia: *Jakarta (Indonesia Stock Exchange Jakarta Composite Index)*, Israel: *TA–25 (Tel Aviv Stock Exchange 25 Index)*, *Italia: FTSE–MIB (FTSE Milano Italia Borsa Index)*, Japan: *NIKKEI–225 (Tokyo Stock Exchange NIKKEI 225 Index)*, Netherlands: *AEX (NYSE Euronext Amsterdam Exchange Index)*, Russia: *RTSI (Moscow Exchange RTS Index)*, South-Korea: *KOSPI (Korea Exchange Composite Stock Price Index)*, Spain: *IBEX–35 (Bolsas y Mercados Españoles Índice Bursatil Español 35 Index)*, Sweden: *OMX–30 (Stockholm Stock Exchange OMX 30 Index)*, Switzerland: *SMI–20 (SIX Swiss Exchange Index)*, UK: *FTSE–100 (London Stock Exchange 100 Index)*, US: *DJIA (Dow Jones Industrial Average)*, *NASDAQ–100 (National Association of Securities Dealers Automated Quotations 100 Index)*, *S&P–500 (Standard & Poor's 500 Index)*, *Russell–2000 (Russell Investments 2000 Index)*;

Commodities; *Brent-Oil* and *Gold*, lastly, currencies: *EUR (Euro)* and *USD (US Dollar)*. Data was captured from Yahoo! Finance and Google Finance [69,70].

While global parameters were investigating it was seen that *currencies* increased during *2008* as expected because of crisis effect. Currencies decreased during *2010*, after *2011*, they started to rise. When important *commodities* investigated seen that *gold* prices increased after *2008* until *2012*. Rise of gold prices means that confidence of Turkish Economy decreased in this time scale. *Brent-Oil* dropped in *2008* and in *first months of 2009*, it started to increase until *2012*. This means oil prices showed reaction to *2008 crisis* and dropped. After this time, prices started to increase and showed that there is an unstable situation in global economy. Some important stock markets were examined and seen that none of them could be oblivious to global crisis. All of them smashed in *2008* and tried to gather strength until *2011*. After this time, they were affected from a new crisis as mentioned before in previous part of thesis.

This section showed that Turkish Economy was shaken especially after *2008*. Expect gross domestic product prices, all economic parameters proved this hypothesis. Also it is known that in *2007*, a global crisis was observed and this lead to *Great Recession 2010–2012* [71,72]. All of these factors pointed that Turkish Economy was greatly influenced from global crisis. Currencies, commodities and stock markets showed negative response to global crisis and were affected, briefly. In these conditions, it could not be reasonable to hope there is not any influence in Turkish Stock Market. Now, firstly, BIST-100 Index opening values in *2003–2013* investigated and tried to determine critical dropping points in Figure 4.21.

As can be seen from polynomial of BIST–100 opening values, there is a certain drop during *2007* and *2008* because of crisis, which affect global markets as well as BIST-100. In *2011* again, a decrease started in global markets and BIST–100 as mentioned above. Second reduction was investigated in first part of thesis by using stocks located in BIST–100 with taking advantage of sectoral and sub-sectoral information, so, in this part, *2007* and *2011* crisis would be demonstrated by using global data in financial network techniques. Finally, we would have a chance to investigate relation among BIST–100 and the other factors during data scale, especially, in crisis periods. To achieve this knowledge, data scale would be divided into ten periods, afterwards, *statistical moments, MSTs, normalized tree length* and *power-law distribution* would be computed. Analysis of these quantities would help us to understand and observe movements and effects of outer agents *(global market data, commodities and currencies)* into BIST–100 during ten years.

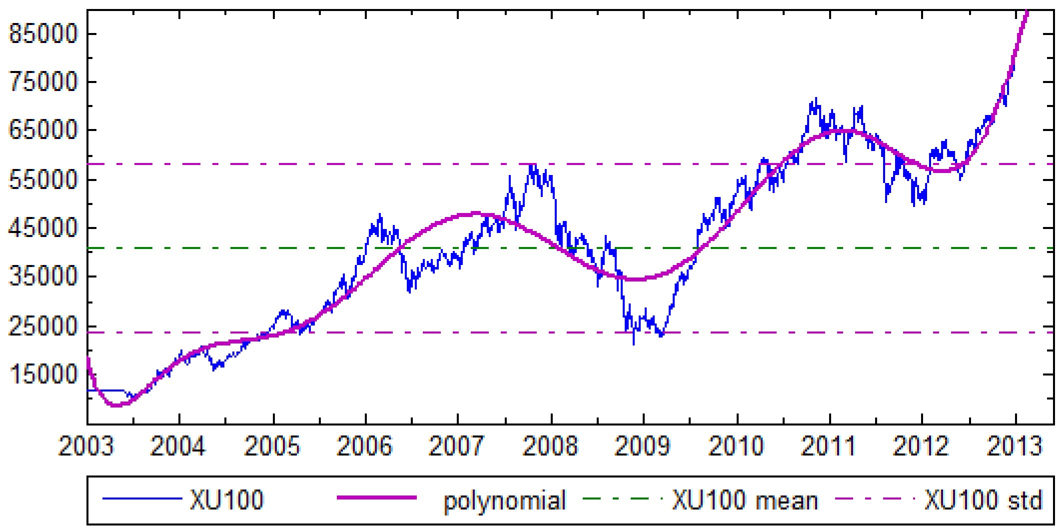


Figure 4.21. BIST-100 opening values for 2003–2013.

4.5.1.1. General View of Global Factors

First, whole data scale was used to plot a MST to analyze situation of BIST–100 among global factors, in Figure 4.22. *N=29* nodes were used to plot MST by using *L=10* years data set. Stock markets, commodities and currencies were separated into three main classes according to their regions as Europe, America and Asia to investigate factors significantly. Thus, with this kind of decomposition, relations among factors would be studied by observing which agents determine trend, major factors canalizing global economy and how BIST–100 was affected from movements of factors.

When the Figure was analyzed, it was seen that stock markets generally form cluster according to their regions as expected. Europe markets, which were colored as *cyan*, were separated into two groups. First group, which consisted of comparatively lower budget markets, were dispersed around *Eurostoxx–50*. Therefore, this index specifies situation of indices existed around it. More budget European indices formed via *S&P 500* index as wells as Asia indices which were colored by *lime*. Thus, this index control information in two clusters can be said. America markets, which were colored by *yellow*, laid generally center of tree and became reference for all. Moreover, it can be said that all clusters were directed by *US* agents due to it is known that *Eurostoxx–50* index is controlled by *NYSE*. Especially, *NASDAQ–100* was important factor among all nodes due to it stands at center of three clusters. Except *Brent-Oil*, currencies *(Euro and US Dollar)* and *Gold* moved together. However, *Brent-Oil* existed among European markets and had close relationship with them.

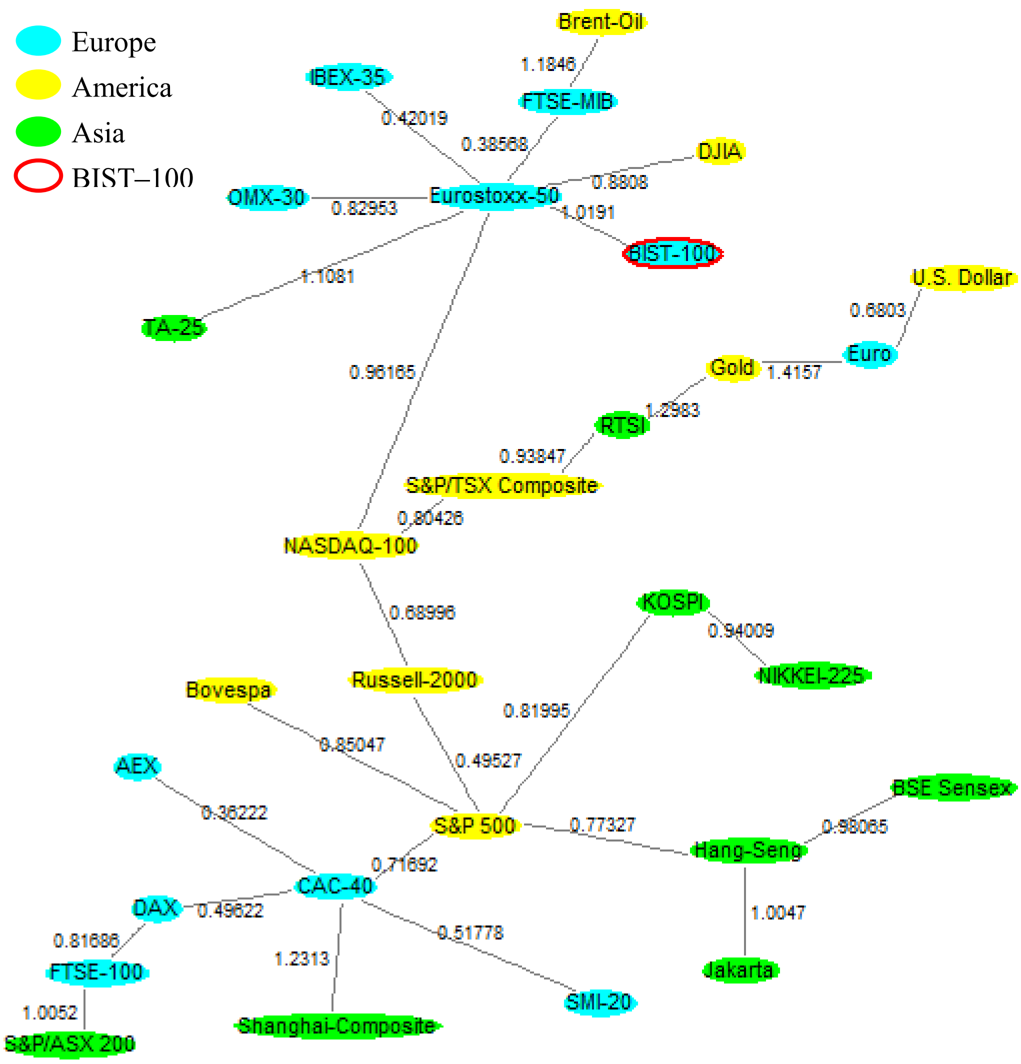


Figure 4.22. MST of BIST–100 and global factors for 2003–2013.

Besides, *BIST–100* was existed among European stocks, which was controlled by *Eurostoxx–50* index. This cluster had *Brent-Oil*, *DJIA* and *TA–25*, also. Therefore, *BIST–100* index generally interacted with European indices can be said. This situation is a summary of ten years of data. Of course, this case would change year by year. Especially, in crisis periods, effects of dominant factors would be observed clearly. In conclusion, data was divided into years and investigated inter-changing of nodes and *BIST–100*, also.

4.5.1.2. Effects of Global Factors on BIST–100

Ten years were divided and investigated in Appendix E, in detail. When, first figure that was plotted by using data in *2003*, it was seen that *"hub"* and *"central"* nodes consist of US Markets. As well as in ten years plot, markets were clustered stick by their regional info. In *2003*, BIST–100 existed in Asia markets and Brent-Oil cluster. Namely, in this period, Turkish Economy was identified by Asia and oil prices. In *2004*, it had direct relationship with NASDAQ–100 that controls European assets. BIST–100 had close relation with currencies; Euro and US Dollars, also. Thus, Turkish market was situated among one of the important stock markets in the world. In *2005*, it was split from Europe and come close to US agents, Russia and currencies, which shaped its status. In *2006*, BIST–100 continued to grow up and became one of the important stocks. Moreover, it became a link between some Europe stocks. Besides, in this year, commodities prices started to affect central nodes and currencies existed near European stocks. In *2007*, Turkish stock became a central node and bridge between *"hub"* (US assets) nodes and comparatively low budget European stocks. While starting the crisis period, BIST–100 was in effects of US Markets, generally.

In *2008*, as can be seen, crisis period affected movements of assets in MST. Regional behavior disappeared and BIST–100 moved away from center of tree and started to act with low budget European stocks. In *2009*, Turkish Stock Exchange spared from its position and continued to stay with European stocks, which was affected by US assets, commodities and currencies. After, in *2010* and *2011*, BIST–100 stayed with European area and lost its importance in world markets due to it was affected from second crisis period. Finally, in *2012*, although, other assets occurred to be in right positions according to both regional and budget states, BIST–100 moved away from central nodes and continued to be affecting from currencies and especially Brent-Oil prices.

In this section, it was seen that while BIST–100 took place in one of the important assets and *"hub"* nodes until *2008*, after this time, Turkish Stock Market and Economy started to decrease with effect of crisis and this action continued through *2012* because of *Global Recession* period. Plotted yearly MSTs and ten years MST showed that Turkish Economy generally moved with second degree assets in Europe. Investigated local factors already showed that these periods affected whole Turkish Economy, so, with taking advantage of financial network techniques, this condition was proved by MST process. These MSTs also showed that central part of global assets formed of US assets. This means that global economy was directed by US, as well as Turkish Economy. Besides, it was seen that commodities, particularly, Brent-Oil prices and currencies had important influence on Bourse Istanbul.

Next, *statistical moments* by using correlation matrices and *normalized tree length* by using distance matrices were computed to analyze condition of assets during time to investigate probability of prediction in crisis periods. Firstly, statistical moments investigated and plotted in Figure Appendix E.11. *Mean of correlation coefficients* showed that it increased generally until *2013*. Significant raises were observed in *2005*, *2007–2009* and *2010–2012* periods. As mentioned before, high mean-correlation values mean sign of crisis. Moreover, skewness of data towards zero in crisis periods, especially, in *2007–2008* period. Low skewness values are signal of crisis, too. Besides, *polynomial* of data changes as exponential, analyzing plots provided clues about before crisis period. Thus, statistical moments proved efficient of financial network techniques in turning point observation.

Afterwards, *normalized tree length* was investigated in Figure 4.23 to analyze distances among assets. As shown, *polynomial* of data decreased greatly in *2007–2008*. It means that distances decreased in this period and assets came closely to each other. Moreover, this investigation helped to gain some clues about prediction crisis period due to it started to fall in *2006* for this example.

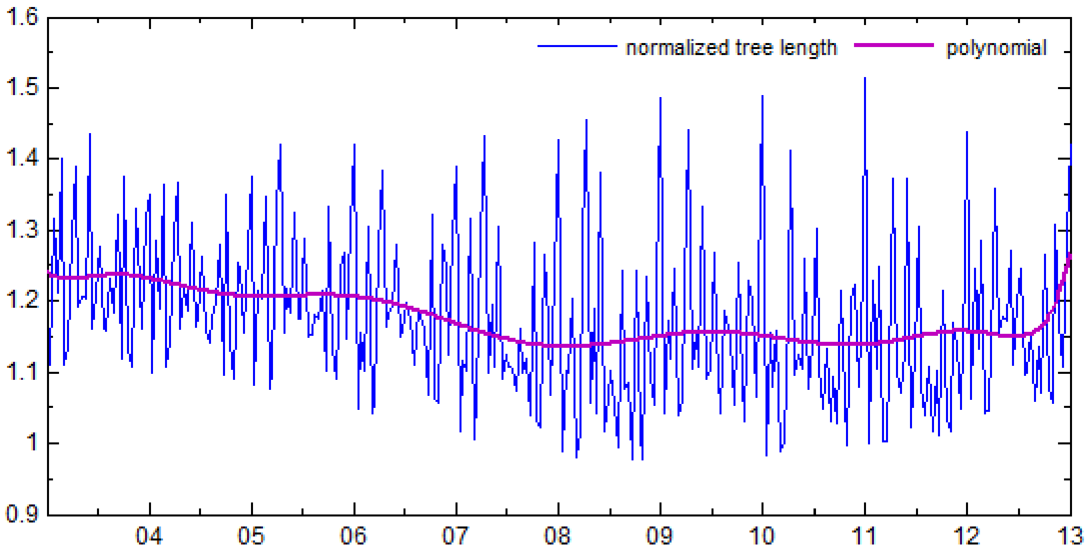


Figure 4.23. Normalized tree length for 2003–2013.

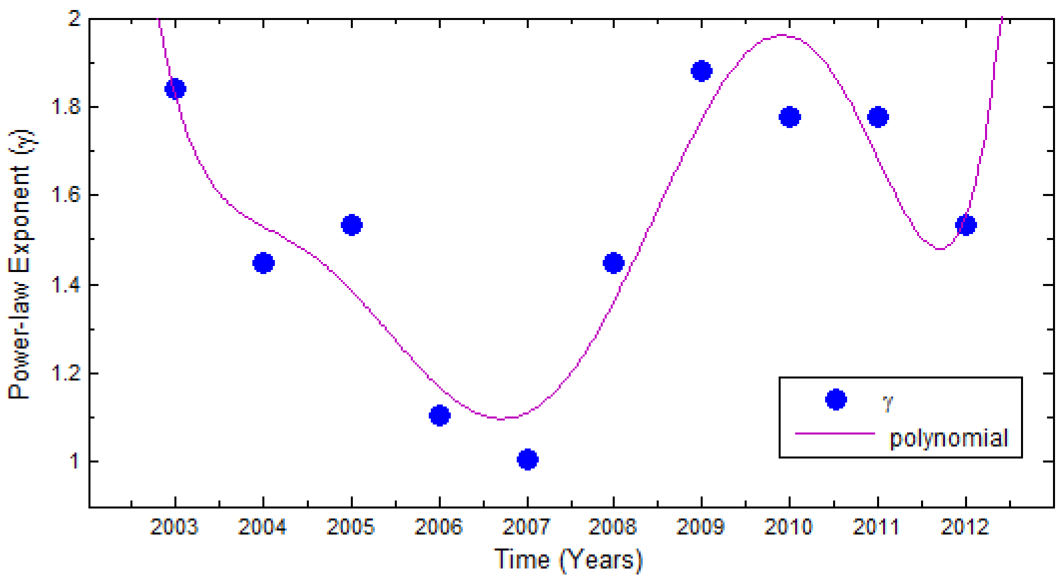


Figure 4.24. Power-law exponent values for 2003–2013.

After that, finally, *the distribution of the degree k* was computed and ability of financial network technique for analysis of certain points in time-period was investigated in Figure 4.24. As mentioned before, low power-law exponent values relate crisis periods. According to plot, after *2004*, γ value started to increase owing to Bourse Istanbul became one of the important assets in the world, as we observed in MST. However, in *2007*, because of global economic crisis Turkish Economy smashed as well as exponent value. After this, stock market and economy began to collect self until *2009*. Next, another crisis period was observed and affected global economy, likewise γ value. This quantity also proved that financial network techniques have considerable skill to investigate certain crisis periods before it appear. Forwhy, as seen in Figure 4.24, before *2007–2008* crisis, in *2006*, exponent value showed quite decrease, as well as, after second part of *2009*, which refer to *Global Recession* period.

In first part of thesis, BIST–100 index was analyzed by using financial network techniques, which are *distribution of correlation coefficients*, *statistical moments*, *normalized tree length*, *minimum spanning tree (MST)*, *hierarchical tree (HT)*, *edge properties* and *node properties* between *2011* and *2013*. This technique proved that it is very useful for stock market analysis, prediction and investigation stock movements and crisis periods. Finally, *Modern Portfolio Theory (MPT)* was used to compare financial network techniques and portfolio optimization; consequently, it was proved that this technique is as capable as *portfolio optimization* for risk analysis.

In second part of thesis, some of these techniques *(MST, statistical moments, normalized tree length and distribution of degree k)* were used to study global financial crisis and effects of the other assets on BIST–100 between *2003* and *2013*. Firstly, local factors affecting Turkish Economy were examined. These were used to determine critical points. After this, global assets were used to investigate world economy, global crisis and movement of assets. This part showed that financial network techniques allow observing not only local economy but also global economy status.

Next part of thesis is a summary, which forms of studies done, results, discussions and some recommendations for new researches, who would study this discipline in future.

PART 5

SUMMARY

Correlation based networks can be obtained from financial markets by investigating time series. *"Filtering procedure"* applied correlation matrix is created by the returns of a portfolio of financial assets provided to obtain distance matrix which selects a topological space for the stocks traded in a market. Therefore, in this study, it was showed how to associate a correlation matrix with a hierarchical tree and correlation based trees or graphs.

The information forms in correlation based trees and graphs provided some clues about the inter-relations among stocks of different *economic sectors*, *sub-sectors* or *indices*. The ultra-metrication in locally MST that was constructed based on stock price fluctuations help to obtain the information concealed in the correlation coefficients of stock price returns. Besides, from the hierarchical tree of the ultra-metric space, it can be viewed more clearly how a stock specifically correlate to one another. It was also studied the distribution of correlation coefficients and its moments by taking advantage of *data mining* and *statistical techniques*. These techniques provide to understand stocks movements better and by using *"normalized tree length"*, it was maintained to investigate and compare *"risk management guide"* abilities of statistical, financial and topological methods. Lastly, to analyze performance of *"stock correlation network"* concept derived information from these techniques used to compare *"Modern Portfolio Theory"* [1,7–9,12,55].

First part of this section form of these results and discussions of findings obtained from financial, statistical and topological techniques used to analyze BIST–100 Index and global crisis. In the other hand, second part aims to be a road map for new researchers and/or formers who want to improve this discipline.

5.1. RESULTS & DISCUSSION

- In this thesis, stock correlation network procedure applied to Bourse Istanbul–100 Index (BIST–100). Investigating data scale for *N=100* stocks was *T=506* days (2011–2013) and opening prices of stocks used as time series. Correlation matrix, which was formed of logarithmic returns of series, was used to compute *D* (distance matrix), which fulfills axioms of Euclidian metric. This matrix allowed to define relationships of stocks in portfolio and provided to compute *MST* and *HT*. Moreover, daily opening values of BIST–100 Index plotted with time and cubic spline of these values showed that portfolio was clearly affected during *2011* because of crisis, especially, in *second six-month period* (July 2011–December 2011).
- Distribution and statistical moments of correlations calculated to strengthen this idea and seen that correlation coefficients exist generally in scale between *0.1* and *0.6* that means this distribution is similar to *"Normal"* which means stocks fluctuate towards the same direction. Moments of correlation coefficients specified that *mean* increased and *variance*, *skewness* and *kurtosis* decreased in crisis period. In particular, high values of mean and skewness towards zero that distributed correlation more Gaussian pointed footprints of crisis.
- *Four six-month periods* of data scale used to plot MST, whose edges were distances among stock, from distance matrices that formed by using *Single Linkage Cluster Analysis (SLCA)* filtering procedure and *Prim's Algorithm* because of its *Big-O Notation* speed. Then, these plots used to calculate *"normalized tree length"* which showed that distances among stocks decreased and MSTs acted like great random clusters before crisis and during crisis period. After the crisis, distances increased and stocks started to move according to their *sectoral* and/or *sub-sectoral* states. This result showed that MST concept is a good guide to observe stock behaves and normalized tree length is successful to obtain crisis periods as well as statistical moments.
- After MSTs for *t=126* days are investigated, a MST that contains full data scale computed and this tree used to analyze network quantities of BIST–100 Index. Small *average clustering coefficient* (0.1667) showed that nodes of this

network passed 1 or 2 triangle and showed meaningful character while form clusters. Computed *diameter* (9.0449) and *average path length* (3.9512) of network was indicated that distances among stocks were small and highly correlated. Node properties *(degree, in-degree, out-degree, node betweenness* and *closeness)* and edge property *(edge betweenness)* of network were also investigated and seen that some nodes of networks had higher degrees than others had. Moreover, it was spotted that those nodes, which had high *node betweenness* and *closeness centrality*, were having high degrees, too. These properties show character of *"hub"* nodes in network and provided to observe controlling information run in network. Besides, *edge betweenness* is a way to show important edges of network and this proved edges between hub nodes had high values and pointed out aspects of network.

- To get more information about BIST–100 from MST, the *distribution of the degree k* computed. *Power-law exponent* was found as 2.68 ± 0.1 with *% 95* confidence bounds. This means BIST–100 network is a *scale-free network* that some of nodes are *"hub"*, which has *"higher status"* and *"stronger market influence power"*. Then, *power-law degree distribution* applied to the other periods of scale and seen that in crisis period, it decreases to *2.0* level as well as the other studies done. This result showed that again MST concept is a good *risk-management* concept.
- Sectoral states of MST was investigated and seen that most of stocks were in *industrial sector* and *service stocks* spread to clusters like extensions. Financial stocks formed in center of network and control nodes behaviors. Besides, it was observed that two technology stocks were in unessential position. After examining sectoral situation of network, *sub-sectoral* behavior was investigated by dividing network to four general clusters. Stock *51 (ISBNK)* existed in central position *(root)* of network and had strong relation with other *XBANK* stocks, which act as *"hub"* nodes, conduct the other clusters. Besides, it was seen that this network is *scale-free network*, also, they generated complete cluster whose all stocks have short distances. This means banks are in strong contact. The other *high-degree* stocks were *12 (ASYAB)*, *28 (ECILC)* and *99 (YKBNK)* form clusters. In summary, financial stocks have important role in network because of the other nodes and clusters spreading via these

nodes. Generally, stocks generated clusters according to their *sectoral* and *sub-sectoral* information. However, sometimes, stocks ignored *sub-sectoral* states and acted according to their *corporative* situations. In addition, in some conditions, stocks moved according to *product manufacturing factor*. *Average daily trading volume* and *liquidity* investigated by grouping stocks according to *indices* and obtained that *XU030* stocks in the center, *XU050* and *XU100* stocks spread to sides of MST in order. Therefore, it was seen that the most liquids exist in central position and become *reference* for the other stocks.

- Again, distance matrix was used to construct Hierarchical Tree (HT). *ALCA* filtering procedure used to clean matrix to create dendrogram because of its stability. Firstly, *sub-sectoral* states were investigated and seen that all bank nodes form cluster in shortest *ultra-metric sub-dominant distances*. This shows banks nodes act together. On the other hand, ALCA captured two Telecommunications stocks together MST was not. This means that HT is successful as well as MST and shows MST cannot. However, although, HT concept shows *sub-asset* acting, it was observed that MST is more skillful for this task. While a sectoral study was being applied to HT, it was seen that *XUMAL* stocks had shortest distances among them. Furthermore, *XUSIN* stocks had short distances means moved together. This clarified those financial stocks in strong connection and control information flow. *Liquidness* was also investigated by dividing HT into four clusters and it was observed most *liquid* stocks *(XU030)* existed in *central point* and more liquid stocks separated from *illiquid* ones. So, stocks moves according to *indices* and *liquidness* situations as well as *sectors* and *sub-sectors*.
- Lastly, *Modern Portfolio Optimization* used to analyze effectiveness of *financial network technique*. *Mean-variance-efficient frontier graphs* were showed that risk rate of portfolio increased as can be seen in *normalized tree length*, in second period. Thus, MST concept is very successful for *risk management*. MPT for all data MST was investigated to define risky stocks. It was clearly seen that *riskless stocks* took place in *central points* of MST or close to *"hubs"*. Besides, *risky stocks* were generally existed far from *"hubs"*, especially the root, *51 (ISBNK)*. In addition, stocks, which ignored *sectoral* and *sub-sectoral* behavior, had more risk than others have. Although, some stocks

were closely to *"hubs"*, they were risky. At this point, HT was used to analyze them and observed that these stocks have long *sub-dominant distances* and form clusters keeping risky ones in *ALCA HT*.

- In second part of thesis, local and global factors *(stock markets, commodities* and *currencies)* affecting BIST–100 index were examined and tried to investigate crisis periods and movements of global assets during ten years *(2003–2013)* by using *minimum spanning tree*. First, local factors were investigated to determine certain points of Turkish Economy during this period. These quantities are: *current account*, *gross and external debt*, *late liquidity and overnight interest rates, current and constant gross domestic product prices*, *growth rates by constant and current prices*, *growth rates in gross domestic product by purchasers in constant and current prices and export and import values*. Briefly, local parameters demonstrated that Turkish Economy was shaken after *2008*. Expect GDP prices, all economic parameters proved this condition. Therefore, with effects of these quantities, a global crisis was examined in *2007* and this lead to *Great Recession 2010–2012*.
- After that, global parameters were investigated to observe effects into Turkish Economy and Turkish Stock Market. These parameters are *worldwide global stock markets*, *commodities* and *currencies*. Currencies rose during *2008* because of crisis. While currencies were decreasing during *2010*, after *2011*, they increased. One of the commodities, *gold prices*, augmented between *2008* and *2011* owing to confidence to Turkish Economy decreased. Besides, *Brent-Oil* prices decreased in *2008* and in *first months of 2009*, it started to rise until *2012*. Thus, oil showed reaction to *2008 crisis*. After, prices started to increase and showed that there is an unstable situation in global economy. When some important stock markets were studied, it was seen that all of them showed negative reaction to global crisis. All markets smashed in *2008*, then, gathered strength until *2011*. Afterward, they were affected from a new crisis period. In these conditions, it could not be reasonable to hope there is not any influence in Turkish Stock Market. Therefore, when BIST–100 opening prices plot was investigated, it was observed that there was a significant decrease in *2007* and *2008*. After that, Turkish stock market started to rise until *2011*, but, after this point, it entered to a new crisis period and began to fall.

- After critical points in time-scale were determined, *financial network techniques* were used to investigate Turkish Stock Market and effects on it. Especially, in crisis periods, it was tried to examine which global factors have ability to form network and direct assets, called *"hub"* nodes. Moreover, with taking advantage of this kind of classification, effects on Turkish Economy could be studied and classified important assets, which have influence on BIST–100.
- Firstly, a MST, which contains *N=29* nodes (assets) and *L=10* years data set, was plotted. Stock markets, commodities and currencies were classified into three categories; *Europe, America* and *Asia*. MST showed that stock markets generally form cluster according to their regions. European stocks were separated into two groups. First group, which keeps lower budget markets dispersed around *Eurostoxx–50* that control information in this cluster. The second *EU* group consisted of more budget indices formed via *S&P 500* index as well as *Asia* indices. Information flowing through these clusters controlled by this *US* index could be said. In addition, this MST showed that American Markets laid generally central points of tree and became reference for all. In summary, it can be said that world economy was canalized by *US* assets; also, *Eurostoxx–50* is controlled by *NYSE*. Especially, *NASDAQ–100* was most important factor among all nodes because of its *"hub"* position. Moreover, except *Brent-Oil*, *currencies (Euro and US Dollar)* and *Gold* moved together. However, *Brent-Oil* existed among European markets and had close relationship with them.
- In addition, it was seen that BIST–100 was existed among *EU* stocks, which was controlled by *Eurostoxx–50* index. This cluster consisted of *Brent-Oil*, *DJIA* and *TA–25*, also. Thus, BIST–100 interacted with European indices can be said. This condition is a summary of ten years. Of course, in crisis periods, this case would change and effects of dominant factors would observe clearly. In conclusion, data was separated into years and investigated inter-changing of nodes and BIST–100.
- Ten years of data was studied and seen that BIST–100 was one of the important factors and *"hub"* nodes until *2008*, after this time, Turkish Economy started to fall because of crisis and this lead to *Global Recession* seen

in *2010–2012*. Plotted MSTs proved that Turkish Economy generally moved with *second degree* assets in Europe. Local factors affecting Turkish Economy already showed Turkish Stock Market was in crisis period in these years. These states proved by using *financial network techniques* and MST process, also. Moreover, MSTs showed that central part of global assets formed of *US* assets. This situation showed that US factors canalized global economy and Turkish Economy. Besides, it was observed that *commodities*, particularly, *Brent-Oil* prices and *currencies* had important influence on Bourse Istanbul.

- Afterwards, *statistical moments*, *normalized tree length* and *the distribution of the degree k* were computed. *Statistical moments* by using correlation matrices were used to analyze condition of assets during time and to study *probability of prediction* in crisis. *Mean of correlation coefficients* increased until *2013*. Main raises were observed in *2005*, *2007–2009* and *2010–2012* periods. Moreover, *skewness* of data felt toward zero in crisis periods, especially, in *2007–2008*. As known, *high mean-correlations* and *low-skewness* values provide signs of crisis periods. Besides, *polynomial* of data changes exponential, analyzing plots provided clues about before crisis. Therefore, it was seen that *statistical moments* showed efficient of *financial network techniques* to investigate crisis periods. Next, *normalized tree length* was analyzed to study distances among assets. As observed, *polynomial* of data decreased greatly in *2007–2008*, so, distances were reduced and assets came closely, in this period. In addition, this investigation helped to gain clues about crisis periods because of falling started in *2006*, in this situation.
- Finally, *the distribution of the degree k* was used to investigate ability of *financial network techniques* in crisis period analysis. While investigating *power-law exponent* values, it was seen that γ value started to increase due to Bourse Istanbul became one of the important assets in world. On the other hand, in *2007*, crisis affected completely global economy and Turkish Economy, *power-law exponent* started to decrease. Although, market started to gather strength until *2009*, another crisis period was observed and γ began to decrease. This quantity also proved that *financial network techniques* have much skill to investigate certain crisis periods before it appears. For example, in *2006*, before *2007–2008* crises, *exponent value* decreased and after *second*

part of 2009, a fall was observed which provided clues about *Global Recession* period.

In summary, the study of correlation based financial networks is a successful method able to gain information from the *correlation coefficient matrix* of a *financial time series*. The MST and the associated *sub-dominant ultra-metric* HT is useful in the examining of stock markets and defining of factors affecting groups of stocks. On the other hand, *stock correlation network concept* is skillful for *portfolio optimization*, *risk management* and *crisis analysis*. These results showed that time series of stocks prices carry valuable (and detectable) economic information and provide an insight into market behavior [1,7–9,12,55].

5.2. RECOMMENDATIONS

- This study can be done by dividing period into <u>less time windows</u> to observe information that is more specific.
- <u>Volatility</u> can be used as time series to investigate effect of price changes for grouping stocks.
- MST can be reduced by defining a <u>threshold</u> value for weights to see edges that are more significant and clusters. In addition, <u>k-MST</u> or P<u>lanar Maximally Filtered Graph (PMFG)</u> can be used to compare methods.
- The other <u>linkage analysis procedures</u> can be used to have another detailed knowledge and compare stabilities of these procedures.
- The other <u>clustering algorithms</u> can be run to investigate different states of stocks in clusters.
- Distances between stocks can be calculated by using other techniques, such as; <u>Euclidean</u>, <u>Manhattan</u>, <u>Chebyshev</u> et al., to investigate importance of distances in stock correlation network.
- The other probability distribution function can be used to gain extra information about network, such as; <u>auto-correlation</u>, <u>Kullback-Leibler distance</u>, et al.
- Risky stock places and their moves can be studied with paralleling to <u>portfolio optimization tools</u>.

- Stocks information; <u>sectors</u>, <u>sub-sectors</u>, <u>indices</u>, <u>liquidness</u>, et al., can be studied in detail by investigating <u>positions in clusters</u> to determine what affects stock moves.
- Lastly, more detail analysis can be run by increasing <u>number of assets</u> or <u>time-period</u> in global economy.

REFERENCES

1. Mantegna, R. N., "Hierarchical structure in financial markets", ***The European Physical Journal B***, 11 (1): 193–197 (1999).

2. Bonanno, G., Vandewalle, N., and Mantegna, R. N., "Taxonomy of stock market indices", ***Physical Review E***, 62 (6): 7615–7618 (2000).

3. Bonanno, G., Lillo, F., and Mantegna, R. N., "High-frequency cross-correlation in a set of stocks", ***Quantitative Finance***, 1 (1): 96–104 (2001).

4. Vandewalle, N., Brisbois, F., and Tordoir, X., "Non-random topology of stock markets", ***Quantitative Finance***, 1 (3): 372–374 (2001).

5. Bonanno, G., Caldarelli, G., Lillo, F., and Mantegna, R. N., "Topology of correlation-based minimal spanning trees in real and model markets", ***Physical Review E***, 68 (046130): 1–4 (2003).

6. Miccichè, S., Bonanno, G., Lillo, F., and Mantegna, R. N., "Degree stability of a minimum spanning tree of price return and volatility", ***Physica A***, 324 (1): 66–73 (2003).

7. Bonanno, G., Caldarelli, G., Lillo, F., Miccichè, S., Vandewalle, N., and Mantegna, R. N., "Networks of equities in financial markets", ***European Physical Journal B***, 38 (2): 363–371 (2004).

8. Situngkir, H. and Surya, Y., "On stock market dynamics through ultrametricity of minimum spanning tree", Departmental Technical Report, ***Bandung Fe Institute***, Indonesia, 1 (2005).

9. Coelho, R., Gilmore, C. G., Lucey, B., Richmond, P., and Hutzler, S., "The evolution of interdependence in world equity markets: Evidence from minimum spanning trees", ***Physica A***, 376 (1): 455–466 (2007).

10. Garas, A. and Argyrakis, P., "Correlation study of the Athens stock exchange", ***Physica A***, 380 (1): 399–410 (2007).

11. Tse, C. K., Liu, J., and Lau, F. C. M., "A network perspective of the stock market", ***Journal of Empirical Finance***, 17 (4): 659–667 (2010).

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

13. Kantar, E., Deviren, B., and Keskin, M., “Hierarchical structure of Turkey’s foreign trade”, ***Physica A***, 390 (20): 3454–3476 (2011).

14. Kantar, E., Deviren, B., and Keskin, M., “Investigation of major international and Turkish companies via hierarchical methods and bootstrap approach”, ***The European Physical Journal B***, 84 (2): 339–350 (2011).

15. Junior, L. S., “Pruning a minimum spanning tree”, ***Physica A***, 391 (8): 2678–2711 (2012).

16. Kantar, E., Keskin, M., and Deviren, B., “Analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods”, ***Physica A***, 391 (7): 2342–2352 (2012).

17. Internet: TechTarget, Inc., “What is Network?”, **http://searchnetworking.techtarget.com/definition/network** (2013).

18. Internet: Hawes, L., “Enterprise Software Architecture”, **http://www.forbes.com/sites/larryhawes/2012/03/14/enterprise-software-architecture-a-network-of-services-not-a-layered-stack/** (2013).

19. Internet: Rodrigue, J. P. and Ducruet, C., “Graph Theory: Definition and Properties”, **http://people.hofstra.edu/geotrans/eng/methods/ch1m2en.html** (2013).

20. Internet: Rodrigue, J. P. and Ducruet, C., “Basic Graph Representation of a Transport Network”, **http://people.hofstra.edu/geotrans/eng/methods/basicnetwork.html** (2013).

21. Acemoglu, D. and Ozdaglar, A., “Graph theory and social networks”, Lecturer Notes, ***Massachusetts Institute of Technology Department of Economics***, Boston, MA, US, 13–18 (2009).

22. Internet: Matteucci, M., “Clustering”, **http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/** (2013).

23. Internet: Irene, M. M., “Hierarchical Clustering”, **http://www.cse.iitb.ac.in/dbms/Data/Courses/CS632/1999/clustering/node3.html** (2013).

24. Internet: TIBCO Inc., “Dendrograms and Clustering”, **http://stn.spotfire.com/spotfire_client_help/heat/heat_dendrograms_and_clustering.htm** (2013).

25. Mantegna, R. N., Lillo, F., Salvatore, M., and Tumminello, M., “Correlation-based networks in finance”, ***International Workshop and Conference on Network Science***, Norwich Research Park, UK, 9, 18 (2008).

26. Internet: Manning, C. D., Raghavan, P., and Schütze, H., “Single-Link, Complete-Link & Average-Link Clustering”, **http://nlp.stanford.edu/IR-book/completelink.html** (2013).

27. Internet: Cohen, W. W. and Rivard, K., "Dependency Network", **http://curtis.ml.cmu.edu/w/courses/index.php/Dependency_network** (2013).

28. Erickson, J., "Minimum spanning trees", Lecturer Notes, ***University of Illinois Department of Computer Science***, Urbana&Champaign, IL, US, 1–6 (2011).

29. Internet: Eppstein, D., "Minimum Spanning Trees", **http://www.ics.uci.edu/~eppstein/161/960206.html** (2013).

30. Bang, Y. W. and Chao, K. M., "Minimum spanning trees", Lecturer Notes, ***National Taiwan University Department of Computer Science and Information Engineering***, Taipei, Taiwan, 1–5 (2007).

31. Suleman, O., "Dynamics of the FX Market: A minimal spanning tree approach", ***Fifth the Oxford Centre for Computational Finance City Seminar***, Oxford, UK (2004).

32. Internet: Averkamp, H., "What is Stock?", **http://blog.accountingcoach.com/what-is-stock/** (2013).

33. Internet: ValueClick, Inc., "Liquid Asset", **http://www.investopedia.com/terms/l/liquidasset.asp** (2013).

34. Internet: ValueClick, Inc., "Illiquid Asset", **http://www.investopedia.com/terms/i/illiquid.asp** (2013).

35. Internet: WebFinance, Inc., "What is Sector?", **http://www.investorwords.com/4430/sector.html** (2013).

36. Internet: WebFinance, Inc., "What is Index?", **http://www.investorwords.com/2426/index.html** (2013).

37. Internet: ValueClick, Inc., "Closing Price", **http://www.investopedia.com/terms/c/closingprice.asp** (2013).

38. Internet: ValueClick, Inc., "Average Daily Trading Volume", **http://www.investopedia.com/terms/a/averagedailytradingvolume.asp** (2013).

39. Internet: ValueClick, Inc., "Opening Price", **http://www.investopedia.com/terms/o/openingprice.asp** (2013).

40. Internet: ValueClick, Inc., "Return on Investment", **http://www.investopedia.com/terms/r/returnoninvestment.asp** (2013).

41. Internet: Mathworks Inc., "Convert Prices to Returns", **http://www.mathworks.com/help/econ/price2ret.html?searchHighlight=price2ret** (2013).

42. Internet: Stockburger, D. W., "Correlation", **http://www2.webster.edu/~woolflm/correlation/correlation.html** (2013).

43. Internet: Weisstein, E. W., “Covariance”, **http://mathworld.wolfram.com/Covariance.html** (2013).

44. Internet: Weisstein, E. W., “Standard Deviation”, **http://mathworld.wolfram.com/StandardDeviation.html** (2013).

45. Internet: Bartlein, P. J., “Statistical Moments”, **http://geography.uoregon.edu/geogr/topics/moments.htm** (2013).

46. Williams, R., “Normal distribution”, Lecturer Notes, ***University of Notre Dame Department of Sociology***, Notre Dame, IN, US, 1–3 (2004).

47. Kiymaz, H., “The initial and aftermarket performance of IPOs in an emerging market: evidence from Istanbul Stock Exchange”, ***Journal of Multinational Financial Management***, 10 (2): 213–227 (2000).

48. Odabasi, A., “ISE indices”, Lecturer Notes, ***Bogazici University Department of Management***, Istanbul, Turkey, 1–5 (2013).

49. Internet: Bourse Istanbul, “Constituent Equities of Indices”, **http://borsa istanbul.com/en/data/data/equity-market-data/index-data** (2013).

50. Mantegna, R. N. and Stanley, H. E., “Taxonomy of a stock portfolio”, An Introduction to Econophysics: Correlations and Complexity in Finance, ***Cambridge Univ. Press***, Cambridge, UK, 105–107 (2000).

51. Huang, W. Q., Zhuang, X. T., and Yao, S., “A network analysis of the Chinese Stock Market”, ***Physica A***, 388 (14): 2956–2964 (2009).

52. Çukur, S., Eryiğit, M., and Eryiğit, R., “Cross correlations in an emerging market financial data”, ***Physica A***, 376 (1): 555–564 (2007).

53. Eryiğit, M. and Eryiğit, R., “Network structure of cross-correlations among the world market indices”, ***Physica A***, 388 (17): 3551–3562 (2009).

54. Internet: Associated Press, “Turkish Stock Market Drops Sharply”, **http://finance.yahoo.com/news/Turkish-stock-market-drops-apf-1945962593.html** (2013).

55. Onnela, J. P., Chakraborti, A., Kaski, K., Kertész, J., and Kanto, A., “Dynamics of market correlations: Taxonomy and portfolio analysis”, ***Physical Review E***, 68 (5): 6110–6122 (2003).

56. Internet: ValueClick, Inc., “Modern Portfolio Theory”, **http://www.invest opedia.com/terms/m/modernportfoliotheory.asp** (2013).

57. Internet: Pettinger, T., “Factors Affecting the Stock Market”, **http://www .economicshelp.org/blog/2841/economics/factors-affecting-the-stock-market/** (2013).

58. Internet: WebFinance, Inc., “The Relationship Between the Stock Market and Forex Markets”, **http://www.investorwords.com/tips/1298/the-relationship-between-the-stock-market-and-forex-markets.html** (2013).

59. Internet: ValueClick, Inc., “Gross Domestic Product (GDP) Definition”, **http://www.investopedia.com/terms/g/gdp.asp** (2013).

60. Internet: WebFinance, Inc., “Gross Domestic Product Definition”, **http://www.investorwords.com/2153/GDP.html** (2013).

61. Internet: ValueClick, Inc., “Economic Growth Rate Definition”, **http://www.investopedia.com/terms/e/economicgrowthrate.asp** (2013).

62. Internet: ValueClick, Inc., “External Debt Definition”, **http://www.investopedia.com/terms/e/external-debt.asp** (2013).

63. Internet: ValueClick, Inc., “Current Account Definition”, **http://www.investopedia.com/terms/c/currentaccount.asp** (2013).

64. Internet: ValueClick, Inc., “Balance of Trade (BOT) Definition”, **http://www.investopedia.com/terms/b/bot.asp** (2013).

65. Internet: Turkish Statistical Institute, “National Accounts”, **http://www.turkstat.gov.tr/Start.do** (2013).

66. Internet: Republic of Turkey Prime Ministry Undersecretariat of Treasury, “External Debt Statistics of Turkey”, **http://www.treasury.gov.tr/default.aspx?nsw=EilDPQez15w=-SgKWD+pQItw=&mid=695&cid=12&nm=765** (2013).

67. Internet: Republic of Turkey Ministry of Economy, “Export Statistics”, **http://www.ekonomi.gov.tr/index.cfm?sayfa=79192159-19DB-2C7D-3D5AE56731D11E50** (2013).

68. Internet: Central Bank of the Republic of Turkey, “Current Account Report”, **http://www.tcmb.gov.tr/research/odemeler/odemeler.html** (2013).

69. Internet: Yahoo!, Inc., “Historical Prices”, **http://finance.yahoo.com/** (2013).

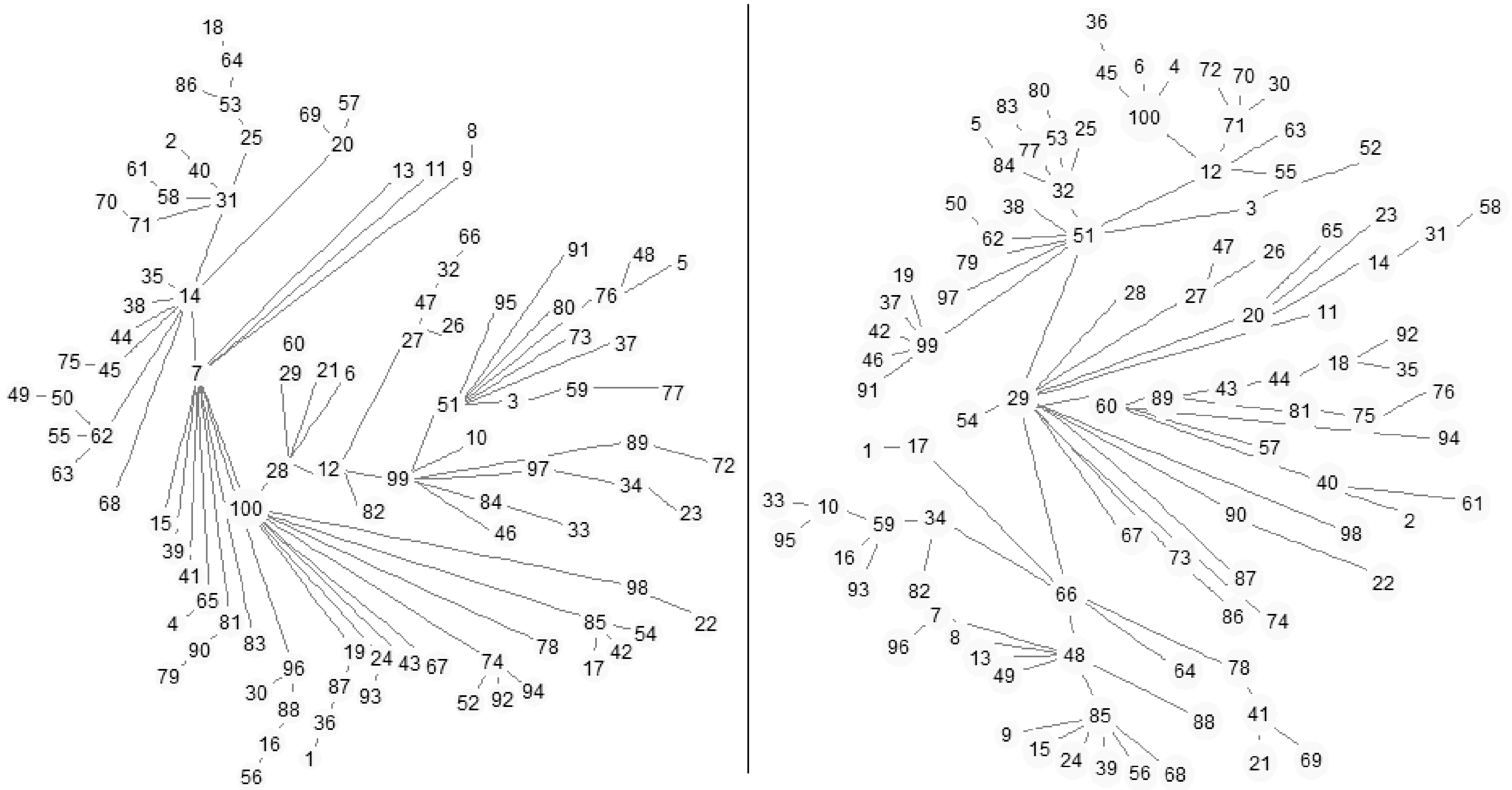
70. Internet: Google Finance, “Historical Prices”, **http://www.google.com/finance** (2013).

71. Internet: ValueClick, Inc., “The 2007-2008 Financial Crisis in Review”, **http://www.investopedia.com/articles/economics/09/financial-crisis-review.asp** (2013).

72. Internet: ValueClick, Inc., “Global Recession Definition”, **http://www.investopedia.com/terms/g/global-recession.asp** (2013).

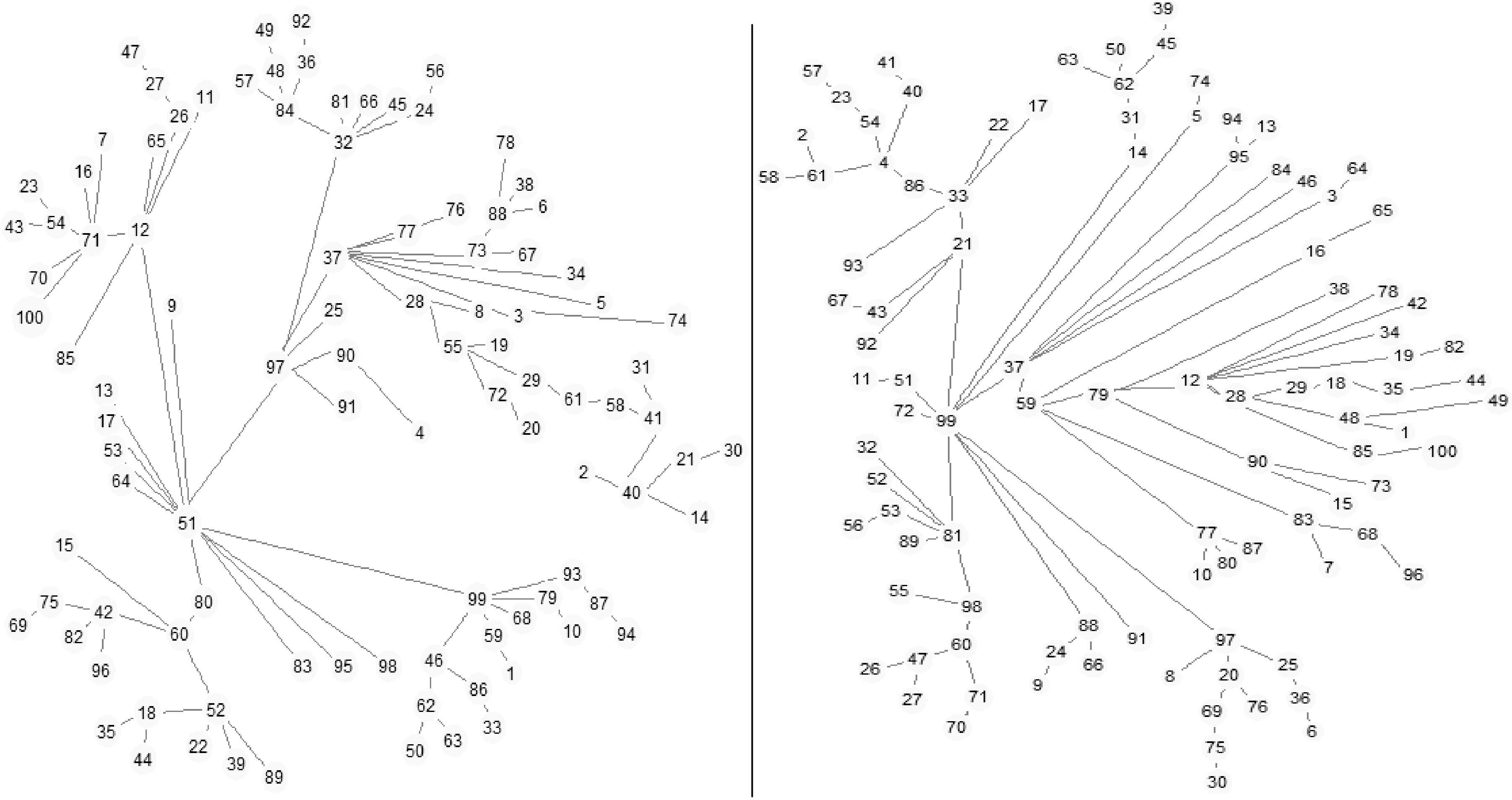
APPENDIX A.

LARGER VIEWS OF MSTs



a) January–June 2011 b) July–December 2011

Figure Appendix A.1. Larger views of six-month periods.



c) January–June 2012 d) July–December 2012

Figure Appendix A.1. (Continuing).

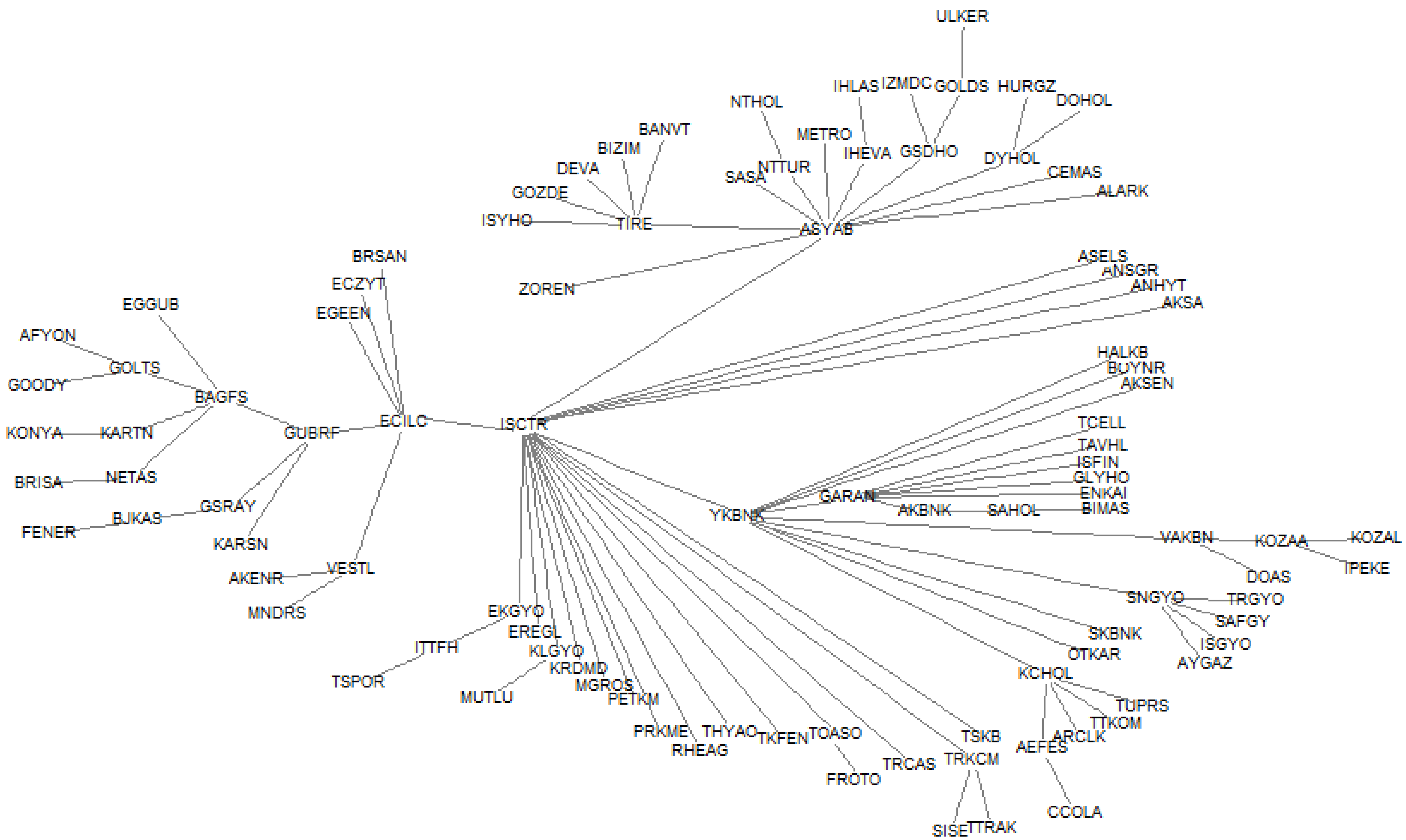


Figure Appendix A.2. Larger view of MST for 2011–2013.

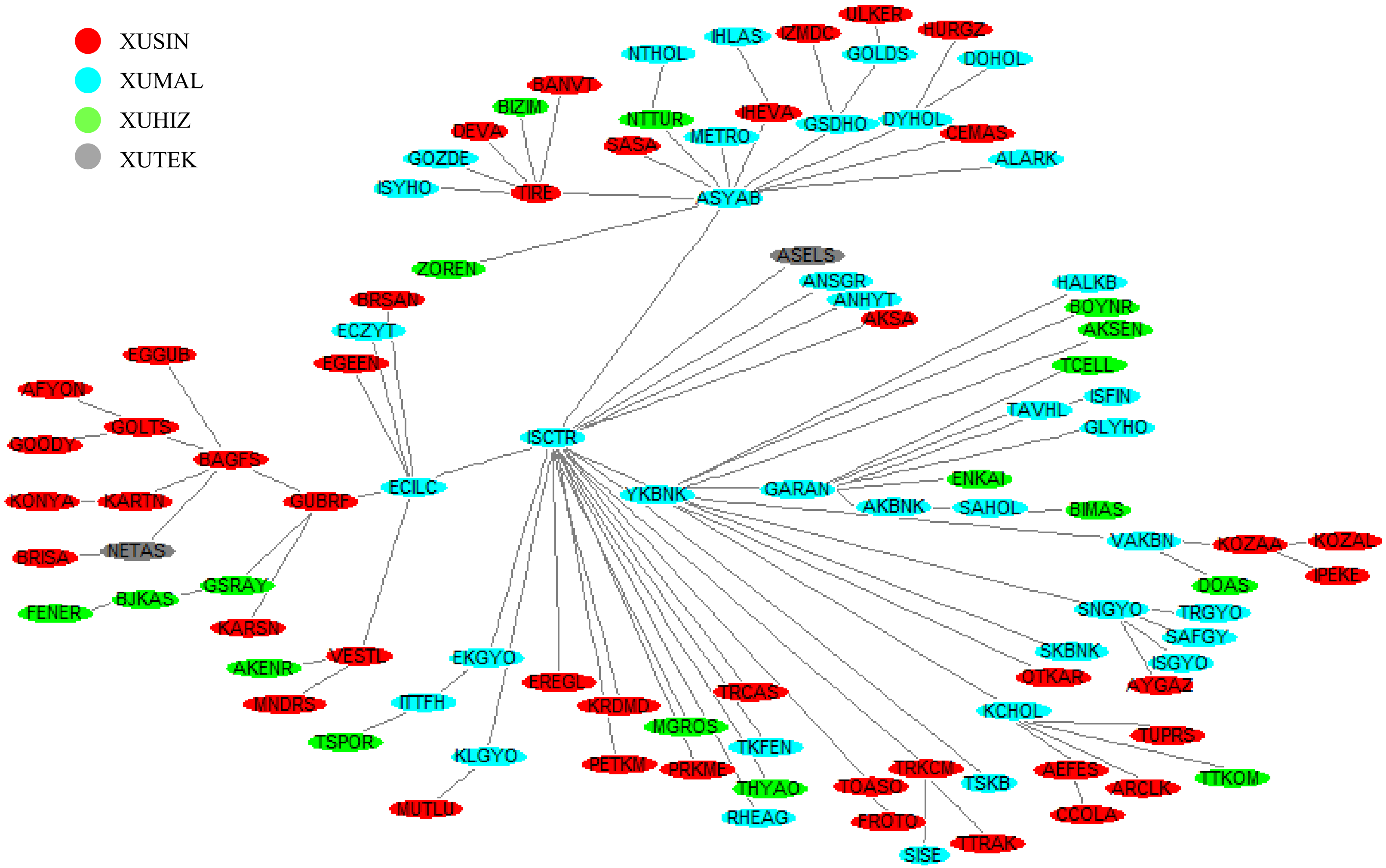


Figure Appendix A.3. Larger view of MST (sector) for 2011–2013.

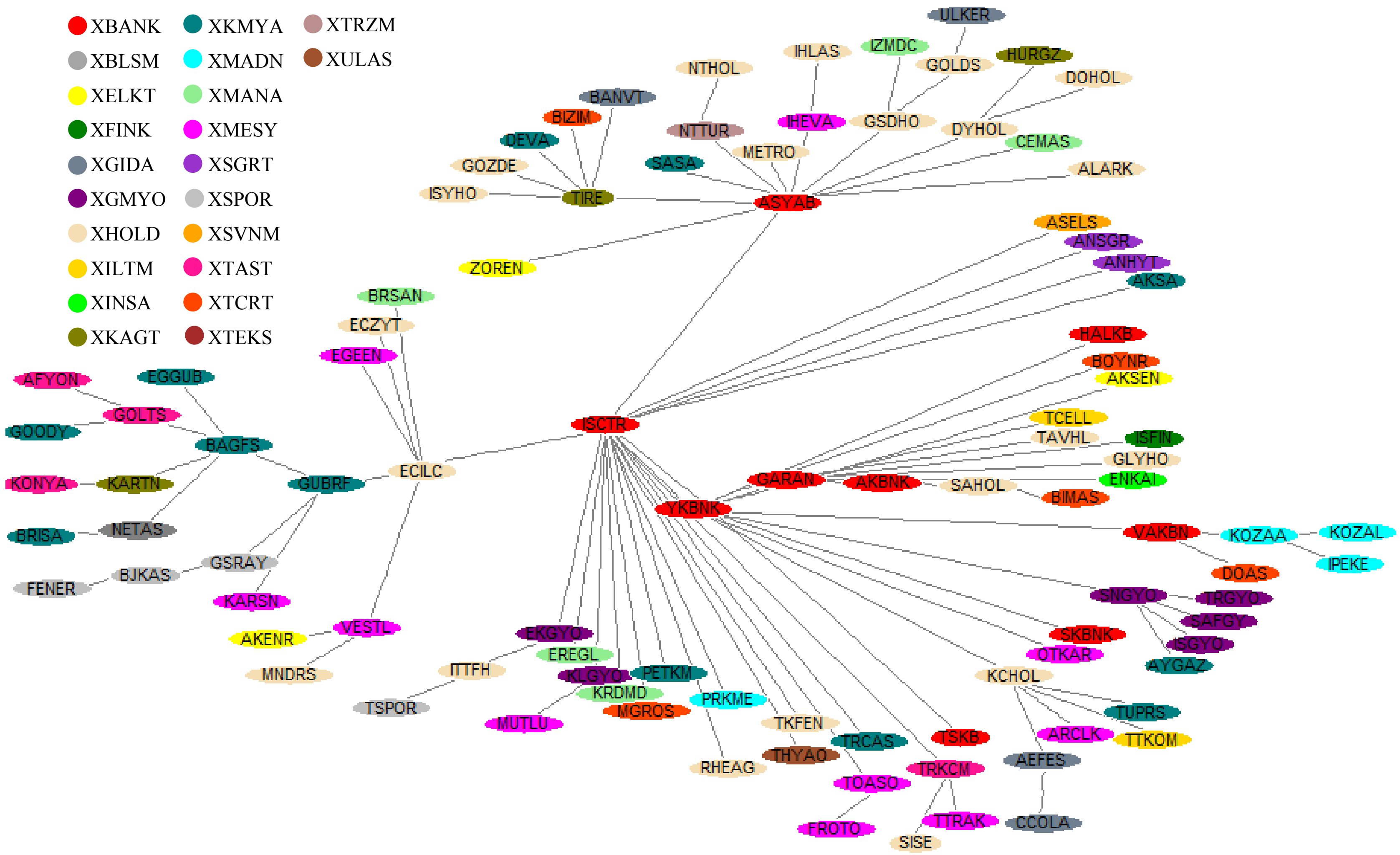


Figure Appendix A.4. Larger view of MST (sub-sector) for 2011–2013.

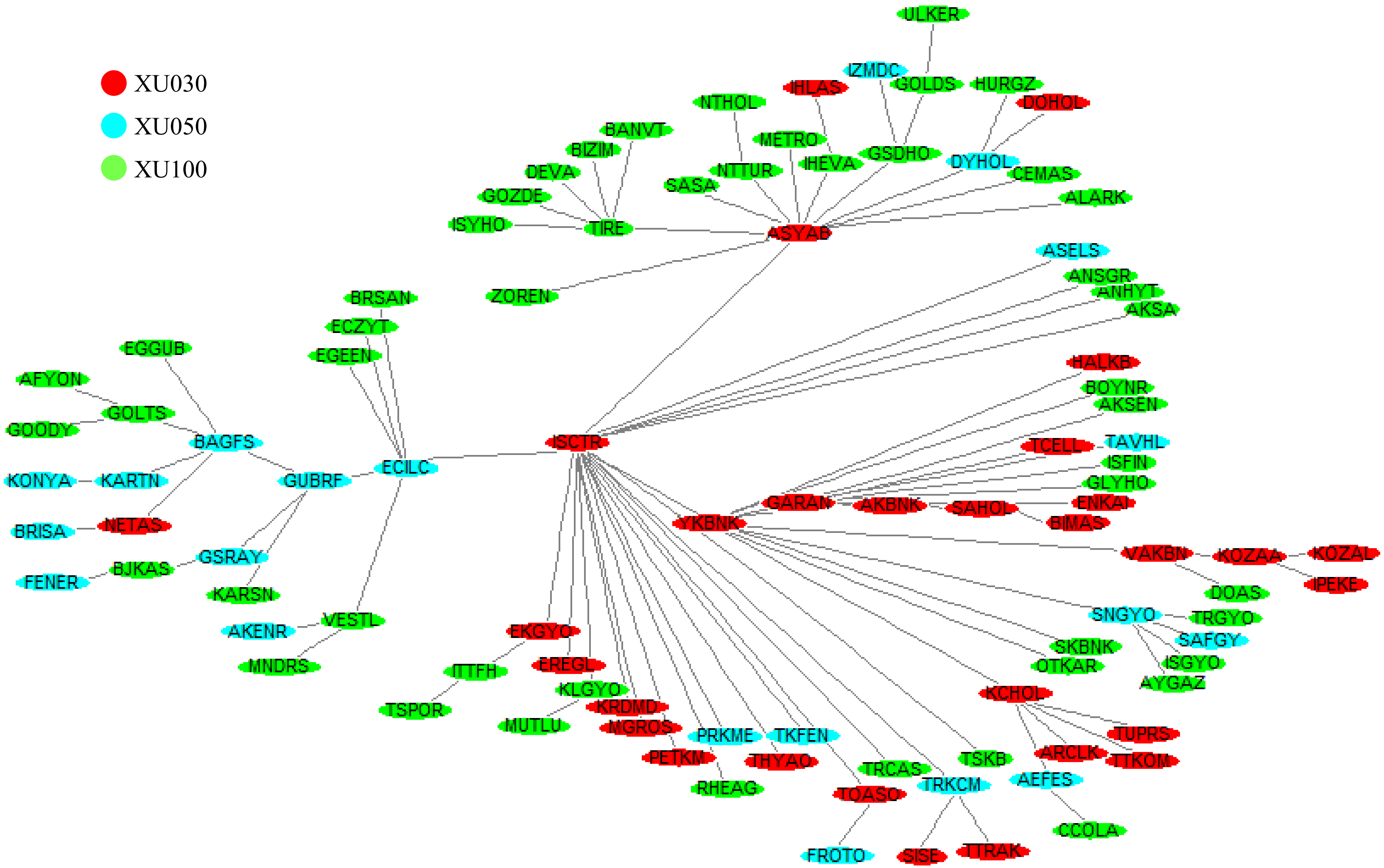


Figure Appendix A.5. Larger view of MST (index) for 2011–2013.

APPENDIX B.

LARGER VIEWS OF ALCA HTs

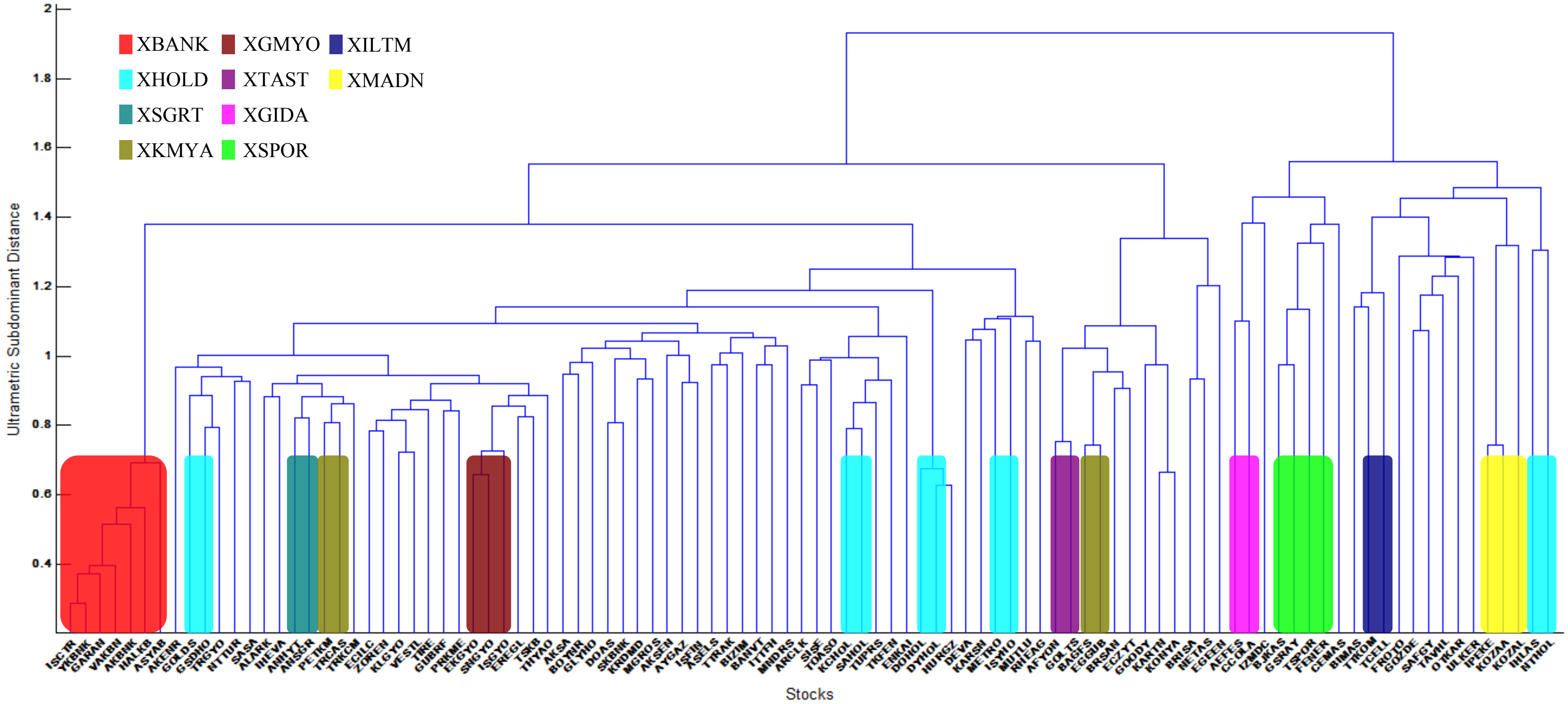


Figure Appendix B.1. Larger view of HT (sub-sector).

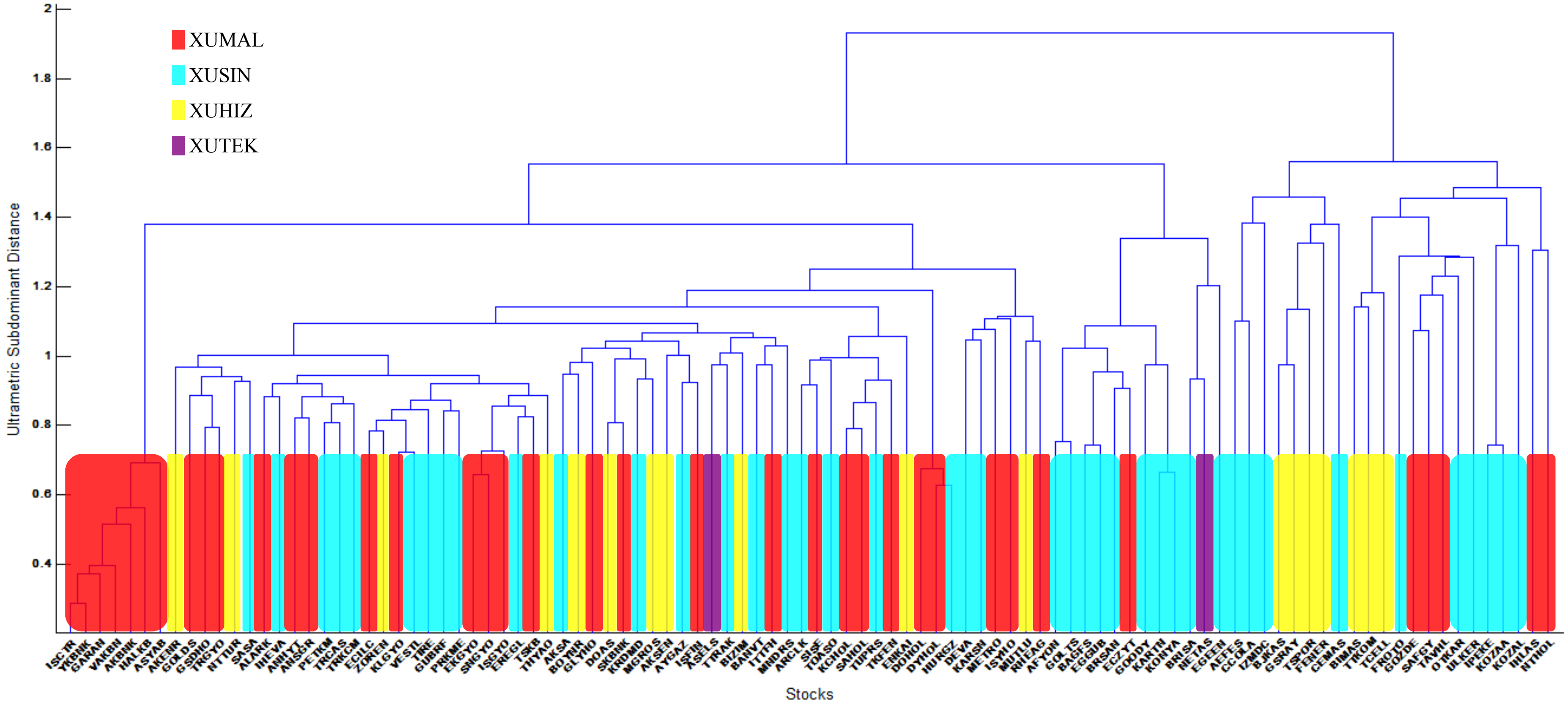


Figure Appendix B.2. Larger view of HT (sector).

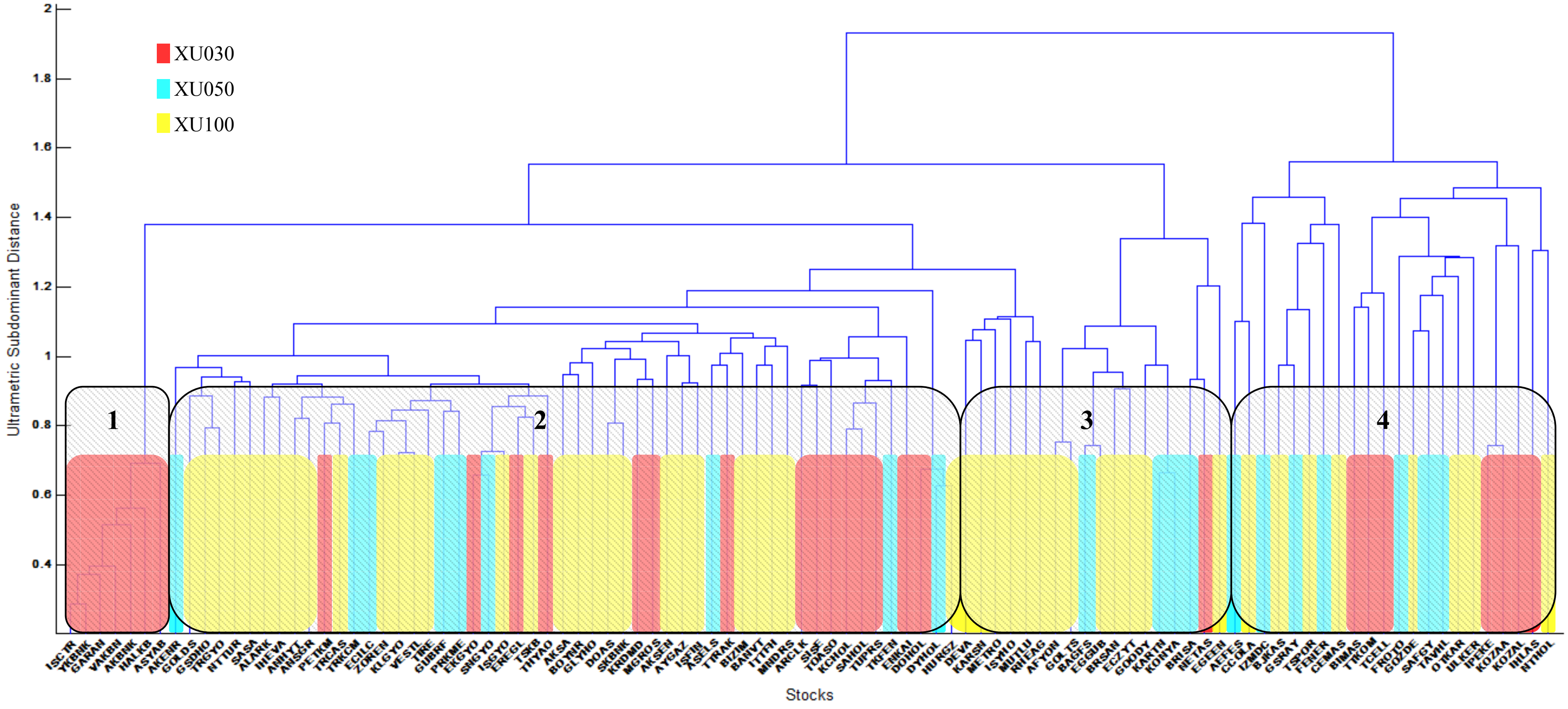


Figure Appendix B.3. Larger view of HT (index).

APPENDIX C.

RISK-VOLATILITY TABLE

Table Appendix C.1. Full-list of risk-volatility.

Stock	Risk	Volatility (%)	Stock	Risk	Volatility (%)	Stock	Risk	Volatility (%)	Stock	Risk	Volatility (%)
AEFES	0.019176	30.41	DOHOL	0.027202	43.14	ISCTR	0.020266	32.14	SAFGY	0.032305	51.23
AFYON	0.029571	46.90	DYHOL	0.02645	41.95	ISFIN	0.019332	30.66	SAHOL	0.021109	33.48
AKBNK	0.02186	34.67	ECILC	0.017361	27.53	ISGYO	0.017561	27.85	SASA	0.024104	38.23
AKENR	0.022026	34.93	ECZYT	0.02131	33.79	ISYHO	0.026774	42.46	SISE	0.022093	35.04
AKSA	0.018918	30.00	EGEEN	0.039405	62.49	ITTFH	0.020959	33.24	SKBNK	0.01933	30.66
AKSEN	0.018895	29.97	EGGUB	0.025461	40.38	IZMDC	0.024761	39.27	SNGYO	0.022328	35.41
ALARK	0.015259	24.20	EKGYO	0.022444	35.59	KARSN	0.024453	38.78	TAVHL	0.018082	28.68
ANHYT	0.018878	29.94	ENKAI	0.018944	30.04	KARTN	0.033321	52.84	TCELL	0.018248	28.94
ANSGR	0.020309	32.21	EREGL	0.019217	30.48	KCHOL	0.019927	31.60	THYAO	0.021246	33.69
ARCLK	0.020651	32.75	FENER	0.037291	59.14	KLGYO	0.019439	30.83	TIRE	0.021471	34.05
ASELS	0.018486	29.32	FROTO	0.018888	29.95	KONYA	0.032594	51.69	TKFEN	0.01895	30.05
ASYAB	0.019936	31.62	GARAN	0.020506	32.52	KOZAA	0.024318	38.57	TOASO	0.023098	36.63
AYGAZ	0.017768	28.18	GLYHO	0.022457	35.61	KOZAL	0.023126	36.67	TRCAS	0.020997	33.30
BAGFS	0.023091	36.62	GOLDS	0.021647	34.33	KRDMD	0.019511	30.94	TRGYO	0.019888	31.54
BANVT	0.018678	29.62	GOLTS	0.029206	46.32	METRO	0.027542	43.68	TRKCM	0.020436	32.41
BIMAS	0.017086	27.10	GOODY	0.034207	54.25	MGROS	0.02508	39.77	TSKB	0.02103	33.35
BIZIM	0.017944	28.46	GOZDE	0.032763	51.96	MNDRS	0.02701	42.83	TSPOR	0.038672	61.33
BJKAS	0.03349	53.11	GSDHO	0.024009	38.08	MUTLU	0.023124	36.67	TTKOM	0.018077	28.67
BOYNR	0.023062	36.57	GSRAY	0.032918	52.20	NETAS	0.038727	61.42	TTRAK	0.023894	37.89
BRISA	0.035654	56.54	GUBRF	0.023281	36.92	NTHOL	0.017	26.96	TUPRS	0.020294	32.18
BRSAN	0.026678	42.31	HALKB	0.022046	34.96	NTTUR	0.020655	32.76	ULKER	0.020033	31.77
CCOLA	0.023587	37.41	HURGZ	0.025601	40.60	OTKAR	0.018482	29.31	VAKBN	0.021833	34.63
CEMAS	0.046516	73.77	IHEVA	0.025672	40.71	PETKM	0.01848	29.31	VESTL	0.017737	28.13
DEVA	0.020919	33.17	IHLAS	0.023401	37.11	PRKME	0.024937	39.55	YKBNK	0.023227	36.84
DOAS	0.021849	34.65	IPEKE	0.02979	47.24	RHEAG	0.028526	45.24	ZOREN	0.018019	28.58

APPENDIX D.

FIGURES OF LOCAL ECONOMICAL FACTORS

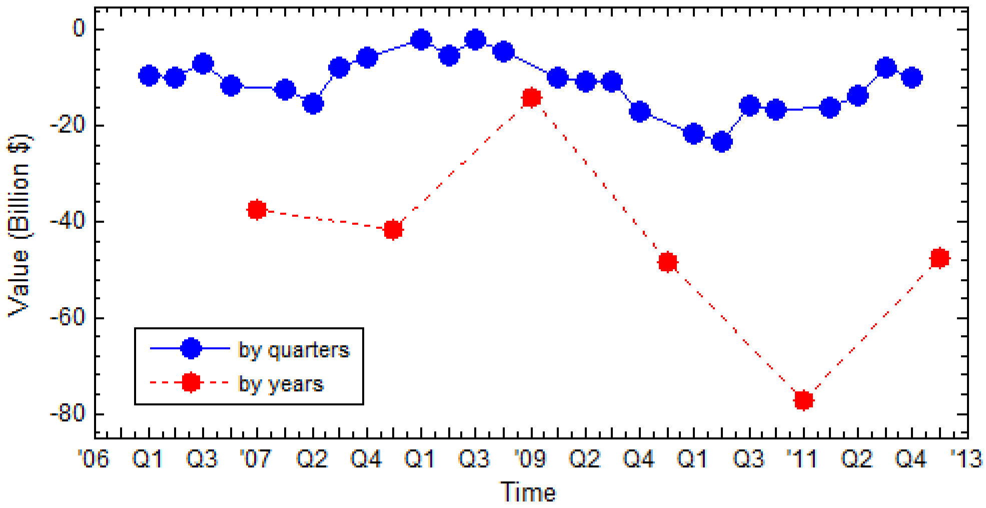


Figure Appendix D.1. Current account.

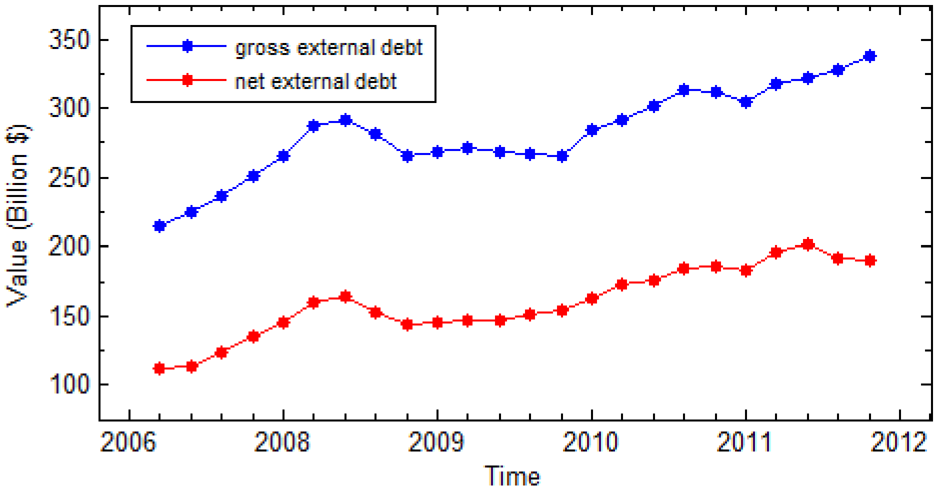


Figure Appendix D.2. Gross and net external debt.

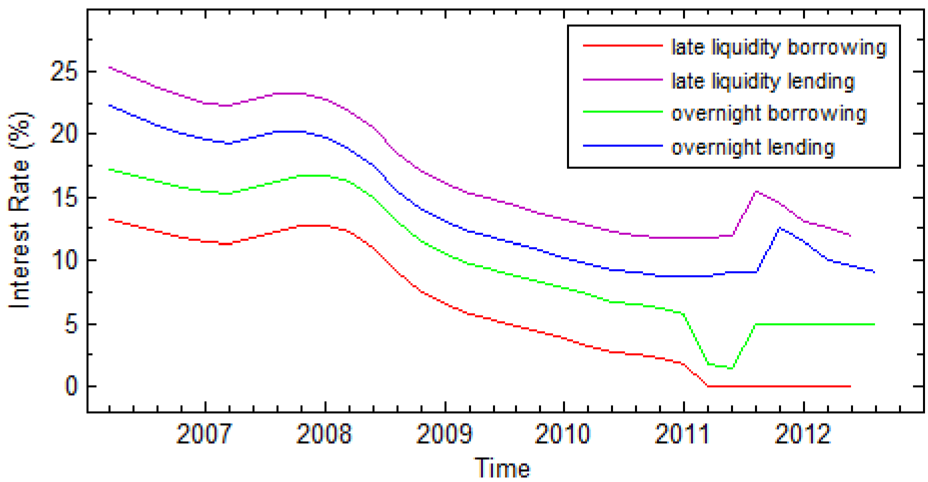


Figure Appendix D.3. Interest rates.

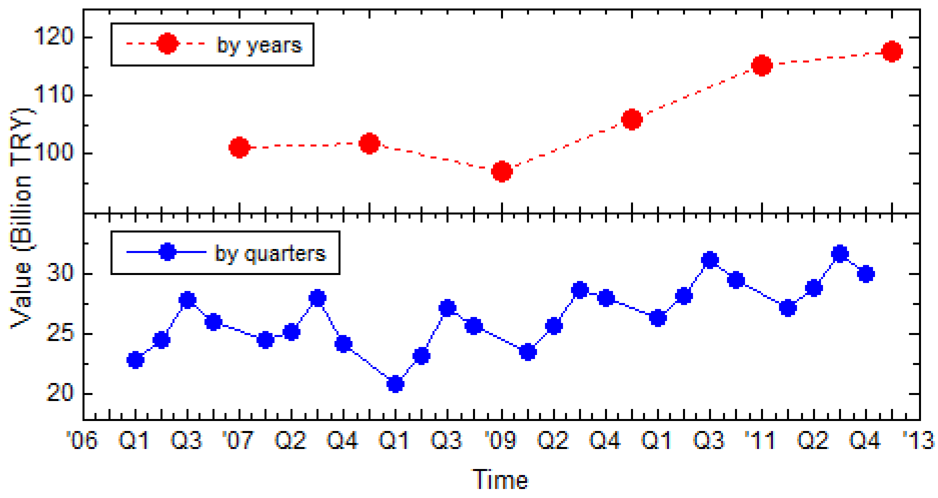


Figure Appendix D.4. Gross domestic product in constant prices.

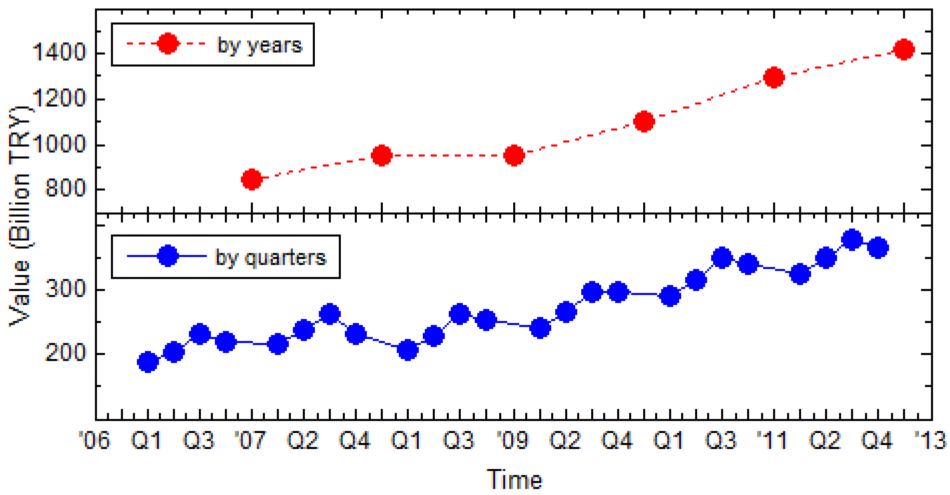


Figure Appendix D.5. Gross domestic product in current prices.

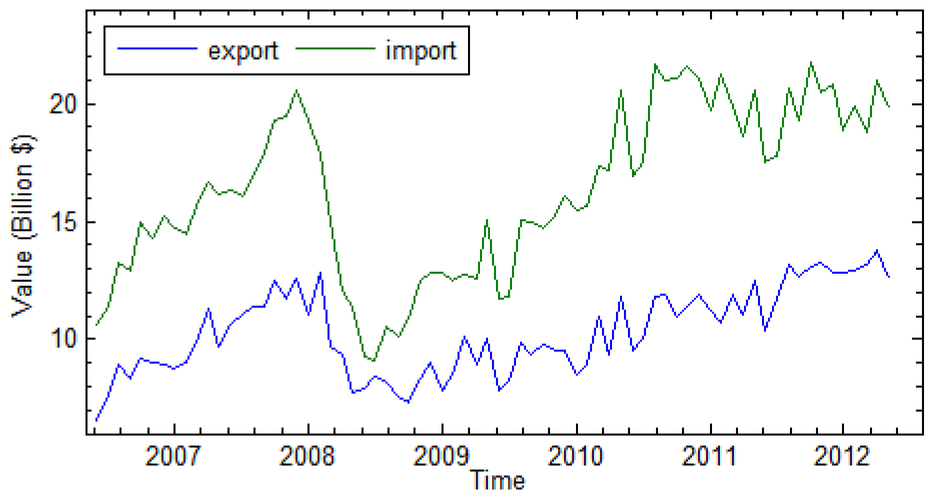


Figure Appendix D.6. 2007-2013 export and import values.

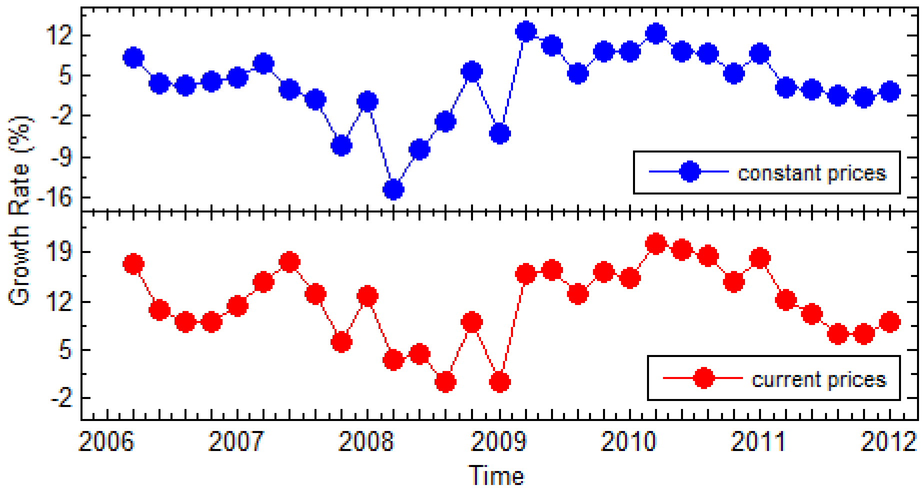


Figure Appendix D.7. Growth rates by constant and current prices.

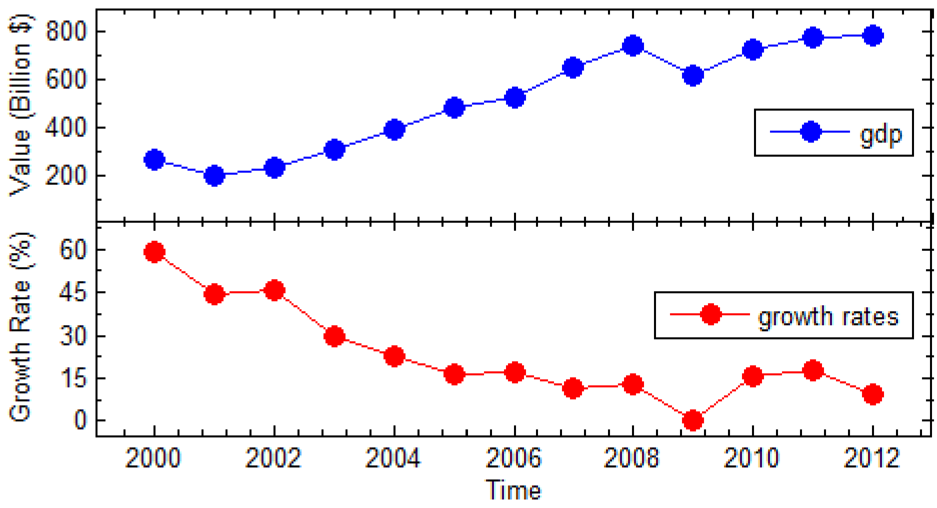


Figure Appendix D.8. Gross domestic product by purchasers current pricers.

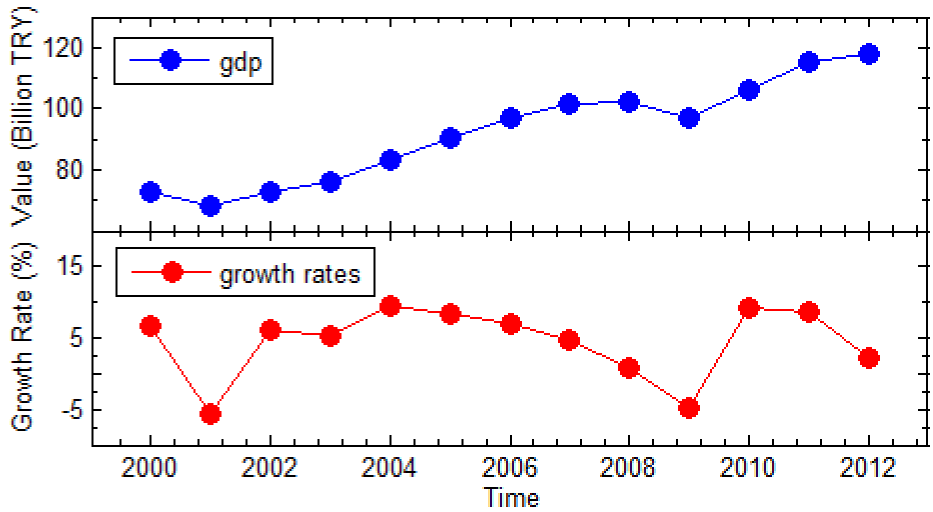


Figure Appendix D.9. Gross domestic product by purchasers constant pricers.

APPENDIX E.

FIGURES OF MSTs FOR 2003–2013

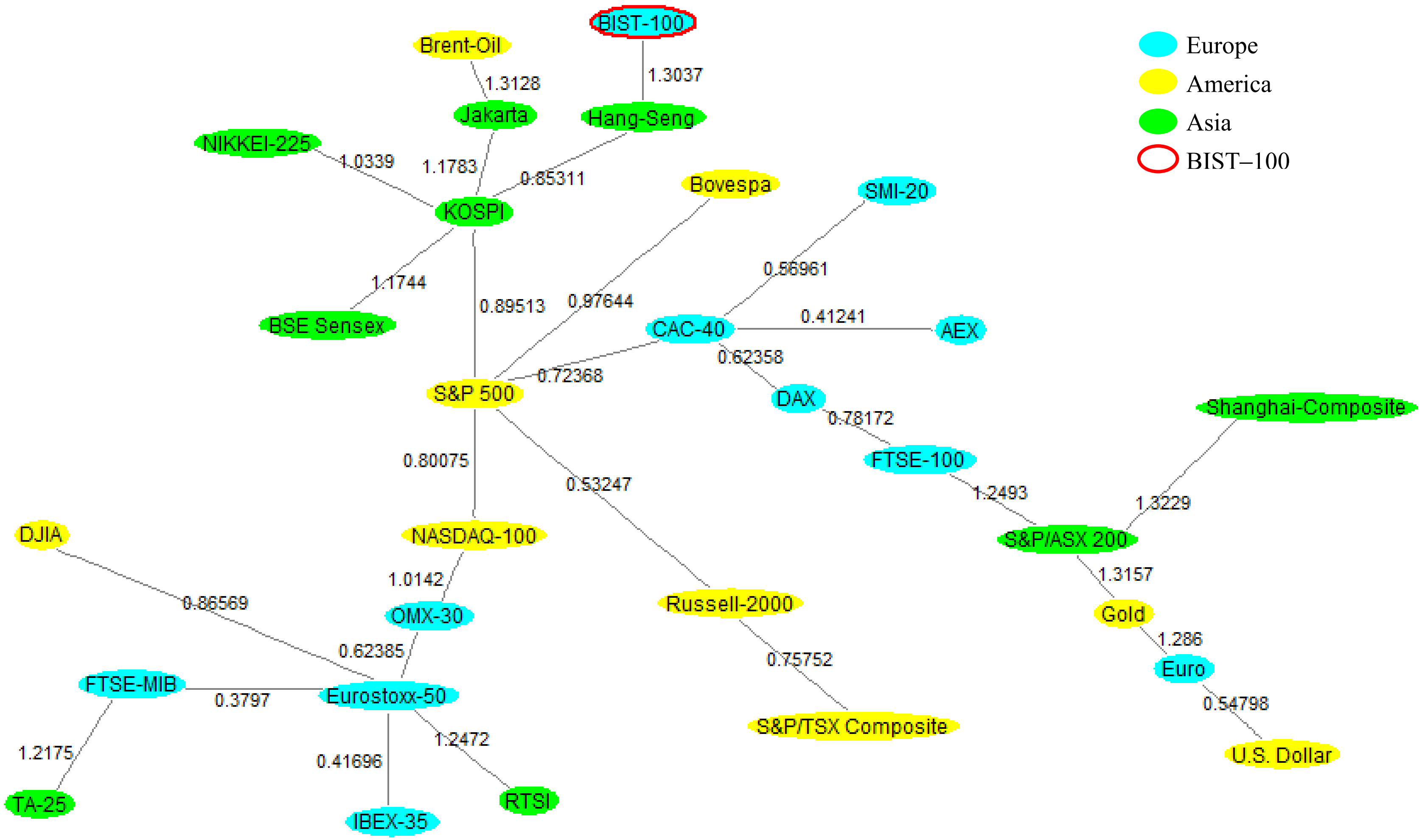


Figure Appendix E.1. MST of global financial assets for 2003.

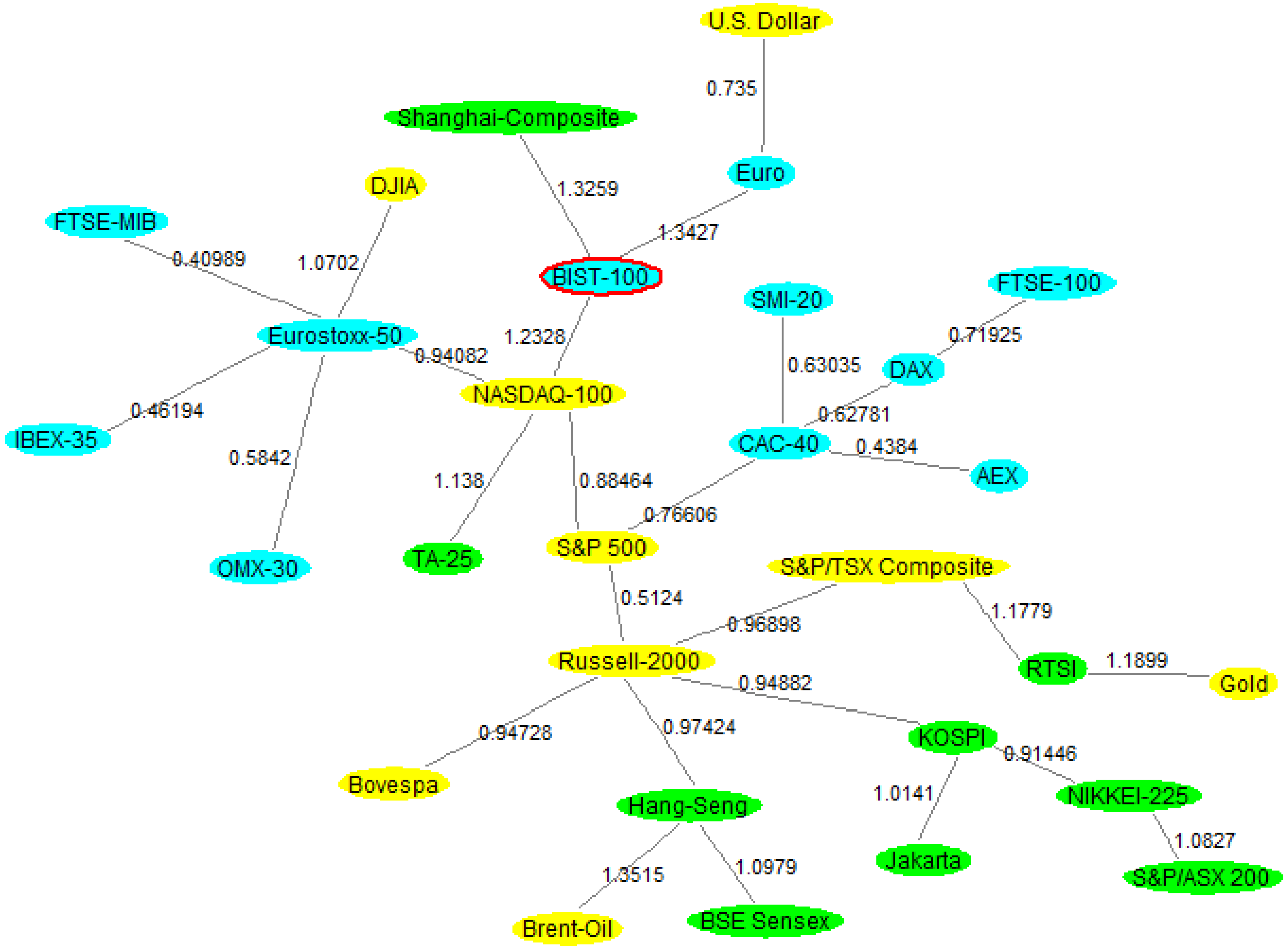


Figure Appendix E.2. MST of global financial assets for 2004.

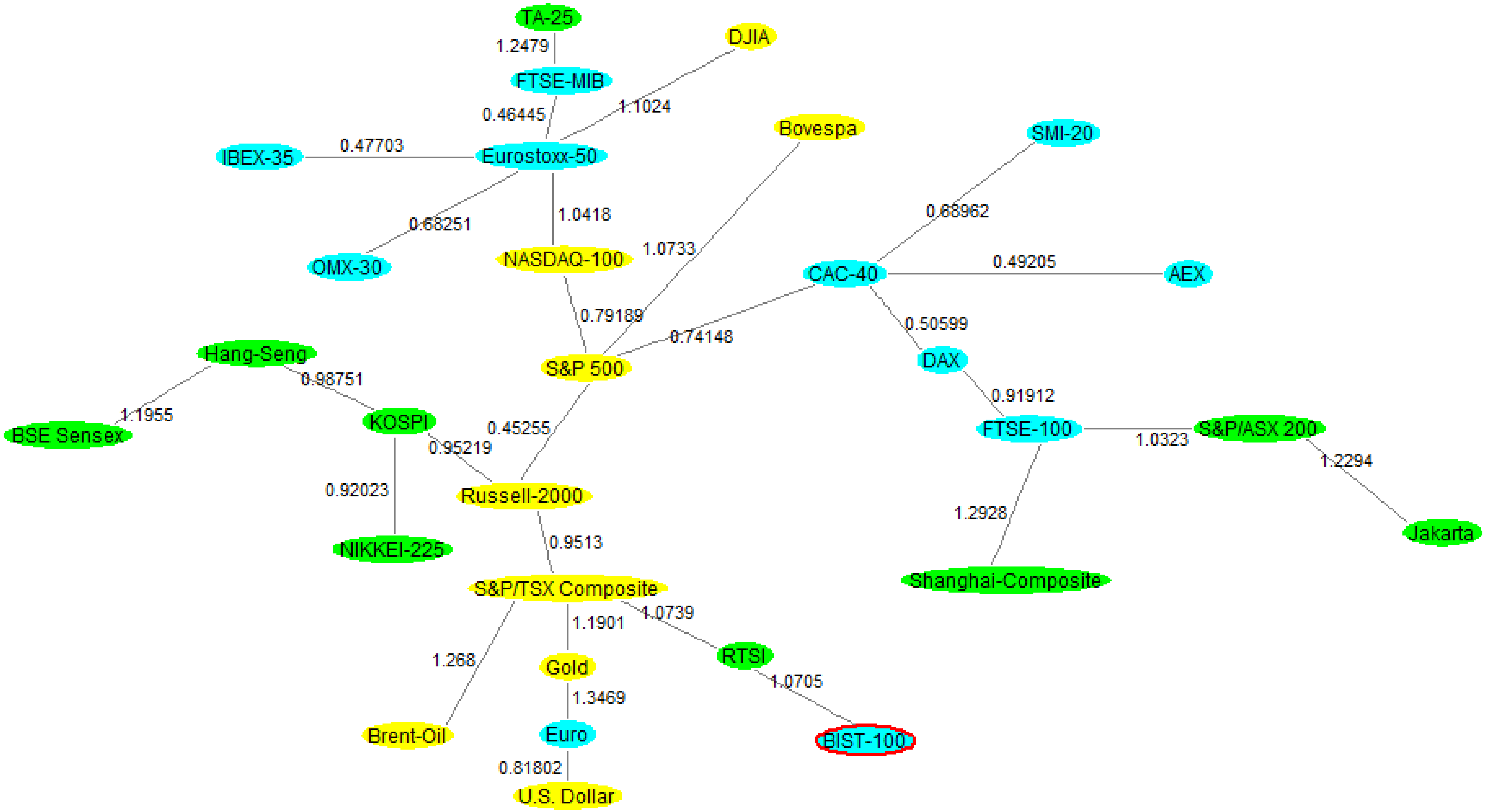


Figure Appendix E.3. MST of global financial assets for 2005.

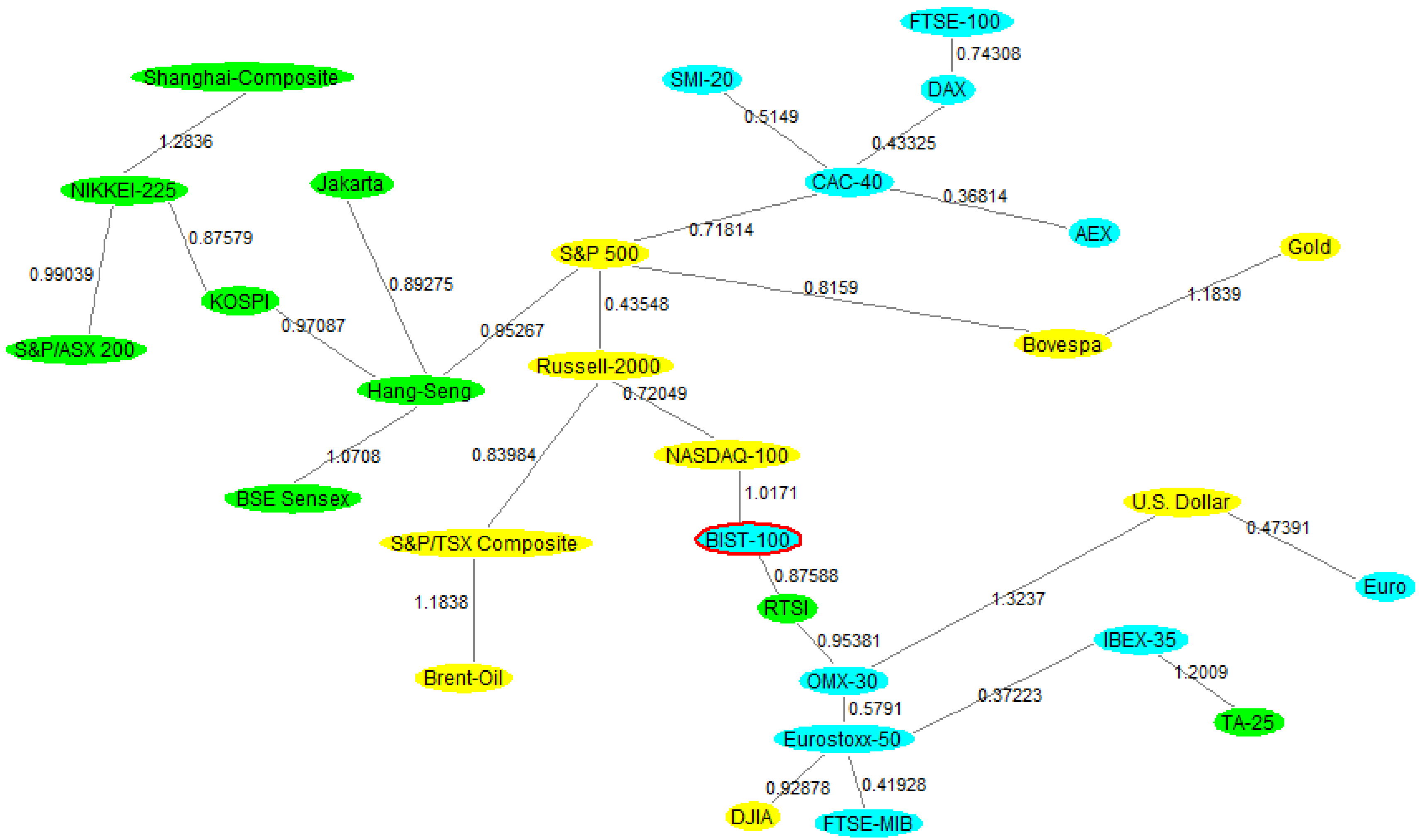


Figure Appendix E.4. MST of global financial assets for 2006.

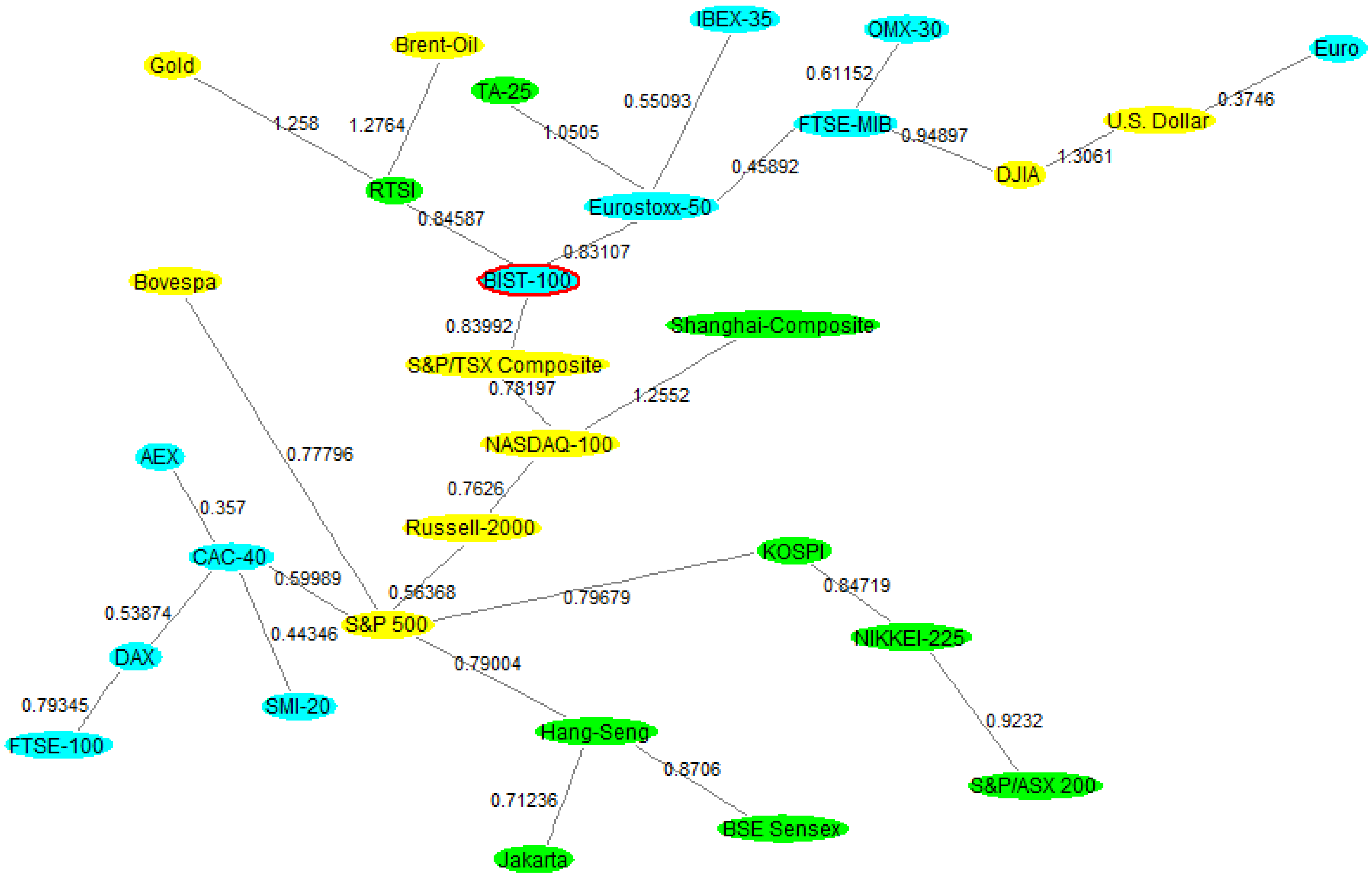


Figure Appendix E.5. MST of global financial assets for 2007.

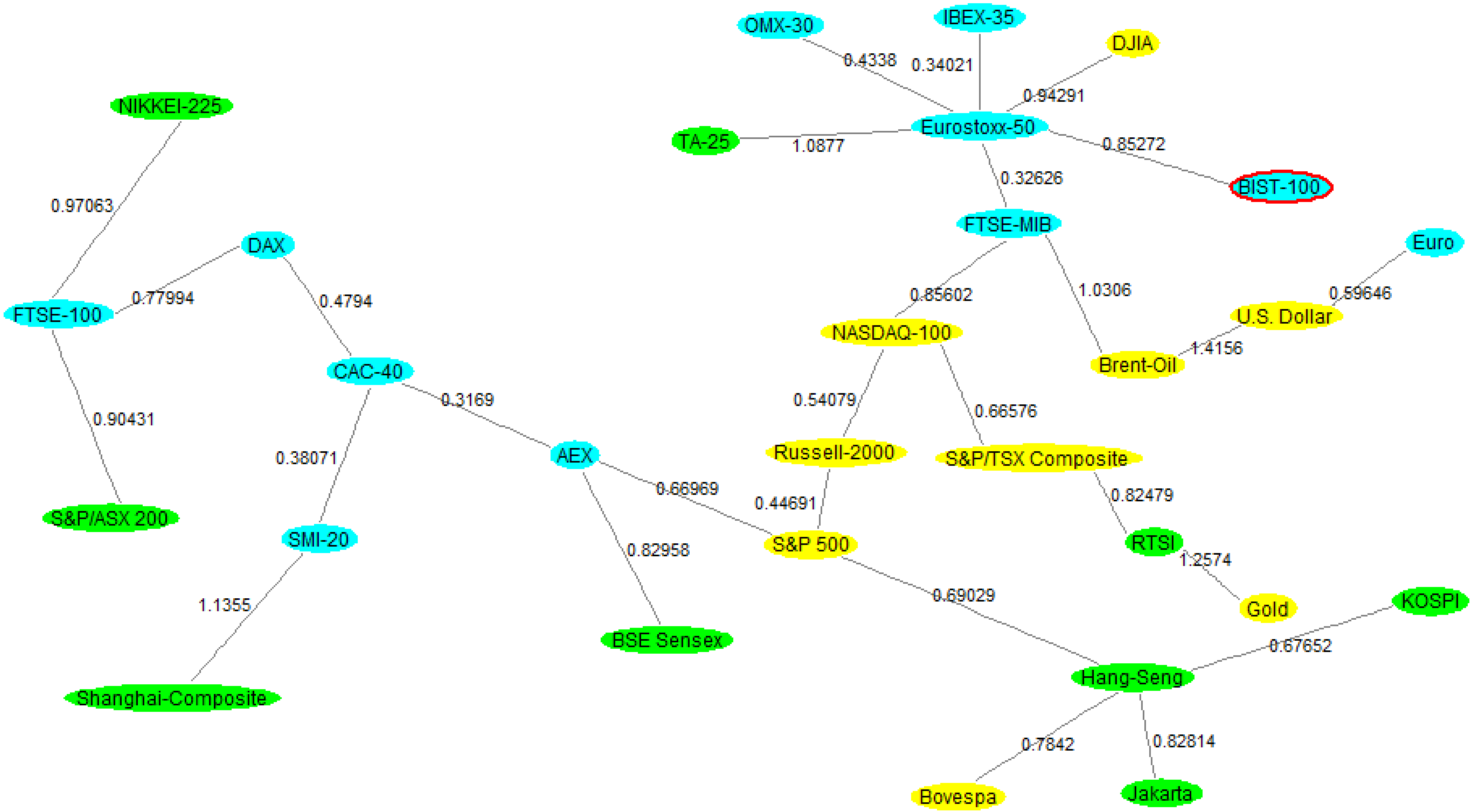


Figure Appendix E.6. MST of global financial assets for 2008.

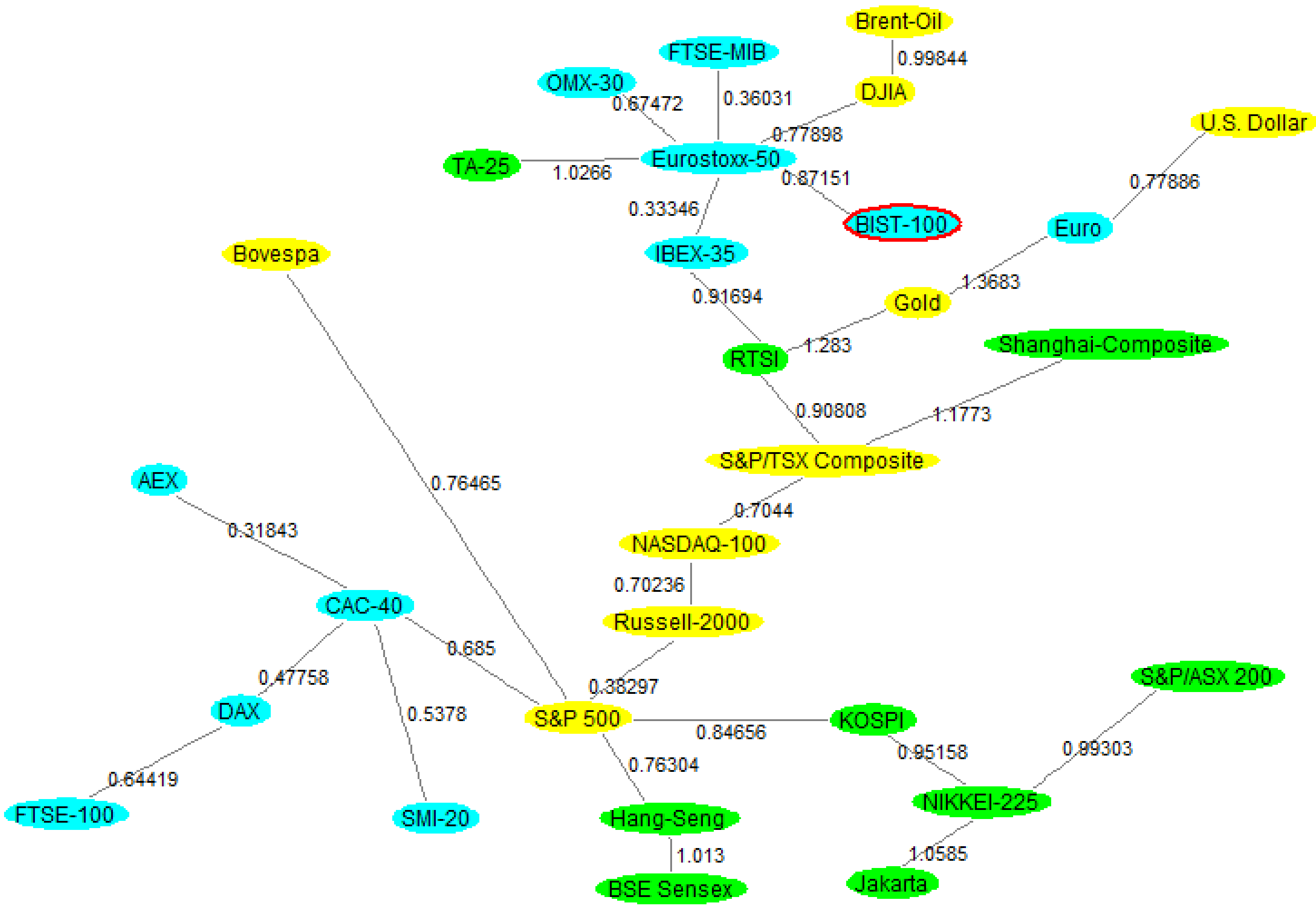


Figure Appendix E.7. MST of global financial assets for 2009.

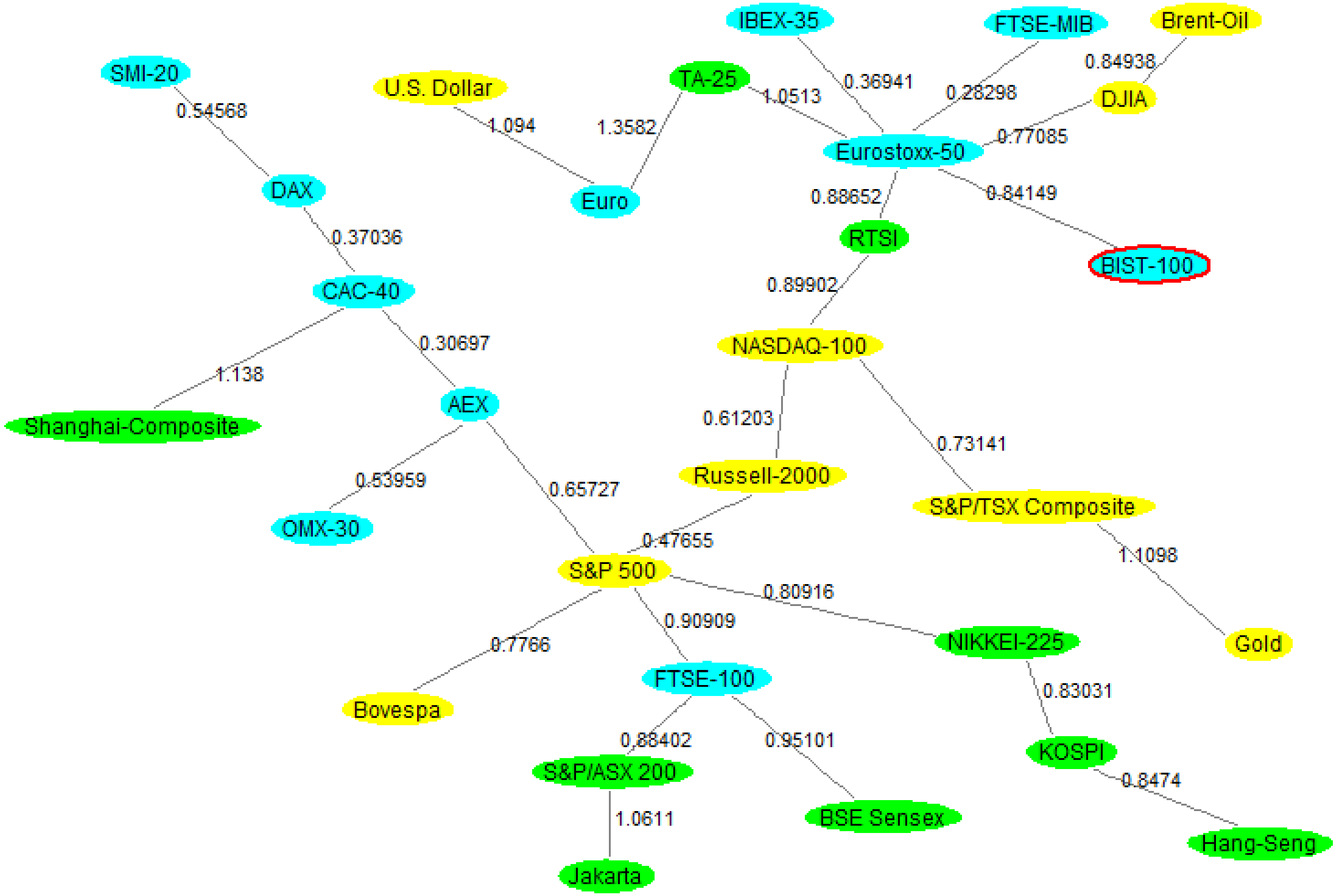


Figure Appendix E.8. MST of global financial assets for 2010.

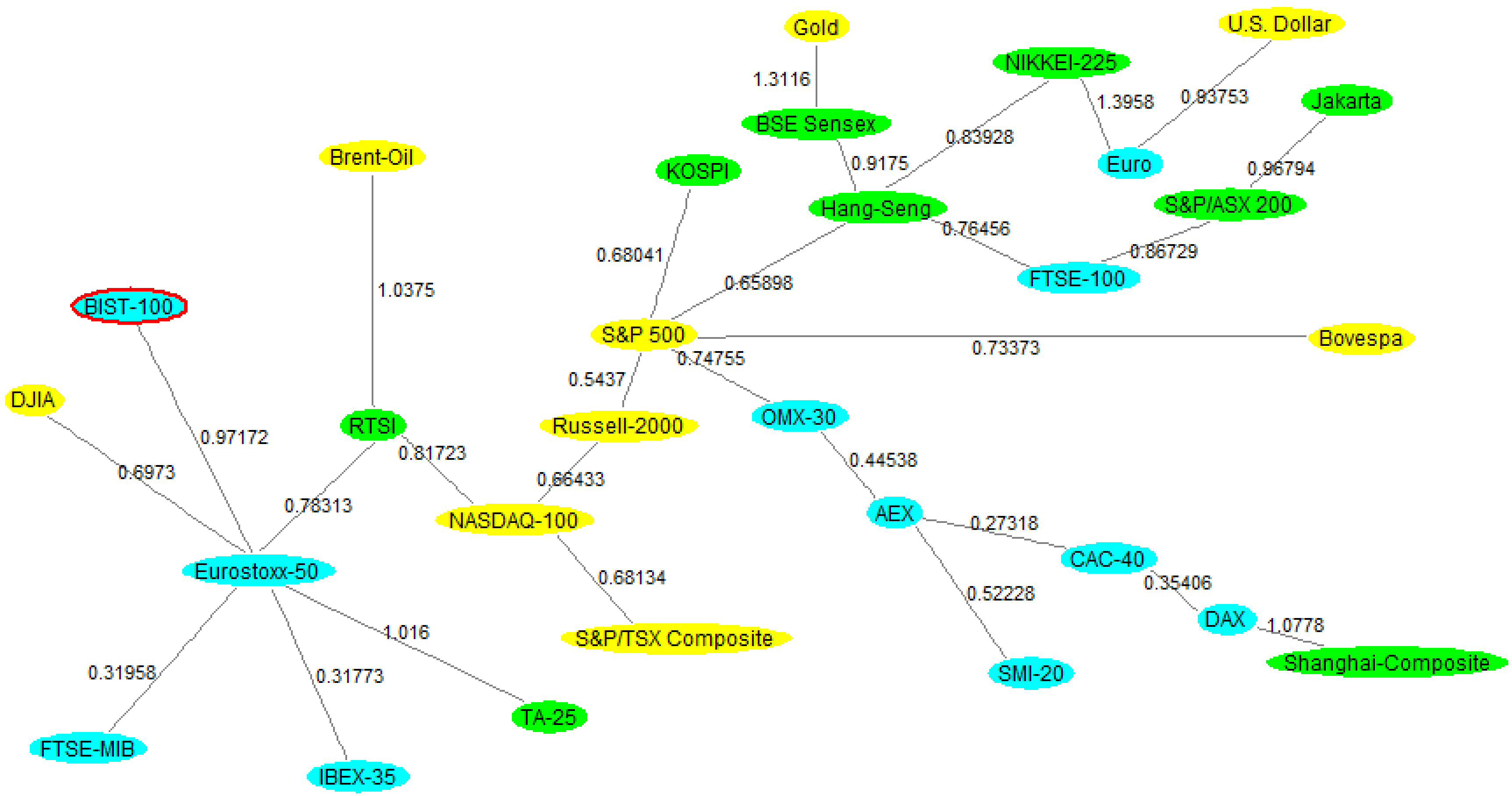


Figure Appendix E.9. MST of global financial assets for 2011.

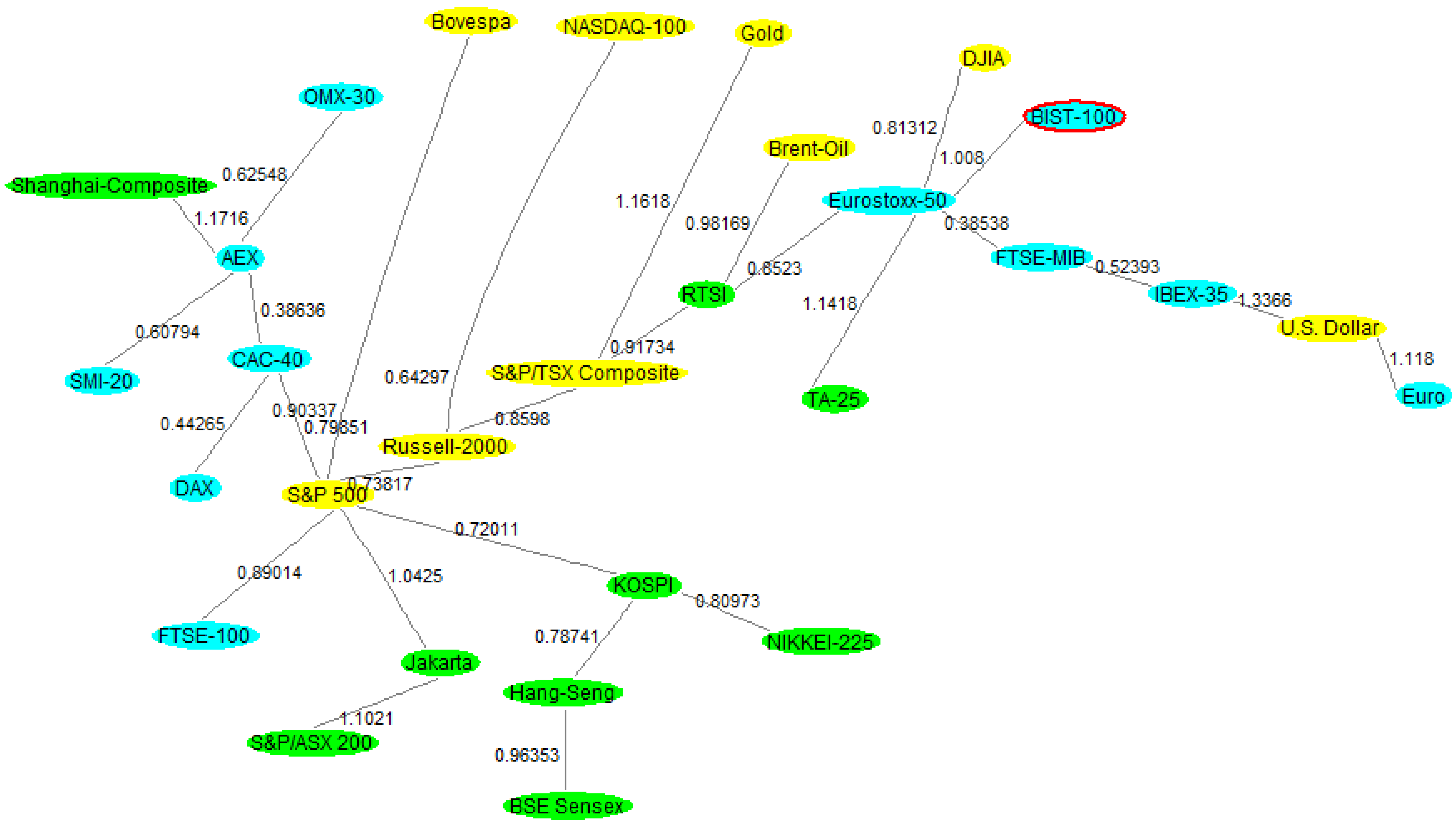
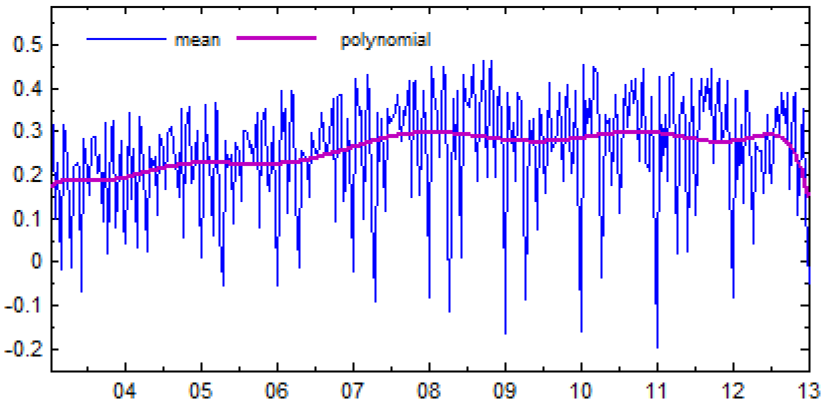
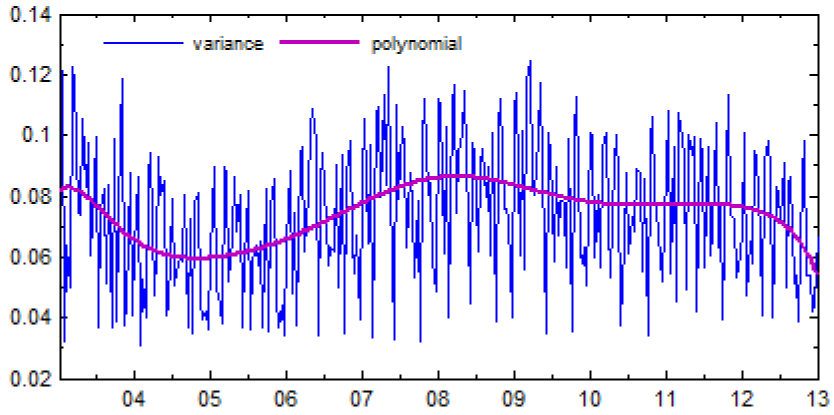


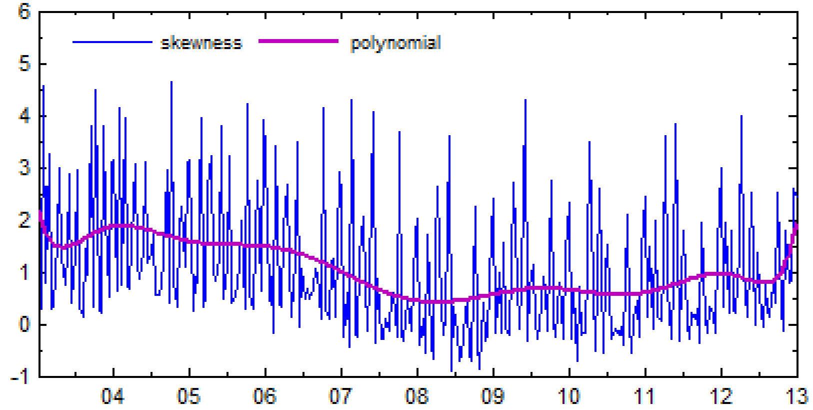
Figure Appendix E.10. MST of global financial assets for 2012.



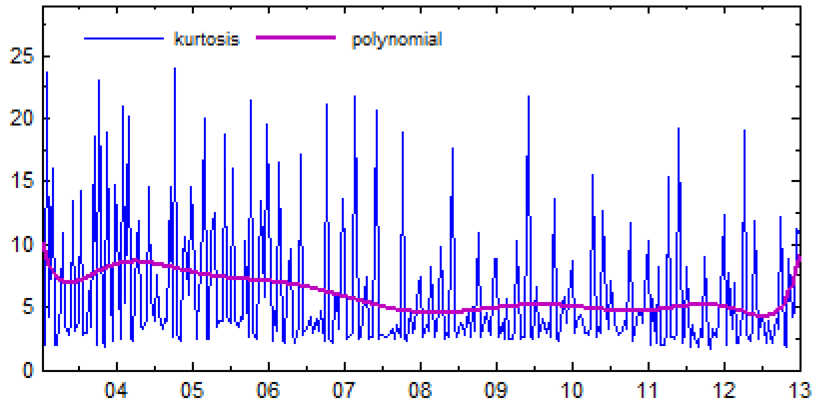
a) Mean of correlation coefficients.



b) Variance of correlation coefficients.



c) Skewness of correlation coefficients.



d) Kurtosis of correlation coefficients.

Figure Appendix E.11. Statistical moments of correlations coefficients of global financial assets for 2003–2013.

RESUME

Yusuf Yargı BAYDİLLİ was born in 1986 and he graduated from first and elementary education in Istanbul. He completed high school education in Üsküdar High School, after that, he started undergraduate program in Abant İzzet Baysal University Department of Physics in 2004. Then in 2010, he started assignment as a Research Assistant in Hakkari University Department of Computer Engineering. In order to complete M. Sc. education, he was charged to Karabük University where he has been still working as a R. A. for.

CONTACT INFORMATION

Address: Karabük University
Graduate School of Natural & Applied Science
Demir-Çelik Campus/KARABUK

E-mail: yusufbaydilli@karabuk.edu.tr; yusufyargibaydilli@hakkari.edu.tr