European Online Journal of Natural and Social Sciences 2013;
vol.2, No. 3(s), pp. 339-347 ISSN 1805-3602

Identification and clustering of effective industrial at Tehran stock market based on minimum spanning tree of distance ultra-metric

Darush Farid[1], Ahmad Farsoudeh[2]

[1]Department of Management, Yazd University, Iran; [2]MA student in Financial Management, Yazd University, Iran

Abstract

In essence, the stock market is a complex network. The price fluctuations among various stocks have complicated relationships. In this paper, the Minimum spanning tree (MST) is used in order to identify, categorize, and evaluate the relationship between different industries in Tehran Stock Exchange in 2012, 35 industries were addressed. Thus, the high degree repetition of cross-correlation among different pair indicators of industries investigated in Tehran stock market, the result is a hierarchical graph based-on metric distance between different pair industries. In following based-on ultra-metric function resulted from metric, a Minimum spanning tree exists from different industries, which showed relation among industries and their influences on each other. The present study in terms of aim of the research is applicable and in terms of the nature of the research, is Analysis –inference. In this paper, in order to analyze the data and obtain a Minimum spanning tree, Excel, Eviews and Matlab are used. Industries are divided to three categories according to the Minimum spanning tree obtained, based on the minimum distance and the effect, are: The first one are industries with five top-grade in the Minimum spanning tree, The second group of industries that have Vertices of degree four in the Minimum spanning tree and The third group of industries that have Vertex third degree in the Minimum spanning tree. The first category are the most effective industries that include Basic Metals, cement, lime and chalk.

Keywords: Correlation, Minimum spanning tree (MST), Metric distance, Ultra metric distance, Tehran Stock Exchange

Introduction

In his book *Index Funds*, Hebner wrote that there is a dark secret in the financial services industry to convince people that somehow it is possible to beat the market. It was first pointed out by Louis Bachelier that the market has followed a random walk for about one hundred years. Since Bachelier's pioneering work, there have been hundreds of academic studies to show that the market has been highly efficient and it is extremely difficult to outperform the market. However, most of the investors do not want to face this disappointing and cold fact (Hebner, 2007). On the other hand, in the stock market, we can clearly see the risk and return interact and achieve a fruitful outcome. With no doubt, the most valuable commodity in this market is information. Information access is the key to profitability and uses of lucrative opportunities in capital markets (Tehrani, Talbnia & Jalili, 2007). There is a high degree of correlation among different industries in financial market simultaneously viewed as an empirical fact (Markowitz, 1959; Elton & Gruber, 1995; Campbell, Lo, & MacKinlay, 1997). Repetition of the correlation among different pairs of industries in Tehran Stock Exchange makes accurate economic forecasting of future industry to face with problems. The sampling period for the covariance or correlation coefficient among the time series of daily returns of different industries in Tehran Stock Exchange, in 2012 is considered. The process of identifying and classifying industries of independent or dependent economy could improve predictions of economic performance and its impact on the industry to other industries. Most common methods for homogeneous categories are analysis key elements of the correlation matrix of the data. Simultaneous classification of financial data sampled at a fixed time horizon with the use of data Correlations, called category Knowledge of financial data. Recently

Corresponding author: Darush Farid, Department of Management, Yazd University, Iran. Email: darush.farid@gmail.com

this method used to find economic data in a correlation matrix with a filtering method based on an ultra-metric function estimate related to metric distance a group of industries have been proposed. Metric distance can be obtained from the correlation among indexes. In fact, a metric distance, the size of path between two vertices in a graph is adjacent. Minimum spanning tree (MST), which is the minimum distance between two adjacent vertices of the tree obtained by minimizing the overall correlation matrix is calculated, thereby can be affected industries and key industries on other industries identified and categorized; the trees often show the effects on each other. However, with increasing or decreasing time horizon, correlation of changes and subsequent a hierarchical graph can be changed. The paper is composed of the following sections: In Section (4), the used data, the correlation and coefficient matrix will be discussed. In section (5), method based on distance metric and subsequent distance ultra-metric will be checked. In section (6), the Minimum spanning tree, identification and classification of effective industries are discussed and at the end will be expressed a summary of the article.

Review of Literature

Several studies on the correlation among financial markets in different countries have been conducted. In 2003, Miccichè *et al*. examined the degree of stability of minimum spanning tree, return and price fluctuations. By using minimum spanning tree, they studied two time series 1) Return on assets 2) their fluctuation. This study showed that minimum spanning tree of return on assets, were described, with using Value of shares that are constant over time. On The other hand, their analysis showed a very slow dynamics in both cases in a time for several years (Salvatore *et al*, 2003).

Thus, in 2004, Skórnik-Pokarowska and Orłowski (2004), use ultra-metric distance to classify the portfolio. They, calculated, ultra metric distance between any pair of share in the same portfolio and they use it for classifying shares to related groups .The research showed that the best strategy for great investment portfolio can be classified portfolio groups which are called Constant Rebalanced Portfolio.

After that in 2006, Jung *et al*. (2006) studied the correlation configuration of stock market of Korea. They studied network structure of stock using tree-Global Minimum and the correlation matrix investigated and found that the stock market of Korea cannot be classified as commercial or industrial groups. When the index MSCI[1] is profitable these categories formed.

In 2007, the research done by Coelho, Gilmore, Lucy, Richmond, Hutzleron (2007), studied the evolution of the dependence among the World Securities Exchange by using of the minimum spanning tree was performed and for this goal they used the international stock index during 1997- 2006. They showed extreme dependence international exchange on each other. Outcome of this study is reducing the potential profit for foreign investors in the portfolio included international exchange .

The researches of Gilmore, Lucy and Boscia in 2008, showed the evolution of exchange Europe by using minimum spanning tree that studied the index of the Stock Exchange for the period 1999-2006 in 21 European countries. The results revealed that the French Stock Exchange is the most important market according minimum spanning tree .

In 2009, Mehmet Eryiğit and Resul Eryiğit (2009) studied cross correlation network structure in the world stock market in dices, using Minimum spanning tree. They studied cross correlation of changes in 143 daily and weekly stocks from 59countries. Asset graph, MST and PMFG[2] analysis shows increase on Correlations among exchanges in recent years. This study showed that North American and European market shave highest correlations among themselves .

However, other foreign researchers have done studies by using this method in other fields that are outside of this research paper and we do not refer them. The purpose of this study is to identify and clustering effective industries in the Tehran Stock Exchange.

The research hypothesis

Considering the above and earlier Subject, the following research hypotheses are examined:

Among available industries in the Tehran Stock Exchange based on minimum spanning tree ultra-metric distance, some effective industries exist.

Methodology

The present study in terms of aim of the research is applicable and in terms of the nature of the research, is Analysis –inference. Research information and data; also have been collected in two methods library studies and field studies. In Section of library studies we used relevant resources, including books and specialized articles, magazines and other internal and external Journals related to subject. And in section of field studies to collect information and data related to present industries in the Tehran Stock Exchange.

[1]Morgan Stanley Capital International Inc

[2]planar maximally filtered graph

For this purpose the reports of Stock Exchange and the websites have been used. For this study, the correlation among pairs of different industries in Tehran stock exchange is calculated according to the formula defined. A matrix of 35 x35 with one core diameter can be produced. Then by using of the defined metric distance, distance among these industries has been calculated, which its result is one complete graph hierarchy among pairs of different industries. Subsequently by using of Function ultra metric obtained from the metric, a minimum spanning tree is created from various industries, which show relationship among industries and their impact on each other.

Description of the data

In this paper, data of daily returns of 35 industries in Tehran Stock Exchange are studied. However, there were other industries, because their symbol is stopped more than half the year, they were eliminated (Such as measuring instruments, medical and optical, other transportation, hotels and restaurants, publishing, printing and proliferation, quarrying other mining and wood products). The Name of the industries under study, the sign industry, Sampling period and sample size, completely expressed in table (1).This information has been obtained from the site of the Tehran Stock Exchange.

Table 1. Demographic information about the sample

Industry Name	Sign industry	Sampling period	Sample size
1-Banks,credit institutions and other monetary institutions and financial	A1	24/03/2012-24/03/2013	239
2-Basic Metals	A2	24/03/2012-24/03/2013	239
3-Chemical products	A3	24/03/2012-24/03/2013	239
4-Construction Devices and communications equipment	A4	24/03/2012-24/03/2013	239
5-Post and Telecommunications	A5	24/03/2012-24/03/2013	239
6 –Industrial Conglomerates	A6	24/03/2012-24/03/2013	239
7-Mining of metallic ores	A7	24/03/2012-24/03/2013	239
8-Vehicle and Parts Manufacturing	A8	24/03/2012-24/03/2013	239
9-Production of coke, refined petroleum and nuclear fuel	A9	24/03/2012-24/03/2013	239
10-Investment	A10	24/03/2012-24/03/2013	239
11-Cement, lime and chalk	A11	24/03/2012-24/03/2013	239
12–Pharma	A12	24/03/2012-24/03/2013	239
13-Computer and related activities	A13	24/03/2012-24/03/2013	239
14-Mass Construction, Real Estate	A14	24/03/2012-24/03/2013	239
15-Technical and engineering services	A15	24/03/2012-24/03/2013	239
16-Food and beverage products except sugar	A16	24/03/2012-24/03/2013	239
17- Transport, warehousing and communication	A17	24/03/2012-24/03/2013	239
18- Other financial mediation	A18	24/03/2012-24/03/2013	239
19-Insurance and Retirement	A19	24/03/2012-24/03/2013	239
20–Sugar	A20	24/03/2012-24/03/2013	239
21-Electrical machineries and equipment	A21	24/03/2012-24/03/2013	239
22-oil extraction except discovery	A22	24/03/2012-24/03/2013	239
23- Machinery and Equipment	A23	24/03/2012-24/03/2013	239
24-Rubber and Plastics	A24	24/03/2012-24/03/2013	239
25- Other non-metallic mineral products	A25	24/03/2012-24/03/2013	239
26-Tiles and Ceramics	A26	24/03/2012-24/03/2013	239
27-Manufacture of fabricated Metallic	A27	24/03/2012-24/03/2013	239
28-Industrial contracts	A28	24/03/2012-24/03/2013	239
29–Financial and monetary intermediation	A29	24/03/2012-24/03/2013	239
30-Tanning and leather polishing and manufacturing of foot gear	A30	24/03/2012-24/03/2013	239
31-Agriculture and Related Services	A31	24/03/2012-24/03/2013	239
32– Textiles	A32	24/03/2012-24/03/2013	239
33-Paper Products	A33	24/03/2012-24/03/2013	239
34-Furniture and other artifacts	A34	24/03/2012-24/03/2013	239
35-extraction of coal	A35	24/03/2012-24/03/2013	239

Reference: www.irbourse.com

Correlation matrix was examined within a framework of random matrix theory by physicists (Laloux *et al.*,1999; Plerou *et al.*, 1999). The correlation coefficient is defined as follows:

$$\rho_{ij}(\Delta t) \equiv \frac{<Y_iY_j> - <Y_i><Y_j>}{\sqrt{(<Y_i^2> - <Y_i>^2)(<Y_j^2> - <Y_j>^2)}}$$

Where "i" and "j" are specific name for a pair of industries, and Y_i and Y_j are value of daily returns for each pair of industries. Statistical average denoted by<.>is displayed.ρ_{ij} Value is located between [1,-1]. When.ρ_{ij} = -1, i.e. industry pairs are negatively correlated with each other, and when.ρ_{ij} = 1, i.e. the pair of industries positively correlated with each other, and when.ρ_{ij} 0, i.e. the pairs of industries considered are uncorrelated, in other words a change in one industry has no effect on another. The correlation coefficient matrix is asymmetric matrixn * n, with the core diameter.ρ_{ij} = 1.

There is $\binom{n}{2} = n(n-1)/2$ a correlation between a pair of industries, which in this paper; there are 595 the correlation between different industries. By using of the obtained data, the minimum, and maximum and average correlation coefficient among pairs of industries return index has been calculated, which the graphics expressed as follows:

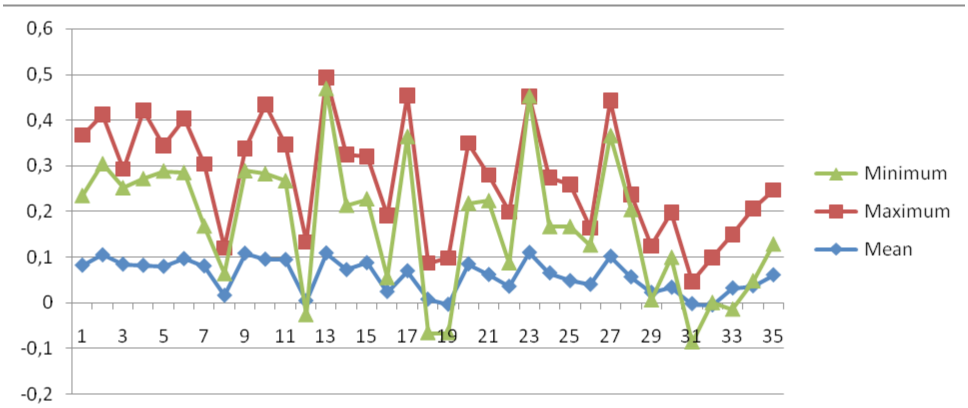

Figure 1. Comparison of minimum, maximum and average

According to the above figure, the regression line in the maximum correlation coefficient is descending. This means that with increasing correlation coefficient, Slope of the line decreases. As the graph shows industries that are at the end of the maximum regression line have fewer correlations with each other. The impact this industries on each other is low, which It will increase the accuracy of the economic forecast. Overall, maximum correlation coefficient is high, which it represents high correlation and low distance these industries from each other. The high correlation among the industries will reduce the accuracy of the economic forecast. On the other hand, average correlation coefficient is close to the maximum coefficient, which shows a decrease in the accuracy of the economic forecast. Also, minimum coefficients have greater volatility than the maximum coefficients. On the other hand, the maximum coefficients in 35 industries are close to1. But, minimum coefficients are not close to-1, and have more volatility. This graph clearly shows what has been presented. This graph shows high correlation and low distance of these industries compared to each other which smallest in attention to it makes defect in economic forecasting and subsequently in the future performance of the industries, and creates irreparable losses. A decrease in economic forecasts leads to difficulty in good analysis of future performance of the industries to be able to invest in these industries with less risk.

Metric-based method

Metric is used for calculating the distance among pairs of financial time series with great precision (Bonanno *et al.*, 2001). This is divided into two categories: 1- financial time series with unequal size 2- financial time series with equal size.

Calculating distance for time series within un-equal size

Assume that $\{X_t, t=1,2,...,n_x\}$ and $\{Y_t, t=1,2,...,n_y\}$ are two financial time series with unequal size ($n_x \neq n_y$). We define the Period gram to the mass follows:

$$P_x(\omega_j) = \frac{1}{n_x}\left|\sum_{t=1}^{n_x} x_t e^{-it\omega_j}\right|^2$$

In Which $\omega_j = 2\pi j/n_x$ foreach $j = 1, ..., m_x$ with $m_x = [n_x/2]$ is defined. Thus the Periodo gram for Y also is defined as above. Wang and Bloste into calculate the distance between two time series with unequal sizes offered following method:

1) Their assumption was based on this factor: the size of series y_t is less than of the series $x_{t.,}$, So to overcome this problem, defined yt as follows:

$$y'_t = \begin{cases} y_t, & t = 1, ..., n_y \\ 0, & t = n_y + 1, ..., n_x, \end{cases}$$

2) Then they defined the Periodo gram $\acute{y}$ as follows:

$$P_{y'}(\omega_j) = \frac{1}{n_x}\left|\sum_{t=1}^{n_x} y'_t e^{-it\omega_j}\right|^2$$

And they called it Zero-padding Periodo gram.

3) At the end, to calculate the distance between the two series, they introduced following equation:

$$d_{ZP}(x,y) = \sqrt{\frac{1}{m_x}\sum_{j=1}^{m_x}\left[P_x(\omega_j) - P_{y'}(\omega_j)\right]^2}$$

This method for calculating the distance between two financial time series with unequal size is frequently used (Caiado *et al.*, 2007).

Calculating the distance for time series with equal size

To calculate the distance between two time series with equal size, the correlation coefficient is used. This metric distance can be calculated with great precision by the following equation (Gower, 1966).

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

This method has three major advantages:

1) It uses the information about auto correlation structure between the time series.

2) It can be used to compare type of movement of the time series related to each other and categories with equal size.

3) For calculating them, information about the fluctuation series relative to each other is not required.

In this paper, the metric with equal size is used to identify and classify industries.

Characteristics of Metrics

Given that X is a non-empty arbitrary set, d: $x * x \longrightarrow R$, is called a metric or meter on x if it has the following properties:

1) For every $i, j \in x$ have $d_j(\Delta t) \geq 0$

2) For every $i, j \in x$ have $d_j(\Delta t) = 0$ If and only if $i = j$.

3) For every $i, j \in x$ have $d_i(\Delta t) = d_j(\Delta t)$.

4) For every $i, j, k \in x$ have $d_j(\Delta t) \leq d_k(\Delta t) + d_k(\Delta t)$. (Triangle in equality).

Distance metric is calculated by using of the correlation coefficients obtained in Section 4. Then distance matrix D(Δ*t*) is formed for it. Because the information of 35 industries is given in Tehran Stock Exchange in this paper, there will be a 35 × 35 distance matrix. By using the Eviews software distance graphs for each of 35 industries by considering the series obtained from metric space which is plotted. These graphs respectively according to (Table 1) are as follows:

The above graphs for each industry compared to other 34 industries are drawn. For example, the first chart from the top left is the comparing of the distance of the first industry, banks and credit institutions and other financial and monetary institutions(using the data described in Table 1) with a distance of every other 34 industries, which is plotted above, and respectively the rest of the graphs as follows are drawn, that represents the distance of each of industry with other 34 industries, i.e. respectively first graph presents the first industry in Table1, the second graph for the second industry in Table1and until the latest graph for the latest industry are presented in Table 1. According to the equation metric distance, it can be seen that the distance is located between [2, 0]. When the correlation coefficient is 1, the distance is zero, and the correlation coefficient is -1, this distance is 2. Thus when the correlation coefficient is maximum, the distance is minimum, and when the correlation coefficient is minimum the distance is the maximum amount, Thus, correlation coefficient has an inverse relationship to the metric distance. According to the above graphs that have yellow color, it can be seen that the industries number twelve, eighteen, nineteen, thirty-one, thirty-two and thirty-three of the A12, A18, A19, A31, A32 and A33 have the greatest distance from other industries. Because these industries have maximum distance from other industries, they will have lowest correlation with other industries, which their functions have no effect on other industries and power Forecast of their future performance, will have significantly high precision. On the other hand, the red graphs above shows industries with the minimum distance from other industries, which it will be explained in the next section about them.

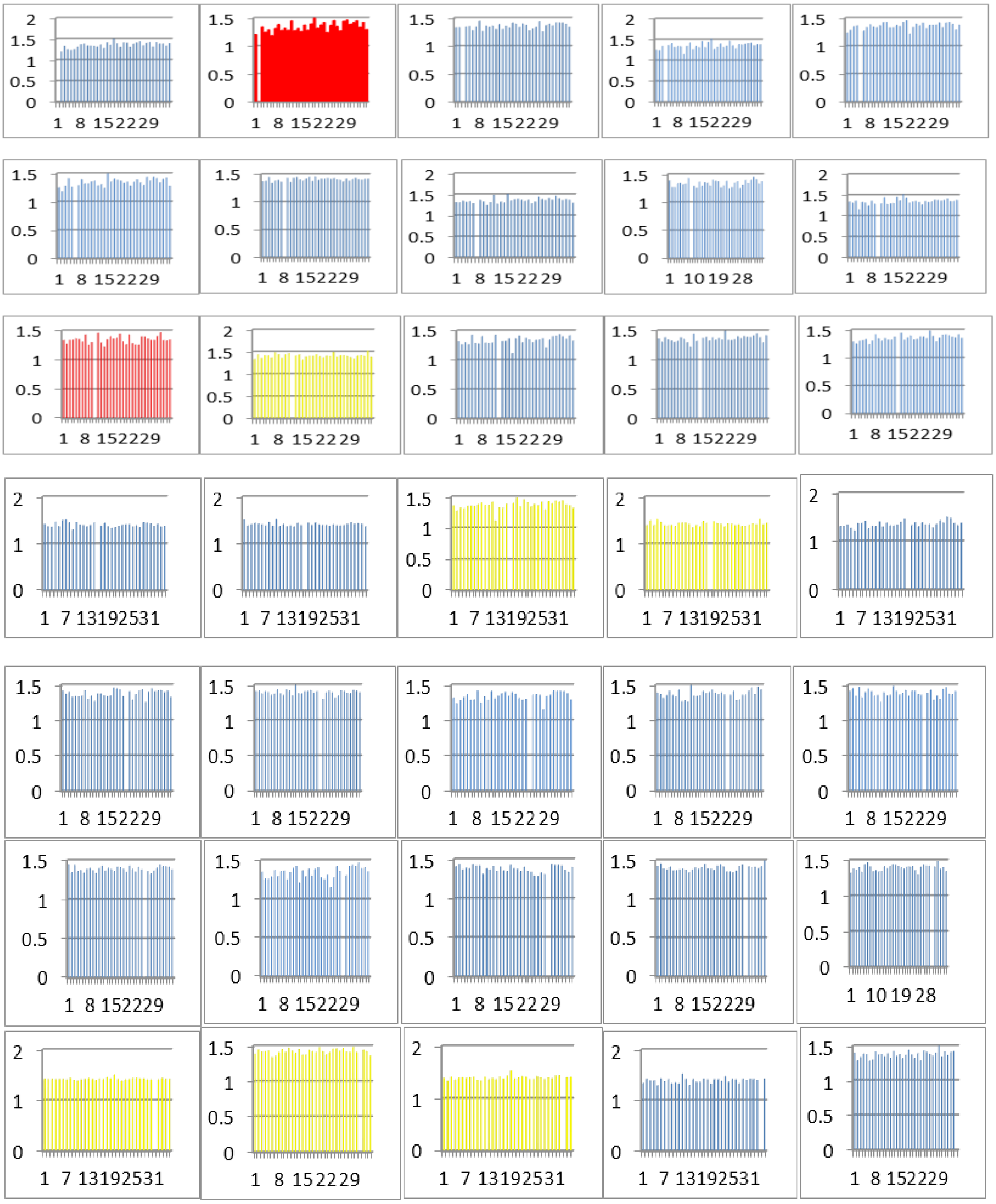

Figure 2. Metric distance graphs

Ultra-metric distance

The knowledge of a distance function makes it possible to, without any prior knowledge of specific groups; decompose a set of n objects into subsets of closely related objects (clusters). Cluster analysis is a common technique in multivariate data analysis and it can be applied in various ways (Johnson, 1998). An ultra-metric space is a space in which the distance among objects is an ultra-metric distance. An ultra-metric distance $d^{<}_{ij}$ must satisfy the first, two and three properties of a metric distance:

1) For every $i, j \in x$ have $d_{ij}(\Delta t) \geq 0$.

2) For every $i, j \in x$ have $d_{ij}(\Delta t) = 0$ If and only if $i = j$.

3) For every $i, j \in x$ have $d_{ij}(\Delta t) = d_{ji}(\Delta t)$.

While the usual triangular inequality (four properties)

4) For every $i, j, k \in x$ have $d_{ij}(\Delta t) \leq d_{ik}(\Delta t) + d_{kj}(\Delta t)$. (Triangle inequality).

is replaced by a stronger inequality, called an ultra-metric inequality:

$$d^{<}_{ij} \leq \max\left[d_{ik}, d_{kj}\right]$$

The above equation puts an even tighter constrain, in an ultra-metric space, the distance between the points i and j is always less than or equal to the maximum of: the distance between i and any other point k and the distance between j and any other point k (Rammal *et al.*, 1986). One of the easiest ways of performing the clustering and finding the ultrametric distances is to obtain the Minimal-Spanning Tree (MST) from the metric distances that link together the objects to be clustered.

Minimum spanning tree (MST)

Since the distance matrix D (t) is a complete graph, it is used to determine the minimum spanning tree. The MST or minimum spanning tree is a theoretical concept from graph theory. A minimum spanning tree of a graph, is a sub graph that, Firstly: it has been tree, Secondly: it contains all the vertices. Minimum spanning tree creates one unique method named matrix ultra metric function. Matrix ultra-metric functions is showed with D <(t)that it will be used for the identification and classification industries. Ultra metric function of the distance between i and j of the matrix $D^{<}(t)$ with $d^{<}_{i,j}$ is shown, from minimum distance metric distance $d_{L,K}$ is obtained from i to j (Mantegna & Stanley, 2000). According to the ultra-metric function matrix obtained in Section 5, the graph minimum spanning tree is drawn, which is as follows:

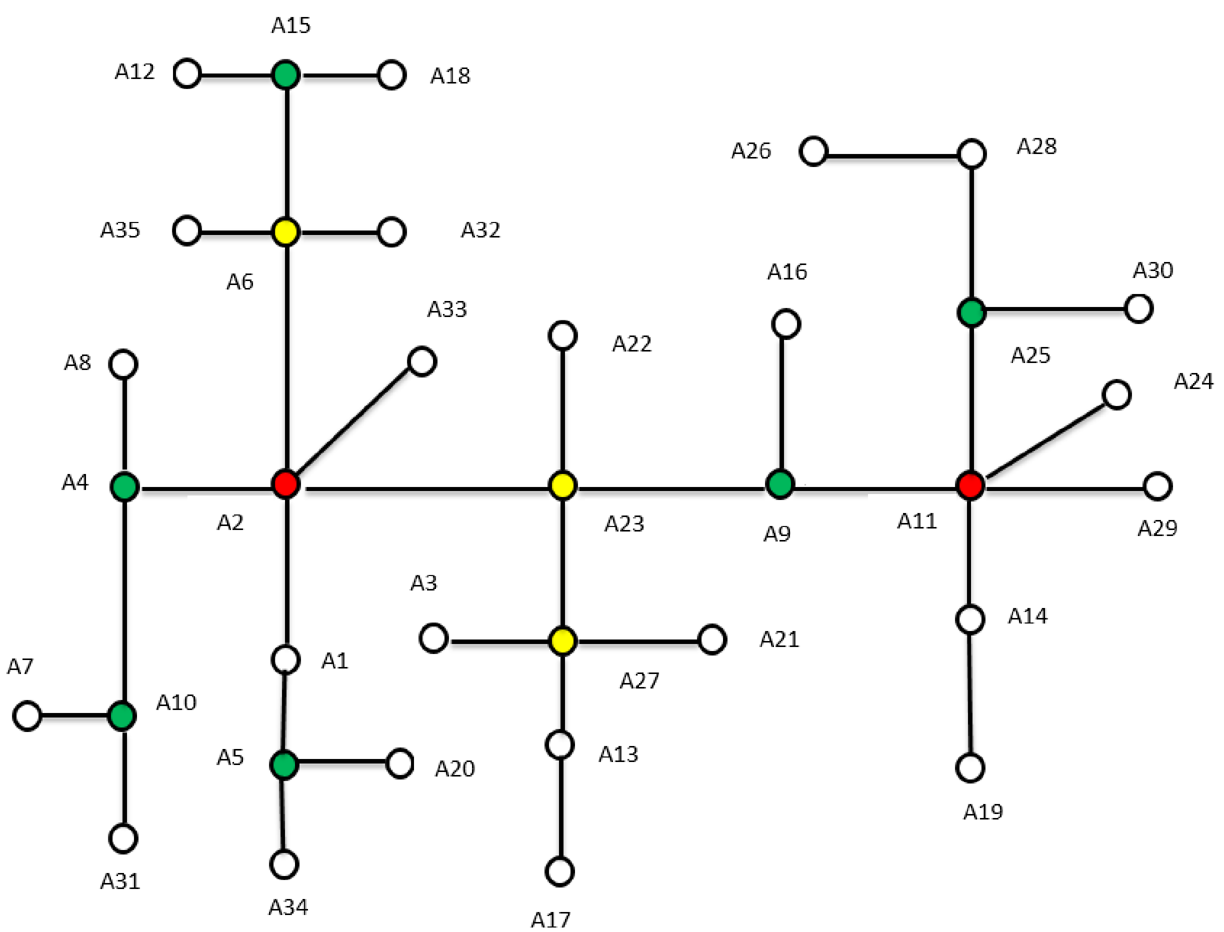

Figure 3. The graph of minimum spanning tree

According to the minimum spanning tree depicted (Notably, this tree is obtained through coding Matlab software), degree five is the highest degree of the industry, which is in the center of other five industries, which is shown in red color. These industries have a minimum distance and maximum correlation with their related industries, which the smallest change in them makes changes in the related industries. These industries that are located in the center and have highest degree are called effective industries their changes can cause changes in adjacent industries and then this wave makes changes in other industries. These industries can also be classified, in this way, that all related industries to them are minimum distance. As it can be seen in the graphs in Section 5, the two graphs are plotted in red, which compared to other industries have minimum distance. It means that they have the maximum correlation with other industries. These industries are located exactly in the center of the minimum spanning tree, which are marked in red. These industries are the most effective industries among industries. By using the minimum spanning tree and the minimum calculated sum of degrees, Industries are classified as follow:

The first category is marked in red. These industries have vertices of degree five and with the sum of minimum distance, whose effectiveness are: 1 -Basic Metals, 2-cement, lime, and chalk. The second category marked in yellow are industries that have vertices of degree four are in the minimum spanning tree. Their sum of minimum distance and effectiveness include: 1-Machinery and Equipment, 2-metal products and 3 –An Industrial Conglomerates and the third category is shown in green; Industries that have third degree vertex in the minimum spanning tree. Their sum of minimum distance and effectiveness are: 1 -petroleum products, coke and nuclear fuel 2- Investment 3 -technical and engineering services 4-Construction Devices and communications equipment 5-Posts and Telecommunications 6-Manufacture of other non-metallic mineral. According to the above categories, when changes occur in the return basic metals industry, the industry affects in return five surrounding industries, namely: financial and monetary banks and credit institutes, Machines and the equipment, Industrial Conglomerates, paper products and Construction Devices and communications equipment. The change in their returns respectively causes a change in the returns related industries and the subsequently the wave are distributed across the minimum spanning tree. Thus, identification of those effective industries in stock market leads to the accuracy of their future economic forecast and their influence on other industries to be more important to be able to make more accurate forecast from their future economy to more accurate decisions about them.

Conclusions

High Repetition of the correlation between different pairs of industries in Tehran Stock Exchange makes accurate economic forecasting of future industry is facie problems. In this paper for solving this problem we use of ultra-metric function and minimum spanning tree, industries which are located in the center of this graph and have the highest degree are part of more effective industries in stock market which every change in their returns lead to change in their related industries. Identification and clustering these effective industries in stock market lead to the accuracy of their future economic forecast and their influence on other industries to be more important to be able to make more accurate forecast from their future economy to more accurate decisions about them. According to minimum spanning tree, the obtained from ultra-metric matrix is divided into three categories. The first category are industries with five top-grade in minimum spanning tree, whose sum of the minimum distance and effectiveness are: 1-Basic Metals, 2-cement, lime and chalk. The second category are industries that have vertices of degree four are in the minimum spanning tree. Their sum of minimum distance and effectiveness includes: 1-Machinery and Equipment, 2-metal products and 3 –An Industrial Conglomerates and the third category are industries that have third degree vertex in the minimum spanning tree. Also, their sum of minimum distance and effectiveness are: 1 -petroleum products, coke and nuclear fuel 2- Investment 3 -technical and engineering services 4-Construction Devices and communications equipment 5-Posts and Telecommunications 6-Manufacture of other non-metallic mineral}.when an economic crisis is caused , this crisis affect the returns of most available industries in stock market , but this effect is not equal to the returns of different industries and have different effects. According to the results of obtained research, industries located in each category have different fluctuations. By identifying of these categories, we can compare the condition of other industries with these industries and use them for forecasting their operation. We can also use the results of obtained minimum spanning tree to produce portfolios. In other words, by considering these results, we can use the benefits of diversification portfolios in the best way. In this way that we use industry stocks in portfolio that are located in different categories. The more study of this subject is one of the actions that we want to do in future.

Suggestions for further research

1. Studying and comparing the fluctuation behavior of available index of industries in Tehran stock exchange with industries of other countries.

2. The Suggestion of doing this research by using the algorithms of artificial intelligence such as genetic, Neural Networks and ....

References

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

Caiado, J., Crato, N. & Peña, D. (2007). Comparison of time series with unequal lengths,manuscript.

Campbell, J. Y., Lo, A. W. & A. C. MacKinlay (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

Ricardo, Coelho, Claire, G. Gilmore, Brian, Lucey, Peter, Richmond, Stefan, Hutzler,(2007).The evolution of interdependence in world equity markets—Evidence from minimum spanning trees", *Statistical Mechanics and its Applications, 376,* 15 March, 455-466.

Elton, E. J. & M. J. Gruber. (1995). *Modern Portfolio Theory and Investment Analysis*, J. Wiley and Sons, New York.

Eryiğit, M. & R. Eryiğit, (2009). Network structure of cross-correlations among the world market indices, *Statistical Mechanics and its Applications*, *388*(17), 3551-3562.

Gower, J. C. (1966). *Biometrika, 53,* 325–338.

Gilmore, C. G., Lucey, B. M. & Boscia, M. (2008). An ever-closer union? Examining the evolution of linkages of European equity markets via minimum spanning trees, *Statistical Mechanics and its Applications*, *387*(25), 6319-6329.

Hebner, M.T. (2007). *Index funds: The 12-step program for active investors* (2nd ed.).Irvine, CA: IFA Publishing.

Johnson, D. E. (1998). *Applied Multivariate Methods for Data Analysts*, 1st ed., Duxbury Press.

Jung, W.S., Chae, S., Yang, J.S. & H.T Moon. (2006). Characteristics of the Korean stock market correlations, *Statistical Mechanics and its Applications*, 361 (1), 263-271.

Laloux, L., Cizeau, P., Bouchaud, J.-P. & M. Potters. (1999). *Phys. Rev. Lett. 83,* 1467-1470.

Maharaj, E. A. (2002). Comparison of non-stationary time series in the frequency domain, *Computational Statistics & Data Analysis*, 40, 131-141.

Mantegna, R. N., & H. E. Stanley. (2000). *An Introduction to Econophysics: Correlations and Complexity in Finance,* Cambridge University Press, Cambridge.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investment*, J. Wiley, New York.

Mirzavaziri, M. (2007). *Metric spaces.* Ferdowsi University of Mashhad Press, second edition.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N. & H. E. Stanley (1999). *Phys. Rev. Lett. 83,* 1471-1474.

Rammal, R., Toulouse, G. & M. A. Virasoro (1986). Ultra-metricity for Physicists, *Rev. Mod. Phys. 58,* 765-788.

Salvatore M., Giovanni B., Fabrizio L., & Rosario N. M. (2003). Degree stability of a minimum spanning tree of price return and volatility, *Statistical Mechanics and its Applications*, 324(1-2),66-73. Skórnik-Pokarowska, U. & A. Orłowski. (2004). Application of the ultra-metric distance to portfolio taxonomy: Critical approach and comparison with other methods. *Statistical Mechanics and its Applications*, *344* (1-2), 81-86.

Tehrani, R., Talbnia, G. & S. Jalili. (2007). Assessing the Tehran Stock Exchange traders rely on accounting and non-accounting information in investment decisions, *Journal of Financial Research*, *21.*

West, D. B. (1996). *Introduction to Graph Theory,* Prentice-Hall,Englewood Cliffs NJ.