Network approach for Stock market data mining and portfolio analysis

Susan George
Department of Futures Studies
University of Kerala
Email: georgesusan@gmail.com

Manoj Changat
Department of Futures Studies
University of Kerala
Email: mchangat@gmail.com

***Abstract*—Stock market dynamics is of great importance to researchers from diverse fields. Network Analysis of stock data can play an important role in the study of stock market. In this paper, network based data mining of stock market is done to identify crucial players. Stock market network in United States created based on dynamics of stocks over one year captured as daily time series, is used for the analysis. Along with structural aspects of the market, our analysis revealed highly influential players based on their relationships with other high influential players. Portfolio analysis of top most of the influential players revealed crucial sectors of the market. Our findings suggest that financial services for specific sectors can reduce the systemic risk without affecting the overall economic growth. The stocks in the finance, banking, insurance, technology, machine, industrials, business services. energy, chemicals, retail, transport, real estate and building sector sectors are heavily dependent on each other and affect each other's performance.**

***Index Terms*- Stock market networks, Market graphs, Financial networks, Portfolio analysis, Network analysis, Lobby index, Stock market dynamics.**

I. INTRODUCTION

A financial market is a complex system composed of many interacting units. Before the financial crisis in 2008, the focus of financial risk analysis has been limited to the balance sheet of individual organizations. The inter-relationships between the organizations had generally not been formally considered. But, this has changed during the 2008 financial crisis. The need to understand the interdependencies in the stock market had increased after the financial crisis. Researchers changed their approach from the earlier isolated way to a more systemic view, i.e. to understand the riskiness of a particular organization; they started considering the riskiness of the organizations that it is connected to. In the real world scenario, there are many other effects that arise and propagate over many intermediate connections.

Network analysis can tackle the complexity of systemic study of stock markets. Its properties can be used to predict market dynamics. For example, when extending emergency funding to Bear Stearns, the Federal Reserve System justified its decision by stating "Board members agreed that, given the fragile condition of the financial markets at the time, the prominent position of Bear Stearns in those markets, and the expected contagion that would result from the immediate failure of Bear Stearns, the best alternative available was to provide temporary emergency financing to Bear Stearns through an arrangement with JPMorgan Chase & Co. also in New York. Such a loan would facilitate efforts to effect a resolution of the Bear Stearns situation that would be consistent with preserving financial stability" [17].

In the language of networks, Bear Stearns was a central node which occupied a very important or prominent position in the network and its failure could have led to a cascade of contagious failures had it been let to fail. So, these chains of events would have got magnified in the finance sector and would have adversely impacted the broader economy. This chain of cascaded effects of related events is termed as systemic risk.

In this work, we intend to know certain structural properties of market and to identify the key players in the market by modeling it as market networks. Importance of the player is not only determined by its association to large number of players but also its association to the highly connected players. Also a portfolio analysis of these crucial players is also intended to identify the sectors that are crucial for the stability the stock market.

II. STOCK MARKET NETWORKS

Complex networks have attracted the interest of many researchers in various fields of the world. Contributions of these researches formed a powerful tool and reference base for understanding many other real world complex systems, such as protein interaction networks in the field of biology, social networks and scientists' collaboration networks in the field of sociology. The theory and tools of complex networks also provide us with a new perspective to study the stock markets. The price value of the stocks is a core indicator that reflects

978-1-5090-6590-5/17/$31.00 ©2017 IEEE

the dynamics of stock market. We construct stock networks based on stock prices and with the help of complex network theory and tools, we analyse the characteristics of the community structure in it. Fluctuations in stock prices are very much intercoupled and they are strongly correlated with the business sectors and industries that they belong to.

Different types of networks can be constructed by giving different definitions to nodes and links. Most of the scholars define ticker names of stocks as nodes [1], [2], [3], [4], while some other scholars define stock indexes as nodes in order to analyze the interaction of the stock markets in different countries [5]. And most literature employed the correlation coefficients to define the links [2], [6], [7], others used Granger-Causality effects as links [8]. In this work, stock network refers to the graph consisting of nodes (vertices) and edges, where nodes correspond to stocks (companies) and edges between them corresponds to the price fluctuation relationships. The later is constructed by computing a correlation coefficient of growth of each pair of stocks over a specific period. Stock network analysis based on stock price correlations was first done by Mantegna [9]. Onnela et al. studied split-adjusted daily closure prices for stocks traded at the New York Stock Exchange and constructed asset graphs and asset trees based on the price correlations and discussed their properties and differences [10]. Vizgunov et al. constructed the stock network based on the Russian stock market [11]. They found that for the Russian market there was a strong connection between the volume of stocks and the structure of maximum cliques during the observation period. Kullmann et al. studied the clustering of companies within a specific stock market index, like the Dow Jones (DJ) or the Standard & Poors 500 (S&P 500), by using the Potts super paramagnetic method [12]. The Topological stability of the China stock market was studied by Huang et al. by constructing a correlation network [13]. Boginski et al. detected cliques and independent sets in it by studying the characteristics of the stock network which represents the structure of the US stock market [14].

III. MINING THE STOCK MARKET NETWORKS

To retrieve information from stock market through network analysis we have to model the stock market as market networks in which nodes represent the ticker name of each stock and edges represents the correlation values of asset returns over a selected time frame. This work examines the time series values of the daily stock prices in the US Stock Market at the year 2016 and establishes connections between pair of stocks. If the cross correlation of the time series of the daily stock prices of two stocks is greater than a threshold of 0.7, we consider that the two stocks are connected. By correlating every stock with every other stock, we end up with a matrix known as the correlation matrix. This matrix plays an important role in modern finance which is especially prevalent in risk analysis and asset portfolio management. Our work aims to study the statistical properties of stock network and more importantly, to investigate the most influential sectors and stocks in the market. One advantage of network analysis is that it enables people to understand the whole network by multiple ways of visualizing the constructed network. It enables to provide some unique attributes for each node (stock) (i.e., degree centrality and lobby centrality). These attributes, defined by different metrics, represent the 'importance' or prominence of stocks within the network.

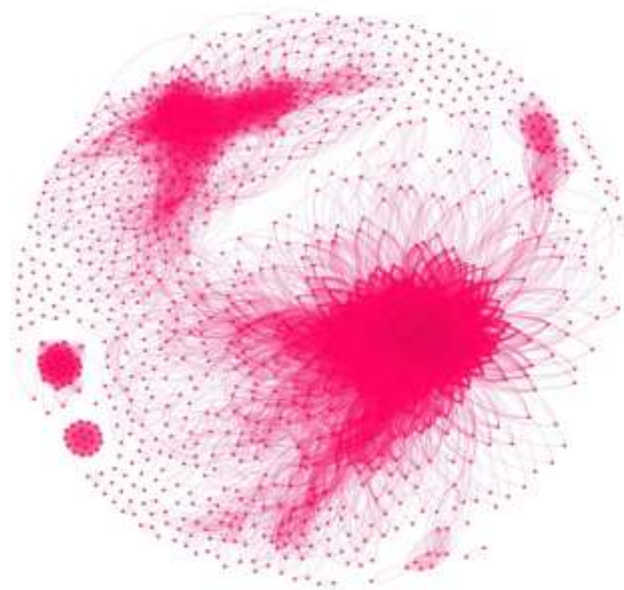

Figure.1. Stock Market Network

A. Degree Centrality

Network approach suggests that power of an individual stock is not an individual attribute, it also interdependent on their relationship with other stocks. Different measures are defined in the literature for the nodes and edges in the network as well as for the network as a whole. It represents the measure of potential importance or influence in the network. Degree centrality measure a stock's central position according to the number of connections to the other stocks. It is the total number of its immediate neighbours. Stocks which may have more ties to other stocks may have greater network of neighbours. This gives them alternative ways to satisfy their needs and hence they are less dependent on a single stock

[1].The degree centrality deg (v) of a vertex v of a network with adjacency matrix A = (a_{iv}) is given by

$$Deg(v) = \sum_{i=1}^{n} a_{iv}$$

B. Lobby Index Centrality

The lobby index of a node x is the largest integer k such that x has at least k neighbors with a degree of at least k. It is clear that a person who has many highly connected neighbors has strong lobbying power and has the ability to influence people's opinions [16]. The degree is denoted by *deg (x)* for nodes and the l-index is defined as follows. Let us consider all y_i neighbours of x so that deg (y_1) ≥ deg (y_2) ...; then,

$$l(x) = \max\{k: \deg(y_k) \geq k\}$$

C. Data

We aggregated daily time series data of 3781 stocks and the descriptive information about every company. A network was constructed by analysing the cross correlations in daily returns in 2016. Previous correlation analysis had showed how correlations arise from external forces across the market [9], [12]. Individual stocks are taken as nodes and a threshold θ is set to create links. At each step of the correlation *Cor*(*i,j*) >0, based on the threshold, we either add a link between i and j, or remove the link. General network measures of stock data for various threshold values are depicted in table I. Out of the 3781 stocks, it is seen that (θ >0.7) exists for 1709 stocks. When the threshold is increased, the number of edges and nodes in the network gets decreased. As a result the average degree, diameter and average path length of the network also gets decreased. In order to reflect the effect of reasonable number of stocks and their interaction in the stock market, we set the threshold at θ > 0.7.

IV. ANALYSIS

The network topology can be interpreted as the inter connections of various nodes of a given system. In case of stock market network, chain-like and star-like topologies are most important. A star-like topology will have a hub as the central connection, to which other nodes will be connected. Chain-like structure indicates lack of powerful companies with high market capitalisation among stocks. Figure 1 shows a star topology with two central hubs, to which all other stocks are connected. Such a network shows the importance of some stocks (which are mainly from banking, insurance and financial sector) that forms the core of the economy and shows the high dependence of other securities on these stocks. When an economic system is robust it can function under a variety of stress and when it is weak, even minor disturbances can cause cascading failures and dislocation of its essential functions.

TABLE I
NETWORK MEASURES FOR DIFFERENT CORRELATION THRESHOLD VALUES

Correlation	0.7	0.75	0.8	0.85
# Nodes	1709	1298	921	560
# Edges	119021	67819	34463	14712
Average Degree	139.287	104.498	74.838	52.54
Network Diameter	19	15	12	9
Average Path Length	4.03	4.319	2.572	2.018
Avg. Clust. Coefficient	0.742	0.748	0.772	0.794

We found that the links between the stocks occurred in five key economic sectors namely technology, oil, basic material, finance and real estate. Banking domain, especially, regional banks are the most linked and connected stocks and can play an important role in the network to prevent the spread of epidemic failures in the stock market.

Top 25 stocks with highest degree are given in table II. All stocks except one in the top twenty stocks in the degree table are regional banks which fall under the generic banking domain. Highest degree is 662 is for WAFD (Washington Federal), the second highest is 641 is for CBU (Community Bank System) and the third is FCF (First Commonwealth Financial Corp & Banks) with degree 638. The stocks which are in the top 25 degree centrality table include FRME (First Merchants Corp, FMBI (First Midwest Bancorp), GS (Goldman Sachs Group), FULT (Fulton Financial Corp), NBTB (NBT Bancorp Inc), PNC (PNC Bank), UCBI (United Community Banks), TRMK (Trustmark Corp), KEY (Keycorp), CHCO (City Holding Company), CVBF (Cvb Financial Corp), TCB (TCF Financial Corp), COLB (Columbia Banking System), ONB (Old National Bancorp Capital), WBS (Webster Financial Corp), IBOC

(International Bancshares Corp), SASR (Sandy Spring Bancorp), BRKL (Brookline Bancorp), RF (Regions Financial Corp), FFBC (First Financial Bancorp), WSBC (Wesbanco Inc). So, it is very clear from the data points that banking domain especially regional banks are the most linked and connected stocks and can play an important role in the network to prevent the spread of epidemic failures in the stock market.

Stocks which topped the lobby centrality of 3781 stocks is shown in Table 2. 112 stocks came at the top most position (lobby centrality = 341) of which 81 stocks were from banking sector, 13 from financial sector, 5 from insurance, 2 each from technology, machine, industrials and business Services.Also, energy, chemicals, retail, transport, real estate and building sector have one stock each in the table. The lobbying power is very much prominent and strong for banking sector.

TABLE II
SECTORS WITH HIGHEST DEGREE CENTRALITY STOCKS IN THE STOCK NETWORK.

Degree	**Ticker**	**Company Name**	**Sector**
662	WAFD	Washington Federal	Banks
641	CBU	Community Bank System	Banks
638	FCF	First Commonwealth Financial	Banks
636	PFS	Provident Financial Services	Banks
631	FRME	First Merchants Corp	Banks
632	FMBI	First Midwest Bancorp	Banks
625	GS	Goldman Sachs Group	Finance
624	FULT	Fulton Financial Corp	Banks
619	NBTB	NBT Bancorp Inc	Banks
622	PNC	PNC Bank	Banks
617	UCBI	United Community Banks	Banks
622	TRMK	Trustmark Corp	Banks
613	KEY	Keycorp	Banks
607	CHCO	City Holding Company	Banks
614	CVBF	Cvb Financial Corp	Banks
617	TCB	TCF Financial Corp	Banks
605	COLB	Columbia Banking System	Banks
613	ONB	Old National Bancorp Capital	Banks
606	WBS	Webster Financial Corp	Banks
605	IBOC	International Bancshares Corp	Banks
606	SASR	Sandy Spring Bancorp	Banks
599	BRKL	Brookline Bancorp	Banks
604	RF	Regions Financial Corp	Banks
599	FFBC	First Financial Bancorp	Banks
595	WSBC	Wesbanco Inc	Banks

V. CONCLUSION

Identification of the dynamics of stock market is much important for players, investors and financial policy makers. Stock market is regarded as a complex system owing to the complex dynamics it exhibit. There are many reasons for this complexity, of which interdependency of various stocks on others is a prominent one. Networks which have the capability to model complex systems and inherent relationships of its various components can be effectively used for stock market mining too. We created stock market network of US stocks by capturing its dynamics over the period of one year (January 1, 2016 to December 31, 2016) in 2016. Identification of the structural properties of the market was one of our objectives.

After studying the properties of the market using various correlation co-efficient thresholds, a large network with stock pair correlation (θ) greater than 0.7 is created. Density of the network itself hints the possibility of heavy dependency of stocks on other stocks. Using that

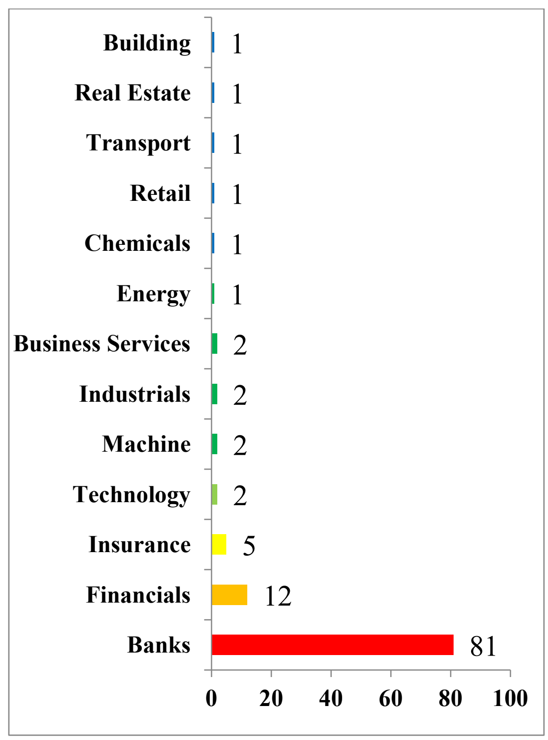

Figure. 2. Sectors with highest Lobby Centrality (341).

network, various key players in the market are identified based on their number of dependencies and also their dependency to other stocks with more dependencies. It is found that stocks with more number of dependencies do not have high dependencies with the ones that have high dependencies. We used lobby index for identification key players with relations with other key players. Players in sectors like banking sector, financial sector, insurance, finance, technology, machine, industrials, business services, energy, chemicals, retail, transport, real estate and building sector are found to have high lobbying power and exercise much control over the market. Thus, network based data mining can reveal important inputs for decision making for instance the portfolio of influential players as well as sectors in the market.

Other portfolio analysis methods such as community detection are intended to be done as an extension of this work. Dynamic study of the structural properties of the market is another interesting that can be studied. Dynamics of the stocks that exercise high and medium lobbying power and correlation dependency of lobbying power is also considered for further research.

REFERENCES

[1] J. Liu, C. Tse, and K. He, "Detecting stock market fluctuation from stock network structure variation," in *International Symposium on Nonlinear Theory and Its Applications*, 2008.

[2] K. T. Chi, J. Liu, and F. C. Lau, "A network perspective of the stock market," *Journal of Empirical Finance*, vol. 17, no. 4, pp. 659–667, 2010.

[3] W. Lan and G. Zhao, "Stocks network of coal and power sectors in china stock markets," *Information Computing and Applications*, pp. 201–208, 2011.

[4] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, "Topology of correlation-based minimal spanning trees in real and model markets," *Physical Review E*, vol. 68, no. 4, p. 046130, 2003.

[5] X. F. Liu and C. Tse, "A complex network perspective to volatility in stock markets," in *2010 International Symposium on Nonlinear Theory and its Applications NOLTA*, 2010, pp. 402–5.

[6] S. Kumar and N. Deo, "Correlation and network analysis of global financial indices," *Physical Review E*, vol. 86, no. 2, p. 026101, 2012.

[7] J. Liu, C. K. Tse, and K. He, "Fierce stock market fluctuation disrupts scale free distribution," *Quantitative Finance*, vol. 11, no. 6, pp. 817–823, 2011.

[8] K. J. Wei, Y.-J. Liu, C.-C. Yang, and G.-S. Chaung, "Volatility and price change spillover effects across the developed and emerging markets," *Pacific-Basin Finance Journal*, vol. 3, no. 1, pp. 113–136, 1995.

[9] 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.

[10] 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, p. 056110, 2003.

[11] A. Vizgunov, B. Goldengorin, V. Kalyagin, A. Koldanov, P. Koldanov, and P. Pardalos, "Network approach for the russian stock market," *Computational Management Science*, vol. 11, no. 1-2, pp. 45–55, 2014.

[12] L. Kullmann, J. Kertesz, and R. Mantegna, "Identification of clusters of companies in stock indices via potts super paramagnetic transitions," *Physica A: Statistical Mechanics and its Applications*, vol. 287, no. 3, pp. 412–419, 2000.

[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] V. Boginski, S. Butenko, and P. M. Pardalos, "Statistical analysis of financial networks," *Computational statistics & data analysis*, vol. 48, no. 2, pp. 431–443, 2005.

[15] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.

[16] A. Korn, A. Schubert, and A. Telcs, "Lobby index in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 11, pp. 2221–2226, 2009.

[17] Minutes of the Board of Governors of the Federal Reserve System, 14 March 2008, https://www. federalreserve.gov/other20080627a1.pdf, accessed on 20th April 2017.