VISUALIZING THE INDIAN STOCK MARKET: A COMPLEX NETWORKS APPROACH

Mohamed Rehan M S, M Hari Haran, Neha Singh Chauhan and Divya Grover
Center of Excellence in Systems Science,
Indian Institute of Technology (IIT) Jodhpur, Jodhpur, India

ABSTRACT

This paper analyses the top-level structure of the Bombay Stock Exchange using a complex networks framework. The market is visualized as a network, consisting of stock prices as its nodes, and their dynamic interactions being reflected in the edges. The edges used here, signify cross-correlations between any two stocks and take into account the time-lagged interdependence on each other. By processing the empirical data in hand, modular communities are identified. These are used to analyze the dynamics of the network, and give an insight to the trends generally observed in the market.

KEYWORDS: *Nonlinear systems, Network theory (graphs), Complex Networks, Stock Markets, Financial Management, Econophysics*

I. INTRODUCTION

A stock market is a public entity for the trading of company shares and derivatives at an agreed price. It need not be always physical. The role of the market is to provide a common platform for the buyers and sellers to exchange their holdings, and for companies to raise capital.

Stock markets are perfect examples of engineered systems, which have come to acquire a life of their own, and slip away from the grasp of conventional techniques. Various techniques have been employed for long, to analyze the non-linear dynamics of stock markets. Various statistical methods, such as regression analysis [1], and moving averages [2] have been used to study this problem. However, they are unable to do proper justice, since they do not often account for the interactions between the entities, which is what actually complicates the scenario, and is the crux of the issue.

The complexity of the interaction between stocks is what makes the market peculiar. To analyze this, we propose using complex networks, aka graph theory. This concept traces its roots back to 1736 when Euler published his findings on the famous problem of Seven Bridges of Königsberg. Complex networks in general, are powerful tools to analyze the interdependencies of a large number of interconnected entities. It has been successfully applied to various natural as well as engineered networks like the food web [3], internet [4], transport networks [5], power grids [6] and more recently, to the study of financial systems.

In a financial market, although there are several parameters involved, such as the prices, volume of trade, valuation, etc., the dynamics are beautifully characterized to a great extent by the percentage change. The price and returns for a particular stock in the market is influenced by numerous factors, both internal and external. The movement of other stocks has a large influence on any given stock, which is why we observe epidemics in the market, such as stock market crashes.

Here, the analysis is done with a different approach, accounting for intra-day fluctuations by using weighted average prices, and using percentage returns as a metric, rather than just returns or logarithmic returns, as it is a normalized metric. Cross-correlation analysis, which takes into account possible time lag between the stocks is also employed, making the model, more intuitive and realistic. Also, community identification was done at different levels, and the threshold is not arbitrarily fixed in the winner takes all model, but instead, is obtained from the data itself.

The organization of this paper is as follows: Section 2 briefly discusses the work done in this area so far. Section 3 presents methods employed in the study. Section 4 is devoted to provide the detailed results of this analysis and draw conclusions based on them. Section 5 discusses directions for the future scope of this work.

II. RELATED WORK

Haldane [7] (2009) first proved that the theory of complex networks provides an insight to the stability of financial systems. However, long before, Mantegna[8] (1999) had proposed a hierarchical structure for financial markets resembling networks, which was later supplemented by the works of Vandewalle et al.[9] (2001) and Onnela et al.[10] (2003). Boginski et al. [11] [12] (2005, 2006) implemented this to study the US stocks markets. This work was continued upon, by Huang et al. [13] (2009), Namaki et al. [14] (2011), Jallo et al. [15] (2013), Vizgunov et al. [16] (2013), in the scope of various stock markets around the globe. Pan et al. [17] (2007) analyzed the stock price movements in emerging markets, using the National Stock Exchange (NSE), India, as an example.

III. METHODOLOGY

The Bombay Stock Exchange, being the world's largest exchange in terms of listed members has over 5000 stocks listed on it. However, the top 500 stocks contribute to more than 90% of the total market capitalization of the exchange [18]. Hence an analysis of these stocks would give us a picture of the network structure of the exchange.

The timespan in consideration is 10 financial years from, 1st April 2003 to 31st March 2013. Any company which has not been in the top 500 companies consistently over these 10 years (including newly listed companies) are excluded from the analysis. This process of filtering identified 181 companies which have been in the top 500 list of the BSE over the past 10 years. The data is collected from the archives of the BSE, as listed on its website [19]. The length of one time tick in consideration is one day, where the Weighted Average Price (WAP) over the period of a day is used as the metric / indicative price for that time tick.

From the time series of WAPs, percentage changes for each day, for every stock is calculated. Due to low volumes, and a variety of other factors, some data points are missing for certain companies. For all those trading days for which data was missing, the percentage change in the WAP is assumed to be 0.

Using this newly obtained time series of percentage changes in the prices, cross correlations are computed for every possible pair, in a time window of ±10 trading days. Cross correlation is calculated by using the relation [20]

$$\frac{E((X(t)-\mu_t)(Y(t-\tau)-\mu_{t-\tau}))}{\sigma_X\sigma_Y}$$

The maximum correlation is identified from the 21 values in the ±10 trading days window. This is deemed to be the correlation between the two given stocks. The time difference/shift between the two time series is called the time lag between the two. Hence, 'A' has a time lag of 'x' with 'B' implies that 'A' leads 'B' by 'x' time ticks.

By iterating this process for all 32761 possible pairs (including self-pairs), two matrices of 181 x 181 dimension are arrived at. The first one contained values of the pairwise correlation, and is, as expected, symmetric in nature. The other, gives values of the time lags for every possible pair, and is anti-symmetric in nature. The correlation matrix is used as the adjacency matrix for the weighted representative graph of the BSE. These correlations form the weight of the edges of the graph.

To make the structure of the graph more relevant, and intuitive, the lengths of the edges are also fixed, in such a manner that they are indicative of the dependencies of the nodes. Those which vary similarly, and have a high value for their correlation should be closer, and vice versa. Hence the edge lengths are required to be inversely related to the correlation. The empirical length of the edge is evaluated by the relation $\sqrt{1-\rho^2}$ where ρ stands for the correlation between the two stocks represented by the nodes between which the edge exists. This relationship is preferred over others as it takes care of all conditions that length of edges should follow, including linear edges, and triangular inequalities.

The calculated parameters are used to visualize the graph. As expected, the raw graph does not provide much of an insight into the network structure, partly due to the fact that being completely connected, the exodus of information partly inhibits extraction of meaningful parameters from the graph. As it is very evident, not all edges are crucial in the market dynamics. Hence, some of the edges are to be eliminated. For this purpose a winner takes all, coarse grain modelling technique is employed. A threshold is to be identified for the same but, the threshold cannot be fixed arbitrarily and has to come from the data itself.

The mean value of the correlations is identified to be around 0.27 with a standard deviation of 0.05. Hence to identify the best possible threshold, the number of meaningful communities [21] are estimated, at different values of the threshold. The threshold is varied from 0 to 1, in steps of 0.05, and communities were identified and enumerated, at modularity resolutions of 0.75 and 1. It is observed that the maximum number of meaningful clusters occur at a threshold value of 0.2 and modularity resolution of 0.75. The 0.15 to 0.25 interval is then further fine-tuned, by decreasing the step size, and subsequently intervals. It is further identified that the maximum number of meaningful communities actually occurs at a threshold of 0.1926.

To validate this observation, similar trials are conducted with two year long time series, for 5 different time ranges, the results obtained, in terms of threshold and modularity resolution are consistent, within experimental limits.

For the initial data set, of 10 years, communities are identified, companies in them are studied, for similarities.

IV. RESULTS AND CONCLUSION

The network parameters identified (Table 1) give an idea of the network topology underlying the Bombay Stock Exchange. High clustering coefficient of the network indicates presence of highly interacting sets of stocks within the network. The diameter of the network is relatively small, indicating, that the network is quite well connected, although the graph is not very dense.

Table 1. List of network parameters

Parameter	Value
Average Clustering Coefficient	0.913
Average Path Length	1.20
Diameter	3
Radius	0
Number of Weakly Connected Components	182
Average Degree	72.28
Average weighted degree	22.158
Graph Density	0.200

Table 2. List of modular communities

Community 1	Community 2	Community 3
HDFC	WYETH	MADRASCEM
BHEL	PGHH	PIDILITIND
HDFCBANK	ABBOTINDIA	BIRLACORPN
HEROMOTOCO	GLAXO	PRISMCEM
INFY	NOVARTIND	SHREECEM
M&M	SANOFI	ACC
WIPRO	GSKCONS	AMBUJACEM
ICICIBANK	PFIZER	ITC
AXISBANK	GILLETTE	KESORAMIND
HCLTECH	3MINDIA	INDIACEM
BHARTIARTL	LINDEINDIA	

Community 4	Community 5	Community 6
ORIENTBANK	PEL	CRISIL

CONCOR	TORNTPHARM	TATAPOWER
BANKBARODA	BHARATFORG	TATAINVEST
BANKINDIA	ORCHIDCHEM	CARBORUNIV
CORPBANK	IPCALAB	MONSANTO
PNB	AUROPHARMA	TNPL
UNIONBANK	GLENMARK	
CANBK		

Community 7	Community 8	Community 9
HINDPETRO	ABB	BEL
MRPL	CROMPGREAV	CIPLA
CHENNPETRO	THERMAX	RANBAXY
BPCL	CUMMINSIND	TATAELXSI
IOC	ASIANPAINT	HINDUNILVR
GAIL	ADANIENT	MARICO

Community 10	Community 11	Community 12
DABUR	LAXMIMACH	BASF
LUPIN	LICHSGFIN	TRENT
TVSMOTOR	MRF	AKZOINDIA
BRITANNIA	NESTLEIND	INGVYSYABK
	APOLLOTYRE	DENABANK

Community 13	Community 14	Community 15
IDBI	BHUSANSTL	BLUESTARCO
DRREDDY	MAHSEAMLES	EXIDEIND
GODREJIND	RUCHISOYA	HOTELEELA
INDUSINDBK	BAYERCROP	VOLTAS
GODREJCP	JINDALSTEL	EIHOTEL

Community 16	Community 17	Community 18
ARVIND	CESC	GUJFLUORO
RAYMOND	HCL-INSYS	MAX
TATAGLOBAL	EICHERMOT	SKFINDIA
APOLLOHOSP	NEYVELILIG	ESCORTS

Community 19	Community 20	Community 21
CENTURYTEX	BAJAJHIND	ONGC
SUNPHARMA	BALRAMCHIN	ASHOKLEY
SBIN	EIDPARRY	BOSCHLTD
GREAVESCOT	WOCKPHARMA	INDHOTEL

Community 22	Community 23	Community 24
GUJRATGAS	SAIL	INGERRAND
RELCAPITAL	SESAGOA	BATAINDIA
RELIANCE	TATASTEEL	BERGEPAINT
RELINFRA	CASTROL	TITAN

Community 25
GTL
ROLTA
CMC
MPHASIS
INFOTECENT
POLARIS

The 25 communities hence detected give valuable insight into the fundamental divisions of the market. Some very interesting features that can be brought to one's notice are:

- Out of the communities identified, the one of the biggest modular component was of stocks generally referred to as 'market movers', with all of them being a part of the BSE SENSEX. The fact that they move in a similar fashion, although some of them lagged others emphasizes on the fact that to a certain extent, the Indian investor is not very rational, and can be driven by sentiments, as the community has a variety of companies from diverse fields.
- Many of the remaining modular components are observed to be largely homogeneous i.e. belonging to the same sector, including the other biggest one. In the homogeneous communities, it is identified that they do not always heed to the 'sector indices' on the BSE. For example, a particular cluster has only companies in the cement industry, while another from the petroleum and petrochemical industry contained companies from the sector which are not a part of the BSE Oil & Gas index. Hence this indicates that there are indeed many factors which affect certain groups of companies only. Moreover they also hint that the existing sectorial indices do not give a clear picture of the market segmentation and sectorial dynamics, since they do not, at times, include all the companies related to directly as well as indirectly to the sector.
- Additionally, heterogeneous communities are also observed which are indicative of the resistance of the market to fluctuations in one industry alone. One of the biggest cluster in itself, comprises of companies cutting across the divisions of industry and cap, thereby resonating the strength and robustness of the system.

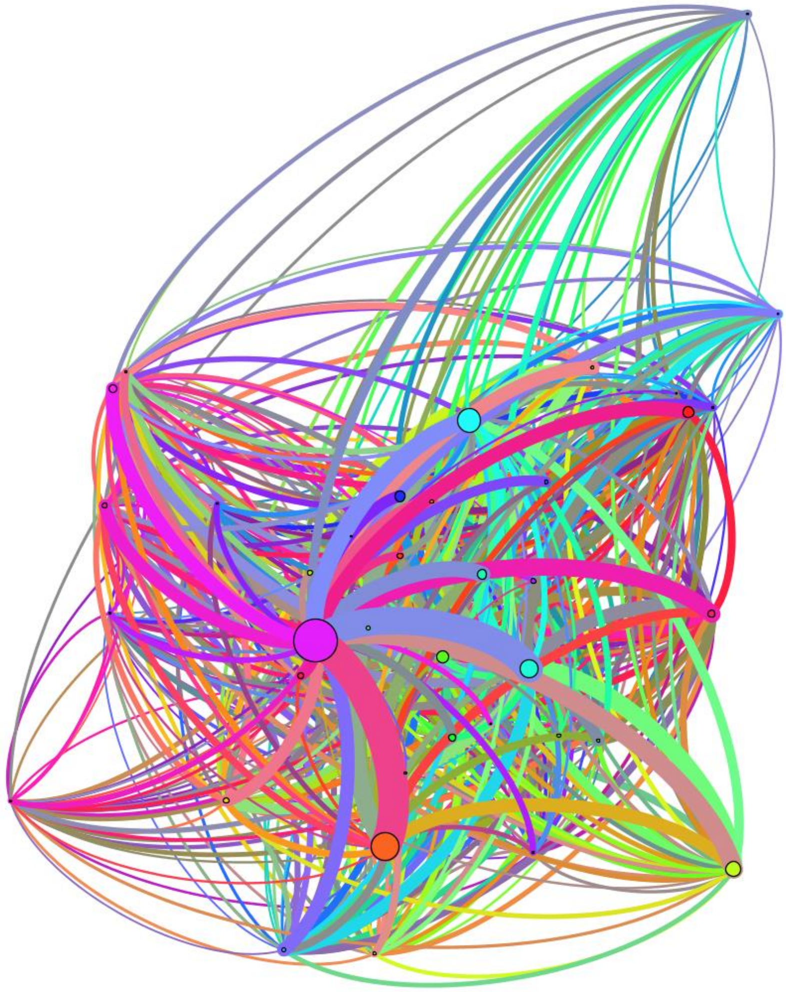

Figure 1. Network structure at a threshold of 0.1926, and a modularity resolution of 0.75. Communities are represented as the nodes and the size of the node is to scale with the number of members in the respective community

V. Future Work

The network topology obtained is an indicator of the dynamics of the market. Although the correlations may vary slightly, the topology remains more or less similar, and hence can be used to predict the movement of stock prices and optimize portfolios. Also, additional research can be conducted to examine the robustness of the network, and the resistance to targeted attacks, as well as attacks on the entire system. The time lags obtained can also be put to use in devising optimum response strategies in reaction to such attacks. The information obtained can also help in predicting the flow of epidemics through the networks, and isolation techniques. In short, these results form a basic framework and foundation, for a wide range of financial engineering applications.

References

[1]. Roberts, Harry V. "Stock-Market “Patterns” And Financial Analysis: Methodological Suggestions." *The Journal of Finance* 14.1 (1959): 1-10.
[2]. Box, George EP, Gwilym M. Jenkins, and John F. MacGregor. "Some recent advances in forecasting and control." *Applied Statistics* (1974): 158-179.
[3]. Dunne, Jennifer A., Richard J. Williams, and Neo D. Martinez. "Network structure and biodiversity loss in food webs: robustness increases with connectance." *Ecology Letters* 5.4 (2002): 558-567.
[4]. Albert, Réka, Hawoong Jeong, and Albert-László Barabási. "Internet: Diameter of the world-wide web." *Nature* 401.6749 (1999): 130-131.
[5]. Banavar, Jayanth R., Amos Maritan, and Andrea Rinaldo. "Size and form in efficient transportation networks." *Nature* 399.6732 (1999): 130-132.
[6]. Albert, Réka, István Albert, and Gary L. Nakarado. "Structural vulnerability of the North American power grid." *Physical Review E* 69.2 (2004): 025103.
[7]. Haldane, Andrew G. "Rethinking the financial network." *Speech delivered at the Financial Student Association, Amsterdam, April* (2009).
[8]. Mantegna, Rosario N. "Hierarchical structure in financial markets." *The European Physical Journal B-Condensed Matter and Complex Systems* 11.1 (1999): 193-197.
[9]. Vandewalle, N., F. Brisbois, and X. Tordoir. "Non-random topology of stock markets." (2001): 372-374.
[10]. Onnela, J-P., et al. "Dynamics of market correlations: Taxonomy and portfolio analysis." *Physical Review E* 68.5 (2003): 056110.
[11]. Boginski, Vladimir, Sergiy Butenko, and Panos M. Pardalos. "Statistical analysis of financial networks." *Computational Statistics & Data Analysis* 48.2 (2005): 431-443.
[12]. Boginski, Vladimir, Sergiy Butenko, and Panos M. Pardalos. "Mining market data: a network approach." *Computers & Operations Research* 33.11 (2006): 3171-3184.
[13]. Huang, Wei-Qiang, Xin-Tian Zhuang, and Shuang Yao. "A network analysis of the Chinese stock market." *Physica A: Statistical Mechanics and its Applications* 388.14 (2009): 2956-2964.
[14]. Namaki, A., et al. "Network analysis of a financial market based on genuine correlation and threshold method." *Physica A: Statistical Mechanics and its Applications* 390.21 (2011): 3835-3841.
[15]. Jallo, David, et al. "Network-Based Representation of Stock Market Dynamics: An Application to American and Swedish Stock Markets." *Models, Algorithms, and Technologies for Network Analysis*. Springer New York, 2013. 93-106.
[16]. Vizgunov, A., et al. "Network approach for the Russian stock market."*Computational Management Science* (2013): 1-11.
[17]. Pan, Raj Kumar, and Sitabhra Sinha. "Collective behavior of stock price movements in an emerging market." *Physical Review E* 76.4 (2007): 046116.
[18]. Highlights of S&P BSE Indices (http://www.bseindia.com/indices/indiceshighlights.aspx)
[19]. Bombay Stock Exchange Archives (http://www.bseindia.com/stockinfo/stockprc.aspx)
[20]. Gep, B. O. X., Gailim M. Jenkins, and Gregory C. Reinsel. "Time series analysis: forecasting and control." *Oakland CA: Holden-Dag* (1976).
[21]. Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (2008): P10008.
[22]. Bastian, M., Heymann, S., & Jacomy, M. (2009, May). Gephi: An open source software for exploring and manipulating networks. In *International AAAI conference on weblogs and social media* (Vol. 2). Menlo Park, CA: AAAI Press.
[23]. Newman, Mark. *Networks: an introduction*. OUP Oxford, 2009.
[24]. Newman, Mark, Albert-Laszlo Barabasi, and Duncan J. Watts. *The structure and dynamics of networks*. Princeton University Press, 2011.

AUTHORS

Mohamed Rehan M S is currently pursuing an undergraduate degree in Systems Science, at Indian Institute of Technology (IIT) Jodhpur. India. His fields of interest include Financial Engineering, Non-Linear Dynamics, Process Optimization and Complex Networks.

Neha Singh Chauhan is an undergraduate student degree at the Center of Excellence in Systems Science, at Indian Institute of Technology (IIT) Jodhpur. India. Her fields of interest are Complex Networks, Optimization Techniques, Machine Learning and Artificial Intelligence.

Divya Grover is pursuing B. Tech in Systems Science, from Indian Institute of Technology (IIT) Jodhpur. India. He is interested in Statistics, Graph Theory and their applications in the fields of Sociology and Finance.

M Hari Haran is a B.Tech student of Systems Science, at Indian Institute of Technology (IIT) Jodhpur. India. His interests lie in theoretical and practical applications of Complex Networks as well as System Dynamics