JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 35, 411-427 (2019)
DOI: 10.6688/JISE.201903_35(2).0010

Analyzing the Dynamics of Stock Networks for Recommending Stock Portfolio*

YUN-JUNG LEE[1] AND GYUN WOO[2,+]
[1]*Research Institute of Computer, Information and Communication*
[2]*Department of Computer Engineering*
Pusan National University
Busan, 46241 Korea
E-mail: leeyj01@gmail.com; woogyun@pusan.ac.kr

Traditional approaches to portfolio management and optimization often rely on the certain statistical properties, such as expected return and price variance. But these properties generally represent the local behavior of the stocks and are thus not able to represent the stock characteristics in terms of the whole stock market. This paper considers the stock market as a complex system, where stocks affect one another due to unknown forces. It may not possible to measure the actual magnitudes and directions of these forces, but it is possible to observe their effects using the external observations such as the correlation between stocks. To forecast the dynamics, this paper proposes a seminal measure, the cohesion of the stock market network induced by the correlation of stocks. The important observations we obtained from the analyses of the real market (S&P500 and KOSPI200) in the past thirteen years are two folds: (1) the cohesion tends to increase more in a bear market than in a bull market, and (2) the cohesion of the stock market Granger causes the stock returns. Based on these observations, we implemented a stock portfolio recommending system, namely *StoPoR*. To evaluate the effectiveness of *StoPoR* system, we conducted the simulated investment based on the portfolio recommended by the *StoPoR*. The result shows that the monthly returns of *StoPoR* portfolio (1.44%) is bigger than that of Markowitz efficient portfolio (1.07%) and further bigger than that of the S&P500 index (0.50%). The similar result also holds for the Korean stock market; the monthly return of the *StoPoR* suggested portfolio (1.83%) gets far bigger than that of Markowitz efficient portfolio (1.61%) and further bigger than that of the KOSPI200 index (0.81%). This result indicates that the characteristics of the stock network are related to the stock returns and its dynamics can be used for constructing a stock portfolio and the prediction of the changes in the stock markets.

Keywords: stock market network, dynamics of stock networks, stock portfolio, cohesion of stock networks, minimum spanning tree

1. INTRODUCTION

A portfolio is a grouping of stocks in which the capital among stocks is invested in such a proportion that the profit is maximum and the risk is minimum. Generally, the portfolio manager can be seen as a function of two parameters, the vector of the expected returns of the risky assets and the covariance matrix of the returns. Markowitz [1] had proposed a mean-variance model for optimizing a portfolio in 1952. This model can be used by investors to achieve desired returns from the portfolio with minimum possible risk and has been used as a practical tool for portfolio optimization.

Received September 21, 2017; revised May 31, 2018; accepted August 21, 2018.
Communicated by Tzung-Pei Hong.
* This study was financially supported by the 2016 Post-Doc. Development Program of Pusan National University.
+ Corresponding author.

There are many approaches to deal with the portfolio optimization problem such as genetic algorithms and particle swarm optimizations. The genetic algorithm has applied to solve the portfolio optimization problems since it is a good tool for optimization problems [2-6]. Sinha *et al.* [2] proposed an algorithm to create an optimum portfolio of a large pool of stocks using a genetic algorithm. The algorithm selects stocks on the basis of a priority index function designed on company fundamentals, and then genetically assigns optimum weights to the selected stocks by finding a genetically suitable combination of return and risk on the basis of historical data. According to their experimental result, the constructed portfolio with optimum weights to stocks by their algorithm beat the market for the considered holding period.

Chen *et al.* [3] proposed a grouping genetic algorithm based approach for dividing stocks into groups and mining a set of stock portfolios, namely group stock portfolio. It uses three parts, namely group, stock, and stock portfolio parts, to form a chromosome. Grouping and stock parts are used to indicate how to divide stocks into groups. Stock portfolio part is used to represent the purchased stocks and their purchased units. Lin *et al.* [4] proposed decision models for portfolio selection problems within minimum transaction lots in portfolio optimization based on Markowitz' model and used genetic algorithms to solve the models. The results of empirical studies showed that the genetic algorithms for these models can obtain near-optimal within a reasonably short time.

Also, an artificial neural network model with the Particle Swarm Optimization (PSO) algorithm has been applied to portfolio management. Zhu *et al.* [7] presented employed a PSO algorithm for portfolio selection and optimization in investment management. The asset allocation in the selected assets was optimized using a PSO based on Markowitz's theory. They tested the model on various restricted and unrestricted risky investment portfolios and compared the study with Genetic Algorithms. The PSO model demonstrated high computational efficiency in constructing optimal risky portfolios. More studies about optimizing portfolio applied POS technique are summarized Ertenlice's paper [8]. In recent years, various algorithms such as fuzzy theory and episode mining used in the field of computer science have been studied to construct a stock investment portfolio [9-13].

Traditional approaches to portfolio management and optimization often rely on the certain statistical properties, such as the expected return and the price variance. But these properties generally represent the local behavior of the stocks and are thus not able to represent the stock characteristics in terms of the whole stock market. Financial markets including stock markets can be regarded as complex systems. A central property of a complex system is the possibility of occurrences of coherent large-scale collective behaviors with a very rich structure, resulting from the repeated non-linear interactions among its constituents [14, 15]. Sometimes, this complex interaction in the stock market causes unexpected huge catastrophic events such as the collapse of the dot-com bubble in 2001 and the 2008 financial crisis [16]. Since the situation of the stock market cannot be interpreted by examining the individual stock behaviors and the external factors, it is necessary to the structure of the stock market and the correlation among stocks in the market.

Network analysis is popular to describe the characteristics or the behaviors of complex networks. Recently there has been some research conducted to model the stock market using networks starting with the Mantegna's work [17, 18]. He represents the stock market as the minimum spanning tree (MST) obtained starting from the matrix of correlation coefficients computed between all pairs of stocks of the portfolio by considering the synchronous time evolution of the difference of the logarithm of the daily stock price.

Onnela *et al.* studied the dynamics of asset trees by analyzing the distribution of correlation coefficients between stocks and their moments over time [19]. From their analysis, the stocks included in the minimum risk portfolio tend to lie on the outskirts of the asset tree. Eom *et al.* investigated the topological properties of stock networks by comparing the original stock networks and the estimated stock networks constructed with a correlation matrix and suggested that the largest eigenvalue has a significant influence on the formation of stock networks [20].

Heiberger has studied the evolution of the S&P500 with a particular focus on the dot-com bubble in 2000 and the financial crisis (2012) [21]. He used a winner-take-all approach to construct a longitudinal network of S&P500 companies and their correlation between 2000 and 2012. From the result, in those times of financial turmoil, the stock network changes its composition but, unlike ecological systems, it tightens and the dis-assortative structure of prosperous markets transforms into a more centralized topology.

Even though many research results on the complex systems of the stock network have been reported, there are few results to apply the stock selection for portfolio management. This is partly because it is not clear how to apply the research findings to the actual stock market. In this paper, we focus on the relation between the dynamics of the stock network and the rate returns of stocks in the stock network. Specifically, we pursue to find the affecting factors on the stock prices analyzing the dynamic change of the stock networks. Further, we verified these factors affect the actual stock prices for the U.S. and the Korean stock markets and show that our findings can be used to construct an efficient stock portfolio.

The remainder of this paper is organized as follows. In Section 2, we describe our data set with a statistical analysis. In Section 3, we analyze the structure and the dynamics of the stock network for the U.S. and the Korean stock markets. In Section 4, we propose a new method for recommending a stock portfolio based on the characteristics of the stock network and show the experimental result. Finally, we conclude this paper in Section 5.

2. DATA SET

Since thousands of stocks are being traded in the stock market of each country, it is difficult to analyze a stock market by considering all shares at the national scale. So we observe the stocks incorporated into a stock index representing a national stock market. We choose two stock indices, S&P500 index and KOSPI200 index, which are different in terms of the scale and the maturity of the stock market. The U.S. and the Korean stock market are considered as the developed market and the developing market, respectively.

The S&P500 index representing the U.S. stock market, maintained by S&P Dow Jones Indices, comprises 505 common stocks issued by 500 mid and large-cap companies and traded on American stock exchanges. It covers about 75% of the American equity market by capitalization. On the other hand, KOSPI200 index (Korea Composite Stock Price Index 200) is a capitalization-weighted index of 200 Korean stocks which make up 93% of the total market value of the Korea Stock Exchange. Both index constituents are updated and checked regularly using rules published by S&P Dow Jones Indices and Korea Exchange (KRX), respectively.

Two stock markets for the U.S. and Korea differ in the size and the composition of industries. S&P500 index includes 10 industries such as 'Financial,' 'Consumer Discretionary,' 'Industrials,' *etc.* The business type of target stocks is summarized in Table 1. The largest industry of S&P500 is 'Financials' comprising 92 items and covers about 18.2% of total shares. KOSPI200 is comprehensive of eight types of businesses including 'Consumer Staples,' 'Consumer Discretionary,' 'Energy & Chemicals,' *etc.* similar to S&P500. 'Consumer Staples' is the largest business type in KOSPI200 which contains 44 items and covers about 22.0% of the total shares.

Table 1. The business types of S&P500 and KOSPI200.

Business type	S&P500		KOSPI200	
	count	rate (%)	count	rate (%)
Financials	92	18.2	17	8.5
Consumer Discretionary	85	16.8	41	20.5
Industrials	71	14.1	–	–
Information Technology	67	13.3	–	–
Health Care	56	11.1	–	–
Energy	37	7.1	32	16.0
Consumer Staples	36	7.1	44	22.0
Utilities	28	5.5	–	–
Materials	27	5.3	18	9.0
Telecommunications Services	6	1.2	–	–
Information & Telecommunication	–	–	24	12.0
Shipbuilding & Transportation	–	–	13	6.5
Construction & Machineries	–	–	11	5.5
Total	505	100.0	200	100.0

We collect the stock price data for S&P500 and KOSPI200 constituents that were current as of July 1st, 2016 for S&P500 and May 13th, 2016 for KOSPI200. The daily closing price of each stock is recorded from January 2003 to June 2016, provided by Google Finance. In this paper, we use the stock logarithmic return as the change of stock price. It is defined as follows:

$$r_i(t) = \ln[s_i(t)] - \ln[s_i(t-1)] \tag{1}$$

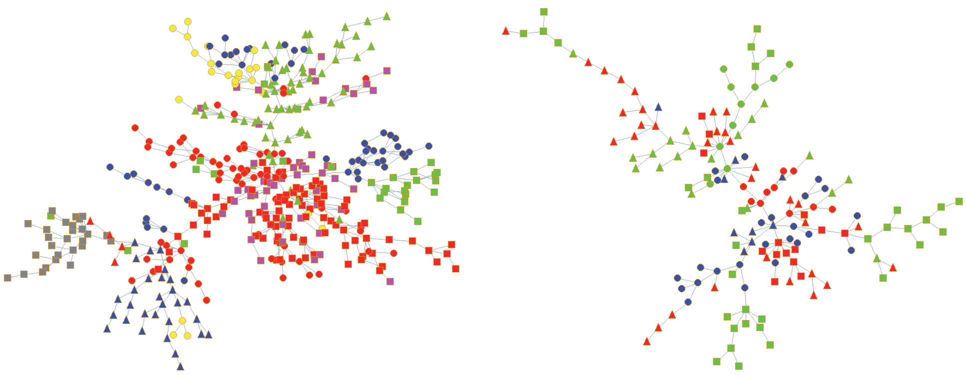
where $r_i(t)$ and $s_i(t)$ represent the logarithmic return and the closing price of stock i on day t. It is used as the time series data to calculate the correlation, and thus is served to construct the stock network.

3. DYNAMICS OF STOCK NETWORK

A network or a graph is a useful tool for representing the structure of a system or a community and grasping the connectivity among the members. A stock network graph is made up of nodes representing stocks which are connected by edges. In the stock network, each stock may be linked directly and indirectly so they affect each other through these paths. To analysis of the stock network, we construct a stock network using daily logarithmic returns of the stocks for three months every month via a sliding window. That is, the window moves to the next month after the network for the first three months is constructed.

3.1 Average Path Length of the Stock Network

To observe the closeness of the stocks by each business type, we construct minimum spanning tree (MST) by calculating the distance matrix of the stocks. Fig. 1 shows the minimum spanning tree of the U.S. stock markets and the Korean stock market represented by S&P500 and KOSPI200 respectively. The networks are made using the daily log returns from January 2003 to June 2016. A node represents a stock in the stock market and it has different shapes and colors according to the business group. In the network graph, the nodes in the same business group are mostly located near to each other. It means that the stocks belonging to the same business group are closely connected.



(a) S&P500 (b) KOSPI200

Fig. 1. The stock relation networks represented by a minimum spanning tree. It shows the relation structure of the stock market for S&P500 and KOSPI200 from Jan. 2003 to June 2016. The shape and color of a node depend on its business group.

A path in a graph is a finite or infinite sequence of edges which connect a sequence of nodes. An edge in a stock network can be considered a path for propagation of the changes of a stock market. The average path length is a concept in the network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes, which is defined as follows,

$$l_G = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}, \qquad d_{ij} = \sqrt{2(1-\rho_{ij})} \tag{2}$$

where n is the number of nodes in G. In stock network d_{ij} and ρ_{ij} denote the shortest distance and the Pearson correlation coefficient between the stock i and the stock j, respectively. $d_{ij} = 0$ if the stock i and j are perfectly correlated and $d_{ij} = 2$ if the stocks i and j are perfectly anti-correlated.

Table 2. The average path length for business types in S&P 500 and KOSPI200.

Business type	S&P500		KOSPI200	
	apl_i	apl_i/apl_{tot}	apl_i	apl_i/apl_{tot}
Telecomm. Services	1.67	0.16	–	–
Utilities	3.24	0.32	–	–
Energy	4.53	0.45	5.23	0.63
Industrials	5.78	0.57	–	–
Financials	6.01	0.59	3.13	0.38
Information Technology	6.94	0.68	–	–
Consumer Staples	7.78	0.77	11.48	1.38
Consumer Discretionary	7.89	0.78	8.97	1.08
Materials	8.38	0.83	4.75	0.57
Health Care	9.62	0.95	–	–
Construction & Machineries	–	–	2.96	0.35
Shipbuilding & Transportation	–	–	4.72	0.57
Information & Telecommunication	–	–	6.68	0.80
Average path length (apl_{tot})	10.13	1.00	8.34	1.00

Shorter the average path length of a stock network means that the stocks composing the network are closely related to each other on average. It also implies that the change of a certain stock could be spread over the whole stock market more widely. Table 2 shows the average path lengths for the business types.

The average path length of the stock network for S&P500 is 10.13. Industries in 'Telecomm. Service' have the shortest average path length. It is 1.67, which is about 16% compared with that of the total network. On the other hand, the business type with the longest average path length is 'Health Care,' which get 9.62. It is about 95% compared with that of the total network.

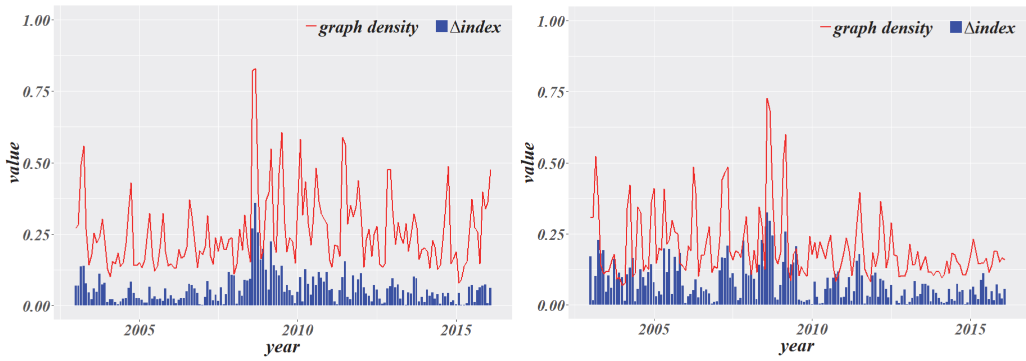
For KOSPI200, the average path length of the stock network is 8.34. The industries with the shortest average path length are in the category 'Construction & Machineries.' Its average path length is about 2.96, which is about 35% compared with that of the whole network. On the other hand, the industries with the longest average path length are in 'Consumer Staples,' which has 11.48 average path length. It is about 138% longer than that of the whole network graph. It implies the stocks in this business type does not have much higher correlation than other industries.

From the result, we found that the correlation between the stocks in the same industry in a short term maybe not high, but in a long term, stocks of the industries in the same category tend to be closely connected to each other than those across the categories. So, the stock prices tend to change similarly according to the industry category. And the average path length of each industry for the U.S. stock market is shorter than that of the Korean stock market. It implies that the U.S. stock market has stronger similarity to the Korean stock market in terms of the changes of the stocks within the same sector.

3.2 Graph Density of the Stock Network

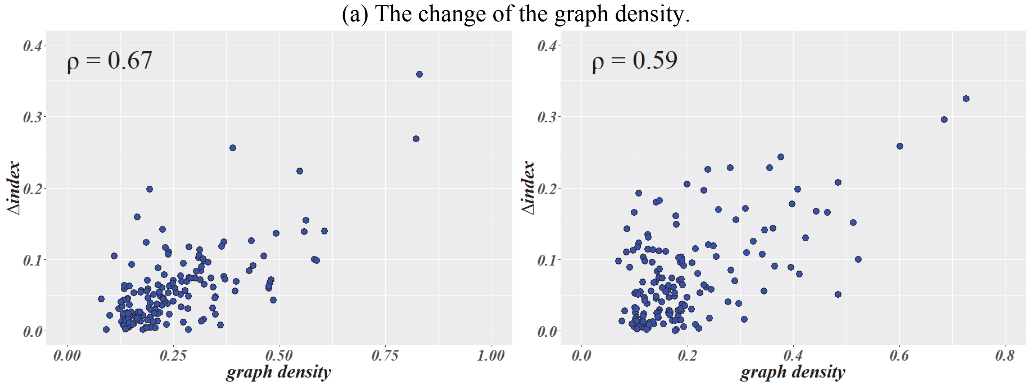
In terms of the stock market, a jumping and plunge in stock prices such as the financial crisis in 2008 is a case of bandwagon effect resulting from herd behavior. The bandwagon effect is the situation in which the tendency of "any individual adopting an opinion increases with the proportion of those who have already done so." [22] During the period of surging and crashing of the stock price, the nodes are connected very densely with each other. It can be measured by graph density how densely the nodes in a graph are connected to each other.

We analyze the correction between the graph density of the stock network and the stock index return for S&P500 and KOSPI200. The graph densities of the stock networks are measured every month. Each stock network is made using the stock log return for three months, shift next one month. The charts in Fig. 2 (a) represent the fluctuation of the graph density (the red line) and the change amount of both indices (the blue bar). The red line and the blue bar represent the change of the graph density and the change amount of S&P500 index (left panel) and KOSPI200 index (right panel), respectively. The graph density fluctuates widely over the whole period and tends to increase when



S&P500 KOSPI200

(a) The change of the graph density.



S&P500 KOSPI200

(b) The correlation between the graph density of the stock network and the return of the stock index.

Fig. 2. The figures on the left (right) panel represent S&P500 (KOSPI200). (a) The charts show the change in the graph density and the change amount in S&P500 and KOSPI200 indices. The *graph density* (the red line) denotes the density of the stock network. The amount of changes in S&P500 and KOSPI200 indices is denoted by Δ*index* in the red line. Both values are measured every quarter. (b) The chart shows the correlation between the graph density (*graph density*) of the stock networks and the range of fluctuation (Δ*index*) in S&P500 and KOSPI200 indices.

the change of the index is great especially. During the 2008 financial crisis, it is about 0.83 (0.73) at its peak of S&P500 (KOSPI200). The Fig. 2 (b) shows the correlation between the graph density and the returns of both indices. From the chart, we can find that the two factors are positively related. The Pearson correlation coefficients (ρ) are about 0.67 and 0.59 for S&P500 and KOSPI200. From the result, we can find that the correlation between two factors is stronger in the U.S. stock market than the Korean stock market.

In Fig. 2, we only consider the graph density and the change amount of the stock index which is the absolute value of the stock index return. To find the correlation of the graph density and the stock index return depending on the state of the stock market, we classify the whole period into two states such as 'bull' market and 'bear' market. Generally, a bull market is a financial market of a group of stocks in which prices are rising or expected to rise. A bear market is a condition in which securities prices fall and widespread pessimism causes the stock market's downward spiral to be self-sustaining. We consider a period with positive (negative) stock index return as bull (bear) market. The correlation between the change amount of stock index and the graph density by the state of the stock market are summarized in Table 3.

Table 3. Correlation coefficients between the change amount of stock index and the graph density by the state of the stock market for S&P500 and KOSPI200.

Index	Correlation coefficient (ρ)		
	bull market	bear market	total
S&P500	0.61	0.75	0.67
KOSPI200	0.51	0.73	0.59

The graph density is positively correlated to the change amount of the stock index in both stock market commonly, but it is a little stronger in the U.S stock market than in the Korean stock market. For the S&P500 (KOSPI200), the correlation coefficient (ρ) is about 0.67(0.59). Considering the situation of the stock market, the correlation of two factors tends to grow strongly during the bear market than the bull market of S&P500 and KOSPI200 in common. For the S&P500 (KOSPI200), ρ is about 0.75 (0.73) of bear market which is higher than about 23% (44%) that of the bear market. In addition, we measure the correlation coefficient between the graph density and the stock index return by business type, which summarized in Table 4.

In both countries, the correlation coefficient between the graph density and the stock index returns is higher in a bear market than that of a bull market. This phenomenon is common for each business types. In the U.S. stock market, the correlation coefficients between two factors of 'Consumer Staples' sector showed the largest difference according to market condition, which increases about 0.366 points from 03 to 0.366. While that of 'Construction & Machineries' sector ranks the highest increasing one in the Korean stock market, which increases about 0.335 from 0.186 to 0.521. Especially, we can see that the correlation coefficient of 'Consumer Discretionary' sector is relatively high regardless of the market condition for both countries. It can be interpreted the stocks in 'Consumer Discretionary' sector relates closely to each other, so the stock prices tend to move in a similar pattern.

Table 4. Correlation coefficients between the change in stock index and the graph density for S&P500 and KOSPI200.

Business type	Correlation coefficient (ρ)					
	S&P500			KOSPI200		
	bull m.	*bear m.*	*total*	*bull m.*	*bear m.*	*total*
Information Technology	0.477	0.676	0.568	–	–	–
Consumer Discretionary	0.471	0.753	0.578	0.579	0.696	0.617
Financials	0.437	0.700	0.542	0.313	0.439	0.357
Industrials	0.418	0.595	0.492	–	–	–
Materials	0.411	0.545	0.464	0.204	0.323	0.245
Consumer Staples	0.300	0.666	0.425	0.224	0.522	0.322
Health Care	0.299	0.645	0.474	–	–	–
Energy	0.193	0.222	0.195	0.305	0.596	0.404
Utilities	0.059	0.268	0.147	–	–	–
Telecom. Services	0.026	0.115	0.063	–	–	–
Information & Telecommunication	–	–	–	0.467	0.520	0.459
Shipbuilding & Transportation				0.308	0.459	0.363
Construction & Machineries				0.186	0.521	0.290
Average	0.309	0.519	0.395	0.323	0.510	0.382

As a result, the more fluctuation of the stock market, the stocks consisting of the stock market are denser connected each other. This is becoming serious especially in decline period such as stock market crises result in 'bandwagon effect.' The bandwagon effect in the selling of stocks is prone to occur when the stock price falls, which results in a larger decline of the stock market.

3.3 Cohesion of the Stock Network and Causality Analysis

Both the path length and the graph density in the stock network are related to the connection intensity between the stocks. In this paper, we define the cohesion of the stock network representing how closely stocks are related to other stocks. The cohesion can be calculated using the average path length and the graph density of the stock network. We measure the cohesion by business type. Consider the set of business types B. Let $coh(b_i)$, where $b_i \in B$ denotes the cohesion of b_i, which can be defined as in Eq. (3).

$$coh(b_i) = \frac{density(b_i)}{apl(b_i)} \tag{3}$$

where $density(b_i)$ and $apl(b_i)$, respectively, denote the graph density and the average path length of the stock network created with stocks in b_i. The cohesion increases if the average path length is shorter and the graph density is higher. The high cohesion of a business type means that the stocks belonging to the business type are closely connected to each other in the stock network.

The cohesion of the stock network base on a graph density is related to the change of the stock price and their correlation tends to be stronger during the bear market. We adopted the Granger causality analysis to show that the cohesion of the stock network is useful in predicting the stock returns. Granger causality is a term for a specific notion of causality in time-series analysis [23]. The concept is that a variable X Granger-causes Y if Y can be better predicted using the histories of both X and Y than it can be done using the history of Y alone. So it can be used as a statistical hypothesis test for determining whether one-time series data is useful in forecasting another. In economics, it has been applied to measure the system risk [24, 25].

We test Granger causality to find whether the cohesion of the stock market affects to stock returns and vice versa. To get the cohesion of the stock market, we have to calculate the graph density and the average path length. And the average path length is calculated with the MST, while the graph density can be calculated with the threshold network which only node pairs with more than a correlation coefficient are connected.

The threshold stock network and the MST are constructed using daily returns for three months and calculate the coherence of the stock network. Each stock networks are built every month by sliding window method and then the time series data of the coherence are made. Likewise, the three-month returns of the stock index is calculated every month. And then we test Granger causality between the coherence series and the three months returns series.

Table 5. The result of the Granger causality test between the cohesion of the stock network and the returns of the stock index.

Index	lag	$c \rightarrow r$		$r \rightarrow c$	
		F-value	Pr(>F)	F-value	Pr(>F)
S&P500	1	23.332	3.23e-06 ***	0.0128	0.9101
	2	16.256	3.965e-07 ***	1.3526	0.2616
	3	10.911	1.592e-06 ***	0.434	0.729
	4	8.015	7.136e-06 ***	1.3218	0.2645
KOSPI200	1	10.256	0.001651 ***	1.2029	0.2744
	2	6.2014	0.002573 ***	2.5121	0.0844 *
	3	4.8225	0.003106 ***	0.6332	0.5947
	4	2.5673	0.04054 **	0.4558	0.7681
Signif. codes:		0.01 ‘***’ 0.05 ‘**’ 0.1 ‘*’			

The result of the Granger causality test is summarized in Table 5. In the table, c and r represent the cohesion of the stock network and the rate return of the stock index, respectively. The significant level is presented in the last row of the table. The header $c \rightarrow r$ indicates that the cohesion of a stock market affects to rate returns of the stocks and vice versa. We test Granger causality between two factors both directions with the lag length from one to three.

In both cases of S&P500 and KOPSI200, it is commonly significant that the cohesion of the stock network Granger cause the returns of the stock index. However, the index returns does not Granger causes the cohesion of the stock network except at lag length two for KOSPI200. The results show that the cohesiveness of the stock market does not directly correlate with the stock returns, but it affects the stock returns and is useful for predicting the state of the stock market.

4. EXPERIMENTAL RESULT

This section proposes a new method for recommending a stock portfolio based on the statistical characteristics of the stock network. In addition, we show and analyze some experimental results based on the simulated investments using the portfolios constructed by the proposed method.

4.1 Construction of Stock Portfolio

Our method for recommending stock portfolio named *StoPoR* system is based on the dynamic characteristics of the stock network. Fig. 3 shows the overview of *StoPoR* system. Our implementation consists of two parts: the stock network module and the portfolio module. Consider a stock market consists of the set of stocks S, which $s_i \in S$ represents time series data with the daily logarithmic return of each stock and they are classified by its business type b_k.

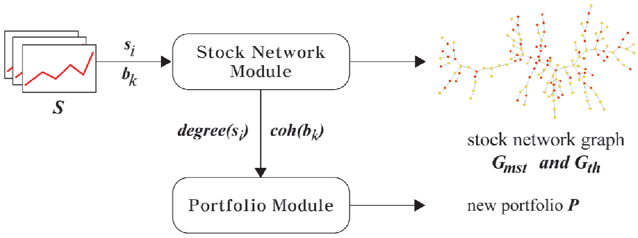


Fig. 3. The flow of the process for generating the portfolio in the *StoPoR* system. Stock Network Module creates G_{mst} and G_{th} with S and measures degree ($degree(s_i)$) of the stock s_i and cohesion ($coh(b_k)$) of the industry b_k. Portfolio Module creates a portfolio (P) containing selected stocks according to our strategy.

The stock network module constructs two type of stock networks, the minimum spanning tree (G_{mst}) and the threshold network (G_{th}) using the Pearson coefficients between stocks. The threshold network, G_{th}, includes only edges with larger coefficient than the threshold value. In this paper, we define the threshold value to the mean Pearson coefficient between stocks for the whole period.

It is calculated to $degree(s_i)$ which indicates the degree of s_i in the G_{mst}. While the cohesion of b_k, $coh(b_k)$, are calculated using G_{th}. because it can be calculated with the graph density and the average path length of b_k. The cohesion, $coh(b_k)$, is calculated by the density and the average path length. We cannot measure the change of a graph density with a minimum spanning tree because the density of a minimum spanning tree is always constant unless the numbers of nodes in the graph.

In the Portfolio module, it constructs the portfolio, P using the $degree(s_i)$ and the $coh(b_k)$ from the stock network module. According to the classic Markowitz portfolio theory, the assets of the efficient portfolio are always located on the outer leaves of the stock market network [8, 11]. It implies that the stocks have a low degree in a stock network. Hence, a stock with a low degree is a candidate to be contained a portfolio. In addition, we consider the cohesion of the business type of s_i. From our statistical analysis result, the cohesion of an industry tends to increase when the change of stock price is large especially in the bear market. Therefore, the stocks belong to the industry with high cohesion is excluded from the possible candidates. Finally, a new portfolio consists of the remained candidates.

4.2 Simulation

We conduct simulated stock investments using two portfolios generated by our proposed method and Markowitz efficient portfolio, and compare the returns on investment with the stock index. The stock data consists of daily closing prices of each firm from January 1, 2003 to June 30, 2016. We build two portfolios for the Korean and the U.S. stock market in the same manner. The stock dataset made up the stock prices of 505 and 200 firms configuring S&P500 and KOSPI200 indices, respectively.

Our trading strategy is shown in Fig. 4. A stock portfolio produced from the *StoPoR* system, $\{s_i\}$, is a set of each stock which is built monthly at the first trading day of every month using the data of the previous three months. The trading strategy adopted is that the stocks of a portfolio are bought and sold at the closing prices $p(t_s)$ and $p(t_e)$ respectively, where t_s and t_e are the first and the last trading days of the month. To show the performance of our method, we compare the stock returns of the proposed portfolio with that of the Markowitz efficient portfolio and the returns of S&P500 and KOSPI200 indices.

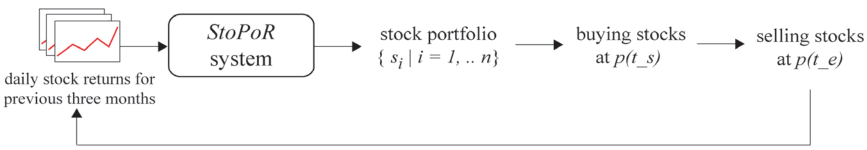


Fig. 4. The procedure of our simulated investment.

The investment returns of our method and Markowitz approach are summarized in Table 6. Due to the lack of space, only a part of the monthly returns is shown. In this table, the *PP* and the *MP* represent the portfolio by the *StoPoR* and the Markowitz efficient portfolio, respectively.

In the case of the U.S. stock market, the monthly average returns of our method is about 1.44%, which it is bigger than 0.37% point and 0.94% point to that of Markowitz portfolio and S&P500 index respectively. The case of the Korean stock market, the monthly average returns of the *PP* is about 1.83%, which it is bigger than 0.22% point and 1.02% point to that of the *MP* and that of KOSPI200 index respectively. This result shows that our method is more efficient than Markowitz approach and investment using the stock market index in the rate of returns.

Fig. 5 shows the accumulated investment returns for the whole period under observation. The line charts show three accumulated rates of returns of the *PP*, the *MP*, and the stock index, respectively.

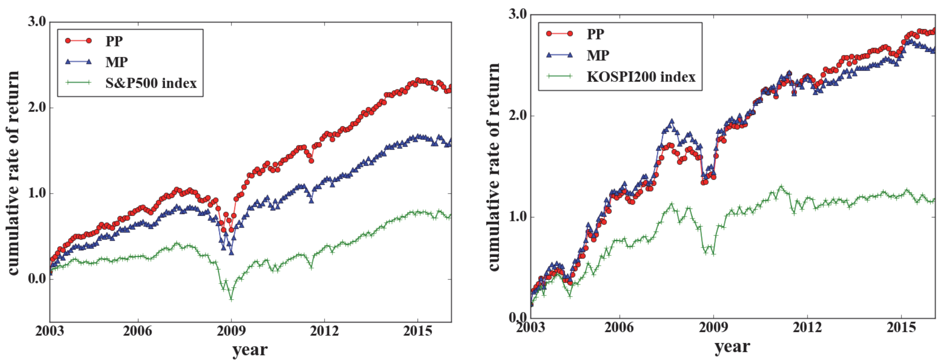
Generally, our proposed trading strategy shows more stable returns in this period than others. From the result, the dynamics of the cohesion of stocks represented by the density and the path length in a stock network affects the change of the stock price and it can be used to establish an efficient strategy for the stock investment.

4.3 Discussion

The stock portfolio recommending system *StoPoR* is based on the dynamic charac-

Table 6. The comparison of the rates of return in our portfolio (*PP*), the Markowitz efficient portfolio (*MP*) and the stock index.

Period	Monthly average return (%)					
	S&P500			KOSPI200		
	PP	*MP*	*index*	*PP*	*MP*	*index*
2003	5.00	3.60	2.32	4.99	5.62	4.00
2004	1.56	1.39	0.31	1.39	0.98	0.37
2005	1.13	0.86	0.08	4.95	4.37	3.02
2006	1.47	1.21	0.81	1.08	1.04	0.32
2007	0.53	0.48	0.16	2.42	3.22	2.12
2008	−2.12	−2.66	−3.22	−1.85	−2.36	−2.96
2009	4.38	3.21	1.86	4.65	3.96	3.30
2010	1.23	0.81	0.15	2.47	1.88	1.18
2011	1.16	1.10	0.69	0.68	−0.25	−1.39
2012	1.68	1.37	0.89	0.97	0.79	0.85
2013	2.53	2.03	1.71	1.08	0.95	0.09
2014	1.81	1.66	1.35	0.07	0.23	0.00
2015	−0.41	−0.42	−0.22	1.81	0.90	−0.23
2016	0.15	0.39	0.11	1.06	1.19	0.72
Average	1.44	1.07	0.50	1.83	1.61	0.81



(a) S&P500 (b) KOSPI200

Fig. 5. The cumulative investment returns of the simulated investment with the proposed portfolio (***PP***), Markowitz portfolio (***MP***), and market index of (a) S&P500 and (b) KOSPI200.

teristics of a stock network, especially the cohesion of a stock network. We found that the cohesion of a stock network Granger-causes the stock price represented by the rate returns of the stock index. It implies that the cohesion of a stock network is useful in predicting a stock price. The high cohesion of a stock network implies that there is more likely to occur a herding behavior resulting in crashing of a stock market. So, *StoPoR* excluded the stocks in the business type with high cohesion in the stock network.

The periods with the largest gap of monthly return on the simulated investment of the *PP* compare to the *MP* are summarized in Table 7. During the simulated investment period, the *PP* achieved the highest returns compared to the *MP* at Apr. 2009 and Dec. 2011 for S&P500 and KOSPI200 respectively.

The case of S&P500 stock market the monthly returns of the *PP* is about 18.99%, which is 5.43% point more than that of the *MP* (13.55%) in this period. In KOSPI200

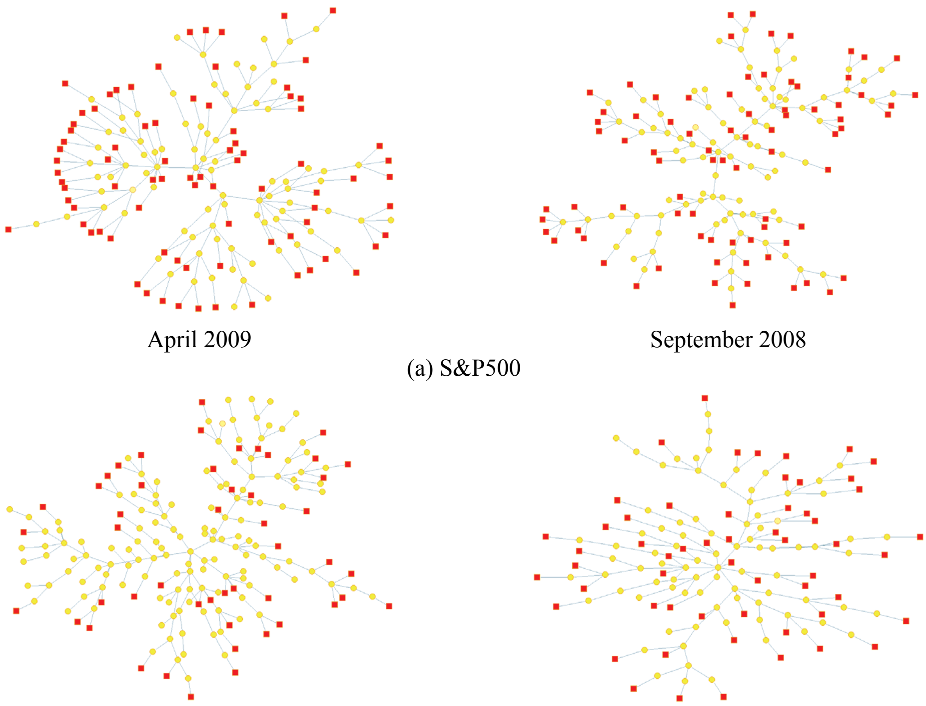
Table 7. The periods of the largest difference of rate of return between S&P500 and KOSPI200.

Market	Date	Stock rate of returns		
		PP (%)	*MP* (%)	Difference (% point)
S&P500	Apr. 2009	18.99	13.55	5.43
	Sep. 2008	−9.61	−7.49	−2.11
KOSPI200	Dec. 2011	3.26	−2.11	5.37
	Oct. 2003	6.31	13.80	−7.49

stock market, the monthly returns of the *PP* is about 3.26%, which is 5.37% point more than that of the *MP* (−2.11%).

On the contrary, the *PP* got the lowest returns compared to the *MP* at Sep. 2008 and Oct. 2003 for S&P500 and KOSPI200 respectively. For S&P500, the monthly returns of the *PP* is about −9.61%, which is −2.11% point less than that of the *MP* (−7.49%) in this period. In KOSPI200, the monthly returns of the *PP* is about 6.31%, which is −7.49% point less than that of the *MP* (13.80%).

Fig. 6 shows the stock networks of the periods with the largest gap between the returns of PP and that of MP. The stocks in the PP are represented red square nodes. In the



April 2009 September 2008

(a) S&P500

December 2008 September 2008

(b) KOSPI200

Fig. 6. The MST of the periods of the lowest rate of return of our portfolio (*PP*) compares to that of the Markowitz efficient portfolio (*MP*). A red square represents a node included in our portfolio (*PP*), respectively. The left panel shows the stock network in the period when the PP had the highest returns compared to that of MP and the right panel shows the opposite case.

figure, the left panel shows the stock network in the period when the PP had the highest returns compared to that of MP and the right panel shows the opposite case.

From these stock networks, we can see that the nodes in the PP are distributed outside the network in the period with the high performance of the PP on the left panel. Otherwise, the stock networks on the right panel, the nodes tend to be located near the center of the network. As a result, it is assumed that the centrality of the nodes in the network also affects the return of the portfolio, and it is expected that the use of these measures will improve the performance of the portfolio.

5. CONCLUSION

Analyzing the characteristics of the stock network, we showed that the cohesion of the stock market network affects the change of the stock prices and can be used for recommending stock portfolios. We focused on how the correlation between stocks affects the variation of the stock price changes. For the statistical analysis, we constructed stock networks using the stock prices traded on the U.S. and the Korean stock markets. Our stock dataset composed of daily closing prices of 505 and 200 stocks configuring S&P 500 and KOSPI200 indices for about thirteen years from January 3rd, 2003 to June 30th, 2016. Each stock network has been constructed based on the correlation between the stocks with daily logarithmic returns.

Analyzing the dynamics of the stock networks for past thirteen years, we found two important facts: (1) the stocks in the same industry category tend to be located close each other, so the average path length between stocks in the same category is shorter than that of the stocks of different categories; (2) the graph density tends to more increase considerably in the bear market than the bull market commonly for both of the U.S. and the Korean stock markets.

Using the average path length and the graph density, we defined the cohesion to measure the closeness between the stocks. And we performed the Granger causality test between the cohesion and the stock returns. According to the result, the cohesion of the stocks Granger causes the stock returns in the U.S. and the Korean stock markets in common but the converse is not. This implies that the cohesion of the stocks can be considered as a sufficient condition for the profitable stock returns.

Based on our network analysis, we proposed the *StoPoR* system for recommending an efficient stock portfolio. To show the efficiency of our method, we conducted a simulated stock investment for the U.S. and the Korean stock market. For the U.S. stock market, the monthly returns of the portfolio proposed by *StoPoR* is about 1.44% on average, which is much better than that of Markowitz (1.07%) and S&P500 index (0.50%). Also, for Korean stock market, the monthly returns of the portfolio proposed by *StoPoR* is about 1.83%, which is also higher than that of Markowitz (1.61%) and that of KOSPI200 index (0.81%).

Though other aspects of the portfolio, such as risks and the stability, are not considered, this result can be used as a starting point to construct a good portfolio. More importantly, it was revealed that the dynamics of the stock network can be modeled and constructed using the relatively simple measures, the density and the average path length in the stock categories. Actually, characterizing the stock networks is the starting point of this research. Mixing up the proposed method with other optimization methods, such as particle swarm optimization and Leverage Space Portfolio Model to construct a further stable portfolio, can be considered as a future work.

REFERENCES

1. H. Markowitz, "Portfolio selection," *The Journal of Finance*, Vol. 7, 1952, pp. 77-91.
2. P. Sinha, A. Chandwani, and T. Sinha, "Algorithm of construction of optimum portfolio of stocks using genetic algorithm," *International Journal of System Assurance Engineering and Management*, Vol. 6, 2015, pp. 447-465.
3. C.-H. Chen, C.-B. Lin, and C.-C. Chen, "Mining group stock portfolio by using grouping genetic algorithms," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2015, pp. 738-743.
4. C.-C. Lin and Y.-T. Liu, "Genetic algorithms for portfolio selection problems with minimum transaction lots," *European Journal of Operational Research*, Vol. 185, 2008, pp. 393-404.
5. Y. Chen, S. Mabu, and K. Hirasawa, "Genetic relation algorithm with guided mutation for the large-scale portfolio optimization," *Expert Systems with Applications*, Vol. 38, 2011, pp. 3353-3363.
6. C.-H. Chen and C.-Y. Hsieh, "Actionable stock portfolio mining by using genetic algorithms," *Journal of Information Science and Engineering*, Vol. 32, 2016, pp. 1657-1678.
7. H. Zhu, Y. Wang, K. Wang, and Y. Chen, "Particle Swarm Optimization (PSO) for the constrained portfolio optimization problem," *Expert Systems with Applications*, Vol. 38, 2011, pp. 10161-10169.
8. O. Ertenlice and C. B. Kalayci, "A survey of swarm intelligence for portfolio optimization: Algorithms and applications," *Swarm and Evolutionary Computation*, Vol. 39, 2018, pp. 36-52.
9. K. P. Wu, Y. P. Wu, and H. M. Lee, "Stock trend prediction by using K-means and AprioriAll algorithm for sequential chart pattern mining," *Journal of Information Science and Engineering*, Vol. 30, 2014, pp. 653-667.
10. K. Lwin, R. Qu, and G. Kendall, "A learning-guided multi-objective evolutionary algorithm for constrained portfolio optimization," *Applied Soft Computing*, Vol. 24, 2014, pp. 757-772.
11. P. Zhang and W.-G. Zhang, "Multiperiod mean absolute deviation fuzzy portfolio selection model with risk control and cardinality constraints," *Fuzzy Sets and Systems*, Vol. 255, 2014, pp. 74-91.
12. Q. J. Zhu, R. Vince, and S. Malinsky, "A dynamic implementations of the leverage space portfolio," *SSRN Electron Journal*, 2012, pp. 1-18.
13. Y.-F. Lin, C.-F. Huang, and V. S. Tseng, "A novel episode mining methodology for stock investment," *Journal of Information Science and Engineering*, Vol. 30, 2014, pp. 571-589.
14. D. Sornette, "A complex system view of why stock markets crash," *New Thesis*, Vol. 1, 2004, pp. 5-17.
15. X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley, "A theory of power-law distributions in financial market fluctuations," *Nature*, Vol. 423, 2003, pp. 267-270.
16. H.-J. Holtrup, "Stock markets as evolving complex systems: Simulations and statis-

tical inferences," Ph.D. Thesis, Department of Economics, University of Duisburg-Essen, Essen, 2006.
17. G. Oh, C. Eom, F. Wang, W.-S. Jung, H. E. Stanley, and S. Kim, "Statistical properties of cross-correlation in the Korean stock market," *The European Physical Journal B*, Vol. 79, 2011, pp. 55-60.
18. R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B-Condensed Matter and Complex Systems*, Vol. 11, 1999, pp. 193-197.
19. J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, and A. Kanto, "Dynamics of market correlations: Taxonomy and portfolio analysis," *Physical Review E*, Vol. 68, 2003, pp. 056110.
20. C. Eom, G. Oh, W.-S. Jung, H. Jeong, and S. Kim, "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*, Vol. 388, 2009, pp. 900-906.
21. R. H. Heiberger, "Stock network stability in times of crisis," *Physica A: Statistical Mechanics and its Applications*, Vol. 393, 2014, pp. 376-381.
22. A. Colman, *Oxford Dictionary of Psychology*, Oxford University Press, UK, 2000.
23. C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, Vol. 37, 1969, pp. 424-438.
24. Y. Hong, Y. Liu, and S. Wang, "Granger causality in risk and detection of extreme risk spillover between financial markets," *Journal of Econometrics*, Vol. 150, 2009, pp. 271-287.
25. X. Zhang, B. Podobnik, D. Y. Kenett, and H. E. Stanley, "Systemic risk and causality dynamics of the world international shipping market," *Physica A: Statistical Mechanics and its Applications*, Vol. 415, 2014, pp. 43-53.

Yun-Jung Lee received a B.S. degree from Pukyoung National University, Korea, in 1995, an M.S. degree from Pukyoung National University, Korea in 1999, and a Ph.D. degree from Pukyoung National University, Korea, in 2008. She is currently with the Research Institute of Computer, Information and Communication in Pusan National University. Her research interests include weblog visualization, social network, big data analysis and stock network analysis.

Gyun Woo is an Associate Professor of the School of Computer Science and Engineering at Pusan National University in Busan, Korea. He received B.S., M.S., and Ph.D. degrees in Computer Science from the Korea Advanced Institute of Science and Technology, in 1991, in 1993, and in 2000, respectively. He worked at Pusan National University as an Assistant Professor. He joined Pusan National University in September 2004. His area of interest includes implementation of programming languages, program analyzes, functional languages, and big data analysis.