

A Novel Population Analysis Approach for Analyzing Financial Markets under Crises – 2008 Economic crash and Covid-19 pandemic

Zahra Hatami Prasad Chetti Hesham Ali David Volkman
 University of Nebraska Omaha University of Nebraska Omaha University of Nebraska Omaha University of Nebraska Omaha
zhatami@unomaha.edu pchetti@unomaha.edu hali@unomaha.edu dvolkman@unomaha.edu

Abstract

Stock markets play an important role in shaping an economic portfolio in many countries and are often used as critical ways to measure economic health and financial status in numerous studies. Financial markets are often volatile and can be influenced by a wide range of direct and indirect variables. The current Covid-19 Pandemic has severely impaired the economic markets in many parts of the world and has negatively affected millions of investors. While some financial markets or stocks are expected to recover or partially recover from this crisis, others may not. With recent crises, such as the 2008 economic crash or the economic impact of the 9/11 event, researchers are looking for innovative ways to analyze the behavior of financial markets under crisis. Can we apply traditional analytical approaches to study the behavior of financial markets under crises, or a different new approach is required to conduct such a study? This paper proposes a new network model and employs a population analysis approach to address such an important research question. We present the basic steps that illustrate how to construct correlation networks of financial stocks and how to utilize graph-theoretic properties of the constructed networks to analyze the behavior of stocks over a given period of time. The proposed population analysis approach allows us to compare the behavior of various groups of companies and their relevant economic sectors in the stock market. We apply the correlation network analysis on different financial data and study the financial implications of two major events, the 2008 economic crash, and Covid-19. In particular, we use the networks to compare the behavior of different economic sectors and uncover the similarities and differences between sectors and their reactions or behavior during these two events. We were able to see certain patterns and extract useful information from the correlation networks. For example, we observed that companies in finance sectors behave in a similar way under the effect of both events and identify

some similarities between the behavior of the energy sector during the current pandemic and the utility sector during the 2008 crash.

1. Introduction

The economic cycle and its financial implications can be impacted by many parameters including various world events, internal developments within companies, inflation and interest rates, and releases of financial reports. Economic cycles can also be impacted by sudden or unexpected events, which may rise to the level of a local, regional or works crisis. In such a case, there will be a significant decline in different economic activities that will continue for a few months. An economic crisis affects different economic sectors, rates of GDP, unemployment, and so forth. The effects of the crises depend on its nature and the governments and industries' reactions[1]. An economic crises can be categorized as a structural crisis such as the 2008 crash, or as a shock or a non-structural crisis such as the Covid-19 pandemic [2, 3, 4]. When an economic crisis happens, all developed and developing countries are affected by the economic crash consequences. The economic crash in 2008 was categorized as a structural crisis since market players were the leading cause. The evidence of the financial crisis in 2008 displayed the weaknesses in the financial system. The problem originated with risky mortgage lending practices by financial institutions. Due to the structural nature of the economic crisis in 2008, the government and relevant financial institutions applied different strategies and pursued different policies that helped to some degree and provided a path for recovery [2, 3, 4]. The current pandemic is widely categorized as a non-structural world crisis. As the virus spread from China, the initial reaction of many countries was to close their borders, and as a result, different economic activities stopped progressing forward including major profit loss in transportation and travel industries as

well as significant increase in unemployment rates. Unfortunately, there is no specific way to predict when a crisis will happen; however, different indicators could help anticipate the coming crisis. The probability of having a crisis increases when all or some of those indicators change. The most common indicators are the rate of GDP, unemployment rate, the devaluation of assets such as stocks. For example, while the rate of unemployment increased for a few months, the rate of GDP decreased consistently for two periods. Lost profit or a low rate of profit in short/long-term deposits can be a sign of an approaching economic crisis [5, 6].

A significant change in stock market movements is one common indicator of a future crisis. In most situations, the decline in the stock market starts months before the crisis begins, and depending on the market, it will impact other markets around the world. As mentioned earlier, depending on the non-structural or structural nature of the crisis, the significant widespread effect of the crisis around the world could be different. For example, in the 2020 pandemic, the Wall Street stock market experienced its worst fall since the financial crisis, and different industries related to airlines, tourism, transportation, and energy were impacted. In this paper, we propose a new model to study the impact of crises on the financial market that is based on the concepts of population analysis and comparative evaluations. We attempt to compare the behavior of different economic sectors during the 2008 economic crash and the Covid-19 pandemic using this proposed model. We applied the correlation network and community/population-based analysis on data sets collected from these two events to find similarities in companies' excess returns. Later, we grouped the companies based on their economic sectors to explain the reasoning behind their behaviors. The remainder of this paper is as follows: First, we review the literature related to the economic crash in 2008 and the current pandemic and briefly review the studies in the financial domain from a big data perspective. The following section looks at the methodology, description of data, and lays out the procedures in the analysis. Then we continue with analytical results, followed by presenting the facts behind the analysis results in the discussion section. In the last section, we present the conclusion.

2. Literature Review

There is no doubt that major activities in the stock market have an impact on the economy. Researchers study the stock market from different perspectives; some studies tried to analyze the market theoretically, and others look at the market empirically. Such studies could

be conducted with statistical analysis, big data overview, financial perspectives, etc. The purpose of these studies is to deeply analyze the market to find the patterns behind the activities, predict the future movements, collect enough information in different periods, and to design proper strategies for any unexpected events in order to help the market recover as soon as possible. Therefore, researchers have tried to conduct different studies applying different methodologies to clarify the effect of different movements either in crises or booms. In [7], the reported study focused on analyzing the impact of the market theoretically and empirically. The result of the analysis showed that during the 1968-70 recession, there was 7% capital loss, 7% in nominal value, and 19% in real terms. Regarding the effect on economic sectors, corporate shareholdings declined by 14% [7].

The United States economic crash in 2008 was the result of a false housing boom, risky financial manipulations in order to grant high-risk mortgage loans, and lack of regulation of stock traders and speculators searching the capital market. The impact of this crisis brought major issues in different sectors: many lost their jobs in the housing sector, the financial market froze, and widespread economic recession occurred. Banks and financial institutions were forced to suspend lending, disrupting the economic cycle in many countries. Following these events, the world financial markets collapsed, the housing market stagnated, and the banking system's problems continued for years. The result of the study [8] showed that besides major economic sectors, such as the financial sector and the housing sector impacted by the crisis, the utility sector had been impacted on three fronts: financing, demand, and expansion. Even though it has been 13 years since the 2008 crisis, researchers are still trying to find similarities and differences in economic sectors' behavior with other crises such as the Great Depression, past pandemics (SARS), and now the Covid-19 pandemic. Another relevant study pointed out that studying simple mathematical models for economics is not enough to prepare for potential corrections to deal with future economic crises [9]. For example, the world was not well prepared to face the consequences of Covid-19 and its wide range of implications. Further, the study argues for the need of economists to work along with social scientists to better understand and face future economic challenges. The researchers in [10] compared various stock market crashes that happened in the 20th century, and concluded that focusing on the stock market is not always a good indicator of economic health, nor is the success of the stock market is an indicator of financial stability.

Another study of the 2008 great crash was reported in [11]. The study concluded that it would be a geopolitical setback to the United States. Other studies highlighted that central banks need to quickly respond to market fluctuations [10, 12].

Since one of the significant ways to invest in the United States is the stock market [13]. Researchers have studied the behaviors in the stock market during unexpected events like the SARS pandemic and Covid-19 pandemic in order to gauge public reaction. The researchers applied a non-parametric Mann-Whitney test and heteroscedastic t-test to examine the impact of the SARS epidemic in 2003 on eight countries' stock market returns. The study showed that there was no significant effect on the economy due to SARS. However, the analysis of the stock market data during the 2020 pandemic demonstrated that stock indices were negatively affected in different countries [14]. Similar to the previous study on the SARS pandemic, researchers applied the same methodology on the weekly panel data to Covid-19 data in order to find any significant impact the pandemic had on the stock market in sixteen countries. When the pandemic started and the time that different health departments verified human-to-human transmission, the stock market had both a short-term and long-term adverse reaction [14]. As of February and March 2020, SP500 lost 1/3 of its value, and a group of researchers conducted a survey to measure respondents' expectations and households' financial prospects. The goal of the study was to find the impact of the 2020 pandemic on the households' decisions and how they adjusted their plan for future investments, savings, and spending [13]. The survey was conducted on the aggregate equity market and dividend futures to investigate how expectations change in a crisis, such as the pandemic. The result of the analysis showed that the investors' behaviors were equally likely to evolve during the crisis and Giglio et al. (2020b) documents that "while respondents on average downward revised their short-run expectations about stock returns and GDP growth during the crash, they remained optimistic about the long-run outlook, and that disagreement across investors increased over the crash" [13].

Besides all statistical and financial analysis, big data analysts have analyzed the stock markets applying different computational and statistical methodologies. One of the methods, computational analysis, builds the correlation network based on stock market data. Companies or stock indices represented the nodes in the network, and edges represented the relationship between nodes. For example, in order to find the similarities in stocks' behaviors, researchers conducted the network analysis on daily prices in different stocks

[15, 16]. The result of the network analysis showed that a small number of companies belonging to the financial sector dominated the market. Correlation network analysis is a powerful method because it can be used for a huge amount of data (such as financial datasets). As an example, a small regional network and a worldwide network on the stock market were constructed and used to make various economic decisions [17, 16]. Correlation-based network analysis has been implemented and validated in multiple research areas such as the bio-sciences (detection of several health conditions), social sciences (identify various stages of disease progression), and various engineering disciplines (safety for the bridges) [18, 19, 20, 21, 22, 23]. The correlation network is identified as a powerful method to apply for complex datasets; accordingly, the researchers need a community detection method to extract the knowledge from the network. Markov clustering algorithm was identified as one of the efficient tools for extracting information [24]. Hence, we applied a correlation network on the complex monthly stock market data and used the same community-based algorithm.

This study uses population analysis to determine which economic sectors behave the same during two crises under study in the United States – the 2008 economic crash and the economic crisis of the current pandemic. In other words, the research question for this study is that: given a crisis, what economic sectors of the stock market will be affected?

3. Methodology

The proposed approach in this study is based on the concept of population analysis, which is to conduct an assessment of an individual element as it compares to a group of peers. The first step in this approach is to build a similarity network based on the desired outcome, in this case, it is based on the similarity in the performance of stocks in a financial market based on their behavior for a certain period, then conducts the analysis based on the structural properties of the constructed network. Graphs are used to model the network, making it possible to take advantage of the numerous tools and algorithms available to extract relevant parameters and relevant information from the graph models. For example, a highly dense subgraph in the obtained network will represent a cluster of stocks with common characteristics as compared to other stocks under study. Such information will be particularly critical when we attempt to assess the behavior of companies under a financial crisis.

When an economic crash or a worldwide pandemic

happens, the immediate effect is always on the stock market, and especially on various sectors of companies in the stock market. However, there are always a few sectors that are resilient to any severe conditions in the stock market. This proposed method aims to study various events and see what sectors are resilient to these events, and why, and how they bounce back very quickly. Identifying these sectors could be an important source of guidance for investors in order to predict pattern in the stock market for financial gain, even in worst case scenarios.

While studying an individual stock or a particular sector gives the information about a specific behavior, grouping/clustering the companies based on their historic behavior will enable us to see the global picture. Once the companies are globally divided, based on their behavior, it is easy to study them at a high granularity level to understand how they are affected and resilient in various events. Comparing one group, with respect to the remaining groups, enables us to identify specific characteristics associated with each group, and what particular parameters are significantly impactful in this group compared to others. The very advantage of graph-based models such as correlation network models in view of stock market related datasets is that they allow us to establish similarities and differences among the companies based on their historical behavior and partition them into groups to assess the overall behavior of each group as related to the rest of the groups [18, 19, 20, 21, 22, 23].

The proposed population analysis method in this study allows us to establish valuable connections among stocks under study and conduct a comparative study of how various stocks or groups of stocks behave in general and under economic crises such as an economic crash or a pandemic. This section explains the series of steps performed right from dataset preparation until identifying various significant sectors using population analysis for both the economic crash 2008 and the Covid-19 pandemic events. The following are the steps in this process (Figure 1): a) Datasets preparation for both the events b) Creating a correlation networks c) Identifying the communities/clusters d) Applying the enrichment analysis to identify the significant sectors.

a) Datasets Preparation

The monthly stock market return data for the economic crash is collected from the Centre for Research in Security Prices (CRSP) database and the Fama-French (FF) databases. The pandemic data is collected from the Bloomberg database. Based on availability, the economic crash data of four years

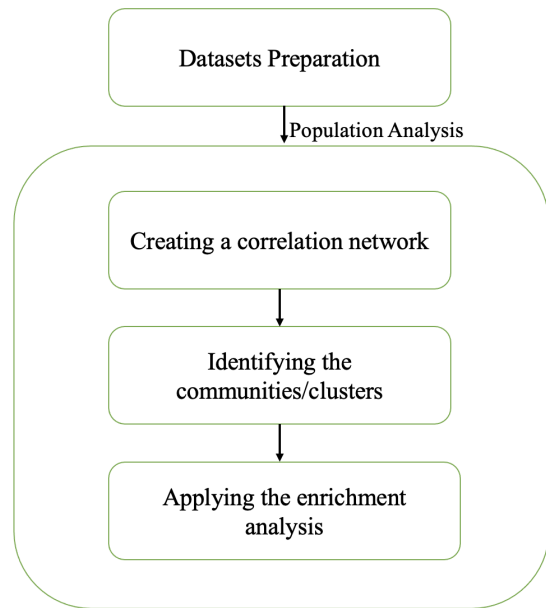


Figure 1. Methodological steps

(48 months - from 2006 to 2009 inclusively) and the pandemic data of eight months (from January 2020 to August 2020) are considered for analysis. The monthly return values from CRSP or Bloomberg databases are subtracted from the risk-free values of the FF database to obtain the parameter Excess Return (ER). The ER values represent the overall behavior of the stock, and it ranges from 1.0 to 2.0, with -1.0 means 100% loss, and 2.0 means 200% gain. This provides the time-series data of the two datasets in terms of their ERs. The time-series data will be used as the input for constructing the correlation networks in the next step. Please note that the 2008 economic crash dataset is referred to as CrashER, and the pandemic dataset as Covid-19ER. There are 1411 companies from 12 sectors in the CrashER dataset and 450 companies from 12 sectors in the Covid-19ER dataset.

b) Creating a correlation network

The correlation network model is based on a graph-theoretic approach that has been used in other application domains, including civil engineering and biomedical informatics as reported in [18, 19, 20, 21, 22, 23]. In this model, each company/stock under study is represented by a node in the graph. Two nodes are connected by an edge in the graph if the behavior of the corresponding companies is similar over the period of time considered in the study, as

shown in Figures 2 and 3. The correlation networks for the two individual datasets were created based on the correlations/similarity patterns of the time-series data of the ER values. Figures 2 and 3 represent the correlation networks ($\rho \geq 0.95\%$) for the CrashER and the Covid-19ER datasets respectively. The time-series data created in the previous step is used to calculate the pair-wise correlations between every pair of companies and similarity is determined based on a correlation value above a certain threshold. We checked the normality of the data and found that it was determined that the data does not follow normality. Hence, we used Spearman ranking correlation coefficients to measure the pair-wise correlation between any two time-series data of the ERs. We have generated networks using different thresholds to create an edge between the nodes in the correlation network including 75%, 85%, and 95%. Different thresholds are tested for the robustness analysis. While different thresholds produce different networks, the overall analysis was not significantly impacted by the threshold variability. In this paper, we are reporting the results obtained using networks generated with correlation parameters equal or above 0.95%.

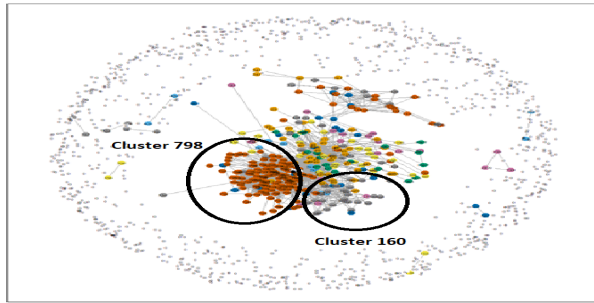


Figure 2. Correlation network of CrashER-dataset

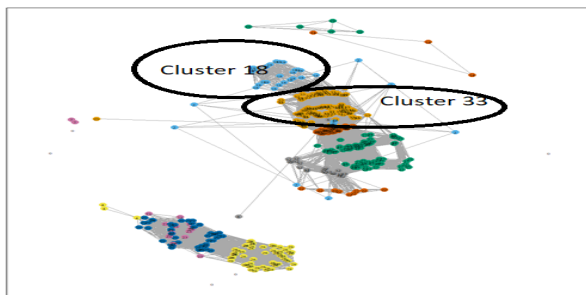


Figure 3. Correlation network of Covid-19ER-dataset

c) Identifying the communities/clusters

The main goal of building a network of stocks that reflects their behavior similarities is to identify groups of stocks that exhibit similar trends over a period time, in particular during disrupting events or crises. To achieve this goal, we used the various clustering algorithms to identify subnetworks of high density in the constructed networks. In this study, we are reporting the results obtained using the Markov Clustering Algorithm (MCL) since it has been shown to perform well on graph-based models [24]. MCL is a graph clustering algorithm based on random-walks that can be tweaked using an input inflation parameter to influence the size of the clusters. The parameter ranges between 1 and 10 with large values producing small clusters. In our experiments, we tested different values of the inflation values on both datasets to identify the values that provide higher degrees of distinguishability among the clusters. In terms of quality of clustering, it is always preferable to produce clusters with high homogeneity of elements in each cluster and higher separation among elements in different clusters. In this paper, we are reporting results obtained using Inflation values between 2 and 4 with 0.5 increments since they provide the best results in terms of homogeneity and separation. We use End Average Excess Return (EAER), which is an average excess return of the last month in the dataset, to assess the quality of the produced clustering, as shown in Table 1. For example, for the CrashER dataset, with the EAER values listed in the last column, the maximum difference between the lowest value of EAER (0.52 for cluster 49) and the highest EAER (5.11 for cluster 160) is 4.59, which is obtained with the inflation parameter value of 2.5. Similarly, different inflation values are tested for the Covid-19ER dataset, and again 2.5 produced the best results. In Table 2, the maximum EAER difference is between cluster 18 (the lowest return value of -12.55) and cluster 30 (highest return value of 103.24) is 115.79. The minimum size of each cluster is calculated based on the average of all the cluster sizes produced. The clusters with at least this minimum size (mean size and above) are considered are called candidate clusters (CC) and that will be considered for the population analysis in the next step.

Row	# Cluster	# Nodes	#Nodes(%)	EAER(%)
[1,]	49	29	2.055	0.52
[2,]	798	132	9.35	1.55
[3,]	160	25	1.77	5.11

Table 1. MCL clusters with the end average excess returns (EAER) (CrashER-dataset)

Row	# Cluster	# Nodes	#Nodes(%)	EAER(%)
[1,]	18	25	5.071	-12.55
[2,]	33	140	28.398	14.81
[3,]	29	47	9.533	16.09
[4,]	32	22	4.462	27.52
[5,]	27	93	18.864	29.02
[6,]	28	59	11.968	30.74
[7,]	30	43	8.722	103.24

Table 2. MCL clusters with end average excess returns (EAER) (Covid-19ER-dataset)

d) Applying the Enrichment analysis

In a process with so many non-deterministic parameters, such as in the assessment and predictions of behavior of stocks under non-structural conditions, it is very difficult to measure performance in absolute values. Hence, the need for assessment by comparison. Stocks that perform better than most can be argued to be performing well as compared to those perform worse than most. Hence, we employ the concept of population analysis, which is a process of assessing performance of one group/community as it compares with the rest of the groups/communities in the domain under consideration. The objective of population analysis in this study is to compare different communities with respect to a certain outcome measure, then identify which input features are enriched in this community. In this study, the outcome parameter is the excess return (ER) and the input parameter is the sector. We apply the population analysis to see the significantly enriched sector(s) in each of the communities with respect to the remaining communities. For finding the significantly enriched parameters by comparing various communities, we use hypergeometric distribution, as shown in Equation 1. The hypergeometric distribution is a sampling without replacement method, used for gene enrichment analysis in earlier studies [25]. We use the hypergeometric distribution with a False Discovery Rate (FDR) of 0.05 to identify the significantly enriched sectors in a certain community of stocks/companies with a common behavior pattern. To apply the hypergeometric distribution, we divide the companies into two sets; a target set and a background set. The background set consists of the universal set; that is all companies used in the study. The target set is any particular community/cluster we want to test if enriched by a particular sector.

$$P(X \geq K) = \sum_{x=K}^{\min(K,n)} \frac{\binom{k}{x} \binom{N-K}{n-x}}{\binom{N}{x}}$$

Where N is the total number of companies in the target set, K is the total number of companies in the same sector, n is the total number of companies in the given cluster, and x is the total number of companies in the given cluster that has the same sector name. $P(X \geq K)$ is the probability of having at least k companies with the same sector name from the given cluster.

Feature	T.S	p_value	R.S	B.S
Basic Materials	1	1	2	3
Utility	5	1	24	29
Finance	122	1.55e-10	25	147
Industrial	2	1	1	3
Real Estate	1	1	1	2
Other	2	1	1	3

Table 3. Significantly enriched parameters for the given cluster #798 of CrashER-dataset.

T.S: Target Set, R.S: Remaining Set, B.S: Background Set

4. Results

The two events, the economic crash 2008 and the Covid-19 pandemic on the stock market data, were compared in this study. As mentioned earlier in the methodology section, we used 48 months of data for the economic crash and eight months of data for the pandemic. Using population analysis procedure, a correlation network is created based on the outcome parameter, and the resultant individual communities are analyzed for the commonly enriched input parameters. In this study, the outcome parameter is the excess return (ER), and the input parameter is the sector. The correlation networks for the two individual datasets were created based on the correlations/similarity patterns of the time-series data of the ER values. Figures 2 and 3 represent the correlation networks ($\rho \geq 0.95\%$) for the CrashER and the Covid-19ER datasets, respectively. A clustering algorithm was applied to the correlation networks individually. Resultant communities/clusters are analyzed further to recognizing the commonly enriched sector(s). The resulting networks are shown with different colors and numbers in Figures 2 and 3. The clusters with at least the minimum size (mean size and above) are the CCs and considered for

the population analysis. The mean cluster sizes for the CrashER-dataset and the Covid-19ER dataset are 15 and 20, respectively. Three CCs were produced with equal and above the mean size of 15 for the CrashER-dataset as shown in Table 1, and seven CCs for the Covid-19ER-dataset as shown in Table 2. In addition, the cluster numbers, number of nodes (or companies), percentage of nodes in each cluster, and their EAER(%) are also shown in tables 1 and 2. Robustness analysis with different thresholds ($\rho \geq 0.75\%$, 0.85% , and 0.95%) was applied to create the correlation networks in order to identify what sectors are significant at different thresholds. Different thresholds are tested for the robustness analysis. While different thresholds produce different networks, the overall analysis was not significantly impacted by the threshold variability. This paper reports the results obtained using networks generated with correlation parameters $\geq 0.95\%$.

The hypergeometric distribution was applied on the resultant clusters to find the parameters' (in this case, sectors) enrichment. Table 1 shows that the clustering algorithm produces three communities/clusters for the CrashER-dataset with a mean size of 15 and above. The three communities are numbered 49, 160, and 798, with the number of nodes 29, 25, and 132, respectively. The total number of nodes is 187. In population analysis, each community is compared against the remaining communities as a target set and a background set using the hypergeometric distribution to identify the significantly enriched sector(s). The background set is the set of all companies/nodes with their respective sector names associated with them, and the target set is the set of nodes/companies from the given community associated with the same sector name. For example, community 798 contains 132 nodes/companies from six different sectors, such as basic materials, utility, finance, industrial, real estate, and others, as shown in Table 3. To identify what sector is significantly enriched out of these six sectors, we apply hypergeometric distribution. For each sector, a target set and a background set is created. For example, there are three companies (Background Set) that are associated with the sector name "Basic Materials" in all three communities together. Out of these three companies, one company (Target Set) is present in community 798. Two companies are in the remaining communities (Remaining Set). The hypergeometric distribution gives a p-value of 1 for the basic materials' sector. Hence, for any significance level (alpha value 0.01, 0.05, and 0.10), basic material is not a significant sector for community 798. For all the remaining sectors of community 798, the similar hypergeometric

distribution is performed, as shown in Table 3. The finance sector is the only significant sector out of all the six sectors with a p-value of $1.55e-10$. This process is applied for both of the remaining clusters with respect to all the sectors. The significantly enriched sectors for all the clusters and the datasets are shown in Table 4.

For the CrashER dataset, finance and utility sectors

dataset	significant parameters	ρ	#CC	p-value
CrashER	Finance	≥ 0.95	CC798	$1.55e-10$
	Utility		CC160	$5.38e-21$
Covid-19ER	Finance	≥ 0.95	CC18	0.0007
	Energy		CC33	0.0002

Table 4. Significantly enriched parameters for both the datasets

are highly significant parameters. Similarly, for the Covid-19ER dataset, finance and energy are the highly enriched sectors. We also observe that the finance sector is commonly enriched for both crises. The third column of Table 4 represents two CCs in each dataset with at least one significant sector. Only these CCs are encircled in the correlation network graphs shown in Figures 2 and 3. For the same CCs, the average excess returns line graphs (behavioral patterns) are plotted as shown in Figures 4 and 5.

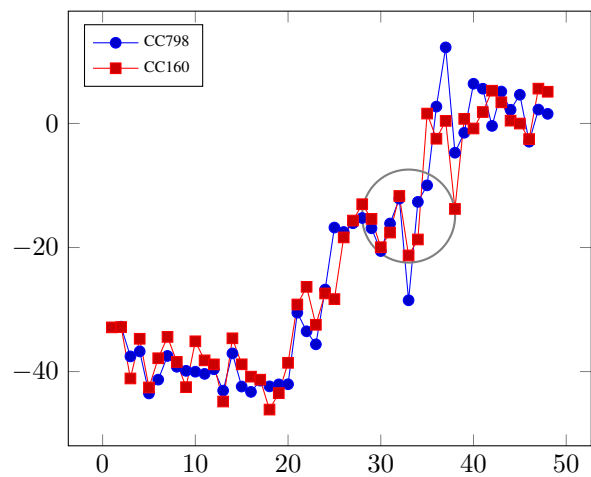


Figure 4. Behavioral patterns of economic crash

From Figure 4, we see that there is a huge crash in September 2008 (the 33rd month and encircled with the event), and from Figure 5, we see that the Covid-19 lockdown started in March 2020 (the 3rd month in the graph encircled with the event). The population

analysis results show that the finance and utility sectors are highly enriched for the CrashER dataset, whereas finance and energy sectors are highly enriched for the Covid-19ER dataset, as shown in Figures 4, 5, and Table 4 (see the encircled patterns), respectively. The reasons behind these sectors being significant for these crises are explained in the next section.

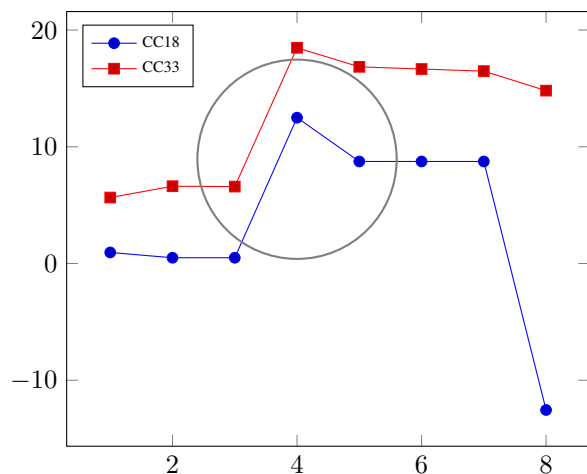


Figure 5. Behavioral patterns of Covid-19

5. Discussion

There are several key findings obtained from this similarity network model. Identifying the financial sector as a sector with common characteristics in both crises is one of the main outcomes of the proposed study, as shown in Table 4. The federal reserve and government interventions played critical roles in the crises. When the economic crash happened in the summer 2008, the government financially assisted Freddie Mac and Fannie Mae, American International Group, and as well as investment bank Merrill Lynch was forced into a merger with Bank of America; in addition to Wachovia was merged into Wells Fargo [8]. Similarly, the federal reserve provided substantial support to the financial sector during the Covid-19 pandemic. The availability of such assistance to financial institutions allowed them more leeway in granting loans to customers and improved their ability to face the challenges that resulted from the pandemic. As a result, banks were able to grant higher loans and decrease their operating costs. The funds provided by the Federal Reserve allowed banks to maintain a good deal of their financial stability, even during economic crisis.

All the sectors were affected by the 2008 financial crisis. However, the utility sector significantly

affected/enriched for the CrashER dataset. Due to the nature of the crisis, the utility sector was affected by financing, demand, and expansion. Since there was decreased trust in the financial market, investors withdrew their investments; thus, this reaction caused more challenges in the market. Although maintaining stability in the market was difficult due to lack of trust, the utility deals decreased. This led to the common characteristics among utility companies that was revealed by our analysis. Similar results were mentioned in previous studies as: "electric utilities are a major presence in the financial markets, particularly in terms of short-term borrowing. When the credit markets froze in mid-September 2008, several prominent utilities took proactive steps to secure access to funds by drawing from their bank credit lines" [8].

The energy sector was one of the sectors recognized as a significant sector from the Covid-19ER dataset. Before the 2020 pandemic, the energy sector suffered from a price battle among key players in this market including Saudi Arabia and Russia. When the pandemic started and social restrictions got put in place, the previous problems in the energy sector accelerated. Based on previous experiences, the energy sector could make a profit if crude oil prices go above \$50 per barrel. However, due to the restrictions in different economic activities, at the end of April 2020, the crude oil prices were \$16 per barrel, which was the lowest price in the last 20 years¹. Therefore, the low-price experience due to the current situation in the economy caused the energy sector to lose its profit. This led to a significant correlation among companies in the energy sector as identified by the network model.

6. Conclusion

In this study, a novel population analysis approach for analyzing the financial markets under crises is introduced and implemented with excess returns as the outcome parameter and financial sector as the input enrichment parameter. The study showed that the proposed network analysis using graph modeling and community-based assessment provides a powerful method for dealing with complex problems utilizing high-volume datasets. In this paper, correlation networks are constructed using financial data associated with the behavior of stocks during two major events: the 2008 market crash and the 2019 global pandemic. For each network, graph clustering algorithms are used to identify groups of companies that behave similarly in terms of their ER values. Each group or community is compared with all the remaining communities using

¹<https://www.statista.com/statistics/326017/weekly-crude-oil-prices>

enrichment analysis to find what sectors are significantly represented in the given community. For each dataset, two communities were identified as significant with at least one significant parameter. Results show that finance and utility sectors are significant for the CrashER dataset, and the finance and energy sector are significant for the Covid-19ER dataset. It was particularly important to observe that the finance sector is a common sector in both crises. The reported analysis showed that no matter whether the crisis is due to a structural or non-structural event, the government always tries to intervene to protect the financial sector. Government support, such as low-interest loans or stimulus packages, is provided to stabilize the economy.

This study represents a starting point for a very promising line of research in analyzing the complex world of financial markets, under normal circumstances as well as during crises. Regarding the limitation of this work, for the CrashER dataset, 48 months of data was available, while for the Covid-19ER dataset, only eight months of data was available to us at the time of testing the network model. Equal length time-series datasets would certainly improve the overall quality of the study, particularly to fully understand the aftermath of any major crisis in the stock market. We also used one enrichment parameter, the financial sector, in the initial study. We plan to include multiple parameters in the enrichment analysis in future studies, including the size and age of the companies. We also plan to consider the intercorrelations among the various parameters. Future studies may also focus on tracking the real-time data to identify the signal/s at the beginning of a stock market crisis based on the comparison with the other crises using this method.

References

- [1] J. E. Stiglitz, "The current economic crisis and lessons for economic theory," *Eastern Economic Journal*, vol. 35, no. 3, pp. 281–296, 2009.
- [2] S. S. Group, *Risk management lessons from the global banking crisis of 2008*. Senior Supervisors Group, 2009.
- [3] J. Crotty, "Structural causes of the global financial crisis: a critical assessment of the 'new financial architecture'," *Cambridge journal of economics*, vol. 33, no. 4, pp. 563–580, 2009.
- [4] M. B. Aalbers, "The financialization of home and the mortgage market crisis," in *The financialization of housing*, pp. 40–63, Routledge, 2016.
- [5] J. Mazurek, E. Mielcová, *et al.*, "The evaluation of economic recession magnitude: Introduction and application," *Prague Economic Papers*, vol. 2, no. 1, pp. 182–205, 2013.
- [6] Y. Karmakar, "Early warning signal system for economic crisis: a threshold and indicators approach," *Pacific Business Review International*, vol. 6, no. 8, pp. 60–70, 2014.
- [7] G. D. Green, "The economic impact of the stock market boom and crash of 1929," in *Federal Reserve Bank of Boston, Consumer Spending and Monetary Policy: The Linkages, Monetary Conference*, pp. 189–220, 1971.
- [8] J. Cannell, "The financial crisis and its impact on the electric utility industry," *Edison Electric Institute*, 2009.
- [9] G. M. Hodgson, "The great crash of 2008 and the reform of economics," in *The Handbook of Globalisation, Third Edition*, Edward Elgar Publishing, 2019.
- [10] F. S. Mishkin and E. N. White, "Us stock market crashes and their aftermath: implications for monetary policy," tech. rep., National bureau of economic research, 2002.
- [11] R. C. Altman, "The great crash, 2008," *Foreign Aff.*, vol. 88, p. 1, 2009.
- [12] E. N. White, "The stock market boom and crash of 1929 revisited," *Journal of Economic perspectives*, vol. 4, no. 2, pp. 67–83, 1990.
- [13] T. Hanspal, A. Weber, and J. Wohlfart, "Exposure to the covid-19 stock market crash and its effect on household expectations," *Review of Economics and Statistics*, pp. 1–45, 2020.
- [14] K. KHAN, H. ZHAO, H. ZHANG, H. YANG, M. H. SHAH, and A. JAHANGER, "The impact of covid-19 pandemic on stock markets: An empirical analysis of world major stock indices," *The Journal of Asian Finance, Economics, and Business*, vol. 7, no. 7, pp. 463–474, 2020.
- [15] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciche, N. Vandewalle, and R. N. Mantegna, "Networks of equities in financial markets," *The European Physical Journal B*, vol. 38, no. 2, pp. 363–371, 2004.
- [16] K. T. Chi, J. Liu, and F. C. Lau, "A network perspective of the stock market," *Journal of Empirical Finance*, vol. 17, no. 4, pp. 659–667, 2010.
- [17] S. Vitali, J. B. Glattfelder, and S. Battiston, "The network of global corporate control," *PloS one*, vol. 6, no. 10, p. e25995, 2011.
- [18] K. Cooper, W. Hassan, and H. Ali, "Identification of temporal network changes in short-course gene expression from *c. elegans* reveals structural volatility," *International Journal of Computational Biology and Drug Design*, vol. 12, no. 2, pp. 171–188, 2019.
- [19] K. M. Dempsey and H. H. Ali, "Identifying aging-related genes in mouse hippocampus using gateway nodes," *BMC systems biology*, vol. 8, no. 1, p. 62, 2014.
- [20] S. Kim, I. Thapa, and H. H. Ali, "A graph-theoretic approach for identifying bacterial inter-correlations and functional pathways in microbiome data," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 405–411, IEEE, 2018.
- [21] E. Rastegari, S. Azizian, and H. Ali, "Machine learning and similarity network approaches to support automatic classification of parkinson's diseases using accelerometer-based gait analysis," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
- [22] P. Chetti and H. Ali, "Analyzing the structural health of civil infrastructures using correlation networks and population analysis," in *Proceedings of the Eighth International Conference on Data Analytics*, pp. 12–19, 2019.

- [23] P. Chetti and H. Ali, "Estimating the inspection frequencies of civil infrastructures using correlation networks and population analysis," *International Journal on Advances in Intelligent Systems*, vol. 13, no. 1&2, pp. 125–136, 2020.
- [24] S. vanDongen, "A cluster algorithm for graphs," *Information Systems [INS]*, no. R 0010, 2000.
- [25] M. L. Hale, I. Thapa, and D. Ghersi, "Funset: an open-source software and web server for performing and displaying gene ontology enrichment analysis," *BMC bioinformatics*, vol. 20, no. 1, p. 359, 2019.