University of Tehran Press

Interdisciplinary Journal of Management Studies (IJMS)

Home Page: https://ijms.ut.ac.ir

Online ISSN: 2981-0795

Graph Theory in Action: Forming an Investment Portfolio (A Case Study: Selected Companies of Tehran Stock Exchange)

Fatemeh Zarei[1*] | Marziyeh Nourahmadi[2]

1. Corresponding Author, Department of Finance and Accounting, Humanities Faculty, Meybod University, Meybod, Iran. Email: zareifatemeh@meybod.ac.ir
2. Department of Financial Engineering, Hazrat-e Masoumeh University, Qom, Iran. Email: m.nourahmadi@hmu.ac.ir

ARTICLE INFO

Article type:
Research Article

Article History:
Received 23 July 2023
Revised 25 April 2025
Accepted 05 May 2025
Published Online 08 December 2025

Keywords:
Investment portfolio,
Hierarchical risk parity,
Stock network,
Adjacency matrix.

ABSTRACT

The research aims to integrate portfolio theory and graph theory to explore the relationship between specific network properties and the diversification problem in forming stock portfolios. The stock portfolio formation process is described using graph theory. The study utilizes stock-adjusted final price data from 138 companies listed on the Tehran Stock Exchange between January 1, 2019, and July 6, spanning 918 trading days. The data is partitioned with 80% as in-sample and 20% as out-of-sample data. A proximity matrix is employed to analyze stock relationships, leading to the construction of diversified and non-diversified portfolios using an optimal threshold. Machine learning techniques and hierarchical risk parity are used to select stocks for the portfolio. The performance is compared with the minimum variance approach and the total index benchmark for both in-sample and out-of-sample periods. The Sharpe ratio evaluates the performance of both diversified and non-diversified portfolios. The findings suggest that the non-diversified approach outperforms during market crashes, while diversified portfolios perform better during other periods. Hence, diversifying the stock portfolio is unsuitable during market crashes due to strong direct correlations among stocks, causing simultaneous declines. Alternative strategies should be considered following market crashes.

Cite this article: Zarei, F. & Nourahmadi, M. (2026). Graph Theory in Action: Forming an Investment Portfolio (A Case Study: Selected Companies of Tehran Stock Exchange). *Interdisciplinary Journal of Management Studies* (*IJMS*), 19 (1), 1-19. http//doi.org/10.22059/ijms.2025.362745.676068

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/ijms.2025.362745.676068

Introduction

One of the primary challenges faced by investors in the stock market is the selection of stocks that will generate profits. Investors typically perceive a direct relationship between the level of risk associated with an asset and its potential return. Consequently, higher levels of risk are generally expected to yield higher returns, and vice versa. To assess the performance of an investment portfolio, investment managers engage in three key activities collectively known as portfolio management, investment policy, market timing, and portfolio selection (Brinson et al., 1986). However, asset management and investment allocation can be arduous and time-consuming tasks. Investment managers often need to devise customized approaches tailored to individual clients or investors.

Portfolio optimization lies at the heart of every successful investment strategy, expertly navigating the delicate dance between potential returns and inevitable risks. It represents the science of crafting a diversified asset allocation that maximizes your desired outcomes, drawing upon a rich tapestry of research and methodologies.

Seminal works, such as Sharpe’s (1963, 1971) groundbreaking papers, laid the theoretical foundation for modern portfolio optimization. These pioneering contributions established the importance of risk-adjusted expected return metrics, shaping the very language of efficient investing.

Various models have been proposed for portfolio construction, with Markowitz's approach being the most renowned among them. Nevertheless, the mean-variance theory associated with this approach presents several practical challenges, particularly in estimating the expected return and covariance for different asset classes. Recent studies have tried to overcome these limitations by applying metaheuristic optimization algorithms. For instance, Sadeghi Moghadam et al. (2022), introduced an asexual reproduction optimization (ARO) method to improve Markowitz-based portfolio selection under cardinality constraints (Sadeghi Moghadam et al., 2022).

In this study, we employ the Hierarchical Risk Parity (HRP) machine learning technique in conjunction with the Markowitz method and compare the outcomes of these two approaches. Notably, predicting the covariance matrix with N stocks necessitates at least (N (N+1))/2 i.i.d covariance estimates (identically and independently distributed) (Dimitrios & Vasileios, 2015).

However, there is substantial evidence suggesting that asset returns exhibit fluctuations, heterogeneous clustered variance, and an unstable correlation structure over the long-term periods. These factors can lead to significant errors that undermine the benefits of portfolio diversification. Moreover, during times of crisis, such as the COVID-19 pandemic, alternative assets, including Bitcoin, have been shown to play a significant role in portfolio management and risk mitigation (Azouzi & Mghadmi, 2025).

López de Prado (2016) was the first to propose a hierarchical method for portfolio construction. López de Prado utilized machine learning techniques and graph theory to develop a unique diversified portfolio using the hierarchical risk parity (HRP) approach, which differs from conventional risk-based portfolio optimization methods. The HRP method addresses the issue of inverting the covariance matrix by organizing the relationship between securities in the portfolio as a hierarchy, creating clusters of similar assets based on correlation coefficients. This hierarchical structure accomplishes three main objectives: First, it fully utilizes the information within the covariance matrix. Second, it ensures stability in weight allocation. Third, unlike most traditional risk-based asset allocation methods, it eliminates the need for inverting the covariance matrix (Bechis et al., 2020; López de Prado, 2016; Nourahmadi & Sadeqi, 2022).

Recent advancements, as highlighted by researchers such as Berouaga et al. (2023), emphasize the ongoing progress in optimization techniques. By delving into innovative tools such as minimum spanning trees, researchers are pushing the boundaries of portfolio construction, ensuring that investors have access to increasingly refined strategies. The analysis of stock networks presents a novel approach to comprehending various aspects of stock markets and has become crucial in recent studies of financial markets. In this research, we employ a combination of portfolio theory and graph theory to illustrate the relationship between graph theory and the diversification problem in stock portfolio formation. Additionally, the process of forming a stock portfolio is described using graph theory (Berouaga et al., 2023).

The primary question driving this research is how to utilize graph theory for stock market investment. Answering this question can assist investors in identifying the genuine risks associated with their investments and help prevent economic downturns, such as the recession of 2008, while also contributing to the enhancement of financial literacy among individuals.

To conduct this research, the theoretical literature on graph theory and relevant research conducted in the field are first reviewed. In the subsequent section, the research methodology is explained. Specifically, the adjusted final price data of 138 companies listed on the Tehran Stock Exchange from 2019-01-01 to 2021-07-06, covering 918 trading days, are utilized. The dataset is split into 80% for in-sample analysis and 20% for out-of-sample evaluation.

To depict the relationships between stocks, the adjacency matrix is employed. Additionally, diverse and non-diverse diversified and non-diversified portfolios are obtained using an optimal threshold. In the third section, the results are compared using the Sharpe ratio for both portfolios, and the outcomes of both methods are compared against the performance of the total index. Finally, the research concludes with suggestions and recommendations based on the findings.

Literature Review

The study of complex systems, particularly their topology structure and the relationships between their elements, has become a significant focus in various social and economic domains. Many of these systems can be effectively explained through complex graphs and associated principles. In the 1970s, as researchers delved deeper into the behavior of complex systems, they recognized that such behavior is not solely determined by the constituent factors of these systems. Instead, these systems demonstrate adaptive and intelligent behavior as they respond to their environments (Kheyrkhah et al., 2016).

During the 1980s and 1990s, scholars began exploring economic models that diverged from traditional approaches. These new models conceptualized the economy as a dynamic and interactive system rather than a static equilibrium-based system. They introduced novel methods that allowed for the simulation of agent interactions within a system, mirroring real-world scenarios. Financial markets were no exception to this trend, as researchers utilized interactive models, such as complex networks, to describe the behavior of agents within these markets (Kheyrkhah et al., 2016). A graph consists of a collection of nodes or vertices that interact with each other through edges or connections (Danko & Šoltés, 2018). This graph-based representation has proven valuable in understanding and analyzing the intricate relationships and dynamics within complex systems, including financial markets.

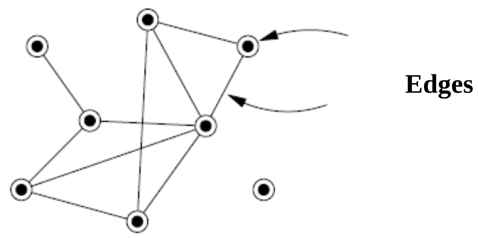


Fig. 1. A Small Network of Vertices and Edges (Newman, 2003)

Networks are generally regarded as adhering to the principles of graph theory, which is a fundamental aspect of discrete mathematics. One of the earliest references to networks and their study dates back to 1735 when Euler provided a solution to the Seven Bridges of Königsberg problem, known as the first problem solved in network theory. Euler's solution is considered a cornerstone of modern graph theory (Attia, 2019).

Following this, a significant contribution to network theory was made by the Irish mathematician, William Rowan Hamilton, who invented a puzzle involving the discovery of a path where each vertex is visited exactly once. This concept is known as the Hamiltonian path. Gradually, this branch of mathematics has developed to the present day, allowing mathematicians and the general public to apply it to the study of relationships and various other fields (Dickson, 2006). Social science researchers have been at the forefront of utilizing graphs, often employing interactive questionnaires to identify and model relationships between individuals within networks (Kim et al., 2007).

In these social networks, individuals are represented as nodes, and their interactions with others form edges. The central focus is often on determining centrality, which involves identifying individuals who occupy a central position in terms of their interactions or possess the most connections with others. The concept of centrality aids in understanding communication patterns within a group (Fortunato, 2010).

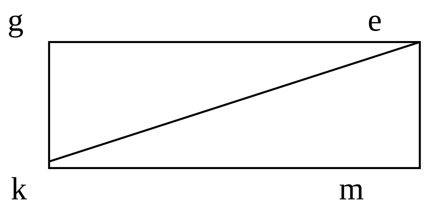
Random graphs, characterized by randomly distributed edges, are among the most crucial types of graphs. They lack distinct and unique patterns, serving as a baseline for comparison against other network structures.

In recent years, the focus has shifted from analyzing small networks at the level of edges and nodes to exploring the statistical characteristics of large-scale graphs. This change reflects the growing interest in understanding the overarching features and dynamics of extensive networks (Attia, 2019).

Network theory encompasses three primary objectives. Firstly, it seeks to identify and analyze statistical characteristics of networks, such as the distribution of path lengths or the degree distribution of nodes. These characteristics determine the behavioral and structural properties of the network, and finding appropriate measures to quantify these properties is an essential aspect of network theory. Secondly, network theory aims to develop methods that allow for a deeper understanding of the network's properties, nature, and the interactions among its nodes. These methods enable researchers to explore the underlying dynamics and mechanisms operating within the network. Lastly, network theory aims to uncover regulatory principles that govern the behavior of networks, particularly concerning the behavior of small communities within the larger network (Kheyrkhah et al., 2016).

In the context of graph theory, graph D consists of two sets: W, which represents the vertices or nodes of the graph, and F, which represents the edges connecting pairs of vertices (Attia, 2019). An example is illustrated in Figure 2 below.

Note: The vertices g and m are not directly connected. Additionally, it is important to clarify whether the edges gk and em are included in the graph.



W= {g, e, k, m}
F= {ge, ek, km, gm}

Fig. 2. D graph

When two points in a graph are connected by an edge, they are considered adjacent to each other. This adjacency relationship is crucial in understanding the connectivity and relationships within a graph. In the case of a simple graph where there is only one possible connection between each pair of points, this connectivity can be represented using an adjacency matrix.

An adjacency matrix is a square matrix that represents a graph by indicating the connections between vertices. The matrix consists of rows and columns corresponding to the vertices of the graph. If two vertices are connected by an edge, the corresponding entry in the adjacency matrix is set to 1, indicating the presence of a connection. If there is no edge between two vertices, the corresponding entry is set to 0, indicating the absence of a connection. By examining the values in the adjacency matrix, one can determine the structure and connections within the graph (Attia, 2019).

	E	F	J	H
E	[[0	1	0	1]
F	[1	0	1	1]
J	[0	1	0	1]
H	[1	1	1	0]]

Fig. 3. Adjacency Matrix

Figure 3 can be represented by the following adjacency matrix. In the adjacency matrix for a simple graph, entries are either 0 or 1, indicating the presence or absence of an edge between vertices.

The diagonal line, starting from the top left and continuing to the bottom right, always contains 0s since a vertex cannot be connected to itself (Dickson, 2006).

Correlation Matrix		**Adjacency Matrix**
[[0 0.2 0.4 0 1]		[[0 1 0 0 1]
[0.2 0 0.2 0.3 0.2]	Diversified &	[1 0 1 0 1]
[0.4 0.2 0 0.6 0.1]	a threshold of 0.21 →	[0 1 0 0 1]
[0.4 0.3 0.6 0 0.1]		[0 0 0 0 1]
[0.1 0.2 0.1 0.1 0]]		[1 1 1 1 0]

Fig. 4. Adjacency Matrix Assuming a Threshold of 0.21

By employing different definitions of nodes and connections, various types of graphs can be created to analyze financial markets. Researchers have adopted different approaches in defining nodes, such as using stock symbols (Bonanno et al., 2004; Lan & Zhao, 2010) or stock indices as nodes to examine the interactions between stock markets in different countries (Liu et al., 2011). In numerous articles, the correlation coefficient has been commonly used to define the connections between nodes (Chi et al., 2010; Kumar & Deo, 2012; Liu et al., 2011;), while others have utilized the Granger causality effect (Huang et al., 2009). In this research, the stock network is defined as a graph comprising nodes representing individual stocks and edges representing the correlations of stock prices between them.

The foundations of modern portfolio theory were established by Markowitz (1952) and Sharpe (1964). They investigated the impact of asset risk, return, correlation, and the variation in expected portfolio returns. Markowitz's portfolio optimization methods are utilized in this article, integrating graph theory into the portfolio optimization approach. Graph theory serves as a valuable tool for understanding and analyzing financial markets, particularly when examining the interrelationships among various financial instruments such as stocks, securities, and indices.

Correlation-based analysis forms the primary field of study in graph analysis, as the correlation coefficient captures the interdependencies among financial instruments. Utilizing graphs based on correlations enables a reduction in the complexity of financial relationships, facilitating a better understanding and prediction of the dynamics within financial markets. In the following sections, several commonly used methodological approaches in the realm of correlation graphs are introduced.

Early studies on correlation-based graphs utilizing graph theory were conducted by Mantegna (1999). In his research on financial markets, Mantegna (1999) introduced the minimum spanning tree method and highlighted its main advantages. Through constructing a minimum spanning tree, he observed that US stocks tended to group based on industry sectors. This suggests that stock prices not only reflect information about a company's financial situation but also provide insights into the structure and topology of the stock market.

Onnela et al. (2003) focused on analyzing financial markets to construct a portfolio of securities. They examined the mean distribution and variance of the New York Stock Exchange from January 1980 to December 1999. Their study revealed that the minimum spanning tree exhibited a fully connected structure and followed a scale-free pattern. Furthermore, they concluded that stocks with minimal risk were located at the periphery of the tree, indicating a greater distance from the central node. They found that the risk of a portfolio was directly related to the normality of the minimum spanning tree. Thus, portfolio diversification should align with the normalization of the tree. They also suggested that assets with the highest potential for diversification were situated at the edges of the minimum spanning tree, hypothesizing that these shares offered significant diversification potential.

These early studies demonstrated the usefulness of graph theory, particularly the minimum spanning tree method, in understanding the relationships and risk characteristics of financial markets. They shed light on the role of correlation-based graphs in portfolio construction and the identification of diversification opportunities.

Bonanno et al. (2004) conducted studies assuming that the returns of stocks traded in the market are influenced by the time horizon used to calculate correlation coefficients. They discovered that as the time horizon decreases, the network structure transitions from a complex organization to a simpler one. They concluded that utilizing network theory is a valuable approach to filter economic data and gain insights into the behavior of stocks during different periods.

Mizuno et al. (2006) analyzed data from the foreign exchange market, focusing on various currencies. They developed a hierarchical classification of currencies using the minimum spanning tree method, which revealed the cluster structure of currencies and identified the main currency within each cluster. The identified clusters often aligned with geographical regions, such as Asia or Eastern Europe, and the major economies were typically represented by the key currencies, as expected. Huang et al. (2009) employed thresholding techniques to construct a correlation network of Chinese stocks and investigated the structure of the stock market based on this network.

In their study, "Correlation, Hierarchy, and Networks in Financial Markets," Tumminello et al. (2010) discussed quantitative models for analyzing characteristics of correlation matrices. The correlation matrix plays a crucial role in stock portfolio optimization and other quantitative aspects of asset price dynamics in financial markets.

In their study titled "Proposal for Creating a Portfolio with Minimum Risk," Šoltés and Danko (2017) presented a model for selecting a portfolio with the minimum expected risk. This model is relevant to our work as it provides a theoretical framework for minimizing risk in portfolio selection, which aligns with our goal of optimizing investment strategies using modern approaches.

Peralta and Zarei (2016) conducted a study to investigate the suitability and effectiveness of financial networks in stock portfolio selection. They focused on the New York stock market and treated each stock as a node, with the correlation between pairs of stocks representing the edges. The study considered two key characteristics of each stock: individual stock performance and systematic stock performance. Individual stock performance refers to the company's performance in isolation from other companies, which is quantified by the standard deviation of stock returns, reflecting the specific investment objectives. Systematic performance is utilized to determine the role of a particular stock in the overall market and is measured by the centrality score of each company in the stock market network. Throughout their research, they examined the interaction between these two dimensions and their effects on optimal investment choices. The primary contribution of their study was to simplify the portfolio selection process by associating groups of stocks with specific centrality areas (Peralta & Zarei, 2016). In a study titled "Network-Based Portfolio Selection: Is It Aligned With the Markowitz Optimal Portfolio?" Hüttner et al. (2018) demonstrated that the relationship between network-based portfolio selection and the Markowitz optimal portfolio does not arise from a structural similarity between the two selection mechanisms but is instead influenced by the specific characteristics of the correlation matrices. Recognizing the significance of financial networks and their role in portfolio selection, the combination of graph theory and portfolio theory is further discussed to explore the connection between network properties and the diversification problem in stock portfolio formation.

Methodology

Complex network analysis is a valuable tool for understanding various aspects of stock markets and has gained prominence in recent studies of financial markets. It originates from graph theory in discrete mathematics and has developed as a theoretical framework for comprehending the structural properties of networks. Complex network analysis provides clear insights into the internal structure of stock exchanges (Dimitrios & Vasileios, 2015).

Investigating and analyzing the structure of the stock market network aids in explaining stock behavior and the interactions between different factors. This approach challenges the assumption of independence underlying traditional linear analysis methods, which focus on identifying the effects of several independent variables on a dependent variable. Many systems in nature can be effectively explained through complex networks and the associated principles (Albert & Barabási, 2002; Newman, 2003; Zhou et al., 2006).

The purpose of this research is to combine network theory and portfolio theory to demonstrate how network features are associated with the diversification problem in forming a stock portfolio. The following steps outline the research methodology.

This research falls within the realm of quantitative and applied research. The data used in this study encompass all companies listed on the stock exchange. The data was collected from the Rahvard

Novin database and processed using Microsoft Excel. Python 3.8 was employed for data analysis and conducting the research.

The data preparation process begins with data extraction. In this step, the adjusted daily data for all companies listed on the Tehran Stock Exchange from January 1, 2019, to July 6, 2021 (approximately 138 stocks) for a total of 918 trading days, was extracted using Rahvard software.

The selection of the period and its division into in-sample and out-sample sets were designed to include both ascending and descending market regimes, incorporating the market fall in 2019. This approach ensures that the data encompasses both regimes, facilitating a comprehensive analysis of the market's impact on the data and ensuring that the data are not biased by a specific market regime.

The next stage of data preparation is data preprocessing. This involves cleaning the data by addressing factors such as noise, outliers, and missing data that could compromise data quality. In this research, the number of trading days for each stock was checked, resulting in 138 stocks remaining in the statistical population, while the rest were excluded due to insufficient trading information (i.e., inadequate number of trading days for those stocks).

In this study, the adjusted closing price data have been employed as the primary input variable for constructing the financial network. The raw price series denotes the stock price of firm *i* at time *t*, while the corresponding price at time *t–1* is utilized to compute returns. Given that stock price distributions tend to exhibit characteristics akin to a log-normal distribution, the logarithmic return formula is adopted for return calculation. The logarithmic return is computed as follows (Nourahmadi & Sadeqi, 2021).

By calculating the logarithmic returns, the researchers can capture the relative changes in stock prices over time, which helps in assessing the dynamics of the stock market network.

$$R_{it} = ln\frac{P_{it}}{P_{it-1}} \quad (1)$$

Where:

$\boldsymbol{R_{it}}$ represents the logarithmic return of firm i on day t,

$\boldsymbol{P_{it}}$ represents the stock price of firm i on day t, and

$\boldsymbol{P_{it-1}}$ represents the stock price of firm i on day t-1.

The general steps involved in conducting the research are as follows:

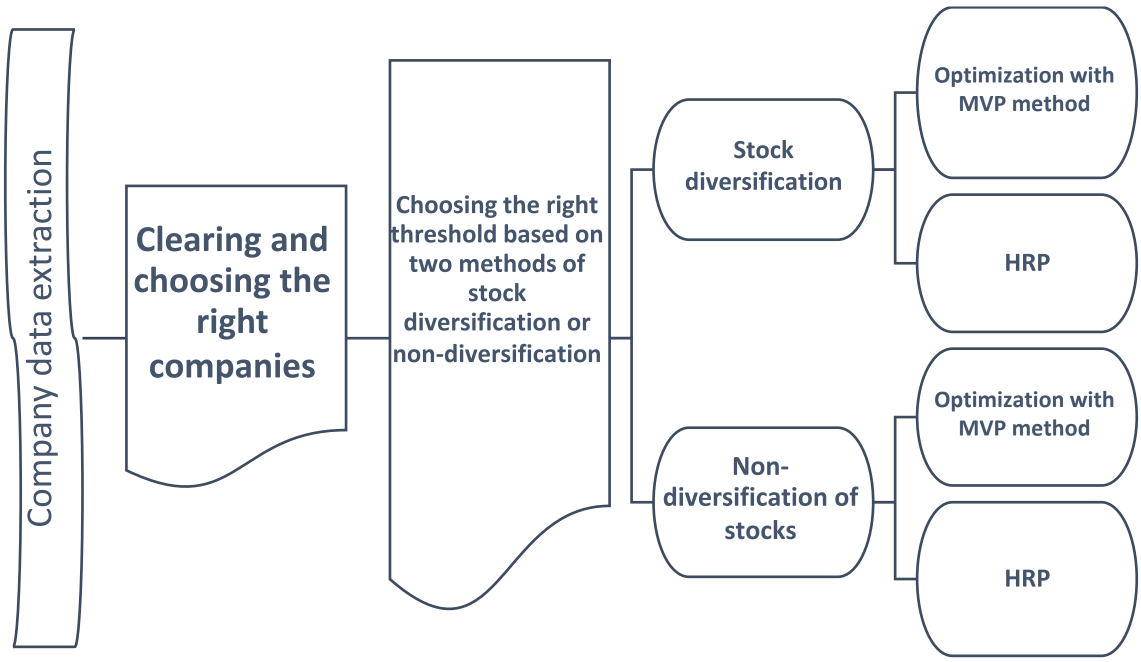


Fig. 5. Research Procedure and Implementation Steps

As shown in Figure 5, diversified and non-diversified stock portfolios were formed after determining the optimal threshold using two methods, MVP and HRP, for two periods: within-sample and out-of-sample.

The HRP model, which combines machine learning techniques and graph theory, involves three main stages: tree clustering, quasi-diagonalization, and recursive bisection. Each step is explained in more detail below.

In the first step, the portfolio assets are divided into clusters using the Hierarchical Tree Clustering algorithm. The correlation matrix is transformed into a correlation-distance matrix D for two assets i and j. (Burggraf, 2021)

$$D\left(X_i\, X_j\right)=\sqrt{\frac{1}{2}\left(1-\rho_{ij}\right)} \tag{2}$$

In the second stage of the tree clustering process, the Euclidean distance between the vectors of two columns in the matrix D is calculated. This computation results in a distance matrix $\bar{\mathbf{D}}$.

$$\bar{D}\left(ij\right)=\sqrt{\sum_{k=1}^{N}\left(D\left(ki\right)-D(kj\right)^2} \tag{3}$$

The main difference between the equation mentioned earlier and the one calculated in the previous section is that the former calculates the distance between two securities, i and j, within the portfolio, while the latter computes the distance between the two assets, which represents a function of the entire correlation matrix.

In the next step, the first cluster is created by identifying the pairs with the minimum distance (Nourahmadi & Sadeqi, 2021).

$$U\left[1\right]=\arg\min_{ij}\bar{D}\left(ij\right) \tag{4}$$

This computation involves taking the average of the pairwise distances between assets in cluster U and assets in cluster i, normalized by the sizes of the clusters (Burggraf, 2021).

$$\bar{D}\left(iU\left[1\right]\right)=\min\left(\bar{D}\left(ii^*\right)\bar{D}\left(ij^*\right)\right) \tag{5}$$

This step is iteratively performed for each stock in the portfolio. With each iteration, a new cluster of assets is formed, and the algorithm updates the distance matrix accordingly. This process continues until only one cluster remains (Bechis et al., 2020).

The final step of the HRP algorithm is the bisection step, which is crucial as it determines the final weights of the securities in the portfolio. In this step, the algorithm leverages the property of the portfolio where "inverse allocation is optimal for the diagonal covariance matrix."

After the clustering process, the algorithm divides each cluster into two sub-clusters $\boldsymbol{V_1}, \boldsymbol{V_2}$, beginning from the end cluster $\boldsymbol{U[N]}$. According to the weight given to the portfolio, $\boldsymbol{w_i}\ \forall \boldsymbol{i} = \mathbf{1} \cdots \boldsymbol{N}$, the variance of Sub-clusters is computed as follows (Burggraf, 2021):

$$V_{12}=w^TVw \tag{6}$$

where:

$$w=\frac{diag\left(V\right)^{-1}}{trace\left(diag\left(V\right)^{-1}\right)} \tag{7}$$

Based on the two weighting factors, the algorithm proceeds to update the portfolio weights for the sub-clusters. The allocation of the portfolio is determined solely based on the assets present within each cluster.

By incorporating these weighting factors, the algorithm updates the sub-cluster weights in accordance with their respective contributions to the total variance, while also accounting for the overall portfolio weight (Burggraf, 2021).

$$w_1=a_1{*}w_1,\ w_2=a_2{*}w_2 \tag{8}$$

The bottom-up weighting method utilized in the HRP algorithm provides a notable advantage compared to other allocation algorithms. Instead of having all the assets in the portfolio competing against each other for allocation, only the assets within each group compete for allocation. By adopting this method, the HRP algorithm effectively reduces the competition and interaction among all the assets in the portfolio, leading to a more efficient and optimized allocation strategy (Nourahmadi & Sadeqi, 2021).

In this research, the stock network is defined as a graph consisting of edges and nodes, where the nodes represent individual stocks and the edges represent the fluctuation of stock prices.

The first step of the analysis involves determining the correlation between pairs of stocks within a specific set and constructing a correlation matrix. Subsequently, an input threshold is applied to the correlation values, converting them to either one or zero based on whether they exceed or fall below the threshold value. This process generates an adjacency matrix, where the nodes correspond to stocks, and the edges indicate the highest or lowest correlations.

Analyzing the correlation between multiple pairs of stocks can be challenging, particularly when establishing appropriate thresholds and interpreting the resulting patterns.

One crucial question that arises is how to select the optimal threshold, which will be elucidated below. To address this, the standard deviation is employed to ensure that a specific percentage of edges is captured in the graph. Specifically, the mean minus one standard deviation is utilized to retain edges between the top 16% of most correlated stocks. Similarly, the mean plus one standard deviation captures the relationship between the top 16% of most correlated stocks (Attia, 2019).

By employing this methodology, the research aims to determine the optimal threshold that best captures the significant correlations within the stock network, facilitating further analysis and interpretation of the relationships between stocks.

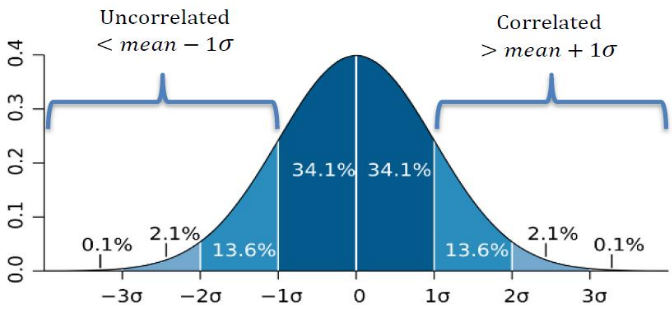


Fig. 6 . Standard Deviation Thresholds for Correlation-Based Portfolio Construction

To determine the optimal threshold, the researcher or user is required to make a choice between constructing a diversified (represented by the letter D) or a non-diversified (represented by the letter U) stock portfolio. This choice is based on the researcher's specific goals and preferences.

In the case of a diversified stock portfolio, the threshold is set using a uniform mean minus one standard deviation. This ensures that the resulting graph retains edges representing the correlations between the top 16% of correlated stocks.

Conversely, for a non-diversified stock portfolio, the threshold is set using a mean plus one standard deviation. This threshold captures the relationships between the top 16% of most correlated stocks.

By offering these options and utilizing the appropriate thresholds, the research aims to facilitate the construction of either a diversified or non-diversified stock portfolio based on the user's preference and specific objectives (Attia, 2019).

```
#If users are searching for correlated stocks or uncorrelated stocks:
user preference=input (Type in 'U' if you want undiversified (correlated) stocks in your portfolio, or 'D' if you want diversified (uncorrelated) stocks)
if user preference == D
   threshold_with_standard_deviation
= (statistics. mean (correlations list)- standard deviation)
if user preference == U
   threshold_with_standard_deviation= (statistics. Mean (correlations list) +standard deviation)
Print ("\nUsing_the_Standard_Deviation,",threshold_with_standard_deviation, "should be the best threshold for the inputted stock tickers")
```

Fig. 7. How to Choose the Best Threshold for Diversified and Non-Diversified Portfolios (Attia, 2019)

In this research, all calculations and data analysis were performed using Python. The implementation of the topic is discussed as follows:

Data Analysis

The first step involved investigating the correlation between different stocks. Positive correlations indicate that if one stock is affected by a negative factor, it can have a similar impact on another stock, causing both stock prices to decline. Therefore, portfolio diversification is achieved by investing in assets that are uncorrelated or have a low correlation with one another.

To calculate the correlation matrix, the dataset was transformed into daily returns. Normalizing the data allows for easy comparison of returns between two assets. To visualize the data, a heat map was created. The heat map represents the correlation matrix, and its color-coded scale spectrum provides a visual representation of the strength of correlations. Strong positive correlations are depicted in dark green, while assets with lower correlations are represented in red.

In addition to the heat map, clustering analysis was performed to examine the data in a more structured manner. This clustering helps identify groups of assets that exhibit similar patterns or relationships.

Figure 8 displays the heat map, which uses a color-coded scale spectrum to represent the correlations between assets, with dark green indicating strong positive correlations and red indicating lower correlations.

By analyzing the heat map and performing clustering, the research gains insights into the relationships and patterns present in the data, facilitating a deeper understanding of the stock market dynamics. (Note: The description of the heat map is provided and Figure 8 is presented for illustrative purposes.)

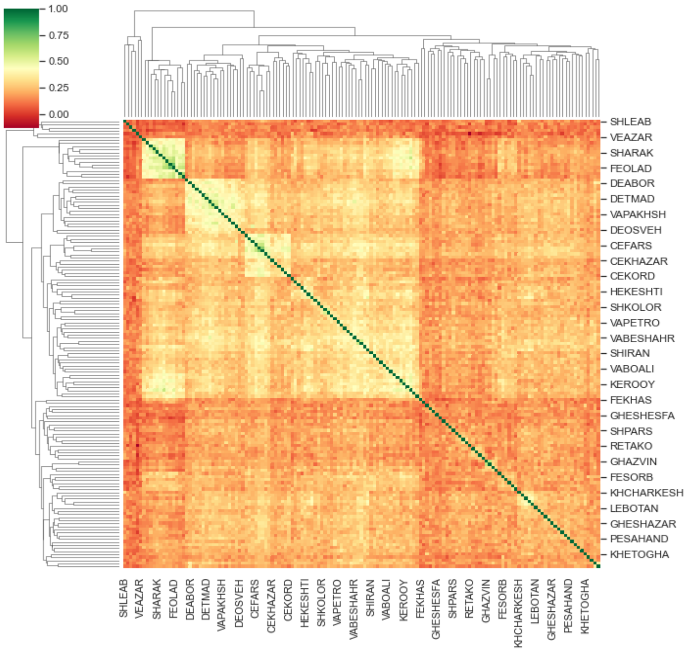


Fig. 8. Correlation Clustering Matrix

Heat maps provide valuable information by visualizing the degree of correlation between assets. However, as an investor making investment decisions, simply relying on a heat map may not be sufficient. In this research, the Network x package in Python was utilized to analyze networks and explore the correlations between assets more comprehensively.

To analyze the correlation matrix as a network, the connections between assets were converted into an edge list. This list contains detailed information about each connection between assets, enabling a more granular analysis.

Investors typically aim to understand the level of correlation among assets to assess the unsystematic risk in their portfolio. They are interested in identifying assets with high correlations and those with low correlations. By introducing a threshold value for the strength of correlation, the number of connections between nodes in the network can be reduced. Additionally, node sizes can be adjusted to indicate the assets with the highest correlations to other assets.

In the research, after calculating the correlations between assets, the average correlation between shares was found to be 0.2357. Furthermore, the average correlation between stocks was determined as 0.2308, with a data dispersion (standard deviation) of 0.08157. These calculations are crucial in determining the threshold, as explained in Figure 7.

In the subsequent analysis, the results of the diversified and non-diversified portfolios using the two optimization methods are presented and compared, shedding light on the effectiveness of each approach. (Note: The description of the calculations and results is hypothetical and provided for illustrative purposes)

Non-Diversified Stock Portfolio

Based on Figure 7 and selecting the term "U" for a non-diversified stock portfolio, the optimal threshold value is determined to be 0.3173. Therefore, the selected shares for the non-diversified basket, as mentioned, are: Fasmin, Fbahner, Fakhuz, Femli, Foulad, Folage, Kechad, Kegel, Pardis, Shepdis, Sharak, Shafen, Shakharek, Webank, Vasanat, Vespe, and Tusum.

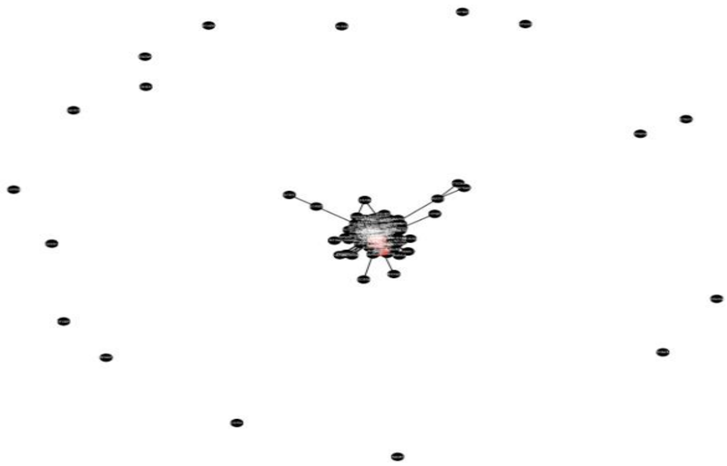


Fig. 9. Non-Diversified Portfolio Stock Network

Diversified Stock Portfolio

Based on Figure 7 and selecting the term "D" for a diversified stock portfolio, the optimal threshold value is determined to be 0.1542. Therefore, the selected shares for the diversified basket, as mentioned, are: Fakhas, Fenward, Gahkamat, Ghpak, Ghishsefa, Kafra, Kasadi, Khagstar, Shalab, and Azar.

Fig. 10. Diversified Stock Network

In the subsequent analysis, the Markowitz and HRP methods have been employed to determine the appropriate weight estimation for the selected stocks and construct a portfolio comprising these stocks. The evaluation of the results involves dividing the period in to two segments: the within-sample period, encompassing 80% of the data, and the out-of-sample period, consisting of the remaining 20% of the data.

For reference, the in-sample period spans from January 1, 2010, to January 3, 2021, comprising a total of 734 days. On the other hand, the out-of-sample period extends from January 4, 2021, to July 6, 2021, encompassing 184 days.

A) Optimization of Non-Diversified Stock Portfolio

Figure 11 depicts the returns of the selected assets for the non-diversified stock portfolio. However, the specific details and values associated with these returns are not provided, making it impossible to draw the figure or conduct further analysis.

Assessing the performance of the non-diversified stock portfolio typically entails analyzing the returns of the selected assets during the in-sample and out-of-sample periods. By examining these returns, one can evaluate the effectiveness of the Markowitz and HRP methods in weight estimation and portfolio formation. It is common to compare the portfolio's performance with benchmark indices or alternative portfolios.

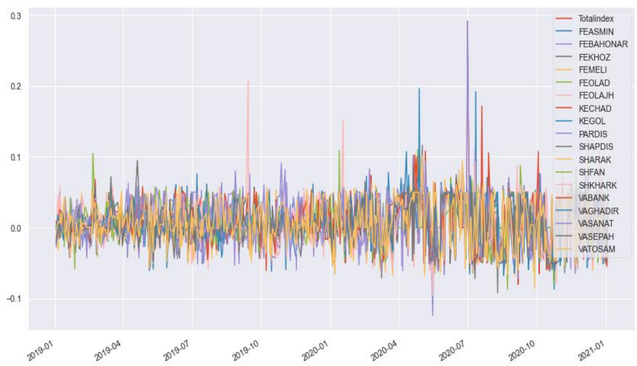


Fig. 11. The Non-Diversified Portfolio Returns

Figure 12 shows the dendrogram of the companies.

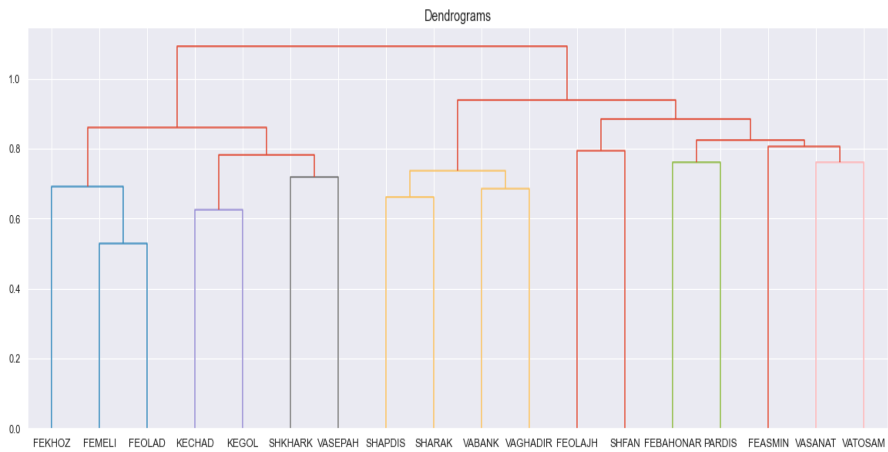


Fig. 12. Dendrogram of Non-Diversified Stock Portfolio

Next, using HRP and MVP methods, the optimal weight of the portfolio has been calculated and these two methods have been evaluated using in-sample and out-of-sample data.

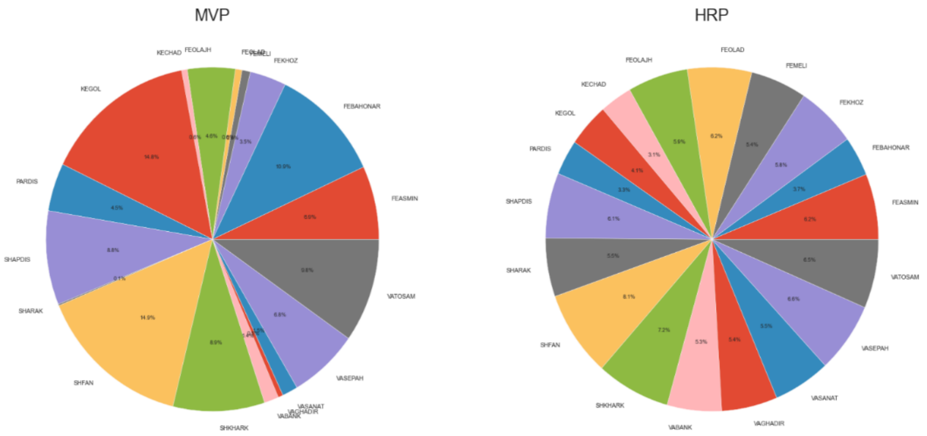


Fig. 13. Optimal Portfolio Weight

In Figure 14, the performance results of the non-diversified stock portfolio within the in-sample data are presented. These results depict the performance achieved using two optimization methods, MVP (Minimum Variance Portfolio) and HRP (Hierarchical Risk Parity). Additionally, the performance of the stock market's total index is included for comparison.

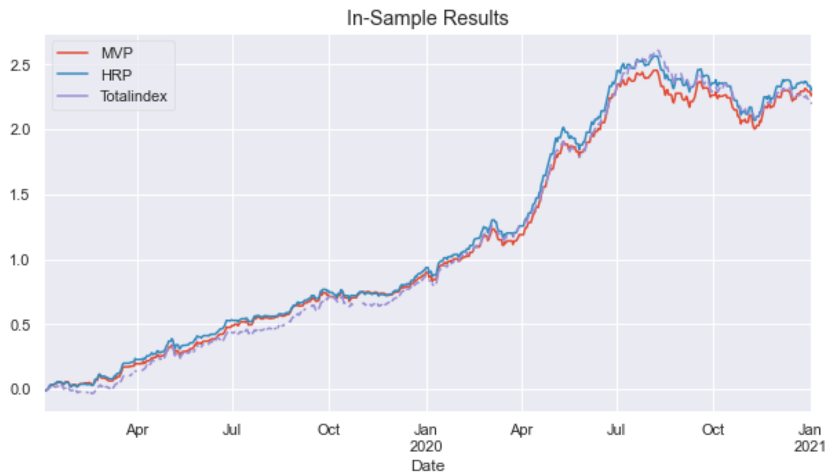


Fig.e 14. In-Sample Performance of Non-Diversified Stock Portfolio

Table 1. Non-Diversified Stock Portfolio

	standard deviation	Sharpe
MVP	0/247	3/165
HRP	0/256	3/113
total index	0/247	4/185

The in-sample results for the non-diversified stock portfolio indicate that the performance of the total index was better than the MVP and HRP approaches, and the Sharpe ratio was equal to 4.18. The out-of-sample results for both optimization methods are plotted against the performance of the total index in Figure 15.

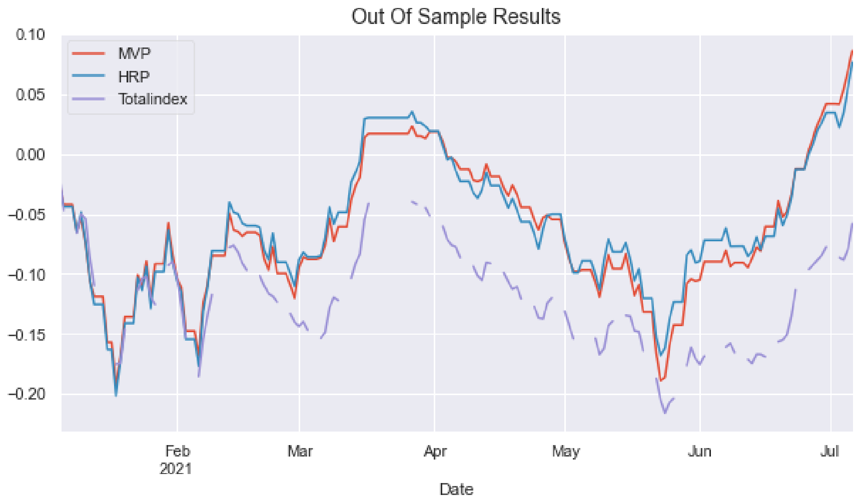


Fig. 15. Non-Diversified Stock Portfolio Out-of-Sample Results

Table 2 displays the out-of-sample results of the non-diversified stock portfolio, comparing the performance of both optimization methods against the total index. The total index performed comparatively weaker during this period in comparison to the other two optimization approaches. This outcome can be attributed to the decline in the stock market during the out-of-sample period.

During market downturns, the diversification of a stock portfolio may not provide significant benefits, as all stocks tend to experience a strong negative correlation and decline together. This observation aligns with the results obtained from the diversification of the stock portfolio during the out-of-sample period, as further elaborated in the subsequent section.

Table 2. The Non-Diversified Stock Portfolio

	standard deviation	Sharpe
MVP	0/227	0/524
HRP	0/240	0/440
total index	0/227	–0/538

b) Optimizing the Diverse Stock Portfolio

In figure 16, the results of stock returns for selected stocks of the diversified portfolio are illustrated.

Fig. 16. The Diversified Portfolio Returns

The dendrogram of the diversified stock portfolio is presented in Figure 17.

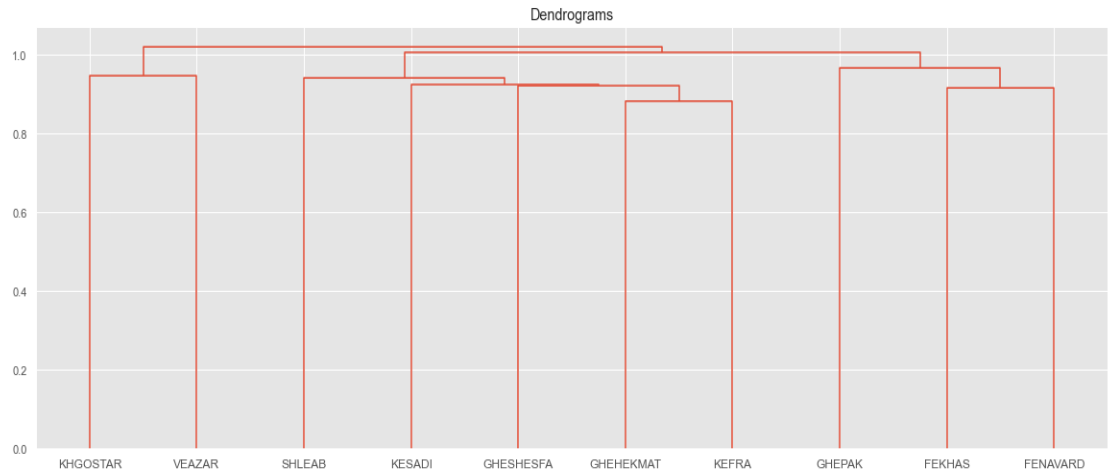


Fig. 17. Dendrogram of Diversified Stock Portfolio

In Figure 18, the optimal weight of the portfolio has been calculated using the HRP and MVP methods. The performance and evaluation of these two methods are assessed based on in-sample and out-of-sample data.

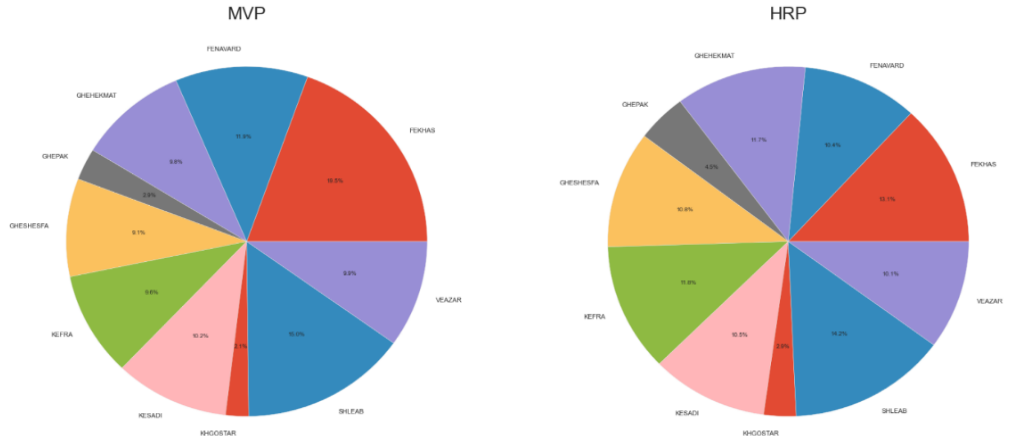


Fig. 18. Diversified Portfolio Weight

Table 3 shows the test results of the diversified stock portfolio sample for two optimization approaches and the total index. As can be seen, the HRP method performs better than the MVP and the total index.

Table 3. Within the Diversified Stock Portfolio Sample

	standard deviation	**Sharpe**
MVP	0/191	4/909
HRP	0/194	4/940
total index	0/274	4/185

In Figure 19, the results of the in-sample shape for both optimization approaches are shown against the performance of the total index.

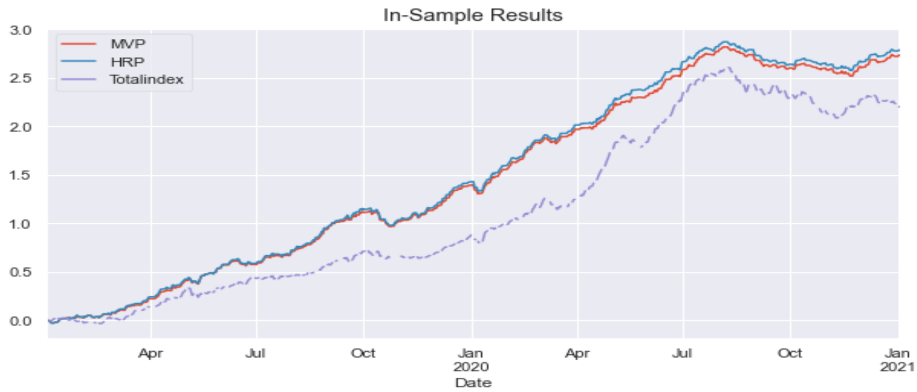


Fig. 19. Within the Diversified Stock Portfolio Sample

Table 4 presents the out-of-sample results for both optimization methods in comparison to the total index. As mentioned earlier, during a declining market, the impact of diversification on yield tends to be limited. Therefore, it is recommended to consider purchasing stocks that have a negative beta with the overall market index during such periods.

Table 4. An Example of a Diversified Stock Portfolio

	standard deviation	Sharpe
MVP	0/161	-2/115
HRP	0/163	-2/179
total index	0/227	-0/538

Figure 20 shows the out-of-sample results of the diverse stock portfolio for both optimization methods against the total index.

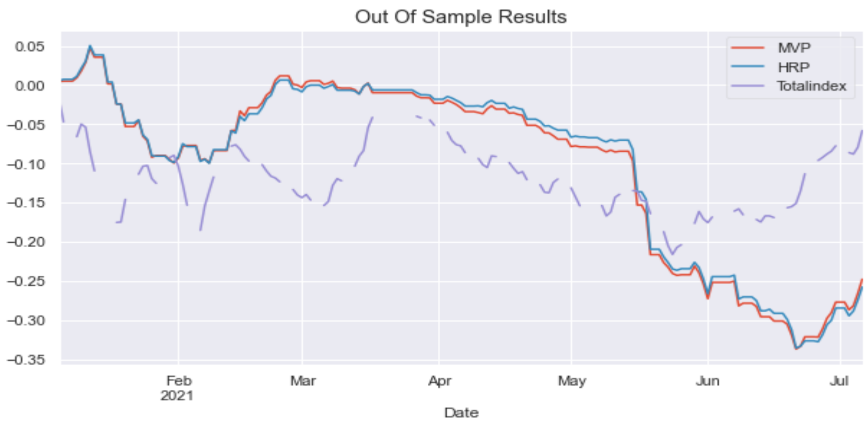


Fig. 20. An Example of a Diversified Stock Portfolio

By comparing the results of diversified and non-diversified stock portfolios, it is observed that, in general, the diversified stock portfolio outperforms the non-diversified stock portfolio during upward and normal market periods. On the other hand, the non-diversified stock portfolio tends to perform better than the diversified portfolio during downward market periods.

This suggests that diversification provides benefits in terms of performance during stable and positive market conditions, where the correlation between stocks is lower. However, in bearish market periods, the benefits of diversification diminish as correlations between stocks increase, and most stocks tend to decline together.

Therefore, it is recommended that investors consider purchasing stocks with a negative beta to the market index during bearish market periods. Even a diversified portfolio may not offer significant advantages during such times. The findings highlight the importance of understanding market conditions and tailoring investment strategies accordingly to optimize returns and manage risk.

Conclusion

Understanding the dynamics of the stock market is crucial for investors, financial policymakers, and market participants. The stock market is a complex system with various factors influencing its dynamics, including the interdependence of stocks.

Portfolio management is a complex process that involves selecting and managing a combination of investments to optimize returns and manage risk. Investors aim to construct portfolios with the highest potential return and the lowest risk. With the increasing complexity of market conditions, portfolio selection has become even more important.

This research explores the use of stock network analysis for portfolio selection, focusing on diversified and non-diversified portfolios. Two optimization methods, MVP and HRP, are used to evaluate and compare the performance of these portfolios within-sample and out-of-sample.

The results show that the non-diversified portfolio approach performs better during market downturns, while the diversified portfolio approach is superior during other market periods. This finding aligns with previous studies that highlight the benefits of diversification during stable economic periods.

It is important to note that non-diversified portfolios can be riskier and more unpredictable, leading to significant gains or losses. Investors need to carefully consider their risk tolerance and investment objectives when selecting a portfolio approach.

This research covers the period from 2019-01-01 to 2021-07-06, aiming to capture different market regimes. Future research could extend the analysis to include additional periods and compare the results. Additionally, researchers can explore trading strategies that can predict market regimes and adjust portfolio optimization methods accordingly.

Overall, this research contributes to the understanding of portfolio selection using stock network analysis. The limited stock selection approach used in this study can be beneficial for investment portfolio managers and investors seeking guidance in their decision-making process.

Practical Implications

The findings of this research have important practical implications for portfolio managers and investors. Specifically, the study indicates that during market downturns, switching to non-diversified portfolios can help reduce losses, whereas diversified portfolios perform better during stable economic periods. This insight can help investors develop strategies that are adaptable to changing market conditions. Moreover, the use of stock network analysis provides a valuable tool for identifying meaningful relationships between stocks, offering investors a new perspective on making more informed investment decisions.

Reference

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47. https://doi.org/10.1103/RevModPhys.74.47

Attia, J. (2019). The applications of graph theory to investing. *arXiv preprint arXiv:1902.00786*. https://doi.org/10.48550/arXiv.1902.00786

Azouzi, M. A. & Mghadmi, N. (2025). The Portfolio management properties of bitcoin during the coronavirus crisis. *Interdisciplinary Journal of Management Studies*, *18(2)*, 333-347. https://doi.org/10.22059/ijms.2024.375551.676692

Bechis, L., Cerri, F., & Vulpiani, M. (2020). Machine learning portfolio optimization: Hierarchical risk parity and modern portfolio theory. *LUISS Guido Carli, Roma, Italy*. http://tesi.luiss.it/28022/1/709261_BECHIS_LUCA.pdf

Berouaga, Y., El Msiyah, C., & Madkour, J. (2023). Portfolio optimization using minimum spanning tree model in the Moroccan stock exchange market. *International Journal of Financial Studies*, *11*(2), 53. https://ideas.repec.org/a/gam/jijfss/v11y2023i2p53-d1104772.html

Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B*, *38*, 363-371. https://doi.org/10.1140/epjb/e2004-00129-6

Brinson, G. P., Hood, L. R., & Beebower, G. L. (1995). Determinants of portfolio performance. *Financial Analysts Journal*, *51*(1), 133-138. https://doi.org/10.2469/faj.v51.n1.1869

Burggraf, T. (2021). Beyond risk parity–A machine learning-based hierarchical risk parity approach on cryptocurrencies. *Finance Research Letters*, *38*, 101523. https://doi.org/10.1016/j.frl.2020.101523

Chi, K. T., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, *17*(4), 659-667. https://doi.org/10.1016/j.jempfin.2010.04.008

Danko, J., & Šoltés, V. (2018). Portfolio creation using graph characteristics. *Investment Management & Financial Innovations*, *15*(1), 180. https://doi.org/10.21511/imfi.15(1).2018.16.

Dimitrios, K., & Vasileios, O. (2015). A network analysis of the Greek stock market. *Procedia Economics and Finance*, *33*, 340-349. https://doi.org/10.1016/S2212-5671(15)01718-9

Dickson, A. (2006) Introduction to graph theory. *CRC Pres*. www.math.utah.edu/mathcircle/notes/MC_Graph_Theory.pdf

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, *486*(3-5), 75-174. https://doi.org/10.1016/j.physrep.2009.11.002

Huang, W. Q., Zhuang, X. T., & Yao, S. (2009). A network analysis of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, *388*(14), 2956-2964. https://doi.org/10.1016/j.physa.2009.03.028

Hüttner, A., Mai, J. F., & Mineo, S. (2018). Portfolio selection based on graphs: Does it align with Markowitz-optimal portfolios? *Dependence Modeling*, *6*(1), 63-87. https://doi.org/10.1515/demo-2018-0004

Kheyrkhah, A., Rahnamay Roodposhti, F., & Afshar Kazemi, M. A. (2016). Using the theory of network in finance. *International Journal of Finance & Managerial Accounting*, *1*(2), 9-23. https://ijfma.srbiau.ac.ir/article_9527.html.

Kim, K., Kim, S. Y., & Ha, D. H. (2007). Characteristics of networks in financial markets. *Computer Physics Communications*, *177*(1-2), 184-185. https://doi.org/10.1016/j.cpc.2007.02.037

Kumar, S., & Deo, N. (2012). Correlation and network analysis of global financial indices. *Physical Review E*, *86*(2), 026101. https://doi.org/10.1103/PhysRevE.86.026101

Lan, W., & Zhao, G. (2010, October). Stocks network of coal and power sectors in China stock markets. In *International Conference on Information Computing and Applications* (pp. 201-208). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16336-4_27

Liu, J., Chi, K. T., & He, K. (2008, September). Detecting stock market fluctuation from stock network structure variation. In *International Symposium on Nonlinear Theory and Its Applications*, NOLTA'08. Budapest, Hungary. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=61758028c040ac09c5f016516cdd9adf16066ebf

Liu, J., Tse, C. K., & He, K. (2011). Fierce stock market fluctuation disrupts scale-free distribution. *Quantitative Finance*, *11*(6), 817-823. https://doi.org/10.1080/14697680902991627

López De Prado, M. L. (2016). Building diversified portfolios that outperform out- of- sample. *The Journal of Portfolio Management*, *42*(4), 59-69. https://doi.org/10.3905/jpm.2016.42.4.059

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, *11*, 193-197. https://doi.org/10.1007/s100510050929

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.12987/9780300191677.

Mizuno, T., Takayasu, H., & Takayasu, M. (2006). Correlation networks among currencies. *Physica A: Statistical Mechanics and its Applications*, *364*, 336-342. https://doi.org/10.1016/j.physa.2005.08.079

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167-256. https://doi.org/10.1137/S003614450342480

Nourahmadi, M., & Sadeqi, H. (2021). Hierarchical risk parity as an alternative to conventional methods of portfolio optimization: A study of Tehran Stock Exchange. *Iranian Journal of Finance*, *5*(4), 1-24. https://doi.org/10.30699/ijf.2021.289848.1242

Nourahmadi, M., & Sadeqi, H. (2022). A machine learning-based hierarchical risk parity approach: A case study of portfolio consisting of stocks of the top 30 companies on the Tehran Stock Exchange. *Financial Research Journal*, *24*(2), 236-256. https://doi.org/10.22059/frj.2021.319092.1007146

Onnela, J. P., Chakraborti, A., Kaski, K., Kertesz, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, *68*(5), 056110. https://doi.org/10.1103/PhysRevE.68.056110

Peralta, G., & Zareei, A. (2016). A network approach to portfolio selection. *Journal of Empirical Finance*, *38*, 157-180. https://doi.org/10.1016/j.jempfin.2016.06.003

Sadeghi Moghadam, M. R., Mansouri, T., & Sheykhizadeh, M. (2022). Markowitz-Based cardinality constrained portfolio selection using asexual reproduction optimization (ARO). *Interdisciplinary Journal of Management Studies*, *15*(3), 531-548. https://doi.org/10.22059/ijms.2021.313393.674293

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, *9*(2), 277-293. https://doi.org/10.1287/mnsc.9.2.277

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425-442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Sharpe, W. F. (1971). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, *6*(5), 1263-1275. https://doi.org/10.2307/2329860

Šoltés, V., & Danko, J. (2017). Proposal of creation of a portfolio with minimal risk. *Invest. Manag. Financ. Innov.*, *14*, 107-15. https://doi.org/10.21511/imfi.14(2).2017.10.

Tumminello, M., Lillo, F., & Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, *75*(1), 40-58. https://doi.org/10.1016/j.jebo.2010.01.004

Zhou, T., Liu, J. G., Bai, W. J., Chen, G., & Wang, B. H. (2006). Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. *Physical Review E*, *74*(5), 56-109. https://doi.org/10.1103/PhysRevE.74.056109