

주식 포트폴리오 추천을 위한 주식 시장 네트워크 분석

Analysis of the Stock Market Network for Portfolio Recommendation

이윤정*, 우균**

부산대학교 IT기반 융합산업 창의인력양성사업단*, 부산대학교 컴퓨터공학과/LG전자 스마트제어센터**

Yun-Jung Lee(leeyj01@pusan.ac.kr)*, Gyun Woo(woogyun@pusan.ac.kr)**

요약

주식시장은 시간에 따라 계속 변하고 특별한 이유 없이 주가가 급등하거나 급락하는 사건들이 발생하기도 한다. 이런 이유로 주식시장은 복잡계로 인식되고 있으며 주가 변동을 예측하는 것은 어려운 일이다. 이 논문에서는 주식시장을 개별 주식들의 네트워크로 이해하고 시간에 따라 변하는 한국 주식시장 네트워크를 분석하였다. 코스피200 지수를 구성하는 137개 회사의 주식들을 대상으로 주식 사이의 상관관계를 측정된 결과 주식 간 상관관계가 매우 높을 때 주가가 급락하는 경향이 있는 것으로 나타났다. 또한, 우리는 이러한 네트워크 분석 결과를 바탕으로 주식 포트폴리오를 구성하는 방법을 제안한다. 제안 방법으로 구성된 포트폴리오의 효율성을 보이기 위해 실제 주식들을 대상으로 모의 투자 실험을 수행하였고, 마코위츠의 효율적 포트폴리오 구성 알고리즘을 이용해 구성된 포트폴리오의 수익률과 비교하였다. 실험 결과 제안 방법으로 구성된 포트폴리오는 평균적으로 약 10.6%의 수익률을 보였으며, 같은 기간 마코위츠의 효율적 포트폴리오의 수익률보다 약 3.7% 높으며, 코스피200 수익률보다 약 5.6% 정도 높게 나타났다.

■ **중심어** : | 주식 포트폴리오 | 주식 시장 | 주식 네트워크 | 네트워크 분석 | 포트폴리오 추천 |

Abstract

The stock market is constantly changing and sometimes a slump or a sudden rising in stocks happens without any special reason. So the stock market is recognized as a complex system and it is hard to predict the change on stock prices. In this paper we consider the stock market to a network consisting of stocks. We analyzed the dynamics of the Korean stock market network and evaluated the changing of the correlation between shares consisting of the time series data of 137 companies belong to KOSPI200. Our analysis shows that the stock prices tend to plummet when the correlation between stocks is very high. We propose a method for recommending the stock portfolio based on the analysis of the stock market network. To show the effectiveness of the recommended portfolio, we conducted the simulated stock investment and compared the recommended portfolio with the efficient portfolio proposed Markowitz. According to the experiment results, the rate of return of the portfolio is about 10.6% which is about 3.7% and 5.6% higher than the average rate of return of the efficient portfolio and KOSPI200 respectively.

■ **keyword** : | Stock Portfolio | Stock Market | Stock Market Network | Network Analysis | Portfolio Recommendation |

* 이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음

접수일자 : 2013년 07월 15일

수정일자 : 2013년 09월 13일

심사완료일 : 2013년 09월 13일

교신저자 : 우균, e-mail : woogyun@pusan.ac.kr

I. 서론

증권 시장이 발달함에 따라 직접 주식을 매매하는 개인들도 점차 증가하고 있다. 일반적으로 개인 투자자들은 기관투자자나 외국인 투자자에 비해 정보력, 기업분석 기술 등이 부족해 변화가 많은 주식시장의 흐름을 파악하거나 투자종목을 선택하는 것은 어려운 일이다. 이렇듯 주식 투자에서 위험 부담을 줄이고 좀 더 높은 수익을 내기 위해서는 효율적인 포트폴리오 선택이 중요하다.

일반적으로 자산운용 회사에서는 직접 개발한 재무 프로그램을 사용하고 있으며 그 결과에 전문가의 조언을 더해 포트폴리오를 완성하고 있다. 하지만 자산운용 회사는 포트폴리오 구성 방식을 공개하지 않기 때문에 자산운용 회사를 이용하지 않는 개인 투자자들은 포트폴리오를 구성하기 어려우며 시장의 흐름을 직접 판단하기 어렵다. 특히 초보 투자자들은 처음 포트폴리오를 구성하는 것도 어려울 뿐만 아니라 어떤 시기에 포트폴리오를 어떻게 수정해야 하는지 몰라 피해를 보는 경우가 많다. 주식 투자에 도움을 주기 위해 개인 투자자들의 의사 결정을 지원하는 알고리즘이나 주가 예측 알고리즘들이 개발되고 있다[1].

최근 주식 시장의 흐름과 특성을 분석하기 위해서 주식 시장을 네트워크로 이해하려는 연구가 활발히 진행되고 있다. 만테냐(Mantegna)는 주식 시장을 구성하는 개별 종목들의 상관관계를 기반으로 하여 주식 시장을 최소신장 트리(Minimum Spanning Tree, MST)로 나타내어 전체 주식 시장의 위상 구조를 이해하려고 시도하였다[2]. 그 이후로 여러 연구자에 의해 주식 시장 네트워크가 경제적인 관점에서 신뢰할 수 있는 것인지에 대한 다양한 분석과 검증이 이루어졌다[3-8]. 현재는 주식 시장을 복잡계로 인식하고 복잡계 네트워크에서 나타나는 특징들을 찾아내고 분석하는 연구도 이루어지고 있다[9-11]. 연구 결과 주식 네트워크에서 노드의 연결선 분포가 거듭제곱 법칙을 따르는 것으로 나타났다.

이 논문에서는 한국 주식 시장을 종목 간 상관관계에 따라 네트워크로 구성하고 사회연결망 분석 기법을 통

해 주식 시장을 분석한다. 또한, 분석 결과를 바탕으로 주식 포트폴리오를 구성하는 방법을 제안한다. 제안 방법의 효율성을 보이기 위해 실제 코스피200을 구성하는 137개 회사의 주가 데이터를 대상으로 포트폴리오를 구성하고, 제안 포트폴리오의 수익률과 코스피 수익률을 비교하여 제안 방법의 효율성을 평가한다.

논문의 구성은 다음과 같다. 2장에서는 주식 네트워크 분석에 관련된 선행 연구들을 살펴본다. 3장에서는 실제 주식 데이터를 이용하여 얻은 주식 네트워크를 분석하고, 4장에서 포트폴리오 구성 방법에 대해서 설명한다. 5장에서는 모의 투자 실험을 통해 제안 포트폴리오의 효율성을 평가하고, 6장에서 결론을 맺는다.

II. 관련 연구

주식 시장에서는 국내외적으로 발생하는 사건들로 인하여 예상치 못한 급락이 발생하는 등 전형적인 복잡계 현상이 나타나고 있다. 이것은 주식 간 혹은 시장 간의 눈에 보이지 않는 연결로 정보에 대한 직접적·간접적 영향 경로가 형성되어 있기 때문이다[12]. 1999년 만테냐의 연구를 시작으로 주식 간 혹은 시장 간의 상호관계를 네트워크 방법으로 이해하려는 연구가 주로 경제물리학 분야에서 진행되었다. 만테냐는 주식 사이의 복잡한 상호작용을 바탕으로 공통요인을 중심으로 주식들이 동질적 집단을 형성하는 집단화 과정을 최소신장 트리 방법을 이용하여 처음으로 주식 네트워크로 시각화하였다[2]. 이 그래프에서 각 노드는 주식 종목이고, 노드 간 연결 거리는 두 주식 사이의 상관관계로 구해진다.

[그림 1]은 다우존스 지수를 계산하는데 사용되는 종목 중 30개 종목을 연결한 최소신장 트리를 보여준다. [그림 1]에서 각 노드의 심볼은 회사명을 나타내는 약어인 티커심볼(ticker symbol)로서 예컨대 'GE'는 'General Electric Company'를 가리킨다. 각 노드 간의 연결 거리는 두 주식의 상관관계를 이용하여 구한 것으로 거리가 짧을수록 두 주식 간의 상관관계가 높음을 의미한다. 이 그래프에서는 'CHV' 주식과 'TX' 주식 사

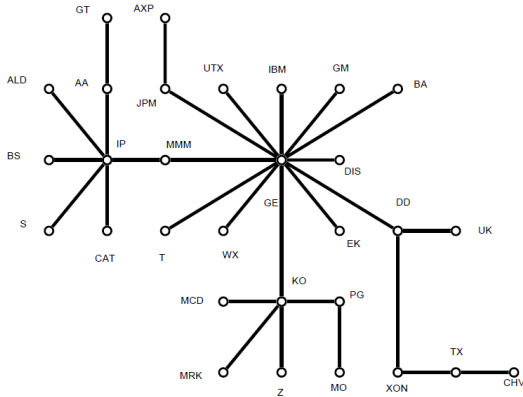


그림 1. 다우존스 지수 계산에 사용되는 30개 주식 종목의 최소신장 트리[2]. 각 노드는 회사의 주식을 나타내고, 두 노드 간의 거리는 두 주식의 상관관계를 이용하여 계산되며, 거리가 가까울수록 상관관계가 높다.

이의 거리가 $0.9 < d(CHV, TX) \leq 0.95$ 로 가장 짧게 나타났다.

만테나가 주식 시장 분석을 위해 네트워크 방법을 도입한 후로 주식 시장 네트워크가 경제적인 관점에서 신뢰할 수 있는 것인지에 대한 다양한 분석과 검증이 시도되었다. Onnela 등은 주식 시장 그래프의 시간에 따른 변화를 관찰하였다[9]. 1980년 1월 2일부터 1999년 12월 31일까지 약 20년 동안의 뉴욕 증시에서 거래되는 주식 종목들의 주가를 대상으로 4년 단위로 연속적인 주식 시장의 최소신장 트리를 구성한 결과 기본적인 트리의 위상구조는 시간에 따라 큰 변화는 없었으나 1987년 10월 19일인 ‘블랙 먼데이(black Monday)’와 같은 주식 대폭락기에는 트리의 평균경로 길이가 감소한 것으로 나타났다. 또한, 이 연구에서는 주식 시장의 최소신장 트리를 포트폴리오 분석에도 적용하였다.

[그림 2]는 S&P500 지수를 구성하는 주식 중 116개의 주식으로 구성된 최소신장 트리를 보여준다. 이 그래프는 각각의 주식을 업종별로 구분하여 나타낸 것으로 같은 업종에 속하는 주식들이 비교적 가깝게 위치하고 있는 것을 볼 수 있다. 따라서 주식 시장 네트워크가 실제 주식 간의 연결 관계를 나타낸다고 할 수 있다. 이 중 에너지 업종은 모든 주식에 같은 클러스터에 포함되어 있다. 이러한 관찰 결과를 토대로 마코위츠(Markowitz)에 의하여 제안된 최적화 함수로 구성된 효율적 포트폴

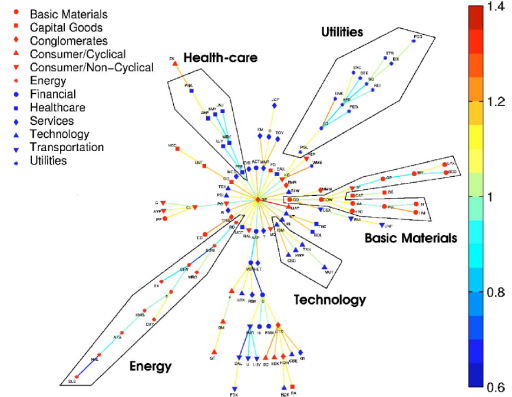


그림 2. S&P500 지수를 구성하는 주식 중 116개 주식으로 구성된 최소신장 트리[9]. 각각의 주식을 업종별로 구분하여 나타낸 것으로 이 중 에너지(Energy) 업종은 완벽한 클러스터로 구분된다.

리오를 분석한 결과 포트폴리오 구성 주식들이 주식 시장 그래프의 단말에 위치하는 경향이 있는 것으로 나타났다.

지금까지 미국 증시에 대한 분석뿐만 아니라 여러 나라의 주식시장 네트워크에서 나타나는 특성을 분석하는 연구도 활발히 이루어졌다. 엄철준 등은 한국 주식 시장에서 관찰 가능한 주식 연결 구조가 어떤 방식으로 형성되는지 연구하였다[5]. 한국 주식시장 그래프에서 대부분의 주식은 다른 주식들과 1개 혹은 2개의 적은 연결을 갖는 반면에 일부 몇몇 주식들은 매우 많은 연결 관계를 가지는 것으로 나타나 주식 네트워크에서 개별주식들이 가지는 연결선 분포가 거듭제곱 분포를 따른다는 공통적인 특성이 관찰되었다. 또한, 많은 연결 관계를 갖는 허브 주식(hub stock)일수록 시장지수와 관련성이 높은 것으로 나타났다.

이와 비슷한 연구로 허화 등은 한국 주식시장의 주식 네트워크와 주식 수익률이 보이는 상관관계에 대해 분석하였다[12]. 이 연구에서는 1980년 1월부터 2003년 5월까지 23년 동안의 연속적인 일별 주식 가격 정보를 갖는 197개의 주식을 대상으로 주식 네트워크와 최적화 함수에 따라 도출된 효율적 포트폴리오를 비교하였다. 포트폴리오에 속한 주식들은 대부분 주식 네트워크에서 하나 또는 둘 정도의 적은 연결선을 가지고 그래프의 외곽에 위치하는 경향이 확인되었다. 이 결과는

앞서 설명된 Onnela의 연구[9]와도 일치한다.

이외에도 중국이나 러시아, 호주 등 여러 나라의 주식 네트워크 특성을 분석한 연구들이 많이 발표되었다 [7][13][14]. Vizgunov 등은 러시아 주식시장의 네트워크 특성을 분석하였다[14]. 2007년부터 2011년까지 러시아 주식시장의 네트워크를 구성하고 분석한 결과 러시아 주식시장에서는 주식 거래량과 네트워크의 최대 클리크(clique) 사이에 강한 상관관계가 존재하는 것으로 나타났다.

지금까지 살펴본 바와 같이 나라마다 주식 시장의 규모나 성숙도 등에 따라 차이는 있으나 주식 사이의 연결선 분포나 업종 간 구분 등과 같은 주식 네트워크에서 나타나는 공통적인 특성들이 관찰되었다. 따라서 주식 네트워크가 실제 주식 시장에서 나타나는 상관관계 등을 의미 있는 수준으로 반영하고 있으며, 따라서 네트워크 방법론이 주식 시장을 이해하는 데 도움이 되는 방법임을 알 수 있다.

III. 한국 주식시장 네트워크 분석

1. 주식 데이터 집합

이 논문에서는 한국 주식시장 네트워크의 시간에 따른 변화를 관찰하기 위해서 코스피200 지수를 구성하는 개별 주식 중에서 2005년 1월 3일부터 2012년 12월 31일까지 7년 동안 연속적인 일별 가격 정보를 갖는 137개 회사의 주식 자료를 수집하였다. 각각의 개별 주식 자료는 해당 기간의 주식 가격을 기록한 시계열 데이터이다.

데이터 집합에 포함된 주식들을 업종 분류에 따라 정

표 1. 데이터 집합에 속한 137개 기업의 업종별 분류현황

번호	업종명	종목 수	비중
1	건설기계	11	8.03%
2	금융	10	7.30%
3	에너지화학	24	17.52%
4	자유소비재	25	18.25%
5	정보통신	16	11.68%
6	조선운송	9	6.57%
7	철강소재	16	11.68%
8	필수소비재	26	18.98%
합계		137	100.00%

리한 것은 [표 1]에 나타나 있다. 이 논문에서 사용하는 주식 데이터를 코스피200 업종 분류에 따라 분류한 결과 전체 137개의 기업 주식이 총 8개의 업종으로 분류되었다. 전체 기업 중 약 18.98%에 해당하는 26개의 기업이 ‘필수소비재’ 업종에 포함되어 가장 큰 부분을 차지하고, 다음으로 ‘자유소비재’ 업종이 전체의 약 18.25%를 차지하고 있다.

2. 주식시장 네트워크 분석

개별 주식들의 상관관계를 이용하여 주식시장 네트워크를 만든다는 개념은 만테냐로부터 발전하였다[2]. 주식시장 네트워크는 개별 주식을 노드로 하고 두 주식 간의 연결 거리는 두 주식 간의 상관관계에 따라 결정된다. 만일 N 개의 주식으로 구성되는 주식시장 그래프는 N 개의 노드와 최대 $N(N-1)/2$ 개의 링크로 구성될 수 있다. 지금까지 여러 연구에서 주식이격의 상호 연관성을 정량적으로 나타내기 위해 주식 i 와 주식 j 사이의 교차 상관계수(cross correlation coefficient) $C_{i,j}$ 는 다음과 같이 정의된다.

$$C_{i,j} = \frac{\langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle}{\sqrt{\langle G_i^2 - \langle G_i \rangle^2 \rangle \langle G_j^2 - \langle G_j \rangle^2 \rangle}} \quad (1)$$

$$G_i(t) = \ln S_i(t + \Delta t) - \ln S_i(t)$$

여기서 $\langle G_i \rangle$ 는 주어진 시계열 자료 G_i 의 평균을 의미하고, G_i 는 주식이격 S 의 로그 수익률을 나타낸다. 주식이격 변화 대신 로그 수익률을 쓰는 이유는 서로 다른 주식에 대해서 수익을 비교할 때 가격 S 는 주식마다 크게 다르지만, 로그를 취함으로써 가격이 싼 주식이나 비싼 주식이나 공평하게 수익을 정의할 수 있게 해 주기 때문이다. 즉, 로그 수익률의 교차 상관관계는 주식이격에 대한 변동폭 사이의 상관관계이다[15]. 이렇게 구해진 주식 사이의 상관관계 계수를 이용하여 주식시장 그래프에서 주식 i 와 주식 j 사이의 거리 d_{ij} 는 식 2와 같이 계산된다.

$$d_{i,j} = \sqrt{2(1 - C_{i,j})} \quad (2)$$

여기서 $C_{i,j}$ 는 -1에서 1 사이의 값이므로 두 주식 간의 거리 d_{ij} 는 0에서 2 사이의 값이며, 상관관계가 클수록 거리는 짧아진다.

이 논문에서는 한국 주식시장의 위상구조를 살펴보기 위해 관찰기간 동안 주식 데이터를 이용하여 최소신장 트리를 구성하였다. 최소신장 트리는 Kruskal 알고리즘을 사용하여 [그림 3]과 같이 생성하였다.

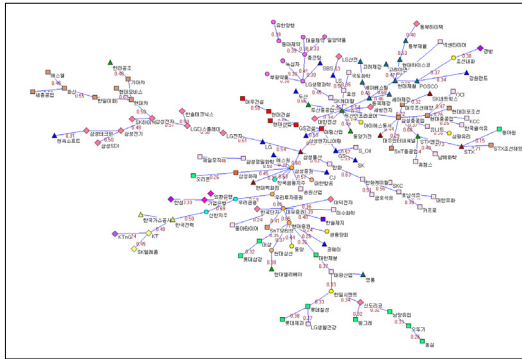


그림 3. 137개 기업의 상관관계를 이용하여 구성된 최소신장 트리. 개별 주식 데이터는 2005년 1월 3일부터 2012년 12월 31까지 연속적인 1,990개의 주식 가격으로 구성되어 있다.

이 그래프는 2005년 1월 3일부터 2012년 12월 31일까지 주식 데이터를 이용하여 그래프 시각화 도구인 Pajek으로 생성한 것이다. 각 노드의 모양과 색상은 업종별로 구분하여 표시하였으며 연결의 방향성은 고려하지 않았다. 주식들의 이름과 연결선의 값을 최대한 겹치지 않게 시각화하기 위해 주식 노드 위치를 수정하였다. 따라서 [그림 3]에 나타난 노드 사이의 물리적인 거리와 실제 계산된 주식 간의 거리가 정확히 일치하지는 않는다. 노드 간 연결선에 표시된 값은 두 주식 간의 실제 상관계수 값을 나타낸다.

먼저 개별 주식들이 서로 어느 정도 연관되어 있는지를 살펴보기 위해 각 노드의 연결선 수를 분석하였다. 기존 연구에서 살펴보았듯이 일반적으로 연결선이 많은, 즉 차수가 높은 주식은 전체 주식시장에 미치는 영향이 크다고 할 수 있으며, 시장지수와 상관관계가 높은 경향으로 나타났다. 즉 차수가 높은 주식은 허브 주

표 2. 한국 주식시장 그래프의 연결선 수 분포

연결선 수	종목 수	연결선 수	종목 수
1	78	6	0
2	26	7	2
3	16	8	1
4	8	9	0
5	4	10	2

식으로 간주할 수 있다. [그림 3]에 나타난 주식시장 그래프의 정확한 연결선 수의 분포는 [표 2]에 정리되어 있다.

다른 주식과의 연결선이 하나인 주식은 78개로 전체에서 약 56.93%로 가장 많았고, 연결선이 두 개인 주식은 26개(약 18.97%)로 두 번째로 많은 부분을 차지하고 있다. 주식 대부분은 연결선 수가 많지 않으나 다른 주식에 비해 월등히 많은 연결선 수를 가지는 주식들도 존재한다. [그림 3]의 그래프에서는 ‘삼성증권’과 ‘두산인프라코어’ 종목이 차수 10으로 가장 높게 나타났다.

[표 2]에서 대부분 주식의 연결선 수가 2 이하로 적으나 일부 주식의 연결선 수가 평균보다 훨씬 많은 것을 알 수 있다. 주식의 수가 많지 않아 연결선 수의 정확한 분포를 말할 수 없으나 주식시장 네트워크의 연결선 수가 거듭제곱(power-law) 분포를 따르는 것은 다른 주식시장에서와 마찬가지로 공통으로 나타나는 현상으로 생각할 수 있다.

주식 시장에서 개별 주식들이 어떤 영향력을 가지는지 살펴보기 위해서 주식시장 네트워크에서 각 노드의 영향력을 측정할 필요가 있다. 사회연결망 분석에서는 노드의 영향 정도를 측정하기 위하여 중심성 개념을 사용하고 있다[16]. 그중에서 연결중심성(degree centrality)은 다른 노드와 얼마나 많은 연결을 가지는지를 의미하는 것으로 네트워크의 지역적 중량을 측정하는 지표로 사용될 수 있다. 또한 인접중심성(closeness centrality)은 다른 노드와의 연결 거리가 얼마나 짧은지를 나타내는 것으로 인접중심성이 높을수록 네트워크의 중심에 있음을 의미한다. 이 논문에서는 연결중심성과 인접중심성을 이용하여 주식시장에서 개별 주식이 가지는 영향 정도를 평가한다.

실제 주식시장은 시간에 따라 계속해서 변하므로 개별 주식의 영향력도 시간에 따라 다르게 나타날 수 있

다. 따라서 이 논문에서는 관찰 기간을 바꾸어가며 주식시장 그래프를 생성하고 개별주식들의 그래프 내의 영향 정도를 분석하였다. 이를 위해 2005년 1월 3일부터 2012년 12월 31일 동안 그래프 생성 단위를 1개월로 하고 하루씩 관찰 기간을 이동시켜 최소신장 트리를 생성하였고, 각 그래프에서 노드들의 연결중심성과 인접중심성의 변화를 분석하였다.

[표 3]은 전체 관찰기간 동안 연결중심성의 평균이 높은 상위 15개 주식의 연결중심성 통계이다. 연결선수의 평균은 직전 한 달 동안의 상관관계를 이용하여 생성한 총 1,968개의 주식 그래프에서 계산한 개별주식의 연결선 수를 평균한 값이다. 표준편차와 최댓값도 같은 방법으로 계산하였다. 연결중심성은 한 노드의 연결선 수를 전체 가능한 연결선 수로 나눈 값이다. 하나의 주식이 평균적으로 다른 회사와 얼마나 많은 연결을 갖는지를 이해하기 쉽도록 연결중심성 값 대신 연결선의 수를 나타내었다.

7년 동안 연결중심성이 꾸준히 높았던 상위 15개의 주식 중에서 '금융' 업종 주식이 6개나 포함되어 있다. 다음으로 많은 업종은 '조선운송'으로 전체 9개의 주식 중 모두 4개가 포함되어 있다. [표 1]의 업종 분류 현황에서 가장 많은 종목이 포함된 '필수소비재' 업종은 연결중심성 상위 15개 종목에 하나도 포함되지 않았다. '자유소비재' 업종 또한 25개의 많은 종목을 가지고 있

표 3. 연결중심성이 높은 상위 15개 회사의 평균 연결 차수 통계

기업명	업종	연결선 수		
		평균	표준편차	최대
대우증권	금융	3.670	1.768	13
삼성증권	금융	3.199	1.469	11
우리투자증권	금융	3.118	1.646	13
현대증권	금융	3.045	1.624	10
현대중공업	조선운송	2.721	1.392	9
삼성중공업	조선운송	2.704	1.473	9
STX조선해양	조선운송	2.616	1.490	9
삼성물산	자유소비재	2.565	1.464	8
기업은행	금융	2.539	1.466	11
대림산업	건설기계	2.517	1.416	8
현대미포조선	조선운송	2.474	1.475	11
현대차	자유소비재	2.467	1.215	7
GS건설	건설기계	2.450	1.454	9
신한지주	금융	2.448	1.417	12
삼성전자	정보통신	2.434	1.382	10

으나 연결중심성 상위 업체에는 '삼성물산'과 '현대차' 두 종목만 포함되어 있다.

이러한 분석 결과를 미루어 볼 때 실험 대상에 포함된 종목들로 주식시장을 한정한다면 종목 수가 많은 업종이라고 해서 주식시장에 미치는 영향력이 반드시 크다고는 할 수 없으며, 한국 주식시장은 '금융'과 '조선운송' 업종의 영향력이 가장 큰 것으로 나타났다.

다음으로 개별 주식들의 인접중심성을 분석하였다. 연결중심성 분석과 마찬가지로 1개월 동안의 상관관계를 이용하여 주식 그래프를 만들고 하루씩 날짜를 이동시켜 연속적인 변화를 살펴보았다. 관찰 기간에 인접중심성이 높게 나타난 상위 15개 주식의 통계는 [표 4]와 같다. 앞서 살펴본 연결중심성 결과와 비슷하게 '금융' 업종에 속하는 4개의 주식이 상위에 위치하고 있고, 전체적으로 '금융' 업종과 '조선운송' 업종이 상위 주식의 대부분을 차지하고 있음을 알 수 있다.

[표 3]과 [표 4]에 나타난 개별 주식의 연결중심성과 인접중심성을 분석한 결과 '금융'과 '조선운송' 업종에 속하는 주식들이 한국 주식시장에서 중심에 위치하고 다른 주식에 영향력이 크다고 할 수 있다. 또한, 업종분류에서 얼마나 많은 기업을 포함하고 있는지는 전체 주식시장에 대한 영향력과 상관관계가 없으며 국내 주식시장에 영향력이 큰 주식은 '금융'과 '조선운송' 업종에 편중되어있음을 알 수 있다.

표 4. 인접중심성이 높은 상위 15개 회사의 평균 연결 차수 통계

기업명	업종	인접중심성		
		평균	표준편차	최대
대우증권	금융	0.199	0.044	0.374
삼성증권	금융	0.194	0.040	0.318
우리투자증권	금융	0.191	0.042	0.323
현대증권	금융	0.190	0.042	0.330
현대중공업	조선운송	0.174	0.040	0.310
삼성중공업	조선운송	0.173	0.037	0.304
삼성물산	자유소비재	0.171	0.040	0.323
대림산업	건설기계	0.169	0.039	0.299
현대미포조선	조선운송	0.169	0.039	0.349
대우조선해양	조선운송	0.168	0.039	0.298
한국금융지주	금융	0.167	0.035	0.320
GS건설	건설기계	0.167	0.039	0.314
신한지주	금융	0.166	0.039	0.302
STX조선해양	조선운송	0.166	0.038	0.307
동국제강	철강소재	0.166	0.037	0.300

이와 같은 주식시장 네트워크 분석 결과는 [표 1]에 나타난 137개 종목을 분석한 것이므로 한국 주식시장 전체로 일반화하기 위해서는 더 많은 종목에 대한 네트워크 분석이 필요할 것이다.

이처럼 주식시장을 최소신장 트리로 표현할 경우 전체 주식들의 계층적 위상구조는 쉽게 파악할 수 있다. 그러나 최소신장 트리의 의미대로 고립되는 노드가 발생하지 않게 모든 주식을 연결하다 보니 다른 주식과 상관관계가 낮은 경우라도 연결선이 반드시 생기게 된다. 예를 들어 [그림 4]의 주식 그래프에서도 <‘두산인프라코어’, ‘아이에스동서’> 그리고 <‘STX’, ‘동아원’>의 상관계수는 각각 0.14와 0.15로 거의 상관관계가 없는 데도 불구하고 트리를 구성하기 위해 연결된 것이다.

이 논문에서는 개별 주식들의 상관관계 정도가 주식의 수익률과 어떤 상관관계가 있는지를 알아보기 위해서 주식 사이의 상관관계에 임계치를 적용하여 임계치

이상의 상관계수를 가지는 주식만을 연결하여 주식 네트워크를 구성한다. 그리고 생성된 주식시장 그래프에서 시간에 따른 그래프 밀도와 코스피200 지수 간의 상관관계를 분석한다.

주식시장의 최소신장 트리를 생성하는 것과 마찬가지로 하나의 주식 그래프는 한 달 동안 주식 사이의 상관계수를 이용하여 생성하고 이를 하루씩 이동시키면서 연속적으로 관찰하였다. 이때 노드 간 연결은 임계치를 적용하여 상관계수 값이 임계치 이상이면 연결선을 생성한다. 그래프 밀도는 전체 가능한 연결선 수에 대한 생성된 연결선 수의 비로 계산된다.

[그림 4]는 시간에 따른 그래프 밀도 변화와 해당 기간의 코스피200 지수의 변화를 보여준다. [그림 4](a)는 관찰 기간에 코스피200 지수의 변화를 나타낸 그래프로 지수가 가장 급격하게 하락한 시기는 2008년 10월에 발생한 금융위기 때를 가리킨다. [그림 4](b)는 동일 기간에 주식 그래프의 밀도 변화를 나타낸 그래프로 가능한 값의 범위는 0에서 1 사이 값을 가지며 그래프 밀도가 높은 것은 전체 주식시장에서 상관관계가 전반적으로 높음을 의미한다.

이 그래프를 보면 그래프 밀도는 주기가 일정하지는 않지만 높아졌다 낮아지기를 반복하는 것을 알 수 있다. 그리고 코스피200 지수와 그래프 밀도 변화의 상관관계를 살펴보면 코스피200 지수가 급락한 시기에 주식 그래프 밀도가 증감하는 것을 알 수 있다. 주식 간의 상관관계가 높다는 것은 주식시장에서 동조화 현상이 나타난 것이라고 할 수 있다. 일반적으로 주식 시장의 동조화 현상은 주가 급락의 전조 현상으로 보고 있다.

IV. 포트폴리오 추천 알고리즘

이 논문에서는 시간에 따라 변하는 주식시장 네트워크의 동적 특성을 이용하여 투자 종목을 추천하는 방법을 제안한다. 제안 방법의 알고리즘은 [그림 5]와 같다. 먼저 관찰 기간의 주식 데이터를 이용하여 각 주식 간 상관관계를 계산한다. [그림 5]에서 ts_i 는 i 회사의 주가 시계열 데이터를 나타내고, $C_{i,j}$ 는 주식 i 와 주식 j 간

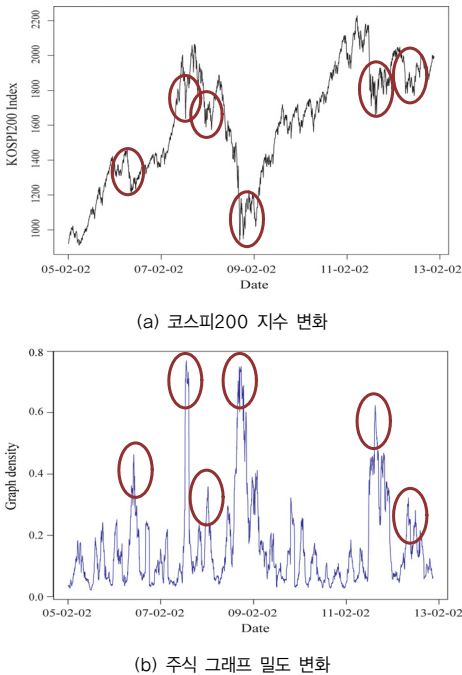


그림 4. 2005년 2월 2일부터 2012년 12월 31일까지 코스피200 지수와 주식 네트워크 밀도 변화

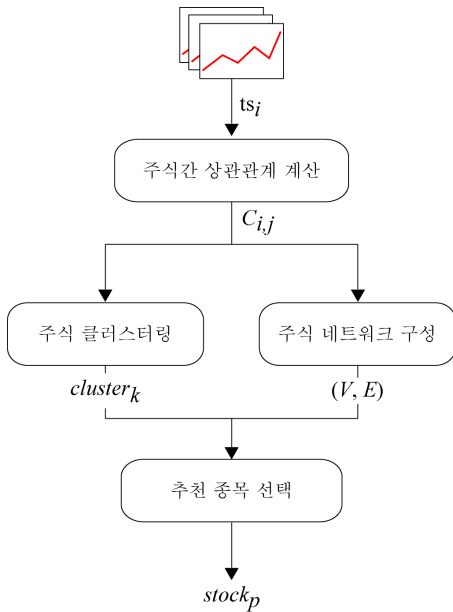


그림 5. 추천 종목 선택 알고리즘

의 상관계수 값을 나타낸다. 전체 주식이 N 개이면 $N \times N$ 상관계수 행렬을 구할 수 있다.

주식 클러스터링 과정에서는 앞서 구해진 상관계수 행렬을 이용하여 서로 유사한 주가 변화 패턴을 보이는 주식들의 그룹으로 나눈다. 이 논문에서는 가장 기본적인 클러스터링 방법인 K-means 클러스터링 알고리즘을 사용하여 몇 개의 그룹으로 주식을 분류한다.

다음으로 주식 네트워크 구성 단계에서는 주식 간 상관계수 행렬을 이용하여 주식 간 거리를 계산하고, 이 값을 이용하여 주식 그래프를 생성한다. 이 논문에서는 Kruskal 알고리즘을 사용하여 주식 네트워크의 최소신장 트리를 생성한다.

마지막으로 전 단계에서 생성된 주식 클러스터 정보와 주식 그래프의 상관계수 정보를 이용하여 최종 추천 종목을 선택하게 된다. 앞서 살펴본 주식 네트워크 분석 결과에 따르면 주식 간 상관관계가 높을 때 주가가 급락하는 경향을 보였다. 따라서 투자 종목을 일부 클러스터에서 집중적으로 선택하는 것보다 주가 변화 패턴이 서로 다른 클러스터에서 골고루 선택하는 것이 더 안정적이라고 할 수 있다. 또한, 기존 연구에서 언급된 바와 같이 효율적 포트폴리오 알고리즘으로 선택된 종

목들은 주식 네트워크에서 외곽에 위치하는 경향이 있다. 따라서 이 논문에서는 이러한 특성을 이용하여 각각의 클러스터에서 차수가 1인 주식들을 최종 추천 대상 종목으로 선택한다.

V. 실험 및 결과

제안 방법의 효율성을 보이기 위해 실제 주식들을 대상으로 하여 모의 투자 실험을 수행하고 수익률을 살펴 보았다. 먼저 실험에 사용된 데이터는 3장에서 설명된 것처럼 코스피200을 구성하는 137개 회사의 2005년 1월부터 2012년 12월까지 약 7년 동안 일별 주식 데이터를 사용하였고, 데이터에 대한 자세한 설명은 3.1절에 나타나 있다.

실험은 다음과 같이 구성된다. 먼저 6개월 동안의 각 주식의 일별 주가 데이터를 이용하여 주식 간 상관관계와 주식 네트워크를 구성한다. 그리고 K-means 클러스터링 알고리즘을 이용하여 전체 주식을 10개의 그룹으로 분류하였다. 투자 기간별 포트폴리오는 그룹별로 차수가 1인 주식을 선택하여 모두 10개의 주식으로 구성된다. 각 그룹에서 차수가 1인 종목이 하나 이상일 경우 무작위로 하나를 선택하였다. 이때 어떤 주식을 선택하느냐에 따라 수익률이 달라질 수 있다. 실험에서는 기간별로 포트폴리오 구성을 100회 수행하고 이때 얻어진 평균 수익률을 해당 기간의 수익률로 간주하였다. 클러스터에 차수가 1인 종목이 없으면 가장 낮은 차수인 종목을 선택하였다.

제안 방법으로 구성된 포트폴리오의 수익률 분석을 위해 2005년 7월부터 2012년 6월까지 6개월 단위로 포트폴리오를 구성하고 이 포트폴리오의 투자 수익률을 측정하였으며 1개월씩 이동시켜 포트폴리오를 재구성하고 수익률의 변화를 살펴보았다. 또한, 제안 방법으로 구성된 포트폴리오가 얼마나 효율적인지를 보이기 위해 마코위츠의 효율적 포트폴리오 구성 알고리즘을 이용하여 구성된 포트폴리오의 수익률과 같은 기간의 코스피200 평균 수익률과 비교하였다. 각각의 수익률은 [표 5]에 나타나 있으며, 지면 관계상 전체 수익률 분석

표 5. 제안 방법으로 구성된 포트폴리오 투자 수익률

투자기간		수익률		
매수일	매도일	제안 포트폴리오	마코위츠 효율적 포트폴리오	코스피 평균
2005-07-01	2005-12-30	35.8%	28.1%	35.5%
2006-01-02	2006-06-30	-7.4%	7.4%	-9.1%
2006-07-03	2006-12-29	19.7%	7.6%	10.8%
2007-01-02	2007-06-29	45.8%	12.3%	22.1%
2007-07-02	2007-12-31	1.0%	10.8%	7.7%
2008-01-02	2008-06-30	-5.9%	-3.1%	-9.6%
2008-07-01	2008-12-31	-36.0%	-16.6%	-32.5%
2009-01-02	2009-06-30	40.3%	52.2%	20.0%
2009-07-01	2009-12-31	20.3%	9.0%	19.2%
2010-01-04	2010-06-30	0.0%	2.6%	0.1%
2010-07-01	2010-12-31	35.0%	20.9%	21.2%
2011-01-03	2011-06-30	5.0%	6.9%	1.2%
2011-07-01	2011-12-30	-18.7%	-19.9%	-14.1%
2012-01-02	2012-06-29	-9.0%	1.6%	1.5%
수익률 평균		10.6%	6.9%	5.0%

결과 중 일부분만 나타내었다.

[표 5]에 나타난 수익률 평균 항목은 현재 표에 나타나지 않은 모든 실험 구간의 수익률 평균을 의미한다. 모의 투자 실험 구간에 제안 방법으로 구성된 포트폴리오의 평균 수익률은 10.6%로 마코위츠의 효율적 포트폴리오의 수익률보다 약 3.7% 높게 나타났으며, 코스피 200 평균 수익률보다는 약 5.6% 높게 나타났다. 따라서 제안 방법으로 구성된 포트폴리오가 효율적으로 구성되었다고 할 수 있다.

VI. 결론 및 향후 과제

이 논문에서는 주식 포트폴리오를 구성하기 위해 주식시장 네트워크를 분석하였다. 주식 네트워크를 구성하기 위해 코스피200 지수를 구성하는 137개의 개별 주식들의 2005년 1월부터 2012년 12월까지 약 7년 동안의 일별 주가 데이터를 사용하였다. 개별 주식의 일별 주가 데이터를 이용하여 상관계수를 구하고 Kruskal 알고리즘을 이용하여 주식시장의 최소신장그래프를 구성하였다.

관찰 데이터를 이용하여 1개월 단위로 주식 네트워크를 구성하고 분석한 결과 대부분 주식은 다른 주식들과 적은 수의 연결선을 가지지만 상당히 많은 연결선을 가

지는 주식들도 존재함을 알 수 있었다. 이렇게 많은 연결선을 가지는 주식들은 주식시장에서 영향력이 큰 주식이라고 할 수 있으면 실험 집합에서는 ‘금융업’과 ‘조선운송’ 업종의 영향력이 큰 것으로 나타났다.

또한, 주식 간 상관관계가 시간에 따라 어떻게 변하는지를 살펴보기 위해 주식 간 상관관계에 임계치를 적용하여 주식 네트워크를 구성하고 그래프 밀도를 분석하였다. 주식 네트워크의 그래프 밀도는 시간이 흐름에 따라 높아졌다 낮아지기를 반복하는 패턴을 보였으며 특히 주가가 급락할 때 그래프 밀도가 급증하는 현상이 나타났다. 그래프 밀도가 증가하는 것은 주식 간 상관관계가 높아진 것을 의미하며 이는 주식시장에서 동조화 현상이 나타난 것이라고 할 수 있다. 일반적으로 주식 시장의 동조화 현상은 주가 급락의 전조 현상으로 보고 있다.

이 논문에서는 이러한 주식 네트워크에서 나타난 특성을 기반으로 주식 포트폴리오를 추천하는 알고리즘을 제안하였다. 포트폴리오 구성 종목은 다음과 같은 두 가지 기준으로 선택한다. 첫 번째 기준은 전체 주식 중에서 주가 흐름이 서로 다른 그룹의 종목을 선택하는 것이다. 주가의 흐름이 비슷한 종목을 선택할 경우 주가가 급락할 경우 포트폴리오 구성 종목들의 주가가 함께 하락할 위험이 크기 때문이다. 두 번째 기준은 주식 간 연결선 수가 1인 종목들을 선택하는 것이다. 기존 연구에서 마코위츠의 효율적 포트폴리오 구성 알고리즘으로 선택한 주식들이 대부분 주식 네트워크에서 외곽에 위치하는 경향이 있다고 알려졌기 때문이다.

제안 방법으로 구성된 포트폴리오의 수익률을 살펴보기 위해 6개월 단위로 포트폴리오를 구성한 후 모의 투자 실험을 실행하였다. 마코위츠의 효율적 포트폴리오 알고리즘으로 구성된 포트폴리오와 코스피200 수익률과 비교한 결과 제안 방법으로 구성된 포트폴리오 수익률이 약 10.6%로 나타났으며, 같은 기간 마코위츠의 효율적 포트폴리오의 수익률보다는 약 3.7%, 코스피 200 수익률 보다는 약 5.6% 높게 나타났다. 따라서 제안 방법으로 구성된 포트폴리오가 효율적으로 구성되었음을 알 수 있다.

제안 방법을 활용한다면 주식시장에 대한 깊은 이해

와 정보가 부족한 일반 개인투자자들도 이전 기간의 주식 네트워크에 나타난 특성을 고려하여 자신의 관심 종목 분석할 수 있으며, 손쉽게 투자 종목을 선택할 수 있을 것이다.

제안 방법에서는 주식 클러스터링을 위해 K-means 알고리즘을 사용하고 있다. 현재 시계열데이터 클러스터링을 위한 알고리즘들이 많이 개발되고 있다[17]. 향후에 클러스터링 알고리즘에 따른 포트폴리오 성능 평가가 필요하며, 포트폴리오의 위험도 분석에 관한 연구도 이루어져야 할 것이다.

참 고 문 헌

[1] 김성동, “주식 투자자의 의사결정 지원을 위한 데이터마이닝 도구”, 한국콘텐츠학회논문지, 제12권, 제2호, pp.472-482, 2012.

[2] R. N. Mantegna, “Hierarchical structure in financial markets,” *The European Physical Journal B-Condensed Matter and Complex Systems*, Vol.11, No.1, pp.193-197, 1999.

[3] G. Bonanno, F. Lillo, and R. N. Mantegna, “High-frequency cross-correlation in a set of stocks,” *Quantitative Finance*, Vol.1, No.1, pp.96-104, 2001.

[4] 김승환, 엄철준, 이운철, “주식간 동적 연결구조의 형성원칙에 관한 연구”, 재무관리연구, 제21권, 제1호, pp.183-204, 2004.

[5] 엄철준, “한국주식시장의 주식 네트워크 속성에 대한 실증연구: 수익률생성모형과 Random Matrix Theory를 이용”, 산업경제연구, 제20권, 제5호, pp.2055-2074, 2007.

[6] J. G. Brida and W. A. Risso, “Hierarchical structure of the German stock market,” *Expert Systems with Applications*, Vol.37, No.5, pp.3846-3852, 2010.

[7] B. M. Tabak, T. R. Serra, and D. O. Cajueiro, “Topological properties of stock market networks: The case of Brazil,” *Physica A: Statistical Mechanics and its Applications*, Vol.389, No.16, pp.3240-3249, 2010.

[8] J. Lee, J. Youn, and W. Chang, “Intraday volatility and network topological properties in the Korean stock market,” *Physica A: Statistical mechanics and its Applications*, Vol.391, No.4, pp.1354-1360, 2012.

[9] 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, No.5, 2003.

[10] G. Oh, C. Eom, F. Wang, W. Jung, H. E. Stanley, and S. Kim, “Statistical properties of cross-correlation in the Korean stock market,” *The European Physical Journal B*, Vol.79, No.1, pp.55-60, 2011.

[11] H. J. Kim, I. M. Kim, Y. Lee, and B. Kahng, “Scale-free network in stock markets,” *Journal-Korean Physical Society*, Vol.40, No.6, pp.1105-1108, 2002.

[12] 허화, 김승환, 강석교, 엄철준, “주식간 연결구조와 효율적 포트폴리오”, 금융공학연구, 제5권, 제2호, pp.65-84, 2006.

[13] W. Q. Huang, X. T. Zhuang, and S. Yao, “A network analysis of the Chinese stock market,” *Physica A: Statistical Mechanics and its Applications*, Vol.388, No.14, pp.2956-2964, 2009.

[14] A. Vizgunov, B. Goldengorin, V. Kalyagin, A. Koldanov, P. Koldanov, and P. M. Pardalos, “Network approach for the Russian stock market,” *Computational Management Science*, pp.1-11, 2013.

[15] 민병원, 김창욱, *복잡계 워크샵*, 삼성경제연구소, 2006.

[16] 김용학, *사회연결망 분석*, 박영사, 2003.

[17] 전진호, 이계성, “시계열데이터의 모델기반 클러스터 결정”, 한국콘텐츠학회논문지, 제7권, 제6호, pp.22-30, 2012.

저 자 소 개

이 윤 정(Yun-Jung Lee)

정회원



- 1995년 2월 : 부경대학교 전자계산학과(이학사)
- 1999년 2월 : 부경대학교 전산정보학과(이학석사)
- 2008년 8월 : 부경대학교 전자계산학과(이학박사)

- 2008년 9월 ~ 2012년 8월 : 부산대학교 U-Port 정보기술사업단 박사후연구원
 - 2012년 9월 ~ 2013년 8월 : 부산대학교 컴퓨터 및 정보통신연구소 기금교수
 - 2013년 11월 ~ 현재 : 부산대학교 IT기반 융합산업 창의인력양성 사업단 박사후연구원
- <관심분야> : 얼굴 애니메이션, 웹 콘텐츠 시각화, 사회연결망 분석

우 균(Gyun Woo)

정회원



- 1991년 : 한국과학기술원 전산학(학사)
- 1993년 : 한국과학기술원 전산학(석사)
- 2000년 : 한국과학기술원 전산학(박사)

- 2000년 ~ 2004년 : 동아대학교 컴퓨터공학과 조교수
 - 2005년 ~ 현재 : 부산대학교 컴퓨터공학과 교수
- <관심분야> : 프로그래밍언어 및 컴파일러, 함수형 언어, 그리드컴퓨팅, 소프트웨어 메트릭, 프로그램 분석, 프로그램 시각화