Graph and Information Theoretic Analysis of Correlation between Stock Market Indices

Gabriel Dominik E. Sison
National Institute of Physics
University of the Philippines, Diliman
Quezon City, Philippines
gsison@nip.upd.edu.ph

Ulyses A. Solon Jr.
Institute of Mathematics
University of the Philippines, Diliman
Quezon City, Philippines
usolon@math.upd.edu.ph

Aimee B. Rarugal
National Institute of Physics
University of the Philippines, Diliman
Quezon City, Philippines
ararugal@nip.upd.edu.ph

Guido David
Institute of Mathematics
University of the Philippines, Diliman
Quezon City, Philippines
gdavid@math.upd.edu.ph

Giovanni A. Tapang
National Institute of Physics
University of the Philippines, Diliman
Quezon City, Philippines
gtapang@nip.upd.edu.ph

ABSTRACT

Most economies today are interrelated. The push towards one globally connected community lead to having countries which depend on one another, not just as governments but also as businesses aiming to improve economic output. This paper concerns itself with identifying countries, being represented by their main stock market indices, as agents determining each other's behaviour. We look at the behaviour of each index in a fixed time interval (1998-2014) and relate it with the behaviour of other indices using correlation, network theory and transfer entropy. From the graphs, we determined communities based on geographical basis and that FTSE is highly influenced by the past values of other stock market indices.

Categories and Subject Descriptors

F.1.2 [**Mathematics of Computing**]: Discrete Mathematics—*Graph Theory*; M.3.11 [**Applied Computing**]: Metrics—*complexity measures, performance measures*

Keywords

Graph Theory, Statistics, Complex Systems

1. INTRODUCTION

Many systems in nature are driven by mechanisms with non-obvious patterns. From the neurons in the brain [16],to patterns of disease incidence [5] , it is often difficult to see how one part influences the others. In addition such systems are often subject to random fluctuations that further muddy analysis and prevent determination of causation and correlation. The analysis of such systems is the subject of the field of complex systems.

One such complex system is the economy. The economies of various countries are highly interlinked in today's age of globalization [10]. A crash in the New York stock market can cause crashes in Europe while causing climbs in Asia. Due to the importance of the economy on people's lives, a keen understanding is vital.

One technique to aid in analysing complex systems is the use of graph theory. Graph theory is the study of graphs, a mathematical representation of a set of objects [12]. In a graph, each object is denoted by a vertex and related vertices are connected by edges. Graph theory has been used to analyze systems as large as the internet [8], to systems as small as school karate clubs [19]. Of interest is the utility of graphs in determining critical members of a system and detecting highly connected components, both of which are important in determining the dynamics of the system.

A time series is a collection of measured observations of one or multiple variables associated with a certain system that evolves over time [9]. The observations themselves are used to characterize the data or if at all possible, the system associated with that time series. When two time series are applied with appropriate techniques and interpretation, one can find important relationships and possibly causal factors between them. The main objective of time series analysis is to seek relations among its measured variables [4].

Data correlation, meanwhile, is a term used to denote dependence of data sets. One may observe an increase in damage caused by typhoons with the increase in the number of typhoons visiting the country, hence even without numerical methods one may infer a positive correlation.

We take inspiration from neurobiology. A recent technique to characterize the brain's functions is take a time series of neuron activations and take their correlation with each other [16]. From this set of correlations, one can then construct a graphs of neurons. This technique has been successful in discovering functional units inside the human brain. In this paper we attempt this technique on another system of signals, a set of 26 stock market indices in order to determine their dynamics. Similar methods have already been employed to study the stock market, however, none have focused on indices [7, 13, 14, 15, 18]. Previous studies have also tended to focus on individual prices in shorter time frames [18]. Instead this work focuses on a longer time scale of 16 years and on indices that account for more of the market.

2. METHODOLOGY

The time series data considered consisted of adjusted closing prices of different stock market indices representative of several countries. Adjusted closing prices, whenever available, were obtained directly from sources. No further changes were made. Yahoo! Finance, the site where most of the data came from, defines the adjusted closing price as the closing price for the requested day, week, or month, adjusted for all applicable splits and dividend distributions [1]. The said adjustments are made so as to reveal the value of the stock (in our case indices) sans the splits and dividends.

A stock market index is a measurement of the value of a section of the stock market. It is typically a weighted average of prices of selected stocks. For this study we looked at stock market indices measuring values, not just of sections but of whole stock markets.

Data required was publicly available [2, 3]. The data was trimmed such that the number of data points in each financial index was the same, and spanned approximately the same time period from about mid-1998 to September/October 2014.

For each pair of indices, the correlation between the data sets was computed. The Pearson correlation coefficient, given by

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

for data sets $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$, with means $\bar{x}$ and $\bar{y}$ respectively, was used as the measure of correlation [4].

Two graphs were constructed. The first graph, shown in figure 1, connects two indices if their correlation is higher than the average correlation minus one standard deviation. The second graph, shown in figure 2, connects two indices if their correlation is greater than the average. The Louvain algorithm was then applied to detect the communities in the graph. The Louvain algorithm detects communities by finding the partition with maximum value of modularity [6]. The modularity $\mathbf{Q}$ is given by

$$Q = \sum_{i=1}^{c}(e_{ii} - a_i^2) \tag{2}$$

where e_{ii} is the fraction of edges with both ends in the same community and a_i is the fraction of edges with at least one edge in community i. The modularity is maximized by detecting small local clusters and then merging them if it increased the global modularity. The communities in a graph represent sets of vertices that connect to each other more than they connect to other vertices in the graph.

In order to determine key vertices, the betweenness centrality and the weighted degree of each vertex was computed. The betweenness centrality is the measure of the number of shortest paths that pass through a given vertex in the graph.

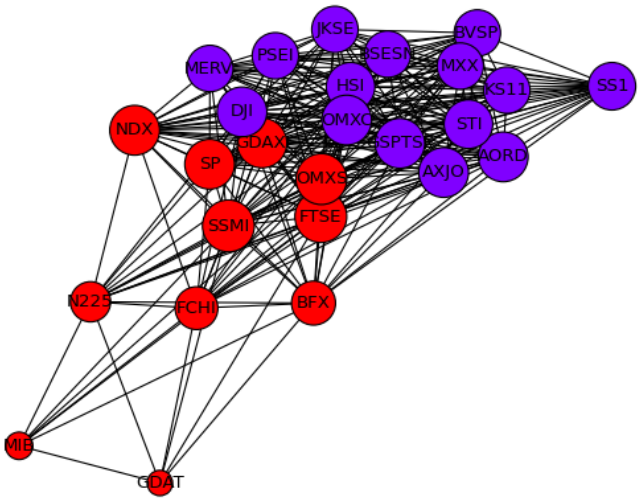

Figure 1: Graph constructed from indices' correlation. Vertices are colored based on community and sized based on weighted degree. Vertices are connected if their correlation is no less than one standard deviation from the average.

It is given by the equation

$$BC(u) = \sum_{v,w \neq u} \frac{\sigma(v, w|u)}{\sigma(v, w)} \tag{3}$$

where $\sigma(v, w|u)$ is the number of shortest paths between nodes v and w passing through node u, $\sigma(v, w)$ is the number of all shortest paths between v and w [11].

Transfer entropy is an information theoretic measure that quantifies the coherence between systems in time by detecting the directed information transport from one system to another. It effectively distinguishes the asymmetry in the relationship of two processes. It also determines the driving and influential mechanism between them thus reflects the causality in these processes [17].

Transfer entropy is calculated using the equation

$$T_{Y \to X} = \sum_k p\left(x_{k+1}, x_k, y_k\right) \log_2 \frac{p\left(x_{k+1} \mid x_k, y_k\right)}{p\left(x_{k+1} \mid x_k\right)} \tag{4}$$

where $p(x|y)$ are conditional probabilities. and X and Y are the two processes

3. RESULTS AND DISCUSSIONS

Figures 1 and 2 show the graph visualizations of the index correlation graphs. We can see that increasing the threshold increased the number of defined communities from two to three. Regardless of the threshold, GDAT (Greece), MIB (Italy), N225 (Japan), FCHI (France), and BFX (Belgium) are all in the same community. This community can be seen as representing the periphery of the graph. This is also reflected in their low weighted degrees. We also see that GDAXI (Germany), OMXS (Denmark), FSTE (United Kingdom), NDX (NASDAQ), DJI (Dow-Jones), SP (Standard and Poor), and SSMI (Switzerland) are always in the same communities. This is to be expected for NDX, DJI,

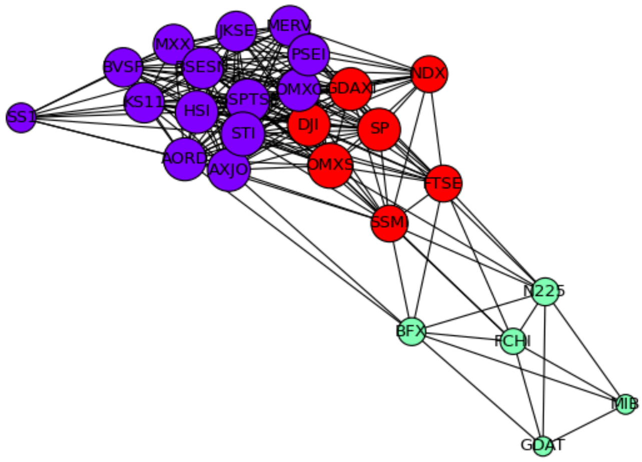

Figure 2: Graph constructed from indices' correlation. Vertices are colored based on community and sized based on weighted degree. Vertices are connected if their correlation is no less than the average. Edge exclusion causes one of the communities in figure 1 to separate.

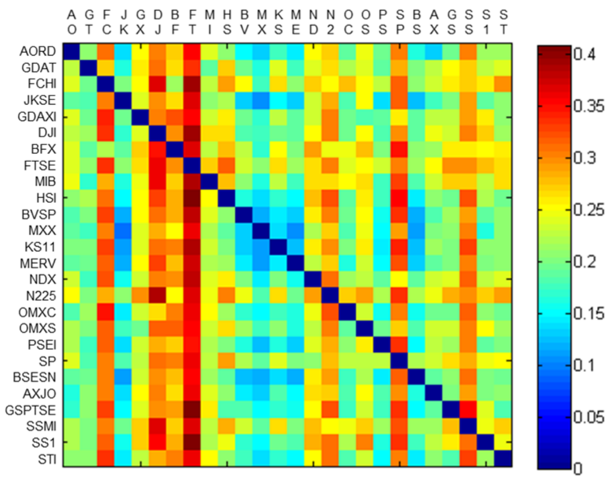

Figure 3: Transfer Entropy between the indices represented as a color map

Table 1: The betweenness centrality of the 5 most central nodes in Graph 1 (columns 1 and 2) and Graph 2 (columns 3 and 4).

Index	Norm. BC	Index	Norm. BC
SSMI	0.041	OMXS	0.091
FTSE	0.041	BFX	0.064
OMXS	0.021	N225	0.051
BFX	0.016	FTSE	0.045
FCHI	0.013	AORD	0.043

and SP as all are United States-based indices. This community structure suggests groupings of stock markets highly related to each other.

From the betweenness centrality in table 1, we see that FTSE, OMXS, and BFX are consistently central despite the threshold. This reflects their importance as links between the communities in the graph. This represents an interesting possibility, that these vertices might be used as guides on whether a change in the core of the graph will affect changes in the periphery or vice-versa.

Transfer entropy is an information theoretic measure that indicates the influential factor between two series of stock market indices. In figure 3, FTSE has the highest transfer entropy values which implies that its distribution is highly influenced by the past values of other stock market indices. It is followed by DJI, SP and SSMI. Moreover, FCHI is also influenced by the other stock market indices except FTSE and N225.

4. CONCLUSIONS

In this paper, we have shown the utility of graph theory in analysing relationships between complex systems by grouping indices into related sets. From the visualizations we can also see clear demarcation between the core and periphery of the system. In addition, the relationships established by the graph can be a useful starting point for investigating how economies are interlinked. On the other hand, transfer entropy allows us to see patterns of influence within the set of indices

References

[1] About historical prices. Yahoo! Finance, Oct. 2014.

[2] Major World Indices. Yahoo! Finance, Oct. 2014.

[3] U.S. Stocks Home - Markets Data Center - WSJ.com. Wall Street Journal, Oct. 2014.

[4] A. M. Albano, P. D. Brodfuehrer, C. J. Cellucci, X. T. Tigno, and P. E. Rapp. Time series analysis, or the quest for quantitative measures of time dependent behavior. *Philippine Science Letters Review*, 1:22, 2009.

[5] R. Anderson, B. Grenfell, and R. May. Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis. *Journal of Hygiene*, 93(03):587–608, 1984.

[6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[7] V. Boginski, S. Butenko, and P. M. Pardalos. Mining market data: a network approach. *Computers & Operations Research*, 33(11):3171–3184, 2006.

[8] K. L. Calvert, M. B. Doar, and E. W. Zegura. Modeling internet topology. *Communications Magazine, IEEE*, 35(6):160–163, 1997.

[9] C. Chatfield. *The analysis of time series: an introduction*. CRC press, 2013.

[10] P. Dicken. *Global shift: Reshaping the global economic map in the 21st century*. Sage, 2003.

[11] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[12] J. L. Gross and J. Yellen. *Graph theory and its applications*. CRC press, 2005.

[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*, 388(14):2956–2964, 2009.

[14] H. Kim, I. Kim, Y. Lee, and B. Kahng. Scale-free network in stock markets. *Journal-Korean Physical Society*, 40:1105–1108, 2002.

[15] L. Kullmann, J. Kertész, and K. Kaski. Time-dependent cross-correlations between different stock returns: A directed network of influence. *Physical Review E*, 66(2):026125, 2002.

[16] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[17] T. Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

[18] C. K. Tse, J. Liu, and F. Lau. A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659–667, 2010.

[19] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.

5. ACKNOWLEDGEMENTS

This research was done at the Instrumentation Physics Laboratory in the National Institute of Physics