Diving into the Greek Stock Market: A Network Analysis Application

Ilias Zosimadis, Ioannis Deliakos

June 28, 2018

Abstract

Social network analysis has become increasingly popular over the past few years. With the emergence of big data, modelling a real-world phenomenon as a network has never been more compelling. The findings of this paper act as supporting evidence to the claim that social network analysis is capable of providing with insights into underlying trends as well as forecasting future behavior within the network.

In this study, we build a network for the Greek stock market based on the correlation of different stock returns. We investigate all stocks in the Greek Stock Market for years 2000 to 2017. These networks are visualized and evaluated using Social Network Analysis methods. Gephi is used in order to detect evolving communities during different time periods and review the structure of the network in the aftermath of the financial crisis. Furthermore, we present a centrality-based portfolio management strategy which is found to be superior to a random portfolio selection.

1 Introduction

Network analysis techniques along with graph theory concepts, have been widely used to model a variety of real-world networks. These networks tend to be naturally dynamic and complex which makes them hard to examine and interpret. In an attempt towards a better understanding of their internal structure and dynamics, we may measure several well-known graph properties which in turn enable us to characterize them by comparing them to benchmark networks. Diving deeper into the network, it is very common to present some of its most notable entities in terms of their influence, centrality etc. Visualization is also a crucial part of the process. Depending on the situation at hand, finding an appropriate way to visualize the network is the key not only to emphasize the numerical results, but also to provide with an intuitive view of the network. Eventually, that may turn out to be an indication of other interesting observations or hidden patterns inside the network, which can set the scene for further research. Applications include multiple kinds of networks formed in biology, finance, physics, sociology, computer science and more. Actually, there is no limit to what can be modeled as a network as long as the nodes are well-defined entities and the edges correspond to interactions among the entities in a comprehensive manner.

In this study, we have applied some of those analytic tasks described above, in order to explore the behavior of the Greek stock market during the last two decades. We are specifically interested in the communities that arise among stocks with similar fluctuations in their price returns. We also focus on the impact of the financial crisis on the market. Last but not least, we investigate the possibility of building and optimizing a portfolio selection using the constructed networks. The rest of this paper is organized as follows. In section 2, we summarize the related work that has already been done both in Greek and foreign literature. Section 3 explains the data collection process and the necessary assumptions made for the purpose of creating the network, which is described in detail in section 4. The results of this work are summarized in section 5 and finally this paper ends in section 6 with some concluding remarks.

2 Problem definition

There are lots of things to consider when dealing with a stock market network. Each and every one of its different aspects, inevitably raises different kinds of questions which can only be answered after thorough case studies and empirical evidence. In recent literature, there is relevant research to be found, trying to shed some light into the following challenging problems. The main approach, followed by most researchers, suggests stocks to be represented as nodes themselves whereas edges are defined based on pairwise correlation among stocks over a certain period of time. Along these lines, a given network would have great potential for knowledge discovery. Some topics worth of investigating are covered in the remaining of this section.

2.1 Community Detection

A community in the context of graphs refers to a group of nodes that are densely connected and share common characteristics. A maximal clique can be thought of as a community since all nodes belonging to the clique are connected to each other. But this condition is not necessary. In real-world networks, communities are formed and evolved through time as certain events occur. Nevertheless, being able to detect separate communities in a graph is of great significance. It gives us a better understanding of its topology as well as its functionality. Additionally, detecting the communities within a network can be exploited in order to monitor the network in terms of finding missing links, identifying fraud, propagating or controlling the spread of information as appropriate depending on certain needs. A variety of algorithmic techniques have been proposed to solve this problem including hierarchical clustering, modularity maximization, Girvan-Newman etc.

Regarding the case of the stock market network, we naturally expect that stocks which belong to the same or relevant business sectors, would be grouped into the same community. In [STY15], researchers were able to prove that assumption for the United States stock market. Even though the topology of their network was entirely based on stock performance correlation, the resulting communities turned out to be consistent with the Standard Industrial Classification system which serves to classify stocks based on their physical functionality. That result demonstrates the fact that fluctuations within the same industry area tend to be positively correlated. Stock performance can be measured by the corresponding stock prices, trading volumes, price returns or a combination of these values. We use Gephi to detect evolving communities in the Greek stock market during different time periods. The results are evaluated and visualized appropriately.

2.2 Financial Crisis and the stock market

Without a doubt, financial crisis had its implications on the stock market. Investors obviously tend to be more reserved during a crisis as risk levels are high and market volatility increases. However, it is quite interesting to study the stocks interactions during such a period and determine the exact transformation in the market behavior. To illustrate this impact, the main idea lies into creating two separate networks of the same market and compare their respective properties. The first network has to correspond to a time period before the crisis in order to depict a healthy market state and the second one is constructed during the crisis. This approach has been applied to the Greek stock market by [DV15], the authors came to the conclusion that at the time of credit crisis, negative price return can spread over the network starting from some initially infected stocks which used to belong to different communities.

2.3 Portfolio Management

The process of allocating funds for investment in different assets (usually stocks) is common for an investor. The selection of these stocks is known as the portfolio. The investor is interested in maximizing the expected return while simultaneously reducing the financial risk as it is possible, which leads to a constraint by definition. One can calculate the expected return of a stock along with the standard deviation but that alone is not sufficient to solve the problem. Network analysis may be able to help in the portfolio optimization problem as suggested in [STY15] by entering more parameters into the equation. It has been proposed that investing in independently fluctuating assets reduces overall risk. In that spirit, selecting the most central nodes from the communities detected in the network may result in a better portfolio. However, centrality can not be defined strictly for a node in a network. Different notions of centralities are popular in literature including degree centrality, betweeness centrality, closeness centrality, eigenvector centrality and more. A combination of these metrics might be the way to go in most scenarios. A proper weight assignment to these centralities has proved to outperform SP 500 index according to [RLL+16].

3 Data Summary

The dataset used in this study has been collected from capital.gr, a popular Greek website providing financial news and articles. In particular, we gathered data related to all Greek stocks starting from year 2000 up to 2017. A different network can be constructed for every one of these years, but we were also interested in experimenting with 9 different values of a parameter named θ, which is explained in the next section. That leads to the creation of $18 \times 9 = 162$ networks. Daily closing prices were mainly used after an initial preprocessing. Not all days were taken into account. Days like weekends, public holidays etc were ignored since the market was closed. We also had to deal with special occasions (less than 5%) during which specific stock prices were not available regardless of the fact that the market was active. In order to maintain a smooth dataset, these closing prices had to be estimated. We used linear regression as a predicting model for the missing value of the closing price (dependent variable). The recent closing history of that stock acted as the observed dataset. Lastly, Greek stock initials were translated to Latin characters for the purpose of being integrated into software packages like Gephi without any possible encoding issues.

4 Network Construction

4.1 Process Followed

Even though there is no unique approach for the construction of a stock market network, the most commonly used one relies on pairwise correlation of the stocks. The basic methodology is similar as described in [HZY09]. Correlation can be measured in terms of stock return, trading volume or other representative attributes of the stock's temporal behavior. We chose to utilize the closing prices of the stocks and compute the logarithmic price return which is shown below.

$$r_i(t) = ln(\frac{p_i(t)}{p_i(t-1)}).$$

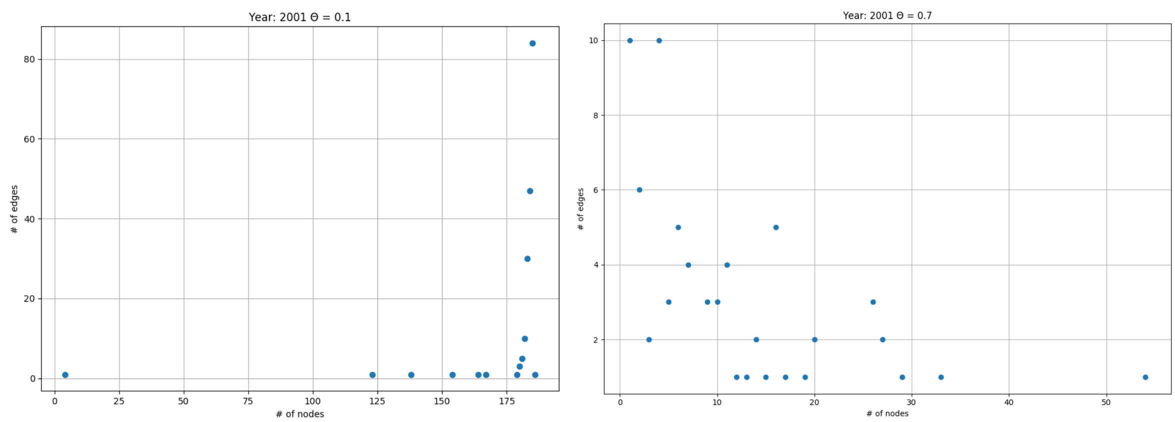
These values formulate a sequence that can be thought of as a time series for every stock within a specific time period. Therefore, different networks may be built by applying the same reasoning for different calendar years. The nodes of the network remain the same, since they always represent the stocks. However, edges are defined by the correlation among the stocks, which means that they are not consistent for every year. Pearson correlation coefficient is used to compute the correlation between a pair of stocks i and j.

$$c_{i,j} = \frac{\sum_{t=1}[(x_i(t) - \bar{x_i})(x_j(t) - \bar{x_j})]}{\sqrt{\sum_{t=1}(x_i(t) - \bar{x_i})^2} \sum_{t=1} \sqrt{(x_j(t) - \bar{x_j})^2}}$$

High values of this coefficient indicate strong correlation between i and j and it makes sense to believe that the two stocks are connected by an edge. We set a threshold variable ($-1 \leq \theta \leq 1$) in order to filter out all weak correlations. Although we are able to adjust the density of the constructed network by configuring the value of θ, there is no point in decreasing the value of θ too much, because that would result in a network almost fully connected, while edges would lose their semantic, keeping in mind that the graph is not weighted. Other approaches to edge definition take into consideration detrended covariance and time-lag correlations extending the assumption that influence in price fluctuations does not have an immediate effect on stocks, but can be observed in the near future.

4.2 Power Law Distribution

Interestingly, we may notice the fact that as we increase the value of the correlation threshold, the constructed network begins to resemble a scale-free network. Such a network is supposed to be characterized by a degree distribution that follows a power law. A power law usually describes a phenomenon according to which a small group of items is responsible for a large amount of occurring events while the rest of the items are connected to just a few events of the same type. Some examples of phenomena with power law distributions include the magnitude of earthquakes, word frequencies in a document, size of cities according to population, distribution of income etc. In our dataset, many nodes with just a few links indicate that a relatively small number of stocks are capable of influencing the market. These stocks have a relatively high degree and this is one of the driving reasons why we decided to use degree centrality in our portfolio management strategy. Similar conclusions have been drawn from other publications in the literature like [TLL10]. To illustrate the point made above, we construct two networks in 2001 and compare the degree distribution of a low threshold value $\theta = 0.1$ (a) and a high threshold value $\theta = 0.7$ (b) accordingly.



(a) Low threshold (0.1)

(b) High threshold (0.7)

Figure 1: Comparing the degree distribution of the Greek stock market network in 2001 for different correlation thresholds

In figure a) we can actually see most of the nodes having a high degree which means that the network has a high density. This network though, is not an accurate depiction of the stocks correlation. In contrast, it looks like a power law holds in the network corresponding to figure b). The latter network is definitely a better representation of reality, since only strong correlations are taken into consideration.

5 Results

5.1 Communities

In this part of the paper, we present the results obtained after importing most of the networks in Gephi, an open source graph exploration and visualization software. During this analysis, we came across an interesting fact according to which most of the members within a particular community, tend to belong to the same or similar industrial area. By all means, this is not revealing as it has already been observed by other researchers in the past.

For instance, we may examine the constructed network for year 2004. Correlation threshold has been set to 0.5 in order to retrieve meaningful results from the network, while keeping its size large enough for the sake of an example. A community detection algorithm is applied to this network. Gephi is able to compute numerous properties of the network. Modularity is measured to be around 0.6 which is considered to be a relatively high value, verifying the existence of communities. A low modularity value would indicate a random link distribution. Every node in the visualization is marked with the color of the community it belongs to. If we take a closer look at the orange community, we may verify that EKTER, AEGEK and ATHINA have been arranged in the same community. All three of these stocks correspond to companies related to construction works and technical offices. Nonetheless, the remaining two members of the orange community (KOB and KOUAL) are both related to computer applications and software development. As a result, we may infer that communities are usually formed by companies of the same business sector, but this is not always the case. Either way, resolution is another parameter that can be configured and produce smaller or larger communities accordingly. In this example, resolution was set to 1.0, but it is possible that setting a lower value would split the resulting communities in a more intuitive way.

5.2 Crisis implications on the market

In order to study at what extend financial crisis was able to impact the Greek stock market, we carried out a couple of experiments on the structure of the evolving networks through the pass of the latest 18 years. More specifically, we detected the resulting communities for all these years and plotted the evolution of the number of communities through time. Evidently, there is a huge drop in that number starting from 2008 almost when financial crisis made its first appearance in Greek economy. This drop does not require a lot of investigation. It can be easily explained by the fact that Greek stock market is considered to be a shallow market, following the herd rule as stated in [DV15].

The same procedure was followed to plot the number of connected components during the same time period and the results are comparable. Both of these network

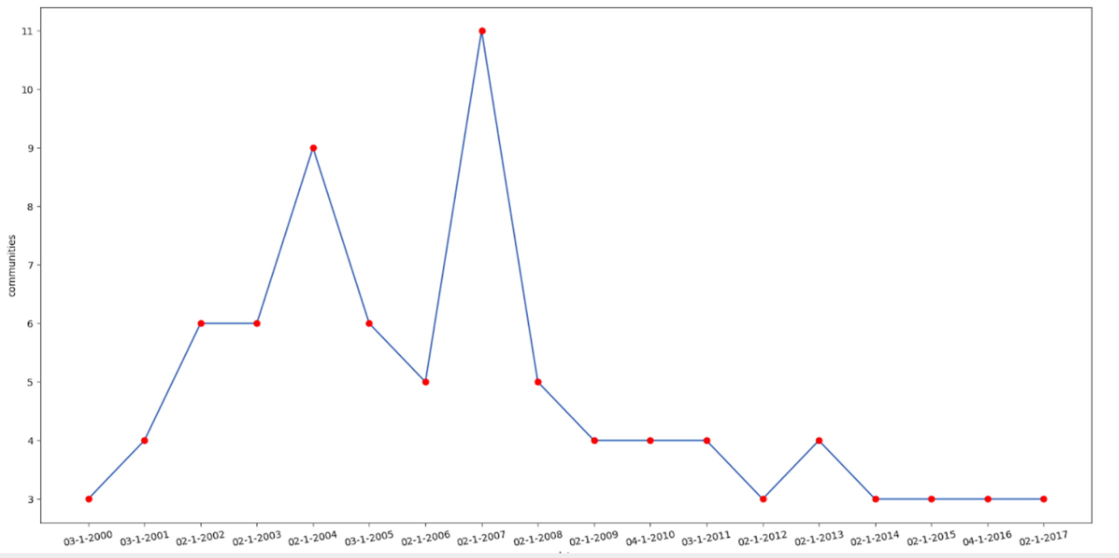
Figure 2: Detected communities in 2004

properties are semantically related after all. There is only one connected component during the first years of the plot which is consistent with a Greek stock crash that happened in year 1999 and lasted until 2003. The following 5 years display a healthier market state, since more connected components emerged, meaning that some stocks were able to function independently of the rest of the market. Once again, financial crisis forced the whole market to merge into one gigantic connected component which means that most of the stocks were heavily influenced by strong investors and the fluctuations of their corresponding stocks.

5.3 Portfolio Management

Portfolio management is an investment strategy that is used by investors while trying to maximize their gain. A naive policy would favor any random partition of the capital into buying shares of multiple stocks without any kind of further sophisticated thinking. A better strategy that is widely adopted by lots of investors today, suggests that not all stocks are equally important but some of them can be proved to be more profitable than others. These stocks are usually included in stock indexes and correspond to large and influential companies. The basic fundamental premise

Figure 3: Number of communities evolving through time



is that a portfolio consisting of these central and stable stocks is more likely to result in reducing the risk and maximizing the profit of an investor in the long term.

In this work, we propose the application of network analysis techniques in this portfolio selection problem. Using the networks described in section 4, we're able to extract centrality features like Betweeness Centrality (Cb), Closeness Centrality (Cc) and Degree Centrality (Cd). The next step would be to reduce the portfolio selection problem into a maximization problem using the following objective function.

$$C = w1 \times Cb + w2 \times Cc + w3 \times Cd$$

where w1,w2 and w3 are the decision variables that need to be fine tuned. C denotes the average centrality of a stock and the sum of all average centralities included in the portfolio has to be maximized. To solve this problem, NSGAII [DPAM02] was used (a multi-objective genetic algorithm) . The proposed portfolio consists of the top K stocks for which C has the maximum value.

In most of our experiments, it turns out that our proposed method is able to build a superior portfolio in comparison with the random strategy. For instance, let us assume that we decided to invest into three separate stocks and our initial capital is equal to 1000 Euros per stock. The results of this experiment are presented below .We constructed a network for year 2007 with $t > 0.5$. Our algorithm selected to invest in ETE, TROP and LYK , whereas a random portfolio consists of DION, INTKA and INKAT.

Figure 4: Number of connected components evolving through time

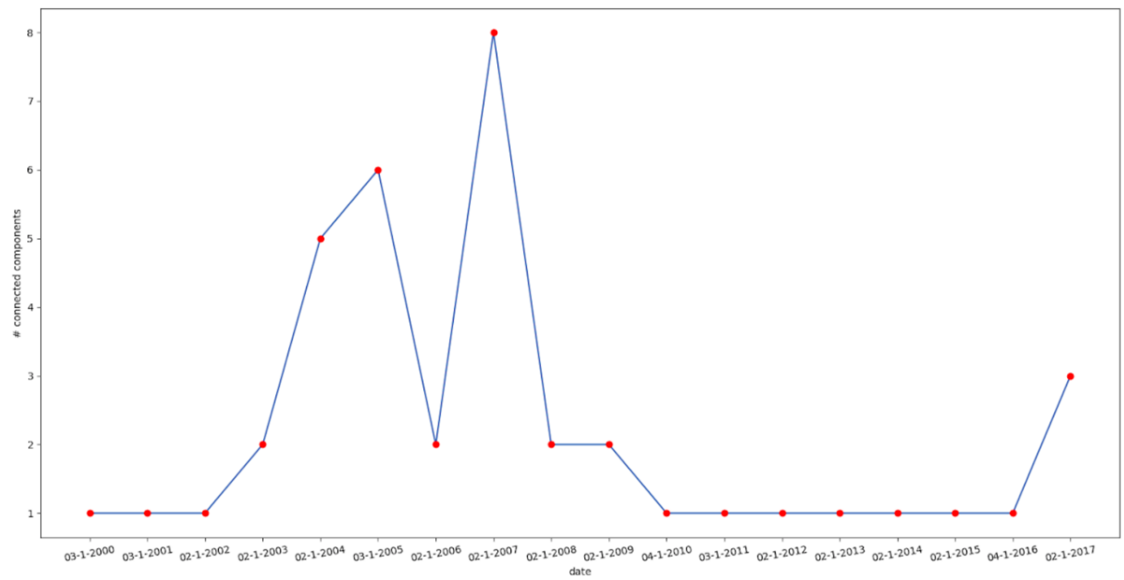
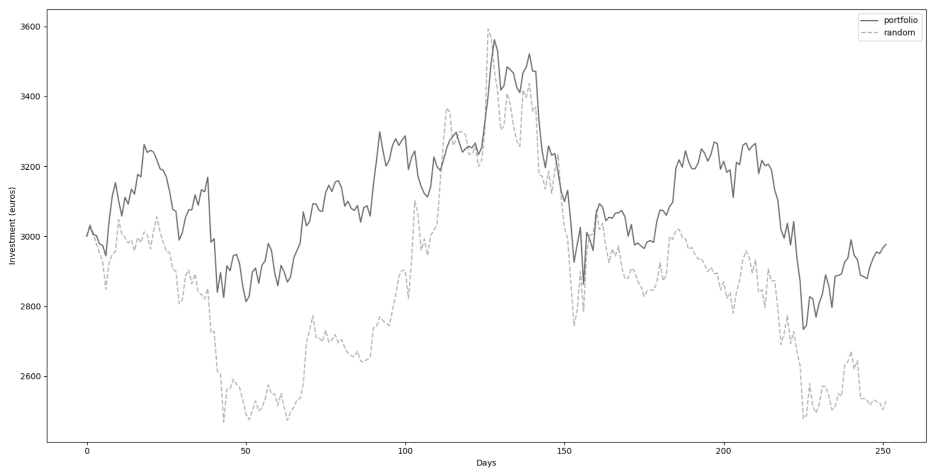


Figure 5: Portfolio selection year 2007



6 Conclusion

As this paper demonstrates, network analysis is a powerful tool when it comes to model a real-world phenomenon. After analyzing the network, one is able to identify some patterns and derive significant conclusions which can result in confident estimations regarding the future behavior of the network. The Greek stock market has been characterized as shallow, since it looks like it is easily affected by a few strong investors. Financial crisis also influenced the market by reducing the detected communities as well as the number of connected components.

Is network analysis the right tool to optimize a portfolio? By all accounts, it may serve as a decent alternative. The suggested method relies on different notions of node centrality. The comparison with a random strategy revealed that it can actually achieve relatively higher performance. It can also be combined with other statistical properties in order to facilitate an optimized investment. Can it compete with stock indexes? Despite new efforts to prove that assumption, more research is needed until a solid network analysis-based method is presented which would hopefully be able to simultaneously outperform the majority of the popular stock indexes.

References

[CLL10] K Tse Chi, Jing Liu, and Francis CM Lau. A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659–667, 2010.

[CM16] Paolo Coletti and Maurizio Murgia. The network of the italian stock market during the 2008–2011 financial crises. *Algorithmic Finance*, 5(3-4):111–137, 2016.

[DPAM02] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, Apr 2002.

[DV15] Kydros Dimitrios and Oumbailis Vasileios. A network analysis of the greek stock market. *Procedia Economics and Finance*, 33:340 – 349, 2015. The Economies of Balkan and Eastern Europe Countries in the Changed World (EBEEC 2015).

[Gre93] George D. Greenwade. The Comprehensive Tex Archive Network (CTAN). *TUGBoat*, 14(3):342–351, 1993.

[HZY09] Wei-Qiang Huang, Xin-Tian Zhuang, and Shuang Yao. A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 388(14):2956 – 2964, 2009.

[RLL+16] Fei Ren, Yanan Lv, Sai-Ping Li, Xiongfei Jiang, Li-Xin Zhong, and Tian Qiu. Dynamic portfolio strategy using clustering approach. 12, 08 2016.

[STY15] Wenyue Sun, Chuan Tian, and Guang Yang. Network analysis of the stock market. 2015.

[TLL10] Chi K. Tse, Jing Liu, and Francis C.M. Lau. A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659 – 667, 2010.