DISTANCES AND NETWORKS: THE CASE OF MEXICO

Linda Margarita Medina Herrera, Tecnológico de Monterrey
Ernesto Armando Pacheco Velázquez, Tecnológico de Monterrey

ABSTRACT

The influence of six different distances on the structure of minimum spanning trees is presented in this paper. Measures of complex networks are built based on the closing prices of stocks of the main companies traded on the Mexican Stock Market. We find that the City block and Chi distances not only match, but also determine more precisely the central vertex, the level of the tree and the clusters formed by the economic sector. The trees formed using Minkowski distances have similar structures and show a disadvantage when classifying the vertices. The construction and telecommunication sectors are most important within the trees, regardless of the distance used.

JEL: C02, C22, C38, C45, C61, C8, D85

KEYWORDS: Stock market network. Econophysics. Distances. Minimum spanning tree.

INTRODUCTION

In 1926 the Czech scientist Otakar Boruvka developed an algorithm that allowed creation of an electrical network in Moravia using a minimal amount of cable. The idea of joining the vertices with cables more efficiently, has been studied extensively in graph theory by spanning trees, which is a plot of N objects (vertices or nodes) connected by N-1 arcs. Among all the spanning trees, the minimum spanning tree is the one that minimizes the weight of the tree (the sum of its arcs).

Minimum spanning trees have been widely used to analyze financial assets behavior. In this type of financial trees the nodes or vertices are assets and the arcs are distances, constructed from the correlation coefficient. One of the most interesting applications of the minimum spanning tree is portfolio optimization. By using special measures in the tree, the central vertex (the center of mass or the vertex with the greatest influence) is selected, vertices are classified according to their distance to the central vertex and finally a function that minimizes portfolio risk is established. Minimum risk assets are located in the outer branches of the tree, while higher returns assets are near the central vertex. Minimum spanning trees have been also used to analyze financial assets behavior at different points in time allowing for extraction of information in times of crisis.

The Euclidean distance has usually been used to build the trees, since the distances are obtained from the correlation matrix in a very simple way. However, there is no evidence that Euclidean distance is the most suitable. One contribution of this paper is that there are distances that distinguish more adequately the central vertex and other important characteristics of minimum spanning trees. An empirical result is obtained from the construction and analysis of six spanning trees whose vertices are the main companies in the Mexican Stock Market (Bolsa Mexicana de Valores).

The Mexican stock market is one of the most important in Latin America and one of the ten big emerging markets. Recently, an a number of Mexican companies have been involved worldwide in mergers and acquisitions.

The main objective of this paper is twofold. First, it is relevant to study the influence of different distances in the structure of minimum spanning trees. Second, the measures of complex networks that are built based on the closing prices of stocks of the main companies traded at the Mexican Stock Market.

LITERATURE REVIEW

Fundamental work in financial networks appears in the late eighties. Mategna and Bonano introduced the concept of graphs in the financial market environment as a method for finding hierarchical arrangements of stocks through the study of clusters of companies; see Mantegna and Stanley (2000), Bonano et all (2000).

Notions of minimal spanning trees as random matrix theory have been of interest in the study of financial correlation matrices. The properties of random matrices combined with the power of minimum spanning trees have made it possible to examine the movements of major stock markets (Bonano et all (2003) in America, Jung et all (2006) in Korea, Medina (2007) in Mexico, Eom et all (2009) in Japan, and Tabak et all (2010) Brazil). Minimal spanning trees also allowed us to analyze global financial indices (Tumminello (2010) and Medina (2012) and other financial assets; Miccichè et all (2003)).

Modeling the correlation matrix of a complex system with tools of hierarchical clustering has been useful in multivariate characterization of stock return time series (Mantegna, 1999; Bonanno et all, 2001, 2003), market index returns of worldwide stock exchanges (Bonanno et all, 2000). The estimation of statistically reliable properties of the correlation matrix is important for several financial decision processes such as asset allocation, portfolio optimization (Tola et all, 2008), derivative pricing, etc. Bonano et all, Medina (2011) and Tabak et all (2010), employ a dynamic approach using complex network measures and find that the relative importance of different sectors within the network varies. Miccichè et all (2003) found that minimum spanning trees of asset returns is characterized by stock degree values, which are more stable in time than those obtained by analyzing a minimum spanning tree computed starting from volatility time series.

Tumminello et all (2010) discuss how to define and obtain hierarchical trees, correlation based trees and networks from a correlation matrix. Tola et all (2008) show that the use of clustering algorithms can improve the reliability of the portfolio in terms of the ratio between predicted and realized risk.

A review of the main topological measures is found in Barthélemy et all (2005), for instance the central vertex, the most strongly connected node of the tree and the Mean Occupation Layer. Onnela et all (2003) show that during crashes, due to the strong global correlation in the market, the tree shrinks topologically, and this is shown by a low value of the mean occupation layer. Another central measure in complex network theory is the clustering coefficient. A large number of networks show a tendency to link formation between neighboring vertices. This tendency is called clustering. Clustering around a vertex is quantified by the clustering coefficient. Ledoit and Wolf (2003) present various generalizations of the clustering coefficient and a comparative study of the several suggestions introduced in the literature.

The empirical tree has features of a complex network that cannot be reproduced, even as a first approximation, by a random market model or by the one-factor model. Several papers have shown that the use of these tools may help in the design of portfolio strategies and risk assessment. Ledoit and Wolf (2003), Tola (2008), Eom (2009). Onnela et all (2003) analyze dynamic trees, that is, with windows in time, to show that the assets of the Markowitz optimal portfolio are virtually all the time in the external branches of the tree. The correlation matrix can be used to extract information about aspects of the hierarchical organization of such a system; Mantegna (1999), Saramaki (2007). The clustering procedure is done by using correlation between pairs of elements as a similarity measure and by applying a clustering algorithm to the correlation matrix (Cukur et all, 2007).

The statistical reliability of hierarchical trees and networks is depending on the statistical reliability of the sample correlation matrix, a bootstrap approach has been used to quantify the statistical reliability of both hierarchical trees and correlation based networks (Tumminello et all, 2007). Ledoit ang Wolf (2003) developed a method to estimate the covariance matrix of stock returns by an optimally weighted average of two existing estimators, the sample covariance matrix and single-index covariance matrix. Tumminello et all (2007) studied topological properties such as the average length of shortest paths, the betweenness and the degree are computed on different planar maximally filtered graphs generated by sampling the returns at different time horizons. Their empirical results show the effect is varying with the sampling time horizon. The more structured network is observed for the intraday time horizon.

In this article we use the correlation matrix to construct a minimum spanning tree on the Mexican Stock Market, calculating the main measures and conglomerates, which will be compared with those obtained using other distances. This article contributes to the literature in at least two ways. First showing that there are distances as Chi Square and City Block that distinguish in clear, clusters and hierarchies as well as building a more suitable Mean Occupation Layer for use in portfolio optimization. Secondly, it shows an empirical study of minimal spanning trees on the Mexican Stock Market that expands the understanding of stock market movements and determines the economic sectors that dominate the Mexican market over time.

The remainder of the paper is organized as follows. Section three deals with distances. The main measures are calculated in order to analyze the structure of the trees. In section four data and methodology are presented, in section five the empirical results are examined. The last section presents conclusions and final considerations.

DISTANCES AND MINIMAL SPANNING TREES

Let N be the number of stocks with price $P_i(t)$ for asset i over time, with $t = 0, 1, \ldots T$. Taking $S_i(t)$ as the log of asset returns. $S_i(t) = \ln P_i(t) - \ln P_i(t-1)$ and standardizing the series $Z_i(t) = Z_{it}$ is obtained. The used distances are as follows.

Euclidean distance: The Euclidean distance between two points A and B is the magnitude of the vector connecting A with B, i.e.

$$d^2(Z_i, Z_j) = \sum_{k=1}^{T} \left(Z_{ik} - Z_{jk}\right)^2 = 2 - 2\,p_{ij}$$

Where ρ_{ij} is the coefficient of correlation between Z_i and Z_j.

Manhattan distance or City Block distance

.

$$d_M(Z_i, Z_j) = \sum_{k=1}^{T} |Z_{ik} - Z_{jk}|$$

$$d_M(Z_i, Z_j) = \sum_{k=1}^{T} |Z_{ik} - Z_{jk}|$$

Minkowski distance

$$d_m(Z_i, Z_j) = \left(\sum_{\mathrm{k}=1}^{\mathrm{T}} |Z_{ik} - Z_{jk}|^m\right)^{1/m}$$

In this article we use m = 3 and m = 4. The Euclidean distance corresponds to m = 2.

Chi-square distance: Chi-square distance between Z_i and Z_j with weights w_i, $i = 1, 2, 3, \ldots, T$ is

$$d_{\chi^2}(Z_i, Z_j) = \sum_{\mathrm{k}=1}^{\mathrm{T}} \mathrm{w_k}(Z_{ik} - Z_{jk})^2$$

The chi-square distance can standardize the distances in terms of ranges or standard deviations. The weights are found in two ways:

Firstly $\mathrm{w}_i = \frac{1}{S_i}$ where S_i is the standard deviation of the stock prices of day i and $i = 1, 2, 3, \ldots, T$.
Secondly, $\mathrm{w}_i = \frac{1}{R_i}$ where R_i is the range of the stock prices of day i and $i = 1, 2, 3, \ldots, T$.

A spanning tree is a graph of N objects (vertices or nodes) connected by N-1 arcs that allow jumps from one vertex to any other. If each arc represents a distance or cost, or in general if each arc is associated with a weight (a real number), the sum of the weights of all sides of a tree will be the total weight of the tree. A minimum spanning tree is a spanning tree that minimizes the total weight of the tree. Let´s briefly review the different measures and parameters that allow a first statistical characterization of trees.

Let a_{ij} the adjacency matrix whose elements take the value of 1 if an edge connects vertex *i* to vertex *j* and 0 otherwise. Weighted trees are usually described by a matrix w_{ij} specifying the weight on the edge connecting the vertices *j and i*. If the nodes *i* and *j* are not connected then $\mathrm{w}_{ij} = 0$.

The degree of vertex i, k_i, is the number of nodes directly connected to node i, that is, the number of elements that comprise the neighborhood of *i*, *V(i)*. Along with the degree of the vertex, a significant property of trees in terms of their weights is obtained by analyzing the strength of the vertex defined as

$$\mathrm{s}_j = \sum_{j \in V(i)} \mathrm{w}_{ij}$$

The strength of the chosen vertex considers the connectivity information and the importance of the weights of the connected vertices. The level V_{ij} is the sum of the arches on the tree in order to go from vertex *i* to vertex *j*. If *i*=*j* then $\mathrm{v}_{ij} = 0$. It is important to characterize the way the nodes are extended in the tree. To this end "mean occupation layer, $\mathrm{l}(\mathrm{v}_i)$" is defined as:

$$l(\mathrm{v}_i) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{v}_j$$

The node with the lowest mean occupation layer v_m is the center of mass of the tree. A high value of $\mathrm{l}(\mathrm{v}_m)$ reflects a thin market structure, while at the other end low values are associated with market crisis.

The central vertex is considered the father of all vertices of the tree or as the root of it. It is used as a reference point in the tree; the position of the other vertices is relative with respect to the root tree. There is some arbitrariness in the choice of the central vertex, however the following criteria can help choosing the best candidate: (1) the vertex of higher degree, (2) the heaviest vertex, (3) the center of mass.

Weighted clustering combines the topological information with the weight distribution of the network considering the fact that some neighbors are more important than others. The weighted clustering coefficient is defined as

$$c^w(i) = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jk}$$

This measure is the counting for each triple formed in the neighborhood of vertex i; the weight of the two participating edges of vertex i. s_i is the vertex strength, k_i is the degree of vertex i and $a_{ij} = 1$ if there is an edge between i and j. The former measure considers not only the number of closed triplets in the neighborhood of a node but also their total relative weight with respect to the strength of the node.

DATA AND METHODOLOGY

The database for this study are time series of daily closing prices of 31 companies traded on the BMV, in the period between 01/03/01 and 2/22/11. To select companies and the length of the series liquidity, capitalization and maintaining were taken into account. The final length of the series is 2553. Among the 31 companies selected for the study all economic sectors are represented. The chosen companies have the highest trading volume in each sector and together represent over 90% stake in the Mexican IPC index (Indice de Precios y Cotizaciones), a capitalization weighted index of the leading stocks traded on the Mexican Stock Exchange. The included stocks were active in the period selected for the study.

The correlation matrix of asset returns is used to construct the Euclidean Distance Matrix. Kruskal's algorithm; Kruskal (1956) is used for joining the vertices (asset) minimizing the total length of the tree, that is, forming a minimum spanning tree. This way we get the "Euclidean tree".

To analyze the topological and hierarchical characteristics of the tree, we calculate the measures of weighted networks, finding the vertex to simulate the center of mass of the tree, the Mean Occupation Layer and the Weighted clustering. These measures help identify clusters by economic sector, shares of major influence on the market and the size of the tree. Similarly, the remaining five distances, distances matrices and corresponding spanning trees are constructed. The last step is to analyze similarities and differences in the measurements obtained in each tree, finding advantages and disadvantages of working with any of these distances.

EMPIRICAL RESULTS

According to the measures Weight, Strength and Mean Occupation Layer, TELMEX, GEO, AMX, CEMEX and TLEVISA are the top five companies in all distances as shown in Table 1. The central vertex was selected between them. These firms belong to the communication and construction sectors.

The Mean Occupation Layer obtained by the Euclidean tree is the largest among all the trees as shown in Figure 1. This is reflected in the extent of the branches of the tree. Due to the characteristics of the Mexican market, lack of diversification and crises, a tree "shrunk" with a minor Mean Occupation Layer, could be more representative. This is the case of trees Chi and City Block.

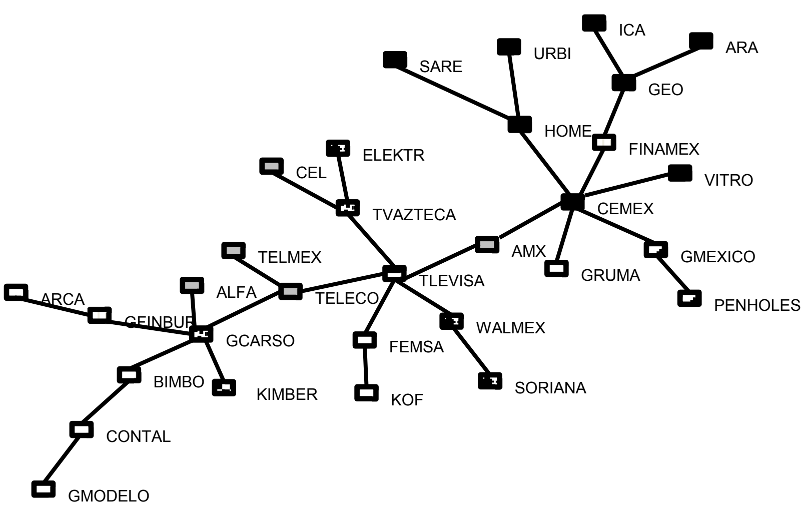
Table 1: Central Vertex

	Weight	Strength	Mean Occupation Layer		Weight	Strength	Mean Occupation Layer
Euclídean				**Chi with standar desviation**			
CEMEX	6	6.541	3.388	TELMEX	20	19.336	1.306
GCARSO	5	5.462	3.881	TELECOM	4	3.383	1.701
TLEVISA	5	4.976	2.979	CEMEX	3	2.485	1.795
TELMEX	3	3.161	3.368	**Minkowski m=3**			
City- Block				AMX	5	3.258	1.801
TELMEX	23	19.499	1.138	TLEVISA	4	2.228	1.704
HOMEX	5	4.113	2.177	CEMEX	3	1.918	3.012
CEMEX	3	2.488	1.593	GEO	4	2.532	2.107
				TELMEX	3	1.756	1.948
Chi with ranges				**Minkowski m=4**			
TELMEX	22	21.427	1.234	AMX	6	2.651	1.068
CEMEX	3	2.600	1.731	TLEVISA	4	1.424	1.260
HOMEX	3	2.723	2.466	GEO	4	1.645	1.288
TELECOM	3	2.303	1.654	TELMEX	3	1.108	1.081

This table Show the estimation obtained for the center of mass, weight and strength of the companies that obtained the best parameters for each distance.

At first sight, the constructed tree using the Euclidean distance cannot distinguish clusters of economic sectors. After analyzing the weighted clustering coefficients it is seen that the most important clusters are construction and telecommunications. The construction coefficient is three times larger than the next cluster, telecommunications. In this tree, CEMEX leads the list of candidates for central vertex. CEMEX is a leading global construction industry with a high share in the IPC and high marketability.

Figure 1: Euclidean Tree

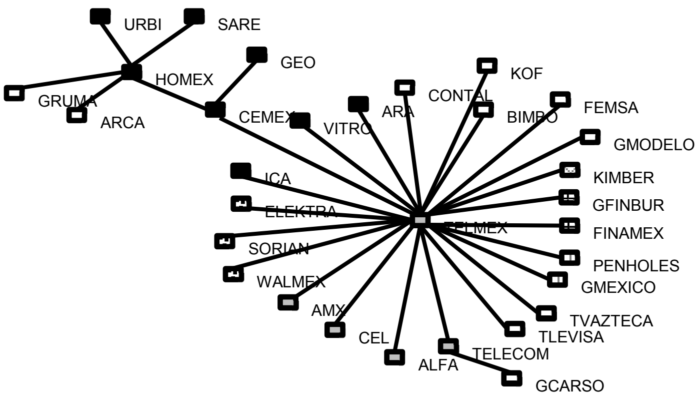


This figure shows the Euclidean Tree. The Mean Occupation Layer is the largest of all trees, therefore the tree appears widespread. A fine look shows the construction conglomerate, led by CEMEX and telecommunications conglomerate led by TELMEX.

Televisa (TLVISA), a Mexican multimedia conglomerate, the largest mass media company in Latin America and in the Spanish-speaking world, has the lowest mean occupation layer. Its mean occupation layer is 2.9, obtained with the Euclidean distance, so it can be compared with that obtained from other countries. In the years researched, the American financial exchanges showed a mean occupation layer between 3 and 9.8, most of the time remaining above 4. The average occupancy of the BMV (2.9) is an indication of the homogeneity of the market and the economic contraction.

City-Block and Chi distances shown in Figure 3, create parameters that separate the stock with greater degree, weight and strength. Moreover the three parameters point towards the same central vertex of the tree. According to these three distances the central vertex is TELMEX, the leading telecommunications company in Latin America and one of the companies with the highest and fastest growth of worldwide audience. This pair of distances separates the stocks of the BMV in two large clusters, one led by TELMEX and the telecommunications companies, and the other led by CEMEX and the construction cluster. No other economic sectors form clusters. This fact suggests the Mexican market is still a developing market, where the behavior of the system is still homogeneous.

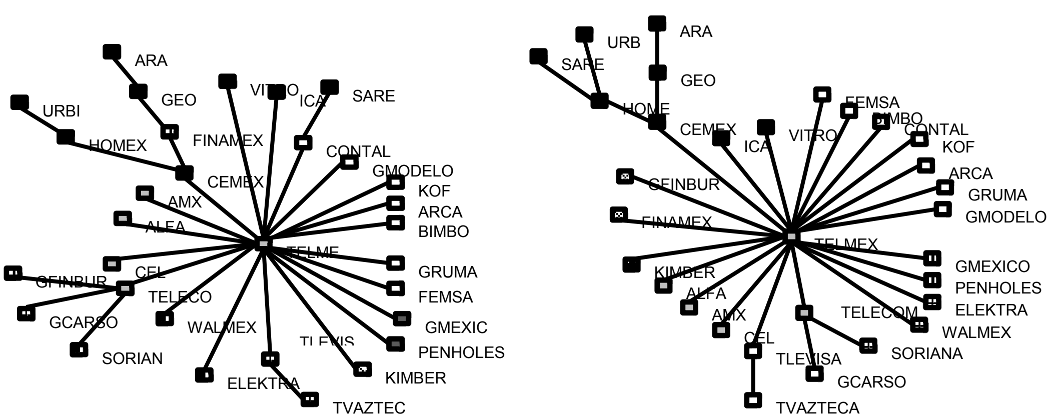
Figure 2: City Block Tree.



This figure shows the City Block Tree. By having Mean Occupation Layer less than Euclidean tree, this tree looks compressed. Clearly identified two clusters: the construction and telecommunications.

The advantage of City Block and Chi distances is that the central vertex is evident. It can easily be located at first sight. All measurements coincide when the central vertex is selected. In addition, they separate the vertices into two major clusters, led by construction and telecommunications sectors. America Movil (AMX) has the lowest mean occupation layer of all trees and also receives the Minkowski distance score required for central vertex.

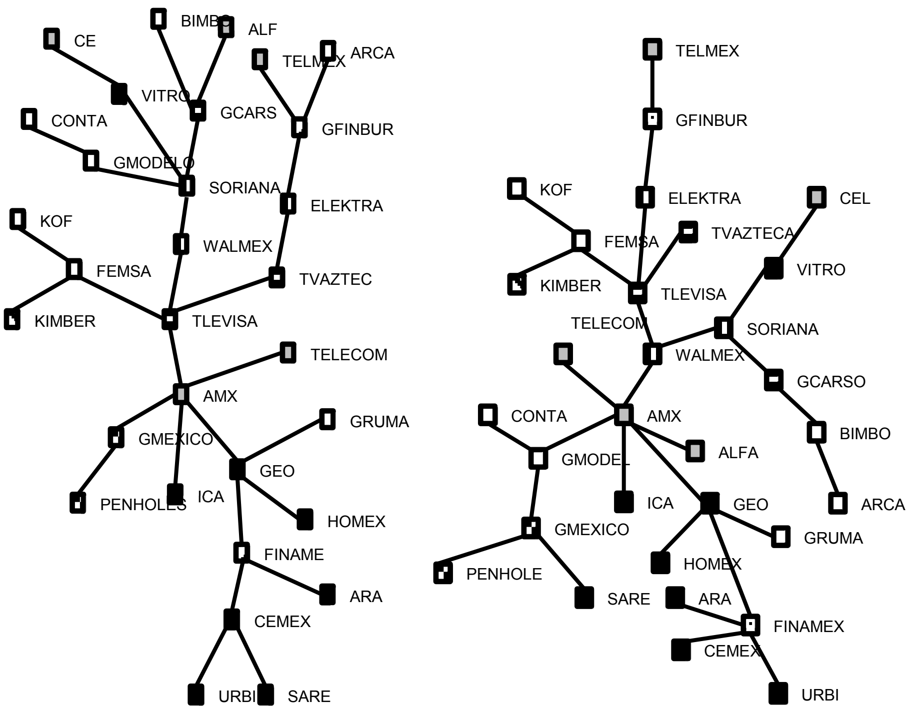
Figure 3: The Chi Trees



Left tree is Chi with standard deviation tree; right tree is Chi with range. The trees are shrunk. Two clusters are clearly visible: construction and telecommunications.

AMX is the leading provider of cellular telecommunications services in Mexico. It has subsidiaries and joint investments in the communications sector in Latin America, Spain and the United States. It is the station with more weight in the IPC. Companies linked to this vertex have high correlations, remain very close to the central vertex and have high returns. With the results obtained by Telmex and AMX, it is not surprising that the telecommunications conglomerate is one of the two major clusters distinguished in all trees. Euclidean tree (Minkowski with m = 2) and Minkowski trees presented in Figure 4 (m = 3 and m = 4) present similar characteristic: spread and not very visible clusters.

Figure 4: The Minkowski Trees



Left: Minkowski tree with m=3. Right: Minkowski tree with m=4. The trees are spread and look similar to the Euclidean tree. The construction and telecommunications conglomerates are hardly visible. More clusters are indistinguishable.

CONCLUSIONS

We have studied the influence of different distances in the structure of minimum spanning trees. Six trees were constructed to study its topological and hierarchical properties. The vertices of the trees are 31 stocks of the Mexican Stock Market and the arcs are distances constructed with historical data (time series of length 2553). The results show City Block and Chi trees better reflect the behavior of the Mexican Market in a much better way than the Euclidean tree. The evidence shows that the Mexican Stock Market form clusters in the construction and telecommunications sectors. The five top rated companies are independent of the selected distances.

The Euclidean tree obtained has the highest Mean Occupation Layer, since this measure shows how extending the branch of the tree, the Euclidean tree is the most widespread of all. Clusters by economic sectors are not evident until the weighted-clustering measure is calculated and two clusters are identified. Construction, with a coefficient three times larger than the next cluster and telecommunications. Minkowski ´s Trees have the same characteristics as the Euclidean tree. This is surprising since Euclidean Tree is a Minkowski tree too (m=2).

The City Block and Chi distances distinguish in a radical way the central vertex and clusters with greater weight. These distances are the most representative of the Mexican market: the average occupancy rate is lower than that obtained with the Euclidean distance, showing a homogeneous market with almost no diversification. The two clusters that remain invariant in all the trees are clearly identifiable.

The network measures suggest that the construction and the telecommunications sectors are the most important sectors in the trees, with a great influence on other sectors. The fact that other economic sectors show very low weighted clustering coefficients suggests that the Mexican market is still a developing market, where the behavior of a system is still very homogeneous and contracted.

Topological characteristics of minimum spanning trees are useful in the theoretical description of financial markets and in the search of economic clusters of stocks. The topology and the hierarchical structure of trees, is obtained by using information present in the time series of stock prices only. This result shows that time series of stock prices have precious and traceable economic information.

Further research could focus on the behavior of measures of the trees in years of crisis, index construction based on Mean Occupation Layer or similar measures to be discovered. Although we focused on networks of stocks prices, this methodology should be similarly insightful for multivariate time series of other asset classes.

REFERENCES

Eom, C. Oh, G. Woo-Sung Jung, H. Jeong, S. Kim. (2009) "Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series" *Physica A: Statistical Mechanics and its Applications* 388 900–906 .

Bonanno, G. Caldarelli, G. Lillo, F. Mantegna, R. (2000) "Taxonomy of stock market indices", *Physical Review* E 62 R7615.

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

Onnela, J., Chakraborti, A., Kaski, K., Kertesz, J., Kanto, A., (2003) "Dynamics of market correlations: Taxonomy and portfolio analysis", *Physical Review* E 68 056110.

Saramaki, J., Kivela, M., Onnela, J., Kaski, K., Kertesz, J., (2007) "Generalizations of the clustering coefficient to weighted complex networks", *Physical Review* E 75 027105.

Ledoit, O., Wolf, M., (2003) "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection". *Journal of Empirical Finance* 10, 603–621.

Barthélemy,M., Barrat, A., Pastor-Satorras, R., Vespignani, A. (2005) "Characterization and modeling of weighted networks", *Physica* A 346 34–43.

Medina, L., Mansilla, R., (2007) "Un árbol de expansion minima en la Bolsa Mexicana de Valores". *Revista de Administración, Finanzas y Economía (Journal of Management, Finance and Economics)* (2007) vol. 1, issue 2, pages 116-124

Medina, L., Díaz, B., (2011) "Caracterización y modelado de redes: el caso de la Bolsa Mexicana de Valores". *Revista de Administración, Finanzas y Economía (Journal of Management, Finance and Economics)* vol. 5, No. 1. 23-32

Medina, L., M., Pacheco, E., (2012) "Caracterización y modelado de redes: indices financieros mundiales". *Global Conference on Business and Finance Proceedings*. Vol 7. No 1. 774-779

Miccichè, S., Bonanno, G., Mantegna, R. (2003),"Degree stability of a minimum spanning tree of price return and volatility". *Physica* A-Statistical Mechanics and its Applications, 324 (2003), pp. 66–73

Tumminello, M., Mantegna, R., (2010) "Correlation, hierarchies, and networks in financial markets", *Journal of Economic Behavior and Organization*, (doi:10.1016/j.jebo.2010.01.004).

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

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

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

Cukur, S. , Eryigit, M., Eryigit, R., (2007) "Cross correlations in an emerging market financial data," Physica *A: Statistical Mechanics and its Applications* 376 555–564.

Tabak, B., Serra, T., Cajueiro, D., (2010) "Topological properties of stock market networks: The case of Brazil," *Physica* A 389 3240-3249

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

Jung, W., Chae, S., Yang, J., Moon, H., (2006) "Characteristics of the korean stock market correlations," *Physica A:* Statistical Mechanics and its Applications
361 263–271.

Kruskal, J., (1956), "On the Shortest Spanning Subtree and the Traveling Salesman Problem," *Proceedings of the American Mathematical Society* 7, 48-50.

ACKNOWLEDGEMENTS

The authors thank the suggestions provided by Rosa María García Castelán, which helped to improve the paper.

BIOGRAPHY

Linda Margarita Medina Herrera. Department of Physics and Mathematics.Tecnológico de Monterrey, CCM. Del puente 222 Col. Ejidos de Huipulco Tlalpan, 14380, México, D.F. México. Phone (55) 54832190. Fax +52(55) 54832163. linda.medina@itesm.mx

Ernesto Armando Pacheco Velázquez. Industrial Engineering Department. Tecnológico de Monterrey CCM. Del puente 222 Col. Ejidos de Huipulco Tlalpan, 14380, México, D.F. México. Phone (55) 54832190. Fax +52(55) 54832163. epacheco@itesm.mx