

# Visualization of the UK Stock Market Based on Complex Networks for Company's Revenue Forecast

Ziyi Wang and Jingti Han

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China  
zy\_wang1008@outlook.com, hanjt@mail.shufe.edu.cn

**Abstract.** As an emerging research field, the complex network theory is able to depict the most daily complex systems' topologies, but in terms of financial market analysis, it still needs more attention. We can apply this theory to construct financial networks and detect them both from macro level and micro level to support a company in forecasting its revenue. This paper aims to explore the macro-characteristics of the UK stock market. We examine the properties of return ratio series of selected components in FTSE100 index, adopt the Kendall's  $\tau$  rank correlation coefficient between series to write adjacency matrices and transform these matrices into complex networks. Then we visualize the networks, analyze features of them at different thresholds and find evidence of WS small world property in the UK stock networks. All these work follow our research framework proposed at beginning of this paper. According to the framework, more future work needs to be done to achieve the goal and make decision support in a company.

**Keywords:** Complex networks, FTSE100 index, WS small world property, Revenue forecast.

## 1 Introduction

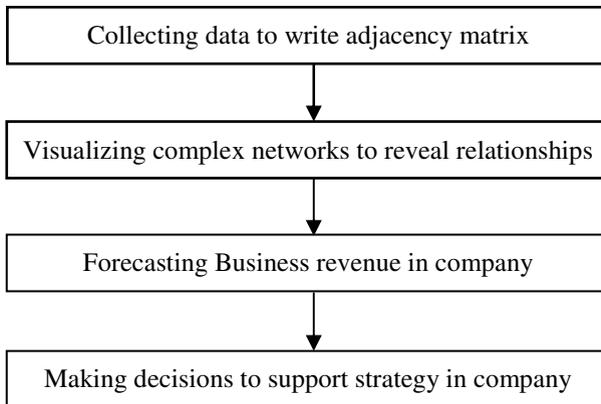
The function of the stock market is to provide a public platform for trading businesses and raise capital for companies. Various information technologies and statistics methods have been engaged in analyzing its dynamics, such as Internet, portable computers, regression analysis and so on. However, the interactions between companies are rarely taken into account, but undoubtedly they play an important role to impact the markets. From inside perspective of a company, forecasting its revenue through the stock market's behavior could be one way to support decision making. In order to achieve this research aim, we propose utilizing complex network theory to visualize and analyze the complexity of the interactions in this paper.

Complex network theory is able to abstract and depict the most daily complex systems' topologies. As an emerging hotspot, this theory has attracted attention of scholars increasingly from various research fields and became one of the most significant tools in social network analysis. Complex network theory has been introduced since decades ago. To date, it has been widely used in the study of social relationships,

transport networks, biology, ecology, emergency management, power grids, etc., and as a credible and powerful tool to characterize the markets.

Up to now, some scholars have done several researches about the major stock markets all over the world. In 1999, Mantegna [1] proposed a hierarchical tree to investigate the common economic factors affecting the stocks' prices. In 2005, Boginski et al.[2,3] constructed a market graph which followed a power-law model and calculated cross-correlations to reflect the market behavior with the data from US stock markets. Till 2009, Haldane [4], the Chief Economist and the Executive Director of Monetary Analysis and Statistics at Bank of England, first convincingly explained the current financial markets with complex networks in an official speech. Caraianie (2012) [5] analyzed the properties of the returns of European stock markets by complex networks, and found that the networks are scale-free and self-similarity. Zhuang et al. [6] constructed a network of stock prices fluctuations in Shanghai stock market and analyzed its topology. Further, Zhang et al.[7] compared the networks extracted from the original series and return series. They showed different characteristics because of different data. The former one fitted with a power law distribution and was a small-world and free-scale network, while the latter one governed by an exponential degree distribution. For Hang Seng index of Hong Kong stock market, Li and Wang[8] found fluctuation patterns based on the network topological statistic. In these researches, networks from various databases and statistic results are visualized to make texts more visually attractive and help audience understand more effectively.

The aim of this research is to apply the complex networks theory to the UK stock market to forecast companies' revenue. The research framework is shown below (Fig.1). The first two steps are contents in this paper and the others are the future work in our research.



**Fig. 1.** A basic framework to forecast company's revenue using complex networks

In this paper, we investigate the relationship between return series with observations from 2004 to 2014 in FTSE100. We find that return series show significant

non-linear correlations that do not meet strict assumption of Pearson's correlation; therefore we adopt Kendall's  $\tau$  correlation as edges' weight to construct and analyze networks more accurately.

## 2 Methodology

FTSE100 index, being one of the most professional stock market exchange indices all over the world, has 101 components from various industries. However, not every company contributes to the market during the period we chose. Hence, an analysis of these long term running companies would give us a big picture of the network structure of the British stock market.

### 2.1 Data Collection and Pre-processing

The weighted network is able to describe the nodes and relations between them much more really and clearly than non-weighted network, therefore, a whole connected, no direction and weighted network is constructed in this paper. Within different thresholds, the research will find their macro-characteristics.

The data are extracted from "Yahoo Finance", and the timespan in consideration of this paper is from 1st January 2004 to 10th October 2014. The dates of the public holidays in UK are deleted as stock market is closed, including New Year's Day, Good Friday, Easter Monday, Early Bank Holiday, Spring Bank Holiday, Summer Bank Holiday, Christmas Day and Boxing Day in each year. To reflect facts as real as we can, companies that do not have data in full period are deleted as well. Consequently, there are 86 components of FTSE100 as the vertices in network, such as BP plc (British Petroleum), LSE (London Stock Exchange Group PLC), and LLOY (Lloyds Banking Group PLC) and so on so forth. The edges are the relationships among the 86 stocks and the correlation coefficient is the weight of each edge.

Pre-processing each company's closing price data series is needed. Denoting the original daily closing price of the stock  $i$  at time  $t$  is  $P_i(t)$ . These data are calculated as

$$r_i(t) = 100 \times \ln \left( \frac{P_i(t)}{P_i(t-1)} \right), (t = 1, 2, \dots, n). \quad (1)$$

Then there will be the return ratio data series  $R_i(t) = \{r_i(t)\}, i, t = (1, 2, \dots, n-1), i = (1, 2, \dots, 86)$ , and the number of data in one series is  $((a_1, \dots, a_d), (b_1, \dots, b_d) \in [0, 1]^d) n-1$ .

The more data we have, the greater extent network can reflect the real UK stock market from results we conclude. Therefore, the components we choose have the most amounts of data during the period. Following this principle, there are 234 178 (2723 per each companies) pieces of efficient data.

## 2.2 Edges and Their Weights

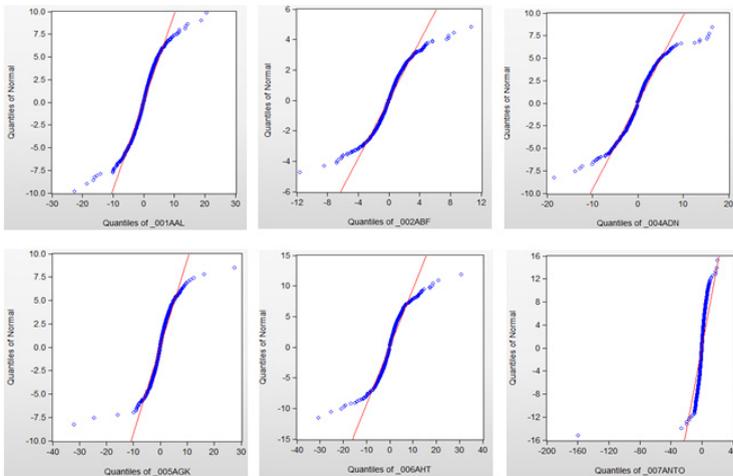
Analysing the statistical characteristics of the return series is the first thing we should do to understand the basic statistic of the return series. Six random return ratio series are shown in table 1.

**Table 1.** Basic statistics of six return ratio series

Company	Mean	Median	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	
						Statistic	Probability
<b>AAL</b>	0.003166	0.040642	2.760246	-0.13138	10.06153	5665.449	0
<b>ABF</b>	0.052881	0.062794	1.341194	-0.07136	10.78818	6884.209	0
<b>ADN</b>	0.072697	0	2.347117	0.146311	10.15676	5820.97	0
<b>AGK</b>	0.082391	0.063191	2.350748	-0.55548	29.70248	81038.36	0
<b>AHT</b>	0.151669	0	3.27993	0.003828	14.71018	15558.36	0
<b>ANTO</b>	-0.01732	0.060223	4.265728	-19.2306	719.1176	58352041	0

From this table, it is clear to see that the means are around 0 and the standard deviations are all over 1, that is to say, there are bigger fluctuations in these return series. Obviously, the fluctuations are depended by the uncertainty of the stock market. Majority skewnesses of data series' distributions, in this table, are smaller than 0, and kurtoses are greater than 6 (twice of 3), which mean the return series show left skewed (thick/fat tail) and high peak distribution. Meanwhile, the probabilities of Jarque-Bera are all 0 and do not accept the hypothesis: the return series are normal distribution. Because of these characteristics, it can be seen that these six return series obey student T distribution, the same as the rest 80 series.

The Q-Q graphs below give evidence of the judgment we made above.



**Fig. 2.** Q-Q graph of return series against normal distribution

As the graphs showed, the return series have the features mentioned above and do not accept the hypothesis of normal distribution.

To examine the correlations among the series, it is calculated the autocorrelation and partial autocorrelation of 21 period-lag of these series. The Ljung-BoxQ value is significant and the results show that there are no significant relations. Therefore, all the return series we chose are independent in the period.

According to the test results, in this paper, we employ the Kendall's  $\tau$  correlation as the edges' weights. A correlation coefficient is a measure of the consistency of changes between the variables. Pearson Product Moment Correlation or PPMC (Pearson's correlation, for short), which is a measure of the linear correlation between two variables, is common to see in the correlation research of the stock exchange markets since it is easier to understand and calculate than others. However, there are some limitations for Pearson's correlation. For example, the variables should be normally distributed and it is for linear relationship between the two variables. Kendall's  $\tau$  correlation measures the coordination degree of variables that based on the rank of random variables. Therefore, Kendall's  $\tau$  correlation coefficient is able to reflect non-linear dependences better without restriction mentioned above.

With these calculations, we can write adjacency matrices. Through the adjacency matrices, we construct complex networks and analyse them with respect to degree distribution, average shortest length path and clustering coefficient.

### 3 Results and Discussion

With Pajek, a social network analysis tool, we transform the adjacency matrices into complex networks.

#### 3.1 Network Structure Analysis from Macro Perspective

Visualising networks at a threshold of 0.36 as below: the nodes are the 86 companies from FTSE100 index and the edges are determined by the Kendall's  $\tau$  correlations between the series. The network is the whole connected weight network.

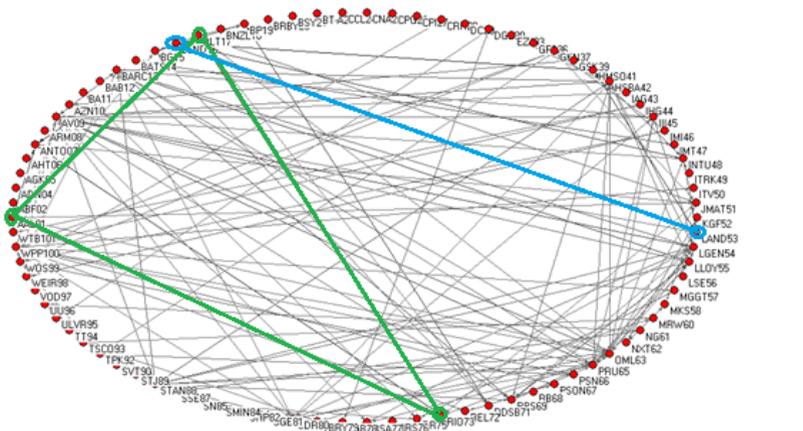


Fig. 3. Complex Network of FTSE100 index at a threshold of 0.36

In Fig.3, the darker edges represent the bigger correlation coefficient between the companies. It is noticed that five companies have the strongest relationships and these five companies are formed into two clusters: AAL01 (Anglo American Plc.), BLT17 (BHP Billiton Plc.) and RIO73 (Rio Tinto Plc.); BLND16 (British Land Co Plc.) and LAND53 (Land Securities Group Plc.). The first three companies highlighted by the green are all multinational mining companies. The last two highlighted by the blue are all the property companies and they all turned into a real estate investment trust (REIT) in January 2007. British Land is the second largest REIT in the UK and Land Securities is the largest one. Undoubtedly, the companies in the same industries would be easy to have strong relations.

In order to describe a network from the macro aspect, it is usual to characterise its density and scale. Clustering coefficient and average path length are the indicators to measure these two features.

Clustering coefficient reflects that how close a vertex's neighbors are being cliques, which is the density of a network. Assume in a network  $G$ , a vertex  $i$  has  $k_i$  edges to connect other  $k_i$  vertices, and these  $k_i$  vertices named neighbourhood of  $i$ . Therefore for each vertex  $i$ , there are at most  $k_i(k_i - 1) / 2$  links that could exist in an undirected graph, whereas real number of edges of neighbours in this undirected graph is  $N_i$ . Then the clustering coefficient  $C_i$  is:

$$C_i = \frac{2N_i}{k_i(k_i - 1)}. \quad (2)$$

The overall level of clustering  $C$  in a network is defined by Watts and Strongatz [9] as the average of all the  $C_i$ :

$$C = \frac{1}{N} \sum_{i=1}^N C_i. \quad (3)$$

Obviously,  $0 \leq C \leq 1$ . If  $C = 0$ , all the vertices in the network are independent. If  $C = 1$ , any two nodes are directly connected in a network, i.e. the network is a global coupling network.

Average path length is defined as the average number of steps for all possible pairs of network nodes. It is a measure of the efficient of information transport on a network, i.e. the scale of a network. Suppose in a network  $G$  the length of any two nodes  $i$  and  $j$  is  $l_{ij}$ , which is the shortest number of steps from node  $i$  to  $j$ . The average length of a whole network  $L$  is

$$L = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N l_{ij}. \quad (4)$$

$N$  is the number of vertices in a network. We do not consider vertices' own distance in this paper.

Clustering coefficients and average path lengths are calculated at different thresholds with data mentioned before, showing in table 2.

**Table 2.** Clustering coefficient and Average path length at different thresholds

Thresholds	Clustering coefficient	Average path length	Number of edges
<b>0.25</b>	0.7848	1.6426	3104
<b>0.27</b>	0.7278	1.8666	2282
<b>0.30</b>	0.6463	2.1320	1206
<b>0.33</b>	0.5772	2.4169	594
<b>0.36</b>	0.6015	2.8870	250
<b>0.40</b>	0.6744	1.3582	86
<b>0.50</b>	0.8571	1.1579	32

In table 2, at various thresholds, the FTSE100 network has a small average shortest path length and a large clustering coefficient. With these two characteristics in a network, referencing the research did by Watts and Strogatz[9,10]in 1998, it is a Watts-Strogatz small-world network (WS small-world network for short). With these features, the fluctuation in this stock exchange market would spread rapidly. Specifically, it is found that sharp changes of some influential companies' stocks will spread at a faster pace to a wider range. These two characteristics, small average path length and large clustering coefficient, are two indicators to measure the small-world property. WS small world network is a transition from a complete regular network to a complete random graph.

### 3.2 Thresholds in Networks

Reasonable threshold selection is a crucial step for constructing networks. If a selected threshold is too small, there would be false associations due to the random noise, resulting in the complete connected network; if the selected threshold is too large, more isolated points will make the existence of the network, the network will contain fewer vertexes, and hence lots of important related information will be lost. Under these situations, it is no need to analyse the network degree distribution and clustering coefficient any more. In addition, according to the same interval, when the thresholds change from small to large, if the number of relations among stocks decreases rapidly, the left companies are more significant than others to contribute to the national economy or local economy. Otherwise, there are no more huge differences among companies to make contribution to a nation or a region. So the research with appropriate thresholds can explore real relationship among stocks to the whole system.

In this paper, the research draws 26 complex networks at 26 thresholds respectively, and the number of edges shows in Fig. 4, which reveals a descending trend of the edges. As we all know, even with the same vertices, various numbers of edges can

construct various networks. Based on the specific networks, researches for parameters could be processed, such as the adjacency matrix of networks, the cumulative distribution of degree, clustering coefficient, average nearest neighbour degree, k-core and partitions and so on so forth. For the thresholds of the financial network we construct, it is steadier to the edges' number between 0.3 and 0.4.

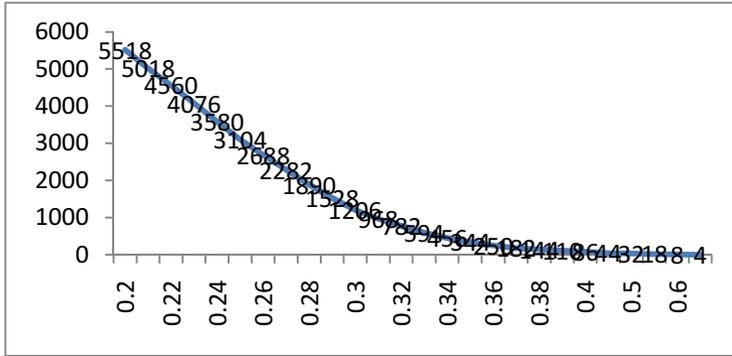


Fig. 4. The number of edges at different thresholds of financial network

#### 4 Conclusion and Future Work

As an emerging method in economic disciplines, complex networks still needs more attention. In this study, we proposed a basic framework to forecast revenue for a company to make decision support based on complex networks theory. We detected data from the FTSE 100 index and write adjacency matrices to construct networks and visualize them. Based on basic statistics of data, we explored features of networks, and found the FTSE100 index network is a WS small world network and thresholds are playing a significant part in constructing networks. We investigated the UK stock market data independently with other research data. Similarly, the WS small world properties show in US stock market, Shanghai stock market[7], international stock markets[11]and so on. Combining data analysis with visualised graphs, decision makers could do predictions much faster and more reliably to support strategies in companies. Other macro features and in-depth micro research will be conducted in the future, for example, clustering communities, detecting topology changes of networks with impacts from internal and external events in companies. Many networks from stocks markets have been studied, but there are still some important questions need to be answered: what are common features in international stock markets, what these markets differ in each country and how we can visualise them clearly to support making decision, as examples.

**Acknowledgments.** The authors would like to thank Prof. Kecheng Liu for engaging discussion. We also thank the editor and reviewers for their insightful comments and helpful suggestions. This research is supported by (1) the Research on Control and Efficiency of Emergency Support Networks (No. 71271126), financed from National Natural Science Foundation of China; (2) Study on Control and Efficiency of Emergency

Transportation Support Networks (No. 20120078110002), financed from Research Fund for the Doctoral Program of Higher Education of China; (3) the Research on Scenario Construction and Decision Model of Unconventional Emergency (No. CXJJ-2013-437), financed from Shanghai University of Finance and Economics; (4) the China Scholarship Council.

## References

1. Rosario, N.M.: Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems* 11(1), 193–197 (1999)
2. Vladimir, B., Sergiy, B., Panos, M.P.: Statistical analysis of financial networks. *Computational Statistics & Data Analysis* 48(2), 431–443 (2005)
3. Vladimir, B., Sergiy, B., Panos, M.P.: Mining market data: a network approach. *Computers & Operations Research* 33(11), 3171–3184 (2006)
4. Andrew, G.: Haldane, Rethinking the financial network (2009)
5. Petre, C.: Characterizing emerging European stock markets through complex networks: from local properties to self-similar characteristics. *Physica A: Statistical Mechanics and Its Applications*, 3629-3637 (2012)
6. Xintian, Z., Zhifeng, M., Shiyang, C.: Characterizing analysis of complex network for shanghai stock market. *Journal of Northeastern University: Nature Science*, 1053-1056 (2007)
7. Jianhua, Z., Huaxi, Z., Lu, J., Yongui, W.: Network topologies of Shanghai stock index. *Physics Procedia*, 1733-1740 (2010)
8. Ping, L., Binghong, W.: Extracting hidden fluctuation patterns of Hang Seng stock index from network topologies. *Physica A: Statistical Mechanics and Its Applications*, 519-526 (2007)
9. Watts, D.J.: Steven Strogatz: Collective dynamics of 'small-world' networks. *Nature*, 440–442 (1998)
10. Xiaofan, W., Xiang, L., Guanrong, C.: *Complex networks theory and applications*. Tsinghua University Press (2006)
11. Dongmei, H., Wen, W.: Analysis of International stock market structure and feature from the perspective of complex networks. *Complex Systems and Complexity Science*, 50-57 (2014)
12. Nick, C.: At last! A coherent account of the financial crisis (2009)
13. <http://www.finextra.com/blogs/fullblog.aspx?blogid=2835>
14. Jukka Pekka, O., Jari, S., Kimmo, K., Janos, K.: Financial Market—A Network Perspective. *Practical Fruits of Econophysics*, 302-306 (2006)
15. Kendall, M.: A new measure of rank correlation. *Biometrika* 30(pt. 1/2), 81–93 (1938)
16. Mohamed Rehan, M.S., HariHaran, M., Neha Singh, C.: Grover.D.: Visualizing the Indian stock market: A complex network approach. *International Journal of Advances in Engineering and Technology*, 1348-1353 (2013)
17. Salter-Townshend, M., White, A., Gollini, I., Murphy, T.B.: Review of statistical network analysis: models, algorithms and software. *Stat. Anal. Data. Min.*, 243-264 (2012)
18. Xiaoping, X., Huan, Y., Dingding, H.: Empirical study on Chinese stock network under financial crisis. *Complex Systems and Complexity Science*, 13-22 (2012)
19. Newman, M.: Fast algorithm for detecting community structure in networks. *Physical Review E* (2004)
20. Chunxia, Y., Yanhua, C., Lei, N., Qian, L.: Cointegration analysis and influence rank—A network approach to global stock markets. *Physica A: Statistical Mechanics and its Applications*, 168-185 (2014)