Student: Zahra Hatami **NUID:** 45988990 **Phone:** 402-210-5259

Title of Project: Big Data Analytics Approaches for Analyzing US Stocks under major events - A focus on the 2008 Crash and the Impact of Covid-19

Faculty member supervising the project: Dr. Hesham Ali, University of Nebraska at Omaha

---

2- Project Description:

a. Description of the project

The behaviors in the financial market are mostly based on two theories--market equilibrium and behavioral theory. The market equilibrium stands for the availability of information for every investor in the market [2]. The behavioral theory seeks the influence in the psychological process in decision making while investors interpret the market based on their information on making investment decisions [9]. Therefore, even though the information is available for everyone, investors' decisions are influenced by the behavior of other investors. These theories tend to work well under normal circumstances, but it is clear whether they hold during unexpected situations or disturbing events like economic crash or Covid-19.

Many unknowns are associated with the emergence of unexpected crises and their economic impact. Hence, this brings new challenges for policymakers in their process of making appropriate decisions. The effect of each crisis depends on the size of the impacted economy and its degree of vulnerability [10]. For example, throughout the economic crash in 2007-2008, primarily due to the housing bubble, the financial stocks shot up, because of the huge amount of fiscal monetary stimulus packages and other governmental support. During the flu pandemic, sectors related to entertainment suffered twice as much compare to companies in the health sector [6].

The current pandemic brought the society to an unprecedent point where people are required to work from homes and use online stores for most purchases. As a result, companies belong to specific sectors, based on the nature of their business, are likely to be affected more than others. Under normal circumstances during healthy economy situations, consumers spend more in consumer discretionary products. On the other hand, during difficult times, consumers tend to focus on spending more on essential products. However, the recent data from stock markets showed that consumer behaviors did not follow their usual pattern. For example, during COVID-19, stocks belong to consumer discretionary companies, such as Amazon, went up and somehow such stocks seem to benefit from the current situation[1]. Such profit does not appear to be related to any economic boosting. The response of the market seems to less predictable and depends on specific conditions that the society and consumers are faced with. Hence, new approaches need to be explored to extract knowledge from available data and allow investors to understand the behavior of the financial markers under crises. The goal of obtaining this new knowledge is to find the hidden patterns that reflect how the companies are performing for a given period and what are the primary factors that impact their performance. Due to the high volume of data and complexity of the relationship between companies, big data techniques, like graph models and network science, are very much needed. With the goal of conducting correlation network analysis, we can structure the data in such a way that facilitate finding common patterns or similar behaviors among companies [8].

In addition, earlier studies revealed that small-size companies in certain sectors such as IT-related sectors tend to group in the same cluster as big size companies in other sectors including the financial sector [8]. In other words, the return pattern for those small size companies followed a consistent pattern like the large size companies. During COVID-19 pandemic, the study reported that the federal stimulus caused large size companies in the US have abnormal returns compared to small size companies [7]. Now, the question is, what are the similarities and differences between companies' behaviors during two crises, Covid-19 and economic crash in 2007-2008? Another question for this study is about which specific sector and its product are affected more by the current pandemic.

[1] https://www.nasdaq.com/market-activity/stocks/amzn/advanced-charting

b. Methodology

This study is an interdisciplinary work that takes advantage of big data associated with the knowledge in business and the financial domain. The computational analysis in this research will take advantage of the amount of the data available nowadays in the public domain. Therefore, this research will utilize the analysis that comes from economics and financial field along with computational analysis that comes from big data analytics.

Correlation networks will be developed to examine similarities in financial parameters and behavioral patterns among various financial entities, including different companies in different sectors. Analysis of the network involves the recognition of companies that are connected to others in a graph. The strength of the relationships, reflected by the weights of the edges in the network, can be measured with different kinds of correlation coefficients such as Pearson's, Kendall, and Spearman's, depending on how data is distributed. We will apply correlation network analysis in order to find hidden patterns and extract knowledge that are difficult to obtain using traditional approaches [11,13]. Companies will be divided into different sectors and their returns will be subtracted from risk free extracted from Fama and French library [5]. The data was collected from "the Center of Research in Security Prices." To fully understand whether this pandemic affects the companies and their volatility behavior, and also to find the similarities and differences between the effect on this pandemic and economic crush in 2007-2008, we divide the data into post and pre crisis/pandemic. The plan is to establish two networks: One network from the year 2006-2009 (inclusively), and another one for 2018-2020. The reason for establishing two networks is to compare the behavior of companies in different sectors for two unique events that have impacted the US market, the 07/08 economic crash and Covid-19. In this regard, the study will focus on identifying the similarities and differences in network correlations, along with network structures such as diameters, density, and other community parameters.

In this study, after constructing correlation networks from input data, cluster analysis will be applied on the network as a data analysis shortcut tool group different companies with high correlations or similar financial behavior over the period under study. The financial market network is one of the most complex networks, which brings significant challenges to visualization. Creating clusters from this complex network consumes considerable time. Since this research relies on population analysis based on clusters extracted from the network, for the robustness of our work, we established the automated process in R program. Therefore, to obtain the candidate clusters that possess high statistical significance for further analysis, the automated R program will run for different correlation coefficients along with different inflation values of MCL clustering [12]. The inflation is the granularity parameter, which is the power used to inflate the correlation matrix [14]. After extracting candidate clusters, the population analysis will be applied to see which parameters are significantly enriched for each candidate cluster. We use hypergeometric distribution for finding the significantly enriched parameters in each cluster [3]. With enrichment analysis applied on each cluster, all different parameters relevant to each node will be identified. We will also apply Kruskal–Wallis test to establish a relationship between various sectors (as the independent variables) and the stock returns (as the dependent variable). [9] Further, the pairwise comparison is done by using Wilcoxon test to see what values of individual sectors have significantly different returns [4].

Expected Results

Companies in different financial sectors can be vulnerable during a crisis. However, we expect the size of their vulnerability to be different depending on economic situations and the type of crises. During a global pandemic, we expect that some specific sectors, due to the nature of their products are affected more than other sectors. Since analyzing data in the Data Insider Consulting reports showed that personal health products such as sanitizer and vitamins and supplements sales increased while sales dropped by more than 40% as compared to other products like food products and luxury goods [1]. Although we expect that companies in finance sectors to be affected more during crises, even though this may not be directly related to crisis but rather due to the associated government reaction to the crisis.

From the analysis, we found that (figure below) finance sector and energy sector are significantly enriched compared to other sectors. In this figure, CC means the candidate cluster. CC33 is finance sector and CC18 is the energy sector [15]. Based on the analysis, we expect that, depending on specific product types and action from the government, some sectors will be affected more than other sectors which bring the conclusion that in crises, economic reaction do

not follow specific theoretical frameworks. Results from the study will be summarized in a manuscript and submitted to one of the top related international conferences.

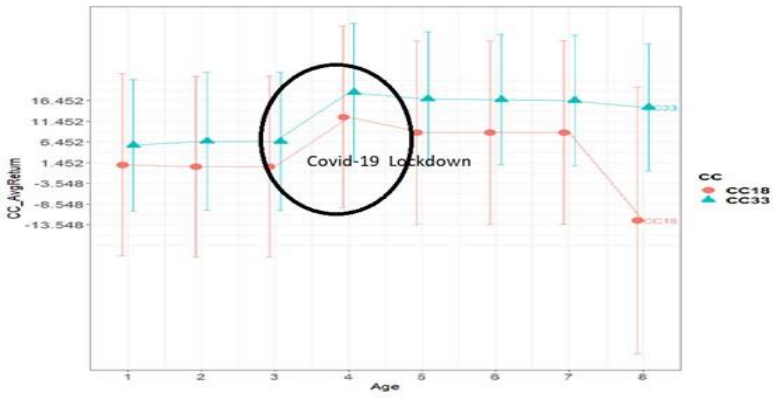

Figure 1-Behavioral patterns of clusters based on excess returns (Covid-19)

c. Project Timeline

This project will be implemented as a summer project. The duration will be two months (June and July).
Schedule of Activities:

- **June 1st to June 15th**: Collecting data from CBRS database and preprocessing the data in order to find any redundancies or inconsistencies.
- **June 16th to June 30th**: Creating a correlation network based on companies' excess returns and applying clustering analysis. Obtained clusters will be analyzed to find common and different behavioral pattern.
- **July 1st to July 31st:** Testing population and enrichment analysis using R and compare the behavioral patterns in different crisis.
- **August 1st – August 15th:** Preparing the project manuscript to be submitted to an international conference.

d. Student/Faculty Mentor Roles:

The student will implement the algorithms associated with the various stages of the project. This includes building a database system to store all collected data, implementing the proposed network model, and employing various data analytical tools to extract knowledge from constructed networks. The role of the faculty advisor is to help in the design of the graph algorithms and the correlation network model as well as provide overall guidance. Regular meetings will be scheduled for sharing results and monitoring progress.

e. Previous Internal Funding: GRACA 2019.

How this project is different than the previous one?

The tile for the previous research was: "Advanced correlation network model for analyzing financial market." The new research is different than the previous ones in multiple aspects. This study focuses on analyzing financial markets under crises. In this research, we will use Hyper-geometric distribution-based enrichment analysis given in [13] to identify the significantly enriched input parameters. This study aims to use a correlation network model (CNM) approach along with enrichment and population analyses to determine if the stock market and economic sectors follow the efficient market hypothesis (EMH) and behavioral theory during crisis periods and identify which input parameters are significantly enriched for each cluster of the correlation network.

3- Budget Justification:

In this project, a summer stipend of $5,000 is requested to support the graduate student, Zahra Hatami. She has a graduate assistantship that supports her during the academic year. The summer stipend will allow her to conduct the research activities over the summer. The project will take approximately 450 hours.

4- References:

[1] Jennifer Bouey. 2020. Assessment of COVID-19's Impact on Small and Medium-Sized Enterprises: Implications from China. (2020).

[2] William A Brock. 1979. An integration of stochastic growth theory and the theory of finance, Part I: The growth model. In *General equilibrium, growth, and trade*. Elsevier, 165–192.

[3] Sylvain Brohee and Jacques Van Helden. 2006. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC bioinformatics* 7, 1 (2006), 488.

[4] Jack Cuzick. 1985. A Wilcoxon-type test for trend. *Statistics in medicine* 4, 1 (1985), 87–90.

[5] Eugene F Fama. 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance* 25, 2 (1970), 383–417.

[6] Thomas A Garrett. 2008. Pandemic economics: The 1918 influenza and its modern-day implications. *Federal Reserve Bank of St. Louis Review* 90, March/April 2008.

[7] Maretno A Harjoto, Fabrizio Rossi, and John Paglia. 2020. COVID-19: Stock Market Reactions to the Shock and the Stimulus. *Available at SSRN 3622899* (2020).

[8] Zahra Hatami, Hesham Ali, and David Volkman, Prasad Chetti. A New Approach for Analyzing Financial Markets Using Correlation Networks and Population Analysis. (Under Review)

[9] John Holland. 2016. A behavioural theory of the fund management firm. *The European Journal of Finance* 22, 11 (2016), 1004–1039.

[10] Bryan W Roberts. 2009. The macroeconomic impacts of the 9/11 attack: evidence from real-time forecasting. *Peace Economics, Peace Science and Public Policy* 15, 2 (2009), 341–367.

[11] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 11 (2003), 2498–2504.

[12] Stijn vanDongen. 2000. A cluster algorithm for graphs. *Information Systems [INS]* R 0010 (2000).

[13] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world'networks. *nature* 393, 6684 (1998), 440.

[14] clusterMaker: Creating and Visualizing Cytoscape Clusters. Retrieved July 22, 2020 from http://www.cgl.ucsf.edu/cytoscape/cluster/clusterMaker.shtml

[15] Zahra Hatami, Prasad Chetti, Hesham Ali, and David Volkman. A Novel Population Analysis Approach for Analyzing Financial Markets under Crises – A Focus on Excess Returns of the US Stocks Under 9/11 and Covid-19. (Under Review)

University of Nebraska Omaha

COLLEGE OF INFORMATION SCIENCE AND TECHNOLOGY

January 15th, 2021

Dear GRACA Review Committee:

I am pleased to write this letter to support Zahra Hatami's proposal titled, "Big Data Analytics Approaches for Analyzing US Stocks under major events - A focus on the 2008 Crash and the Impact of Covid-19."

Zahra is a fourth year Ph.D. student in the college PhD in IT program. With a good foundation in various areas in computing and business, Zahra brings an interdisciplinary perspective to our doctoral program, which is particularly designed for advancing the impact of IT in other scientific areas. Such foundation is perfect for addressing the complex problem addressed by the proposed project. The project attempts to tackle the difficult issue of analyzing the unpredictable financial markets. With the increase in the unpredictability levels during unexpected local, national, or global events, such analysis would be even more difficult. However, with the availability of rich data related to financial industries, advanced big data analytics can be utilized to provide in-depth analysis of financial markets under different circumstances and allow researchers to understand, and potentially predict, their behaviors.

The proposed study is based on utilizing graph theory and network science to develop correlation networks models for the analysis. The network analysis has the potential to provide answers to key questions such as: Which sectors are likely to struggle under unexpected events and which ones are likely to survive. In addition, the proposed analysis can provide new insights on which financial parameters have a bigger impact on the performance of different financial markets/sectors? Such information can be used to predict financial behaviors based on the extracted patterns from the available data and help investors in building stable financial portfolios.

Again, with her business background, Zahra has the perfect background to conduct the proposed project. She is well positioned to take on the many challenges associated with the project. We have been working very closely with Dr. David Volkman from CBA and he has been supporting our efforts as the domain expert in the project, and he continues to be very excited about the potential of the proposed study. We have recently submitted a joint paper that contains the early results of this project.

Zahra is an excellent student with great potential. She is one of the top doctoral students in the college and has been one of the best teaching assistants. She has taught various classes with great success and has been receiving excellent student evaluations in the classes she has been teaching. She works very well in interdisciplinary settings and has the ability to connect ideas from various scientific fields. I strongly recommend her for the GRACA award. Please feel free to contact me if you require further information.

Sincerely,

Hesham H. Ali, Ph.D.
Professor of Computer Science
College of Information Science and Technology
University of Nebraska at Omaha