7th International Conference, The Economies of Balkan and Eastern Europe Countries in the changed world, EBEEC 2015, May 8-10, 2015

A Network Analysis of the Greek Stock Market

Kydros Dimitrios, Oumbailis Vasileios

Department of Accounting and Finance, School of Business and Economics, TEI of Central Macedonia, Terma Magnisias, Serres, 621 00, Greece

Abstract

In this paper we analyse stock relationships in the Greek Stock Market. We propose a model that can depict such relationships and create networks of stocks. We investigate all stocks in the Greek Stock Market for years 2007 and 2012 (one year before and during the current economic crisis). Different networks are created according to the degree of correlation of stocks. These networks are visualized and evaluated, using methods from Social Network Analysis. A number of metrics, mainly centrality measurements, are calculated and interpreted. We discuss the hypothesis that Greek stocks follow the "herd" rule and investigate the role of important actors (stocks) in these networks. Our results show that the Greek Market is a "shallow" market, easily affected by a few big investors or the economic climate.

Keywords: Social Network Analysis, Greek Market; Athens Stock Exchange; Cross Correlation Coefficients

1. Introduction

The need to understand the complex relationships in the stock markets has increased since the spread of the recent financial crisis on world markets and financial institutions. The existence of a large number of heterogeneous components which do not vary smoothly leads to complicated behaviour of the stock markets. The behaviour of individual stock indices are commonly monitored using time series data. The diachronic analysis of such data is very difficult since a growing number of variables to be monitored exist.

* Corresponding author. Tel.: +30-321-49268; fax: +30-321-49136.
E-mail address: dkydros@teiser.gr
Recent literature has focused on an approach based on networks as a viable alternative to study the complex interplay in stock markets, presenting hidden information in the huge volume of time series of stock markets in a handy way. Such an approach can reveal the internal structure of various stock markets around the world. This in turn can help the international portfolio diversification and enhance global financial systems (Roy R., Sarkar U., 2011).

The concept of networks has already been applied to study the behavior of the US (Boginski V., et al., 2006), Korea (Kim K., et al, 2007), China (Huang W., et al., 2009), India (Pan R., Sinha S., 2007), and Brazil (Tabak B., et al., 2009) stock markets. In this paper we will present a similar study to depict and analyze the Greek stock market through social networks.

More specifically we seek to build network models that demonstrate relationships between stocks. As raw data we use share’s closing prices from the Greek stock market and compute the related interaction between them (cross correlation). Consequently, by using data mining techniques we create various network models as alternative presentation of the associated shares. We focus our analysis on two years, a year before the economic crisis and a year during the crisis, in order to explore if the crisis has affected the interaction of the shares. Techniques from Social Network Analysis (SNA) will be used in order to draw conclusions regarding the topology and in finding the most central shares. For this purpose we used the open-source social network analysis software "Gephi" by Bastian et al. (2009), along with other data management programs.

2. Theoretical and Methodological Issues

2.1. Elements of Graph Theory

A graph $G$ is an ordered pair of two sets, $N$ and $A$ and is denoted as $G = (N, A)$. The elements in $V$ are called nodes, vertices or simply points. The total number of nodes is denoted by $n$ ($n = |N|$). Set $A$ contains ordered or unordered pairs of elements on $V$, $(i,j)$ where $i,j \in V$, are called arcs or links respectively. A graph can be easily drawn in order to depict useful relationships. In such a visualization, a node is drawn as a point, circle, rectangle containing the nodes name (when nodes are not named, it is useful to use integer numbers). Arcs are drawn as directed arrows and links as simple lines that join the corresponding nodes. For example, arc $(i,j)$ is drawn as an arrow emanating from node $i$ and ending on node $j$. In recent literature, terms Graph and Network mean the same thing.

A subgraph $F = (N', A')$ is a Graph that is produced by another Graph $G = (N, A)$ by removing node and/or arcs. In such a case, $N' \subseteq N$ and $A' \subseteq A$.

A chain in $G = (N, A)$ is an alternating sequence of nodes $i_k$ and arcs $a_k$, such as $i_1 - a_1 - i_2 - a_2 \ldots i_r - a_r$, that starts and ends in different (or the same) nodes, where for each $k, 1 \leq k \leq r - 1, a_k = (i_k, i_{k+1}) \in A$ or $a_k = (i_{k+1}, i_k) \in A$. In a chain a node or arcs can be met twice or more times. A path is an open chain where each node (hence each arc) is found exactly once. The length of the path is the number of arcs needed to complete a walk on it (Kydros D., et al., 2012, p. 3). When more than one paths between two nodes exist, then the shortest path (or distance) is the one(s) with the smallest length.

Two nodes $i$ and $j$ of $G = (N, A)$ are called to be connected when there exists at least one way (not necessarily oriented) having $i$ and $j$ as its ends. $G$ is connected when every pair of its nodes is connected, otherwise it is called disconnected. Maximally connected subgraphs of a disconnected Graph are called components (Paparrizos K., 2003).

The density of a graph is the total number of present arcs divided by the maximum number of potential arcs in this graph. In a complete graph, all possible pairs of arcs are present, so its density equals to 1. In a completely disconnected Graph (no arcs present) the density equals to 0. When a graph’s density approaches 1, then this graph is dense, otherwise it is sparse. In an undirected Graph with $N$ nodes, the maximum possible number of links equals to:

$$\frac{N(N - 1)}{2}$$

Hence, if this graph has $L$ links, then its density, $S$, is calculated as:
According to Scott, “density cannot be used to compare two networks with significantly different sizes” (Scott J., 2000). Instead the notion of “local density” or Clustering Coefficient can be used. The Clustering coefficient of each node is calculated as the number of arcs between a node’s immediate neighbours divided by the maximum possible number of arcs in this neighbourhood. This metric can be used as an index on whether some nodes tend to create strongly connected areas. It has been shown that a large number of real-life networks, especially in social networks, nodes tend to form closely connected communities, bearing high local densities. This situation happens in much larger probability that the average that comes from randomly created networks with similar numbers of nodes and links (Watts J., Strogatz H., 1998) (Holland P., Leinhardt S., 1971). The average on all clustering coefficients is the global clustering coefficient of the network, a metric that gives a general sense of a networks clustering (Nettleton D., 2013).

A graph $G = (N, A)$ can be stored in a computer’s memory in various ways. The simplest one is through the use of a rectangular, $N \times N$ matrix, called Adjacency Matrix ($A$), where $A_{i,j} = 1 \text{ iff } (i, j) \in A$ and 0 otherwise.

2.2. Centralities

When analysing a network’s data, it is useful to rely on well-established metrics and especially the ones under the general notion of centrality. There exist different measurements of centrality both on nodes and links. These measurements can show the most important variables in a network, according to the exact definition of the metric (Newman M., 2002, pp. 1-12).

The simplest centrality metric is degree centrality and it is quite intuitive on the impact of one node in the network. The degree centrality of a node (or simpler, its degree) is the total number of its immediate neighbours. Degree centrality can be normalized when it is divided by its maximum possible number.

The notion of eigenvector centrality relies on the same idea, but in a more subtle way: instead of merely counting the number of links emanating from a node, the idea is that not all links are of the same importance. Thus, links coming from more important nodes will offer more centrality in a node than links coming from less important nodes. The idea is definitely recursive and this centrality measurement is calculated using the eigenvalues of the adjacency matrix.

Other notable measurements of centrality are closeness and betweenness centrality. These metrics rely on the notion of paths within the network and can also be normalized. Closeness centrality for one node can be calculated as the average distance of all distances from this node to all other nodes in the network. (Newman M., 2002). Some scholars define closeness as the inverse of this average, so that larger numbers would mean better performance. We use this latest definition in the following section.

A node has high betweenness centrality if it serves as an intermediate between many other nodes, that is, it lies between them with respect to their shortest path. In other words if we calculate the total set of shortest paths, then a node with high betweenness centrality is present with a large proportion of this set. Again, the higher this metric is, the more important the node is, since it controls the flow of information between many other nodes (Newman M., 2002). Closeness centrality and it’s study has special meaning when studying fault tolerance or attacks in networks. If a node with high closeness centrality is affected, then the overall impact on the connectivity and distances on the network is severe. Although this metric is quite important, it is rather expensive to compute, especially in large networks, since it requires global information for the network examined. Recent algorithms try to minimize this complexity (Brandes U., 2001) (Newman M., 2001), so closeness centrality and its study provides us with very important results.

The study of distances on networks can be also used in order to classify these networks. It has been shown that many real life networks belong in the “small-world” or “scale-free” class. A small world is a network with a small number of “central” nodes (with many links and high clustering coefficients). Furthermore, the average distance in a small-world is much lower in this network than the one computed in a randomly created similar network. Finally, in a small-world, the average distance is increased proportionally to the logarithm of the number of the nodes in this network. A large number of social networks, wikis, Internet connectivity, gene networks are small-worlds (Watts J., Strogatz H., 1998) (Barthelemy M., Amaral L., 1999).
3. Methodology and Analysis

3.1. Data Collection and Network Creation

The initial plan was to collect all daily closing prices for all stocks in the Greek Stock Market. We chose to investigate years 2007 and 2012 (one year before the economic crisis and one during the economic crisis) in order to explore possible different behaviours. All these closing prices are stored in a spreadsheet and processed there.

In order to find the most credible data sources we chose to use a Greek site with economic orientation (www.capital.gr). A Greek site was chosen in order to avoid problems with data discrepancies that could be induced to our data because of differences in holidays or other days with closed stock markets. 239 active stock closing prices were mined for each day that the market was open (about 250 days). For reasons of simplicity we use the code name of each stock, using English characters, since some SNA software has difficulties coping with Greek characters. From the overall 600,000 initial records, we removed all companies that were introduced in the market after (or during) 2007 to avoid data loss (7 stocks). We also removed some stocks that fall in special categories that could distort the data. Among them are stocks in suspension (39), stocks to be deleted (2) and low dispersion stocks (10). These stocks are extremely stable so they would influence the correlation measurements. After all the pre-processing, 181 stocks were finally selected and all data was inserted in a spreadsheet where rows correspond to stocks and columns correspond to their respective closing prizes.

Furthermore, it was realized that for some of these stocks there has been a short stop in negotiation and had missing closing prices. This could have happened because during those days a merge offer was set up or some other important event happened that had this effect. Again, in order to avoid data and results distortion we calculated the average closing price for these specific days (based on the previous and next actual days), so that the transition should be as smooth as possible. For smoothing purposes, we also calculated the daily logarithmic changes of every stock, I, for every day of the year, through the following equation:

\[ r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1) \]

where \( P_i(\tau) \) is the closing price of stock \( i \) on day \( \tau \) (Mantegna R., 1999).

Following, we calculate the correlation coefficient \( \rho_{ij} = \frac{\text{Cov}_{ij}}{\sigma_i \sigma_j} \) for every pair of stocks \( i, j \) where \( \sigma_i \) is the standard deviation of stock \( i \) and \( \text{Cov}_{ij} \) the Covariance of stocks \( i \) and \( j \), using \( \text{Cov}_{ij} = \frac{\overline{(P_i - E(P_i))[P_j - E(P_j)]}}{n-1} \), where \( E(P_i) \) is the average of stock \( i \) and \( n \) is the count of the population.

In this way we created a cross correlation, rectangular, one-diagonal and symmetric matrix between all stocks, with dimensions 181 X 181 and values varying from -1 to 1. When the correlation value of two stocks approaches 1, then these two stocks move similarly, that is, when the price of one increases, then the price of the second increases similarly, while when the correlation coefficient approaches -1, then the process of the two stocks vary inversely (increase for the first means decrease for the second). When the correlation coefficient is zero, then we cannot make any prediction, since the two stocks vary in a random way.

In order to create networks from the cross correlation matrix, at first we consider that every stock is a node in a network. We set a threshold \( \theta_i \), \((1 \leq \theta \leq 1)\) and create a link between nodes \( i \) and \( j \) if the correlation coefficient of the corresponding stocks, \( \rho_{ij} \), is greater than \( \theta \). Obviously, for different values of \( \theta \) we obtain networks with the same number of nodes but different number of links (Huang W., et al., 2009). Transforming a correlation metric to a binary format undoubtedly leads to some information loss. However, this type of transformation has been extensively used in the literature, as already noted.

In order to transform elements from the cross correlation matrix in (-1, 0, 1) form, we used the following function and applied it to every cell of our spreadsheet (in our example, cell B2):

\[
\text{If(IsBlank(Sheet1!B2);"";}\text{If(Sheet1!B2<>SA$1;1;(If(Sheet1!B2<-SA$1;-1;0))})
\]

where cell A1 contains the different values of \( \theta \). In this way we transform the cross correlation stock matrix to an adjacency matrix. Finally, we used add-on Kutools to change the array’s dimensions so that Gephi could recognize changes in the matrix’s dimensions and software Excel2pajek in order to change spreadsheets to .net files (network files). Finally we calculated all metrics with Gephi and used the same software to produce or following visualizations of our networks, through the use of Force Atlas (Golbeck J., 2013) and Label Adjust algorithms.
3.2. Data Analysis

We now proceed to analyse correlation networks of stock for different values of threshold \( \theta \), as mentioned in the previous section. A large number of networks were produced for all centrality metrics, for both years of study and for different \( \theta \) values. All these networks were analysed, using all centralities mentioned before, however we will present the most important (to our opinion) aspects of this analysis. We will present year 2007 together with the corresponding network for year 2012, for comparison reasons.

In each of the following Figures, according to the centrality metric used, most important nodes (stocks) are shown with larger size and/or darker gray levels. Thus, in Figure 1 we present stock correlations for \( \theta > 0.1 \), where betweenness centrality is used, according to the algorithm from (Brandes U., 2001). In Figure 2 we use closeness centrality, with a raised \( \theta \) value (\( > 0.2 \)), while in Figure 3 we use \( \theta > 0.3 \) and clustering coefficient which actually shows in how many closed triangles does a node belong, according to the appropriate algorithm from (Latapy M., 2008). Figure 4 present the network of stocks with size and grey-levels according to degree centrality, while in Figure 5 (and all subsequent ones) we use eigenvector centrality, a metric similar but more advanced than degree centrality. In the last Figure (7), instead of raising \( \theta \) by 0.1, we present the network created by negative values of \( \theta \) (\( < -0.1 \)).

From Figure 1 it is easy to deduce that the threshold value of 0.1 is too small to produce meaningful results. The networks produced are extremely dense, especially for year 2007 where density is calculated to be 0.748. On the left side we see year’s 2007 correlations, with most important nodes to be Alapis Ind.& Comm. SA, MLS Informatics SA and Alpha Bank SA, while on the right side (2012), the most important nodes are Elgeka SA, Etem SA and Elval SA.
In the following figures, densities are reduced since larger $\theta$ values correspond to smaller numbers of links. Figure 2 was created using closeness centrality, and most important nodes lie in the center of networks. For 2007, these nodes were National Bank SA and Dionic I.&C. SA, with no significant differences in centrality values than the following in the re-ranking. In 2012 it seems that since the initial network is split in subnetworks, more important nodes belong to smaller subnetworks. These nodes-stocks are calculated to be Mytilinaios SA and Chalkor SA.

No significant results come when we increase $\theta$ to 0.3, regarding clustering coefficient. There exist 16 to 17 stocks with a coefficient equal to 1 and many more with coefficient close to 1.
In Figure 4, we keep increasing $\theta$ by 0.1 and show most important nodes according to eigenvector centrality. For 2007, the most important nodes were Chalkor SA and Etem SA, while for 2012 Mytilinaios SA and Sidenor SA.

When increasing $\theta$ values, many subnetworks are created, especially for year 2007. This is a “healthy” situation, since it is reasonable to expect few highly interconnected clusters in a healthy market. Most important stocks for 2007, according to eigenvector centrality were Sidma SA and (again) Chalkor SA (both in constructions industry). Instead, during 2012, with the economic crisis in its high impact, stocks keep clustering together, behaving as a flock or herd. Mytilinaios SA and GEK Terna SA, were calculated to be more important in 2012.
For $\theta > 0.6$, densities keep reducing. Only 11 stocks were found to be connected for year 2007, with most important according to eigenvector centrality to lie to the bottom left of the network. For 2012 we still see the herd phenomenon, with most important nodes to be National Bank SA and Alpha Bank SA.

In an attempt to explore the situation for negative values of the correlation threshold ($\theta < -0.1$), meaning that a line is drawn when stocks move in opposite directions with respect to their closing prices, we produced networks in Figure 7, with respect to eigenvector centrality. In these networks, for year 2007 most important nodes are Elviemek SA and Proodeytiki SA, while for 2012 we found Imperio-Argo Group and Kordellou Bros SA. Clearly, the “herd” phenomenon is present, since there exist too few stocks with negative correlation, compared to the positive correlation ones. In 2007, the ratio of negative to positive correlations is merely 0.3%.

In Tables 1 and 2 that follow, a synopsis of the data is shown.
Table 1: Synoptic Topology Chart Table 2007

<table>
<thead>
<tr>
<th>Correlation</th>
<th>&gt;0,1</th>
<th>&gt;0,2</th>
<th>&gt;0,3</th>
<th>&gt;0,4</th>
<th>&gt;0,5</th>
<th>&gt;0,6</th>
<th>&lt;0,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>181</td>
<td>176</td>
<td>150</td>
<td>92</td>
<td>30</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>Edges</td>
<td>12,179</td>
<td>7,058</td>
<td>2,383</td>
<td>342</td>
<td>33</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>Average Degree</td>
<td>134,575</td>
<td>80,205</td>
<td>31,773</td>
<td>7,435</td>
<td>2,2</td>
<td>2,182</td>
<td>2</td>
</tr>
<tr>
<td>Network Diameter</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Graph Density</td>
<td>0,748</td>
<td>0,458</td>
<td>0,213</td>
<td>0,082</td>
<td>0,076</td>
<td>0,218</td>
<td>0,05</td>
</tr>
<tr>
<td>Avg. Path Length</td>
<td>1,253</td>
<td>1,585</td>
<td>2,008</td>
<td>2,55</td>
<td>1,491</td>
<td>1,294</td>
<td>2,622</td>
</tr>
<tr>
<td>Number of Communities</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2: Synoptic Topology Chart Table 2012

<table>
<thead>
<tr>
<th>Correlation</th>
<th>&gt;0,1</th>
<th>&gt;0,2</th>
<th>&gt;0,3</th>
<th>&gt;0,4</th>
<th>&gt;0,5</th>
<th>&gt;0,6</th>
<th>&lt;0,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>181</td>
<td>132</td>
<td>68</td>
<td>44</td>
<td>34</td>
<td>23</td>
<td>175</td>
</tr>
<tr>
<td>Edges</td>
<td>4,007</td>
<td>1,583</td>
<td>850</td>
<td>482</td>
<td>198</td>
<td>57</td>
<td>484</td>
</tr>
<tr>
<td>Average Degree</td>
<td>44,276</td>
<td>23,985</td>
<td>25</td>
<td>21,909</td>
<td>11,647</td>
<td>4,957</td>
<td>5,531</td>
</tr>
<tr>
<td>Network Diameter</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Graph Density</td>
<td>0,246</td>
<td>0,183</td>
<td>0,373</td>
<td>0,51</td>
<td>0,353</td>
<td>0,225</td>
<td>0,032</td>
</tr>
<tr>
<td>Avg. Path Length</td>
<td>1,862</td>
<td>2,417</td>
<td>1,585</td>
<td>1,569</td>
<td>1,825</td>
<td>2,128</td>
<td>3,079</td>
</tr>
<tr>
<td>Number of Communities</td>
<td>3</td>
<td>15</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

4. Conclusions

In this paper, our motivation was to analyse the interconnections in the Greek Stock Market for years 2007 and 2012 (before and within the economic crisis). Furthermore, this analysis uses techniques from Social Network Analysis, include visualization algorithms and calculations of well-established metrics from this scientific paradigm. In order to ensure concrete resulting, we used five different centrality metrics for different degrees of stocks correlations, namely degree, closeness, betweenness, eigenvector centralities, together with clustering coefficients (a type of local centrality).

We first created cross-correlation tables for a large number of stocks and then filtered them using an increasing threshold value. Each table became an adjacency matrix, which in turn was regarded as a network. A total of about 70 different networks were analysed, while some of them were presented in the body of the paper.

Our results had significant variability, with respect to the metric used and the threshold on correlation applied, especially when the correlation threshold was small (but positive), hence the number of correlated stocks was large. If we want to rank stocks in order to distinguish the most important for each year of study (meaning that their fluctuations had large similar effects for many other stocks), these would be “National Bank of Greece S.A” and “Chalkor S.A”. (for 2007), while for 2012 these would be “Mytilinaios S.A.” and “GEK TERNA S.A.”.

From Tables 1 and 2, we can easily see that in 2007 there has been a larger number of correlated stocks compared to 2012, regarding small correlation thresholds. On the other hand, in 2012 there have been stocks with more concrete common fluctuation. An example of this result can be seen in Banking stocks, which in 2012 were very closely connected for correlation thresholds larger than 0.6 (Figures 5 and 6). This shows that the economic crisis had a small effect in the common fluctuation of many stocks but has large effect in creating communities of stocks belonging in the same sector.

When examining Figure 7, where a network of negative correlations is shown, we can clarify which ones were the stocks with opposite variations in their closing prices. For year 2007 it seems to be “Elviemek S.A.” and for 2012 it is “Imperioc-Arge Group”. In normal markets, it should be expected that there should have been many more stocks that would fall in this category. This could be used by a conservative investor who would like to minimize risk in his portfolio. However we should point out that this result should be carefully studied, since it is very well known that the historical data cannot make us certain on future gains. Minimizing risk in such investments should take into account many other factors.

From Figure 7 again, we can observe that there have been a very small number of negatively correlated stocks in the Greek Market, just 41 stocks. If we see this number in percentage over all positive correlations, then the result becomes 0.3%. This is a concrete result that means that the Greek Stock Market follows the “herd” rule to its
extreme. Furthermore it is a proof that this market is “shallow”, since it means that a small number of strong investors have large impact in the common fluctuation of many stock prices.

References


