ISSN 2395-8618

A Complex Network Approach to Identify Potential Financial Scandals: The Colombian Market Case

Jaime Humberto Nino-Pena, Cesar Garcia-Diaz, and German Jairo Hernandez

Abstract—**Financial data is abundant, diverse and generated in large volumes at any time worldwide. Finding fast and reliable ways of analyzing it is key for market actors (regulators, market makers, brokers and investors). In this work, we intend to use graph theory as a vehicle to analyze relationships among owners of publicly traded companies in order to extract latent dynamics that are very difficult to discover otherwise. As a case study we took the bankruptcy of the largest brokerage firm in Colombia in 2012. Network theory uncovered latent dynamics in the relationships among key owners and companies during the period of analysis (2009 - 2012) that could be used in the future as an early warning tool for market participants.**

Index Terms—**computational finance, graph theory, knowledge discovery, time series mining, financial data mining**

I. Introduction

Financial data is being generated almost instantly across the world. It comes in two basic forms: fundamental and market data. On one hand, market data reflects the market dynamics when agents trade any financial asset. On the other hand, fundamental data refers to news, press releases, financial statements and other economic, social or political events that may affect prices of financial assets. Given the fact that prices respond to actions generated by market agents, time becomes a key element since the price formation process happens as agents interact over time. As a result, financial data constitutes an attractive source for researchers, since it exhibits complex systems' properties[13] which are difficult to analyze and understand[14][16][17]. Finding ways to summarize financial data is fundamental to support market agents' decision making process.

a) : Graph theory has proven to be very useful for analyzing complex dynamics in a wide variety of fields[1][4], including financial markets' behavior [2][19][20][21][22], ownership analysis of publicly traded companies[3][9] and systematic risk by contagion models[5]. It facilitates the analysis of inter-dependencies analysis among graph components, and synthesizes huge amounts of data such as those observed in financial markets.

Manuscript received on January 16, 2017, accepted for publication on May 3, 2017, published on July 1, 2017.

The authors are with the Universidad Nacional de Colombia (email: jhninop@unal.edu.co, gjhernandezp@unal.edu.co), and the Universidad de los Andes (email: ce.garcia392@uniandes.edu.co).

b) : This paper extends ownership structure analysis by specifically analyzing behaviors in owners' sub-networks. Our interest is focused in the preceding time lapse of a bankruptcy by aiming at finding possible patterns in the ownership structure. As a case study we take the bankruptcy of INTERBOLSA, the largest security brokerage firm in Colombia. The paper continues with the following sections: a background of the financial case of analysis, a brief description of key graph theory definitions, the experiment, discussion and conclusions.

II. Financial Context

Market actors use a wide variety of informational sources to negotiate financial assets such as fundamental data (e.g., macroeconomic news, industry analysis, and companies' financial statements), technical information (e.g., statistical analysis over past data) and insights obtained from application of modern techniques such as machine learning) [16][18]. Data processing for any of these categories is a key component for the market agent decision making process. In this paper, we focus our attention on analyzing fundamental market data, particularly ownership dynamics of the top 20 owners of major companies listed in the Colombian Stock Market. Our motivation lies in the 2012 bankruptcy case of the biggest brokerage firm in Colombia (INTERBOLSA). Given the fact that this firm managed a significant part of the money flow in this market, its bankruptcy caused a very negative impact in market confidence. Moreover, the collapse was caused by its own top management team, which in collusion with another important investors in the Colombian market (private funds and individuals), tried to gain control on other public traded company (FABRICATO) via repurchasing agreements (REPOS)[1] using their own money as well as money from INTERBOLSA's clients. Under this scenario we want to apply network based principles to detect patterns from data, not easy observable by other means.

III. Graph theory definitions used

– Bipartite Graph: A graph in which links relate two independent set of nodes (U, V). That is, an element in U could only be linked to an element in V[1].

[1]A REPO is an acronym for Repurchase Agreement. In Colombia they are used to facilitate leverage operations on stocks.

ISSN 2395-8618

– Eigenvector centrality: It is a recursive measure that determines how central is a node within a network. The node's centrality is based on the number of, and the quality of, its connections[12].
– Community detection (modularity): It is a measure that determine modules or communities depending on how nodes are interconnected. It allows to analyze the overall community structure within a graph[1][7].

IV. Experiment setup

Network theory has been used in the extant literature to analyze the dynamics between owners and public traded companies[3][9]. We want to apply specific network concepts such as community detection, bipartite graphs and eigenvector centrality, in order to extract patterns of ownership behavior of firms and individuals with the largest involvement in INTERBOLSA collapse[2] (i.e., INTERBOLSA, VALORES INCORPORADOS, INVERTATICAS, ALESANDRO CORRIDORI, HELADOS MODERNOS DE COLOMBIA, GITECO SAS, MANANTIAL SVP and RENTAFOLIO BURSATIL).

A. Data acquisition

Colombian public listed companies must report information to the market regulator (Superfinanciera de Colombia)[3]. This information is publicly available and it includes information about the Top 20 owners. As a result, data was collected quarterly basis from 2009-01-01 to 2012-09-30 for the major companies of the Colombian Market, including INTERBOLSA.

B. Data pre-processing

Data pre-processing was carried out using Talend Data Integrator and Python. Original data was downloaded in Excel format and includes company name, quarter reported, owner id, owner name, number and class of shares owned, and the percentage of ownership. Data only includes the top 20 of the owners. Given the original formatting of numbers, it was necessary to verify companies' id's, number format (decimal character and thousands separator). Because Gephi was used to build the graphs, the pre-processing output were CSV files containing columns required by Gephi[4].

C. Graph construction

Two types of graphs were used:

1) A bipartite graph having, on the one hand, owners as one group of nodes, and on the other hand the companies as the othe group. Links represent the ownership relation, whereas link weights are the percentage owned by owners in companies. The following conventions are used for these graphs:
   a) Node color is given by the in-degree value, being red the lowest value, and blue the highest one.
   b) Node size is given by the out-degree value.
   c) Node label size is given by the weighted degree; it means that the biggest the font the larger the number of connections for a particular node n_i.
   d) Link thickness is given by the ownership $\%$ of owner A in company X.
2) A projected owner graph, derived from the relationship observed on the bipartite graph. Conventions are as follows:
   a) Node color is given by the modularity class (detected community).
   b) Node size is given by the eigenvector centrality measure.
   c) Node label size and label color are given by the degree, being red the lowest degree value and blue the highest one.

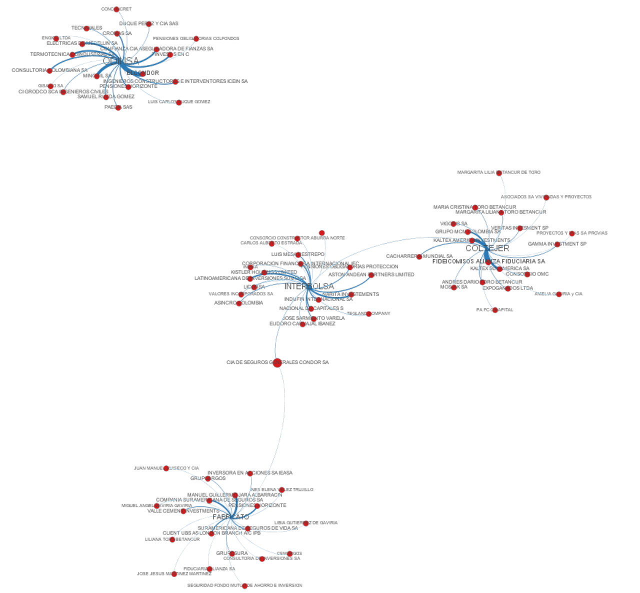

Fig. 1. Bipartite graph for the first quarter of 2009

In particular, we were interested in the following publicly listed companies: INTERBOLSA, FABRICATO, COLTEJER, ODINSA and BIOMAX, which were the ones that presented stronger declines in their stock prices by the time the mismanagement allegations became public. We were also interested in the following list of owners: INTERBOLSA, VALORES INCORPORADOS, INVERTATICAS, ALESSANDRO CORRIDORI, HELADOS MODERNOS DE COLOMBIA, GITECO SAS, MANANTIAL SVP and

[2]http://www.elespectador.com/noticias/infografia/actuaciones-de-autoridades-el-caso-interbolsa-articulo-429718

[3]http://www.superfinanciera.gov.co

[4]One CSV file for nodes, which are ids of owners and listed companies; another one for edges, which included a tuple (owner,owned) and the ownership percentage

ISSN 2395-8618

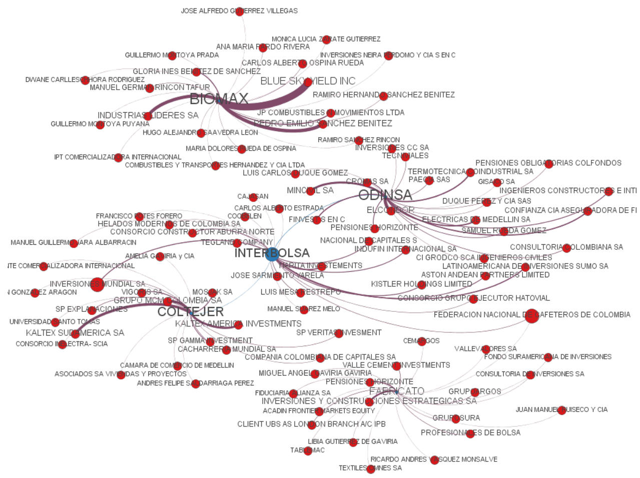

Fig. 2. Bipartite graph for the third quarter of 2009

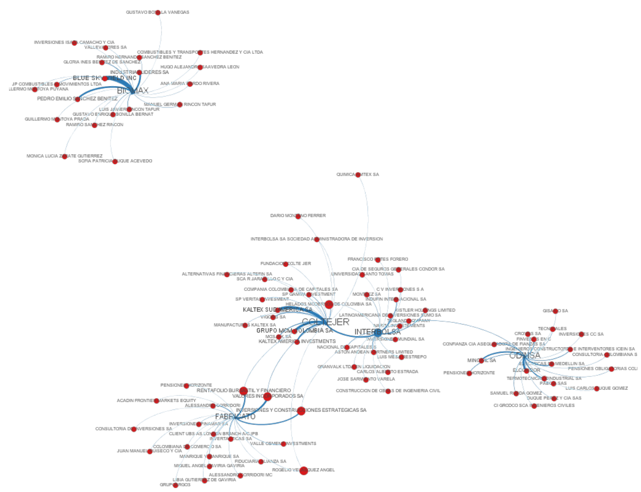

Fig. 3. Bipartite graph for the first quarter of 2010

RENTAFOLIO BURSATIL, since they were accused of collusion for trying to take ownership of FABRICATO using others investors' money, with consent of INTERBOLSA.

V. Results

Results indicate that network theory uncovers relationships difficult to identify by other means, which are different from the observed ownership dynamics of companies that did not go bankrupt.

In order to present the results, Figures 1 - 5 illustrate the bipartite graphs of the top 20 owners and companies for different quarters, over time. As the reader can observe, all the listed companies previously mentioned share the same owners. This feature is kept by all of the companies during the period of analysis.

Visualizations reveal a strong relationship among names involved in the allegations of FABRICATO takeover (i.e., INTERBOLSA, VALORES INCORPORADOS, INVERTATICAS, ALESANDRO CORRIDORI, HELADOS MODERNOS DE COLOMBIA, GITECO SAS, MANANTIAL SVP and RENTAFOLIO BURSATIL).

Link thickness represents owner's stake A in company X. As a result, visualizations confirm that involved owners had indeed an important stake in FABRICATO stock, and it was increasing over time. Figure 5 shows that all of the names previously mentioned account for a large interest in FABRICATO.

The second part of the results about ownership structure dynamics are the most interesting. Owners projected graph are analyzed. Figure 6 reveals the community structure of the involved names. For most of the time periods analyzed these

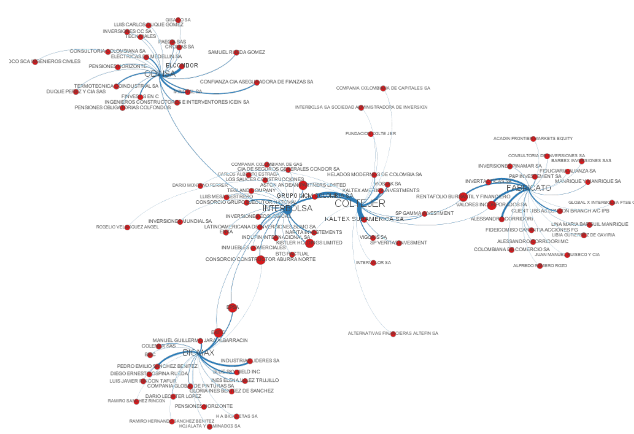

Fig. 4. Bipartite graph for the first quarter of 2011

ISSN 2395-8618

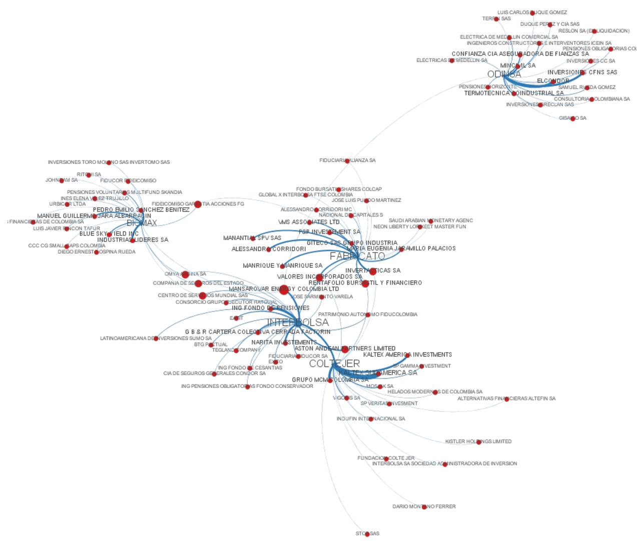

Fig. 5. Bipartite graph for the third quarter of 2012

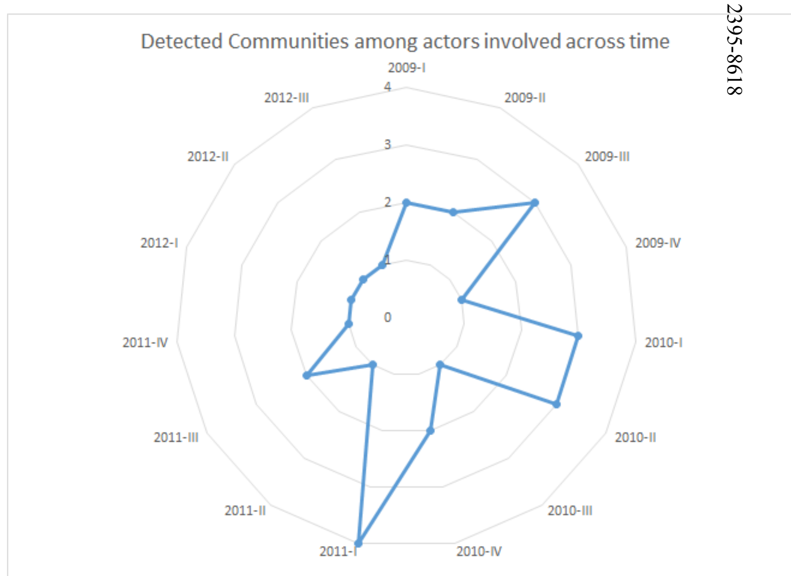

Fig. 6. Summary of number of communities detected across quarterly data

names are grouped under a few number of communities. In fact, they are grouped under just one community by the time allegations went public (2012-III).

Analysis of eigenvector centrality measures (Figures 7 to 10) revealed that VALORES INCORPORADOS, RENTAFOLIO BURSATIL and INVERTACTICAS, were the most influential nodes within the analyzed community structure. In fact, these three companies were the most compromised in the scandal.

Ownership structure strongly changed for the three public traded companies most involved in the scandal (COLTEJER, FABRICATO and INTEROLSA). Figure 11 evidences how ownership concentration among the top 20 owners builds up over time, reaching similar levels by the time the scandal went public.

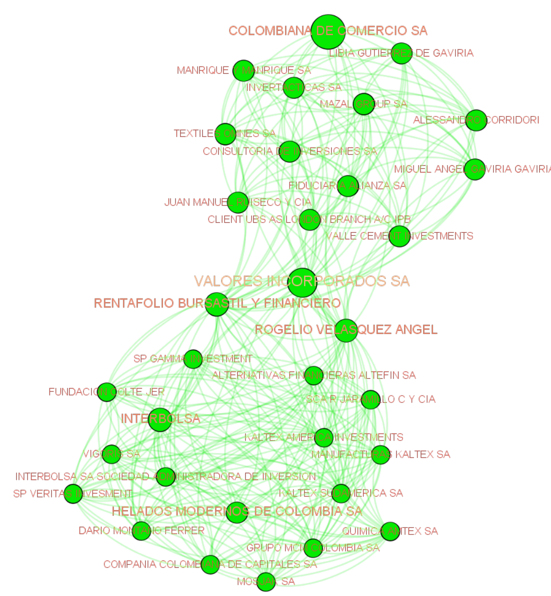

Fig. 7. Projected graph for the fourth quarter of 2009, colored by community

VI. Discussion

Network theory facilitates the analysis of ownership interdependencies for the case proposed. In fact, community detection and centrality measures allowed fast identification of key players. Visualizing results for each quarter revealed the influence of these owners through the whole period of analysis. Also, by considering the weights of the graph, it was possible to observe how the concentration of ownership grew over the different quarters for the three companies which shared the most owners (INTERBOLSA, FABRICATO and COLTEJER). See Figure 11. While the ownership concentration of these firms increased over time, for other companies not involved in ownership manipulation their concentration pattern proved to be completely different: some of them appeared to be almost unchanged over the same period of time.

VII. Conclusions

Network theory concepts applied to financial data yield compelling evidence regarding the particularities of complex ownership dynamics. Data analyzed in a timely manner could serve as early warning system for market participants, particularly when reflecting salient changes in ownership structure. As we saw in the case of this paper, activities that derived into INTERBOLSA bankruptcy during the third quarter of 2012, seemed to have started back in the four quarter of 2009. As a result, it is plausible to apply network

ISSN 2395-8618

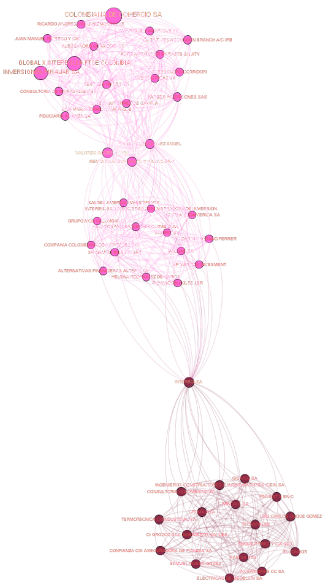

Fig. 8. Projected graph for the fourth quarter of 2010, colored by community

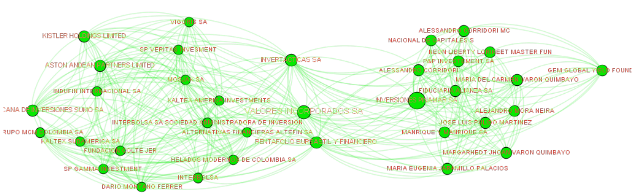

Fig. 9. Projected graph for the forth quarter of 2011, colored by community

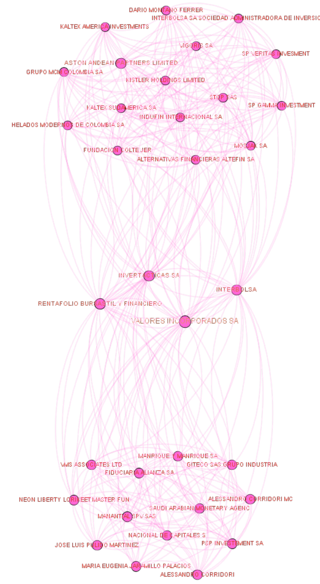

Fig. 10. Projected graph for the third quarter of 2012, colored by community

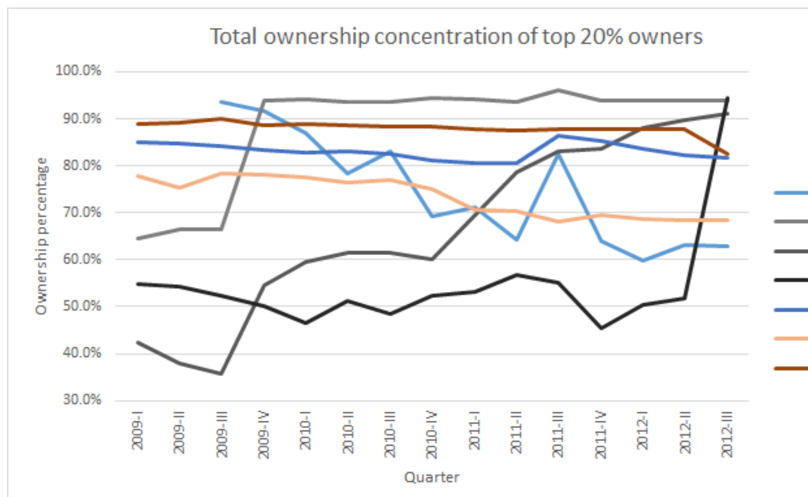

Fig. 11. Graph showing behavior of total ownership for companies involved from 2009 to 2012

theory concepts to ownership reporting of public companies to look for possible changes in communities, ownership concentration and centrality of players. This information could timely uncover undesired dynamics that might trigger negative market impacts in the future. Given the fact that network theory has been widely used to analyze contagion, approaches such as [5] could complement our case in order to analyze the stress that this bankruptcy brought to the Colombian security markets. Moreover, incorporating other sources of information such as financial indicators (Net profit, D/E ratios, etc.) could be helpful in order to complement our analyses and should be considered in future works.

ISSN 2395-8618

a) Acknowledgment.: The authors of the paper thank infoCapitales.com for the expertise provided.

REFERENCES

[1] Barabasi, A.: Network Science. Cambridge University Press, Cambridge (2015)

[2] Bezsudnov, I., Snarskii, A.: From the time series to the complex networks: The parametric natural visibility graph. Physica A: Statistical Mechanics and its Applications, vol 414, pp 53-60 (2014)

[3] Bohlin, Ludvig: Network Analysis of the share ownership structure on the Swedish market. Umea University, Faculty of Science and Technology (2012)

[4] Easley, D., Kelinberg, J.: Networks, Crowds and Markets: Reasoning about a Highly Connected World. Cambridge University Press, Cambridge (2010)

[5] Elliot, M., Golup, B., Jackson, M.: Financial Networks and Contagion. American Economic Review, Vol 104, issue 10, pp 3115-3153 (2014)

[6] Bouveret, A., Guillaumie, C., Roqueiro, C. A., Winkler, C., Nauhaus, S.: High-frequency trading activity in EU equity markets. ESMA Report on Trends, Risks and Vulnerabilities, issue 1, pp 41-47 (2014)

[7] Fortunato, S.: Community detection in graphs. Physics Reports Vol 486, issue 3, pp 75-174 (2010)

[8] Geetha, N., Sekar, P.: Graph Theory Matrix Approach A Review. Indian Journal of Science and Technology, vol 9, issue 16, pp 1-4, (2016)

[9] Vitali, S., Glattfelder, J. B., Battiston, S.: The network of global corporate control. PloS one, vol 6, issue 10, e25995 (2011)

[10] Luca, C.: Technical Analysis Applications in the Global Currency Markets, 2nd Edition, New York Institute of Finance, New York (2000).

[11] Naish, J.: Managing chaos, Nursing management, vol 2, issue 1, p 3 (1995).

[12] Newman, M.: The mathematics of networks. The new palgrave encyclopedia of economics, Vol 2, pp 1-12 (2008)

[13] Ohlsson, S.: Deep Learning: How the Mind Overrides Experience. Cambridge University Press, New York (2011).

[14] Ortega, L.: A neuro-wavelet method for the forecasting of financial time series. Proceedings of the World Congress on Engineering and Computer Science, vol. I, pp 24-26 (2012)

[15] Ruhnau, B.: Eigenvector-centrality: a node centrality? Social Networks, vol 22, issue 4, pp 357-365 (2000)

[16] Stasinakis, C., Sermpinis, G.: Financial forecasting and trading strategies: a survey. Computational Intelligence Techniques and Trading and Investment (2014)

[17] Tay, F., Cao, L.: Application of support vector machines in financial time series forecasting. Omega, vol 29, pp 309317 (2001)

[18] Treleaven, P., Galas, M., Lalchand. V.: Algorithmic trading review. Communications of the ACM, vol 56, issue 11, pp 76-85 (2013)

[19] Yan, S., Wang, D.: Time series analysis based on visibility graph theory. 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, Vol 2, pp 311-314 (2015)

[20] Yang, Y., Wang, J., Yang, H., Mang, J.: Visibility graph approach to exchange rate series. Physica A: Statistical Mechanics and its Applications, vol 388, issue 20, pp 4431-4437 (2009)

[21] Yang, Y., Yang, H.: Complex network-based time series analysis. Physica A: Statistical Mechanics and its Applications, vol 387, issue 5, pp 1381-1386 (2008)

[22] Zhuang, E., Small, M., Feng, G.: Time series analysis of the developed financial markets' integration using visibility graphs. Physica A: Statistical Mechanics and its Applications, Vol 410, pp 483-495 (2014)