Filomat 32:5 (2018), 1931–1936
https://doi.org/10.2298/FIL1805931H

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Applying Data Visualization Techniques for Stock Relationship Analysis

Jie Hua[a], Mao Lin Huang[b], Guohua Wang[c], Mouataz Zreika[d]

[a] *Shaoyang University*
[b] *University of Technology Sydney*
[c] *South China University of Technology*
[d] *Western Sydney University*

Abstract. Decision making in stock investment is often made based on current events in the market and the analysis of historical data on specific stocks. Besides, similar rates of price changing over a long-term period on different stocks may indicate potential connections between those listed corporations. The proposed methodology applies the force-directed algorithm and time-series chart to offer stakeholders capability to gain deeper insights initiative on potential relationships between stocks comes with less human interventions. Hence to assist in future decision making on stock investment via graph layouts.

1. Introduction

Stock investment decisions require time, knowledge and awareness including historical data, the stock market contains a huge amount of data that varies over time [5]. Stock prices are influenced by various factors ranging from the performance of the company itself to the conditions of the economy in general. Thus, to manage investment portfolios, stock market data has to be analyzed regularly to identify potential relationships between various stocks, hence to adjust investment based on related stocks trends.
A picture is worth a thousand words presenting data in visual form can assist human in exploring deep insight of vast amount of complex raw data, especially when people have limited knowledge of the data [22, 23, 24]. Visual representation is one of the most efficient ways to assist investors to have a clear overview of movements of the stock market, as well as providing a deeper understanding of each individual stock. The application of graph drawing method can offer visualized data with specific attributes such as weight information, comes with graphical connections between each data element. The key issue here is not only to provide users with a comprehensive display of large graphs on the screen, but also a user-friendly navigable visual structure for users browsing through the structure to find a particular detail of relevant data [1].
In this study, an approach is proposed which applies the force-directed algorithm and time-series chart to represent a clear overview of the entire structure on relevant stocks, plus detailed stock connection representing. The proposed method is applied to draw weighted graphs, comparing with time-series charts on targeted stocks. The early outcome of our approach indicates deep relationship overview/detailed view and match the traditional time-series charts results.

2010 *Mathematics Subject Classification*. 76M27.
Keywords. Force-directed; time-series chart; graph drawing; data visualization; stock market analysis.
Received: 21 October 2017; Accepted: 7 December 2017
Communicated by Hari M. Srivastava
Email addresses: Jie.hua@alumni.uts.edu.a (Jie Hua), Mao.Huang@uts.edu.au (Mao Lin Huang), GHWang@scut.edu.cn (Guohua Wang), M.Zreika@westernsydney.edu.au (Mouataz Zreika)

2. Existing Work

Treemap concepts are designed to facilitate financial decision-making, provide an overview of large amount of hierarchical financial data, such as FolioMap (Jungmeister & Turo 1992), Self-Organizing Map (Kohonen, 1997; Joseph & Indratmo, 2003), FundExplorer (Csallner et al. 2003) and ordered treemap (Shneiderman & Wattenberg 2001) etc. [2, 3, 5, 7, 9, 16, 17]. And 3D treemap was proposed to monitor real-time stock market performance and to identify a particular stock that produced an unusual trading pattern (Huang et al. 2003) [10]. Treemap enables an easier and faster visual approach, and it has already been utilized successfully in the visualization of stock portfolios and markets [18]. Although using the pivot-by-size treemap algorithm is very unstable [3].
Parallel coordinates yield graphical representations of multi-dimensional relations rather than just finite point sets. It is a system for visualizing analytic and synthetic multi-dimensional geometry [19]. Its adoption in the stock market is either through sharing the typical parallel layout of axes or via the mapping of data samples to lines [14, 15].
Wyer and Eades proposed a system that visualizes the movements of fund managers within the UK Stock Market in terms of their share ownership over time [4]. Simunic conducted a way of showing the principal distribution of chart shapes, from the whole chart data set to generate representative charts representing clusters of similar charts [6].
Although most current stock visualization systems only visualize the users portfolio and therefore do not show the potentially huge number of stocks available in the market [3], and the relationship analytics is simply based on users interests, there is no initiative capability for the large-scale datasets in the stock market. In this study, aspects were emphasized on:

- discovering potential relationships among massive stocks.
- exploring the correctness of visualized relationship representation.

3. Proposed Approach

The approach proposed in this study is represented as:

1) Stock raw data collecting from the Australian Securities Exchange (ASX);
2) Data cleansing on raw data collected, including data formatting and data filtering etc. Eventually finalized raw data and graph models have been produced accordingly;
3) Experiments on imported graph models, which involves force-directed algorithm and time series chart;
4) Stocks relationship analytics based on comparisons between finalized graph layouts and time series charts;

3.1. Data Cleansing

All raw data were collected from the ASX, including attributes such as Open Value and Close Value and Volume etc. In experiments, we only applied raw data of stocks which still exist nowadays (till 30/06/2017) and the total absence days of each stock were less than 40% of the entire stock open days in last 20 years. For producing graph model purpose, edges need to be finalized as follows (sees [21] for data processing details).

- Individual changing rate
  Suppose the costs of an independent stock in two continuous business days are c_1 and c_2, the rate $r_1 = (c_2 - c_1) / c_2$, then the rate array of stock k is $r_k = \{r_1, r_2, \ldots, r_k\}$;
- Related changing rate
  Suppose the rates of two independent stocks in two same continuous business days are r_{mi} and r_{ni}, the rate $r_{mni} = (r_{ni} - r_{mi}) / r_{ni}$, then the rate comparison array of stock m and n is $r_{mn} = \{r_{mn1}, r_{mn2}, \ldots, r_{mnk}\}$, all the different time periods are dropped, only those rates changes happened within same time period on both stocks are taken into account.

- Date matching rate
  The rates $rd_{m,n}$ of total dates matched between every two stock d_m and d_n, the rate array of stock k is $rd = \{rd_{0,1}, rd_{0,2}, \ldots, rd_{k-1,k}\}$;
- Weight computing
  Edge thickness is applied for displaying how close the connection is between stocks based on rates computing from above, the transmission from rate to weight in experiments is: If related average changing rate r_{mni} * $rd_{m,n}$ is smaller than 0.6, it is considered no connection; or it is assigned with weight values 4(>=0.9), 3(0.8-0.9), 2(0.7-0.8), 1(0.6-0.7).

3.2. Data Visualization

To provide visualized results for analysis purpose, two types of graph drawing methods have been adopted in experiments: force-directed algorithm and time series chart.
The force-directed algorithm aims to position nodes with a few crossing edges as possible by assigning forces among the set of nodes and edges for drawing graphs in an aesthetically pleasing way. Noack explains that the most important difference among force-directed algorithms is the role played by distance in graph spatialization [13], he defines the energy model or "(attraction, repulsion)-model" of a layout as the exponent taken by distance in the formulas used to calculate attraction and repulsion (the log being considered as the 0^{th} power). As a side effect, Linlog algorithm tends to group nodes with large degrees in the centre of the layout, where their distance to the remaining nodes is relatively small [20]. Here, the ForceAtlas (FA) layout algorithm (Bastian, Heymann & Jacomy, 2009) is applied, which is a spatial layout algorithm under the category of force-directed algorithms, it aims at giving a readable shape to a real-world network [12]. FA algorithm offers real-time settings including speed, gravity, repulsion, auto-stabilize, inertia or size-adjust [11]. ForceAtlas2 (FA2) is based on FA but offers more options and innovative optimizations that make it a very fast layout algorithm. Its implementation of adaptive local and global speeds brings good performances for a network of fewer than 100000 nodes. It was empirically observed that FA2 is at its best with strongly clustered networks. The ability of FA2 to show clusters is better than FR algorithm and worse than Linlog [12]. FA2 was applied in the evaluation.
A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. Time series control charts are popular methods for statistical process control of autocorrelated processes [8]. In our study, time series chart matches long-term raw data collection analytics requirements, which could perfectly offer reader another angle of view on stock data as well.

4. Evaluation

In experiment evaluation, 5088 stocks in Australia were collected, ranged from 02/Jan/1997 to 30/Jun/2017, around 6.400 million raw data entries were gathered (before formatted/filtered); after cross-comparison on related rates between every two stocks, nearly 194.000 million data were finalized further; eventually 2203 stocks have been kept for testing after data cleansing steps; and connected/undirected graphs were created artificially based on cleansed raw data, to test the proposed approach (Gephi was adopted to represent graphs).
The final layouts are shown in Figure 1 and 2. They provide overviews of the entire/selected structure of relevant stock data (vertices representing each stock and edges showing connections between stocks); interaction features such as zoom in/out; vertex selection etc. Thicker/shorter edge indicates a stronger connection, larger vertex presents more degree, edges and vertices in the same colour are put into the same cluster which is grouped by vertex degree. On the other hand, relevant time series charts of the chosen stocks were built up based on price rate changing accordingly; then comparisons were made between force-directed layouts and time series charts, to analyze the visualized relationships, furthermore to find out if they are matched.
Stakeholders might be offered the capability of analyzing connections among stocks based on long-term analytics, and a new angle of views come with less human disturbance, only based on pure math calculation on changing rates, and this may assist stakeholders to handle investments other than advice from stock experts.

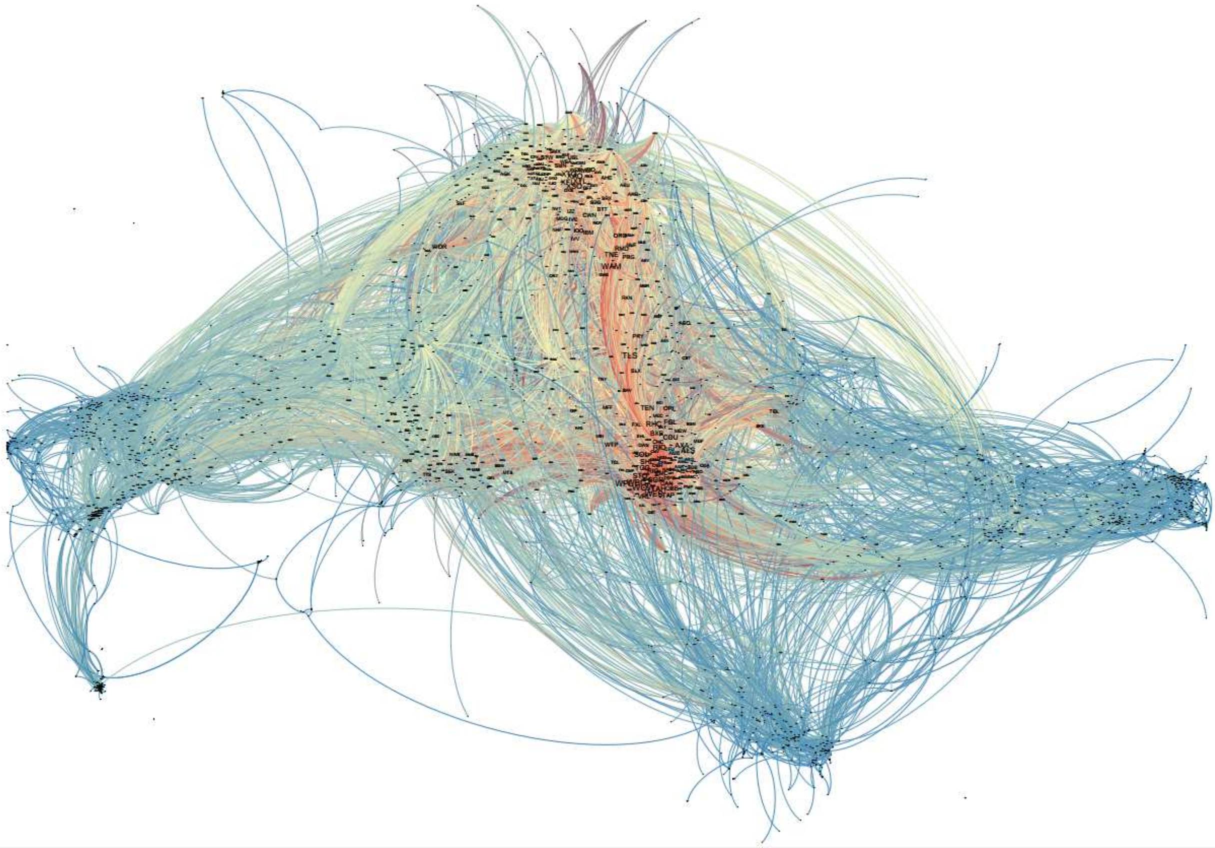

Figure 1: Entire layout after applying FA2 algorithm

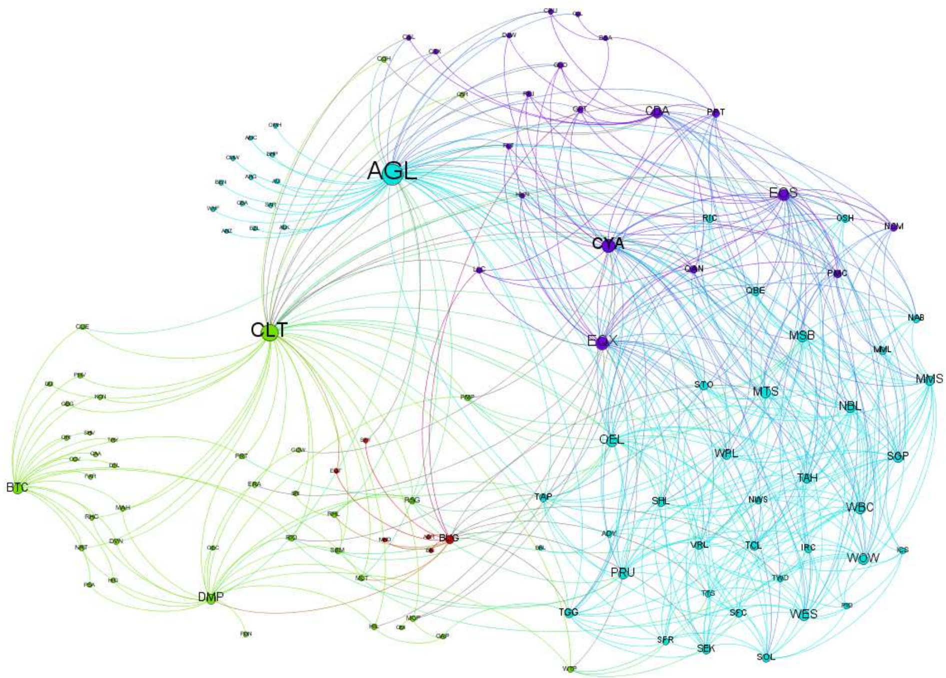

Figure 2: Selected stock layout after applying FA2 algorithm

5. Case Study

A case is provided here to explain details of our proposed methodology. In this case study, relevant stocks connected to AGL have been presented for analysis, including AGL: AGL Energy Limited; NAB: National Australia Bank Ltd; WOW: Woolworths Limited; SAR: Saracen Mineral Holdings Limited; ALK: Alkane Resources Limited; and ANZ: Australia and New Zealand Banking Group etc.

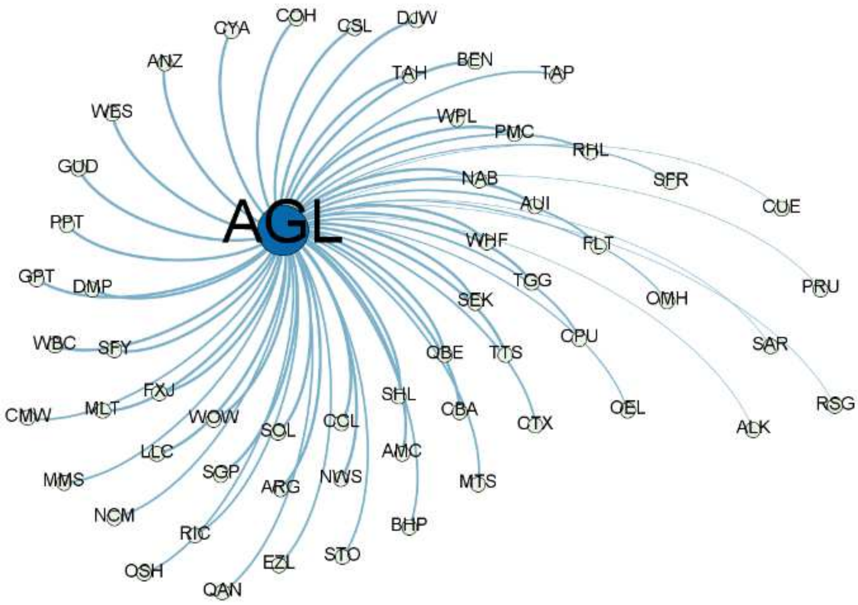

Figure 3: Detailed layouts of selected stocks (close-value based calculation)

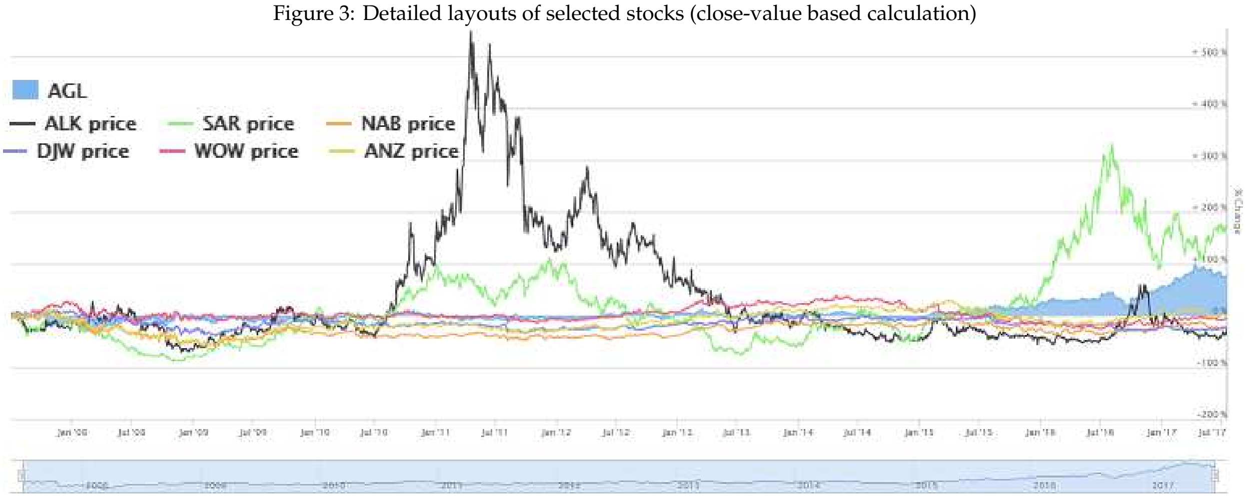

Figure 4: Time series chart on selected stocks (price changing rate 07/2007-07/2017 http://www.asx.com.au)

As shown in Figure 3 (filtered/zoomed in the graph from Figure 2), details of stocks selected are shown that:

- AGL and NAB, ANZ, WOW have similar closer connections;
- There are weak connections between AGL and SAR/ALK etc.
- in the closest 28 stocks to AGL:
  50% companies belonged to the financial sector; 14% were in the real-estate/development area; 17% were related to the healthcare/medical field; Only 10% were in energy industry like AGL was.

Besides, those visualized results match the trends on price changing from the time series chart in Figure4 (Generated from http://www.asx.com.au) Hence, it can be seen that the force-directed algorithm adopted provides similar relationship representation as time series charts in this case.

6. Conclusions and Future Works

The stock market is one domain comes with thousands of companies where complex data arise, and the performance is traditionally measured by chart diagrams representing price fluctuation over time, as well as applying treemap for showing volume and relationship etc. Here we propose a method that adopts force-directed algorithm, to offer stakeholder a new angle of view on relationship analytics between stocks, comes with an overview and detailed graphs, based on pure math calculation. Comparing to existing methods, it could help users to discover potential connections between stocks. Hence, stakeholder could gain deeper insights from the potential relationship of stocks in the market and may adjust investments accordingly based on related stocks trends.

In our experiments, only limited stocks/factors have been tested, and some data have been dropped during data cleansing process, which may affect the final accuracy of experiment outcomes. In our future work, more factors on stock data will be considered and more data visualization techniques will be revised and adapted to satisfy specific needs in stock data analysis areas.

References

[1] J. Hua, M.L. Huang and Q.V. Nguyen, Drawing large weighted graphs using clustered force-directed algorithm, 18th International Conference on Information Visualisation (2014) 13-17.

[2] W.A. Jungmeister and D. Turo, Adapting treemaps to stock portfolio visualization, (1992)

[3] C. Csallner, M. Handte, O. Lehmann and J. Stasko, Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps. IEEE Symposium on Information Visualization (2003) 203-208.

[4] T. Dwyer and P. Eades. Visualising a fund manager flow graph with columns and worms. Sixth International Conference on Information Visualisation (2002) 147-152.

[5] J. Joseph and I. Indratmo, Visualizing Stock Market Data with Self-Organizing Map. Florida Artificial Intelligence Research Society Conference, North America (2013)

[6] K. Åimuni, Visualization of stock market charts (2003).

[7] T. Kohonen, The self-organizing map. Neurocomputing, 21(1) (1998) 1-6.

[8] D.W. Apley, Time series control charts in the presence of model uncertainty. Transactions-American Society of Mechanical Engineers Journal of Manufacturing Science and Engineering, 124(4), (2002) 891-898.

[9] M. Wattenberg, Visualizing the stock market. In CHI'99 extended abstracts on Human factors in computing systems, ACM (1999) 188-189.

[10] M.L. Huang, J. Liang and Q.V. Nguyen, A visualization approach for frauds detection in financial market. 13th International Conference In Information Visualisation (2009) 197-202.

[11] M. Bastian, S. Heymann and M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks, Third International AAAI Conference on Weblogs and Social Media (2009). doi: 10.1136/qshc.2004.010033.

[12] M. Jacomy et al. Force Atlas 2, a graph layout algorithm for handy network visualization, (2011) Available at: http://scholar.google.com.

[13] A. Noack, Energy models for graph clustering, Journal of Graph Algorithms Appl., vol. 11, no. 2, (2007) 453-480.

[14] J. Heinrich and D. Weiskopf, State of the Art of Parallel Coordinates, Eurographics Conf. Vis., (2013) 95-116.

[15] J. Alsakran, Y. Zhao, and X. Zhao, Tile-based parallel coordinates and its application in financial visualization, in IS &TISPIE Electro Imaging (2010).

[16] B. Shneiderman and M. Wattenberg, Ordered treemap layouts, IEEE Symp. Inf. Vis. 2001. INFOVIS 2001., vol. (2001) 2-7.

[17] B. B. Bederson, B. Shneiderman, and M. Wattenberg, Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies, ACM Trans. Graph., vol. 21, no. 4, (2002) 833-854.

[18] W. Jungmeister and D. Turo, Adapting Treemaps to Stock Portfolio Generalization, Human-Computer Interaction Laboratory, Univ. of Maryland, College Park, MD, Sept, 1992, CS-TR-2996.

[19] A. Inselberg and B. Dimsdale, Parallel coordinates: a tool for visualizing multi-dimensional geometry, Vis. 1990. Vis. 90., Proc. First IEEE Conf., (1990) 361-378.

[20] S. G. Kobourov, Force-directed drawing algorithms, Handb. Graph Draw. Vis. Discrete Math. Its Appl., (2013) 383-408.

[21] M. Zreika, J. Hua, and G.H. Wang, Applying Data Processing Method for Relationship Discovery in the Stock Market, in the International Conference on Data Science and Business Analytics (ICDSBA 2017). In Press.

[22] D. A. Keim, Information visualization and visual data mining, IEEE Trans. Vis. Comput. Graph., vol. 8, no. 1, (2002) 1-8.

[23] G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis, Graph drawing algorithms for the visualization of graphs, (1999).

[24] B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations, Proc. 1996 IEEE Symp. Vis. Lang., (1996) 336-343.