Journal of Money and Economy
Vol. 20, No. 4, Fall 2025
pp. 575-600

Original Research Article

Unveiling Financial Market Structure through Network Filtering: MST and PMFG Approaches

Fatemeh Lakzaie* **Alireza Bahiraie†**
Ali Arshadi‡

Received: 28 Sep2025 Approved: 18 Nov 2025

Stocks are fundamental instruments in capital markets, and understanding their interdependencies is crucial for effective investment and risk management. This study adopts a network-based framework to examine the structure and dynamics of the U.S. stock market by constructing correlation networks through analytical methods rather than purely data-driven approaches. Specifically, the Minimum Spanning Tree (MST) and Planar Maximally Filtered Graph (PMFG) techniques are employed to model the relationships among 371 constituent stocks of the S&P 500 index, using daily closing prices from January 11, 2013, to December 12, 2022. The key empirical findings indicate that MST and PMFG effectively capture the hierarchical organization of financial markets. These network structures facilitate the identification of stock clusters and central nodes, providing insights into systemic risk and potential contagion during periods of financial turbulence. Furthermore, the identified network topology supports enhanced portfolio diversification by enabling the selection of stocks across different clusters or based on centrality metrics. Temporal analysis of the network evolution also reveals shifts in market conditions and sectoral rotations. Additionally, the study integrates Markowitz's portfolio optimization framework by applying the Sharpe ratio as the performance criterion. The resulting optimized portfolio achieves an expected annual return of 37.61%, with a volatility of 14.15%, yielding a Sharpe ratio of 2.6581 demonstrating robust risk-adjusted performance. In conclusion, MST and PMFG offer valuable tools for capturing market structure, informing portfolio construction, and enhancing risk assessment, thereby contributing to more resilient investment strategies.

*Department of Mathematics, Faculty of Science, Semnan University, Iran
†Department of Mathematics, Faculty of Science, Semnan University, Iran; alireza.bahiraie@semnan.ac.ir (Corresponding Author)
‡ Faculty Member, Monetary and Banking Research Institute, CBI, Iran

Keywords: Portfolio Optimization, Minimum Spanning Tree, Stock Market, Planar Maximally Filtered Graph, Networks
JEL Classification: G11, G12, G32, G38

1 Introduction

In the realm of investment management, understanding the intricate relationships among stocks is essential for effective portfolio optimization and risk management. Traditional approaches, such as the Markowitz model, focus on balancing risk and return, while network-based methods like the MST and PMFG provide tools to visualize and analyze market structures. However, a key challenge lies in integrating these network techniques with performance metrics like the Sharpe ratio to enhance practical portfolio construction, particularly in dynamic markets such as the U.S. stock market. This integration is crucial because it allows investors to not only identify hierarchical stock interdependencies but also evaluate risk-adjusted returns in a method-driven framework.

Despite advancements in financial network analysis and portfolio theory, there remains a gap in studies that combine MST and PMFG with Sharpe ratio optimization for the S&P 500 index using daily closing prices over extended periods. Previous research has often relied on high-frequency data or focused on specific markets like China, overlooking the potential of these networks to inform diversification strategies and systemic risk assessment in mature markets like the U.S.

This study addresses this gap by proposing a framework that applies MST and PMFG to correlation networks derived from S&P 500 stocks, followed by an empirical evaluation using the Markowitz model and Sharpe ratio. The primary objectives are: (1) to uncover the hierarchical structure and clusters within the U.S. stock market; (2) to assess how these network insights can optimize portfolios for better risk-adjusted performance; and (3) to identify central nodes and diversification opportunities that mitigate systemic risks. The key research questions include: How do MST and PMFG reveal market dynamics in the S&P 500? What is the impact of these structures on Sharpe ratio-optimized portfolios? And how can they enhance investor decision-making during periods of market stress?

The novelty of this work lies in its method-driven approach, shifting from high-frequency correlations to robust, filtered networks applied to a decade-long dataset of 371 S&P 500 stocks. This contributes added value by providing actionable insights for portfolio diversification and market monitoring, extending beyond theoretical models to practical applications in investment management.

The remainder of the paper is organized as follows: Section 2 reviews the relevant literature on financial networks and portfolio optimization. Section 3 describes the data and methodology. Section 4 presents the empirical findings. Section 5 discusses implications and limitations, and Section 6 concludes with recommendations for future research.

2 Literature Review

The literature on financial networks and portfolio optimization can be categorized into three main areas: (1) network-based analysis of stock markets, (2) portfolio optimization models, and (3) integrated applications of networks in risk and return assessment. This review systematically examines key studies, highlighting their methodologies, findings, and limitations, before discussing how the present study advances this body of work.

Network-Based Analysis of Stock Markets

Financial networks have emerged as powerful tools for modeling complex market structures. Temporal networks, for instance, capture the dynamics of stock markets and detect instability through topological changes (Zhao et al., 2018). Stock correlation networks represent stocks as vertices and Pearson correlations as edge weights, illustrating interdependencies in price movements (Wang et al., 2018; Birch et al., 2015). To simplify these dense networks, filtering algorithms such as the MST (Mantegna, 1999; Guo et al., 2018), PMFG (Wang & Xie, 2015; Tumminello et al., 2005; Tumminello et al., 2007), and Correlation Threshold Method have been developed (Boginski et al., 2005; Namaki et al., 2011; Chi et al., 2010). MST, in particular, is favored for its simplicity and ability to visualize hierarchical interconnections without cycles, selecting M-1 key edges. PMFG extends this by preserving the MST hierarchy while retaining additional edges 3(M-2) to capture more information, thus providing a planar graph that filters out noise (Wang et al., 2017). Empirical applications include Výrost et al.'s use of MST, PMFG, and threshold graphs to study network dynamics, incorporating centrality measures like eigenvalue, betweenness, and expected force to identify influential nodes (Výrost et al., 2019). Similarly, Guo et al. (201٨–202٢) analysed 100 Chinese stocks using PMFG to assess market risk and pinpoint systematically important firms (Guo et al., 2022). These studies demonstrate that networks reveal clusters and central hubs, aiding in systemic risk identification, but often focus on emerging markets or short-term data, limiting generalizability to mature markets like the U.S.

Portfolio Optimization Models

Portfolio optimization seeks to maximize returns for a given risk level, with no universal solution due to varying investor preferences (Ivanova & Dospatliev, 2017). Harry Markowitz's 1952 model laid the foundation by emphasizing diversification through mean-variance analysis (Markowitz, 1952). Despite criticisms, it remains a cornerstone in modern portfolio management (Deng et al., 2012; Ozyesil, 2021; Chen et al., 2021; Iqbal et al., 2019; Bower & Wentz, 2005; Stempien & Chan, 2017; Barroso et al., 2021; Yusan & Riyadi, 2024). Stock investments aim to generate future financial returns, making financial data analysis crucial for investment decisions. Decision-making models based on such analyses help investors select the best options (Pratama et al., 2020). Recent extensions include Lakzaie et al.'s (2024) optimization of a 50-stock portfolio from the Tehran Stock Exchange using the Markowitz model integrated with MST, based on synchronous correlations of daily returns. Their findings underscored MST's role in extracting economic insights from time series (Lakzaie & Bahiraie, 2024).

However, these models often overlook network structures, relying solely on statistical correlations without filtering for robustness.

Integrated Applications of Networks in Risk and Return Assessment

A growing body of research integrates networks with portfolio theory. Li et al. (2020) combined GARCH-BEKK and PMFG to examine Sino-US trade friction's impact on China's stock market, revealing influences from global commodities, indices, and macroeconomic factors on fluctuations (Li et al., 2020). This highlights networks' utility in understanding interdependencies but is context-specific to trade events.

Overall, prior studies provide robust tools for market analysis and optimization, yet they predominantly use high-frequency data or isolated applications of MST/PMFG without linking to Sharpe ratio evaluations in large-scale indices like the S&P 500. The present study builds on this by employing MST and PMFG on a decade-long dataset of 371 S&P 500 stocks, integrating these with Markowitz optimization and Sharpe ratio assessment. This adds value by offering a comprehensive framework for identifying hierarchical structures, clusters, and diversification strategies, thereby enhancing risk management and portfolio performance in the U.S. market.

3 Data and Methodology

3.1 Data set

The S&P 500, established by Standard & Poor’s in 1957, is an American stock market index that reflects the market capitalization of 500 large companies with common stock listed on NASDAQ or the NYSE. It is widely regarded as one of the most accurate indicators of the U.S. stock market and serves as a key index representing a broad segment of the American economy. Due to its diversity and size, the S&P 500 enables investors and analysts to assess overall market performance and monitor economic trends effectively. The empirical analysis is based on the daily closing prices of 371 constituent stocks of the

S&P 500 Index, covering the period from January 11, 2013, to December 12, 2022. The selection of these 371 stocks for the study was made because they are the only stocks for which complete data is available throughout the specified period of analysis. The daily return of stock i on day t is defined as:

$$r_i(t) = ln\, p_i(t) - ln\, p_i(t-1) \tag{1}$$

where $p_i(t)$ and $p_i(t-1)$ are the closing prices of stock i on days t and $t-1$, respectively, for $i = 1. \dots . n$ and $t = 1. \dots . n$. A total of 2,498 observations are recorded for the returns of each stock.

3.2 Methodology

3.2.1 Cross-correlation and distance matrices

Before construing the MST and the PMFG networks of the S&P 500 Index, we have to calculate the correlation coefficients between any two daily returns in the data set. The Pearson's correlation coefficient between each pair of stocks i and j is computed by the PCC as follows (Mantegna, 1999):

$$\rho_{i.j} = \frac{\sum_{i=1}^{n}(r_i(t)-\bar{r}_i)(r_j(t)-\bar{r}_j)}{\sqrt{\sum_{i=1}^{n}(r_i(t)-\bar{r}_i)^2(r_j(t)-\bar{r}_j)^2}} \tag{2}$$

where $r_i(t)$ and $r_j(t)$ denote the daily returns of stocks i and j, respectively, and $\bar{r}_i$ and $\bar{r}_j$ represent their mean values over the observation period. For both stocks, I calculate the $n \times n$ matrix of correlation coefficients based on daily logarithmic price differences, which closely resemble returns. By definition, the value of $\rho_{i.j}$ ranges from-1 to 1. A negative correlation coefficient ($\rho_{i.j} < 0$) indicates that the two stocks exhibit non-correlated behavior, meaning that when one stock declines, the other tends to rise. Conversely, a positive correlation coefficient ($\rho_{i.j} > 0$) signifies that the stocks move in a positively correlated manner, fluctuating in the same direction. If $|\rho_{i.j}| \approx 0$, the two stocks are considered uncorrelated. In cases where $|\rho_{i.j}| \approx 1$, the stocks demonstrate perfect correlation or perfect non-correlation. The correlation coefficient matrix is symmetric, with $\rho_{i.j} = 1$

along the main diagonal. Subsequently, we convert the correlation coefficient into a metric distance, as the correlation coefficient does not satisfy the three axioms that characterize a metric. The definition of metric distance is as follows:

$$d_{(i.j)} = \sqrt{2(1 - \rho_{(i.j)})} \qquad (0 \leq d_{(i.j)} \leq 1) \tag{3}$$

with $d_{(i.j)}$ satisfying the three axioms of a metric distance: (i) $d_{(i.j)} = 0$ if and only if $i = j$; (ii) $d_{(i.j)}= d_{(j.i)}$; (iii) $d_{(i.j)} \leq d_{(i.k)} + d_{(k.j)}$. The distances calculated between each pair of N stocks result in a symmetric distance matrix of size $N \times N$, referred to as D, which forms the basis for constructing the MST using Prim's algorithm. By using the adjacency matrix D, we can then create the network $N(V.E)$ that represents the stocks, where each stock s_i corresponds to a vertex $v_i \in V$, and an edge $e_{ij} \in E$ indicates the connection between vertices vi and v_j with a corresponding distance $d_{(i.j)}$. The network $N(V.E)$ is defined by the presence of a single edge between any pair of vertices. In scenarios where the portfolio includes a significant number of stocks, the number of edges is huge. Consequently, a simplification of the network is attained by eliminating less significant edges.

3.2.2 Minimum Spanning Tree (MST)

A Minimum Spanning Tree is a concept from graph theory that, when applied to stock market analysis, creates a tree-like structure connecting all stocks in the market with the minimum total correlation distance. This method effectively reduces the complexity of the full correlation matrix while preserving the most important connections. Mantegna (1999) proposed a nonlinear transformation of correlation coefficients to derive weights, defined as $d_{(i.j)}= \sqrt{2(1 - \rho_{(i.j)})}$. This supports the concept of preserving only the most critical edges in a graph. In a network comprising N vertices, MST retains exactly $N - 1$ edges. The hierarchical organization derived from the MST, in conjunction with the distance matrix D, holds considerable economic relevance [2]. We aim to build a connected graph that reflects the correlations among stocks. This network should represent all significant relationships while maintaining simplicity. To achieve this, we will extract a minimum spanning tree (MST) using Prim's algorithm to construct the stock networks.

3.2.3 Planar Maximum Filter Graph (PMFG)

The correlation of stock returns is a fundamental principle of the stock market. Price fluctuations in a few stocks can lead to price fluctuations in other stocks, thereby affecting the overall stability of the market. Consequently, constructing a network based on the correlation of stock returns has gradually become one of the most commonly used methods (Yao & Memon, 2019). The PMFG proposed by Tumminello et al. (2005) maintains the hierarchical structure of MST while incorporating more intricate topological configurations and a greater number of connections (correlation information) between vertices than those found on the MST. While the MST is a widely utilized subgraph, its inherent simplicity may impose certain limitations. To address the need for a more diverse array of network structures, the PMFG has been proposed. This approach relaxes the constraints of a spanning tree in favor of those required for constructing a planar graph (Tumminello et al., 2005). A key distinction between the PMFG and the MST is the number of edges: the PMFG contains $3(N-2)$ edges, while the MST has $N-1$ edges. Additionally, the PMFG incorporates the concept of cliques.

3.2.4 Portfolio Construction

Suppose a portfolio comprises N stocks, and S_0 represents the set of initial values for each stock in the portfolio, denoted as $S_0 = (s_1^0 . s_2^0 . \dots . s_N^0)$. The number of shares for each stock within the portfolio is represented by $X = (x_1 . x_2 . \dots . x_N)$. The initial value of the portfolio, V_0, is calculated as follows:

$$V_0 = x_1 s_1^0 + x_2 s_2^0 + \cdots + x_N s_N^0 = \sum_{i=1}^{N} x_i s_i^0 \quad (4)$$

Our primary concern is capital allocation. The number of shares for each as set is determined by this allocation, expressed as weights $W = (w_1 . w_2 \dots . w_N)$, where

$$\sum_{i=1}^{n} w_i = 1 \tag{5}$$

Each weight is defined as

$$w_i = \frac{x_i s_i^0}{V_0}. \quad i = 1.2. \dots. N \tag{6}$$

At the end of period t, stock values change to $S_t = (s_1^t. s_2^t. \dots. s_N^t)$. This results in a final portfolio value, V_t, which is a random variable:

$$V_t = x_1 s_1^t + x_2 s_2^t + \cdots + x_N s_N^t = \sum_{i=1}^{N} x_i s_i^t \tag{7}$$

The actual return of a portfolio $R_P = (r_1. \dots. r_N)$ represents the set of random returns for each stock in the portfolio. The expected return vector is denoted as $\mu = (\mu_1. \dots. \mu_N)$, where $\mu_i = E(r_i)$ for $i = 1. \dots. N$. The actual return of a multi-asset portfolio over a specific time period is calculated as follows:

$$R_P = w_1 r_1 + w_2 r_2 + \cdots + w_N r_N \tag{8}$$

where w_i represents the weight of each asset in the portfolio. The expected return of a portfolio is determined as the weighted average of the expected returns from each asset within the portfolio. The weight assigned to the expected return of each asset corresponds to the ratio of the market value of that asset to the total market value of the portfolio. Consequently, the expected return $E(R_P) = \mu_P$ for the portfolio at the end of time period t is calculated using the following equation:

$$E(R_P) = w_1 E(r_1) + w_2 E(r_2) + \cdots + w_N E(r_N) = \sum_{i=1}^{N} w_i \mu_i \tag{9}$$

The variance of the portfolio return is expressed as follows:

$$Var(R_P) = E[(R_P - \mu_P)^2] \tag{10}$$

The covariance between asset i and asset j is denoted by $\sigma_{ij} = Cov(r_i. r_j)$, with a special case for $\sigma_{ii} = \sigma_i^2 = Var(r_i)$. These covariances form the entries of the $N \times N$ covariance matrix, denoted by $\Omega_{(n \times n)}$:

$$\Omega_{n\times n}=\begin{pmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2\end{pmatrix} \quad (11)$$

The covariance is calculated as:

$$Cov(r_i.r_j)=E[(r_i-\mu_i)(r_j-\mu_j) \quad (12)$$

3.2.5 Portfolio Optimization

The Markowitz Model, developed by Harry Markowitz, is a security selection process that maximizes returns while managing risk [13]. The fundamental assumption underlying this model is that the returns on securities over a given period are random variables. Consequently, both the mathematical expectation and standard deviation can be calculated, with the standard deviation serving as a measure of investment risk (Ivanova & Dospatliev, 2017). The expected return on the portfolio, denoted as $\mu_P = E(r_P)$, is determined as a linear combination of the expected returns of the individual assets it comprises, influenced by the relative weights of these assets within the portfolio. Investment risk is quantified by the standard deviation, σ_P, which is influenced by the non-linear standard deviations and covariance's of the returns on the individual assets. The explicit form of parametric optimization in the Markowitz model can be represented mathematically as follows:

$$\begin{aligned}&max\,E(r_P)=max\sum_{i=1}^{n}w_i\mu_i\\ &min\sigma_P=min\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}w_iw_j\sigma_{(i.j)}}\ ,\\ &0\le w_i\le 1. \quad i=1.\dots.n\\ &\sum_{i=1}^{n}w_i=1\end{aligned} \quad (13)$$

where w_i represents the percentage of capital to be invested in asset i; r_i denotes the return on asset i; μ_i indicates the expected return on asset i; μ(i,j) refers to the covariance between the returns of assets i and j; $E(r_P)$ represents the expected return of the portfolio; and σ_P signifies the risk associated with the portfolio and possibly additional constraints, such as non-negativity of the weights or specific risk thresholds. The optimization problem is solved using Python, leveraging libraries such as NumPy, SciPy, etc. The result is an optimized weight vector that defines the proportion of each asset in the portfolio.

3.2.6 Sharpe Ratio

The Sharpe ratio, introduced by William Sharpe in 1966, is a risk-adjusted performance metric that is widely employed to assess the efficacy of stock investments. It quantifies the relationship between a portfolio's excess return over a specified target and its overall standard deviation, calculated using the following formula (Sharpe, 1964):

$$S_P = \frac{E(r_p) - r_f}{\sigma_p} \quad (14)$$

In this context, σ_P represents the overall standard deviation of the portfolio, r_f denotes the risk-free interest rate, and S_P signifies the Sharpe ratio. S_P indicates the maximum excess return that an asset portfolio can generate for each additional unit of risk. The metric effectively balances both returns and risks. Furthermore, this indicator will be incorporated into the empirical research segment of this study as a benchmark for assessing the strengths and weaknesses of investment portfolios (Ma, 2023).

3.3 Software and Computational Environment

All computational analyses in this study were conducted using the Python programming language (version 3.12.4) within the Anaconda distribution (2024.06) on a Windows 11 (64-bit) operating system. Data preprocessing, correlation analysis, and network construction were performed using the following Python libraries and their corresponding versions:

Table 1

Computational environment and Python libraries

Methodological Step	Library / Package	Version	Description / Application
Data import and preprocessing	Pandas	2.2.2	Loading Excel files, managing missing values, and handling time-series data of S&P 500 stocks
Numerical computation	Numpy	2.0.1	Performing matrix operations, calculating returns, correlations, and covariance matrices
Visualization	Matplotlib	3.9.2	Plotting network structures (MST, PMFG) and comparative visualizations of correlation layouts
Portfolio and network analysis	riskfolio-lib	6.2.3	Construction and visualization of financial networks (MST, PMFG, DBHT), portfolio modeling, and optimization
Warning management and display formatting	Warnings, pandas. options. display	—	Suppressing runtime warnings and formatting numerical output

Source: Research Findings

4 Discussion and Analysis of Result

4.1 MST and PMFG results

Figure 1 illustrates the results of the Minimum Spanning Tree (MST) analysis employing the Pearson Correlation Coefficient (PCC), as delineated by Mantegna (1999), within the context of the United States stock market over the period from 2013 to 2022. This analysis underscores the interconnectedness of market entities during this timeframe, providing insights into the structural relationships among various stocks based on their comovement. The application of PCC in this context enhances the understanding of market dynamics and offers a robust framework for assessing financial correlations. To analyze the correlation of stocks in the aforementioned figure, one may examine the positioning of companies within clusters and their relative proximity. The MST illustrates the clustering of companies based on various industries, denoted by different colors. As depicted in Figure 1, it is evident that stocks within the same sector, denoted by uniform color, exhibit a clustering pattern.

To analyze the diverse industries and their interconnections based on the colors and positions of circles in the provided MST, the information can be categorized as follows: Specifically, we find that (1) The green cluster primarily includes companies from: public services, consumer products, and energy, (2) Blue cluster of industries: Information Technology (IT), financial services, health and treatment, transportation, (3) Orange cluster of industries: retail, raw materials, industrial services. The Blue Cluster is recognized as the largest cluster due to the presence of major technology and financial companies, and also the Green Cluster also includes large companies in the energy sector.

As presented in Figure 1, AAPL (Apple) in the Blue Cluster is often recognized as a central company due to its strong correlations with several other major companies in the technology and financial sectors. Reasons: (1) High Connectivity: Apple tends to have numerous partnerships and relationships with other companies in its sector. (2) Market Influence: As a leading technology firm, it often has significant market influence, leading to strong correlations with related companies. (3) Network of Collaborations: Companies like Microsoft (MSFT) and other tech firms in the same cluster typically collaborate or compete closely with Apple.

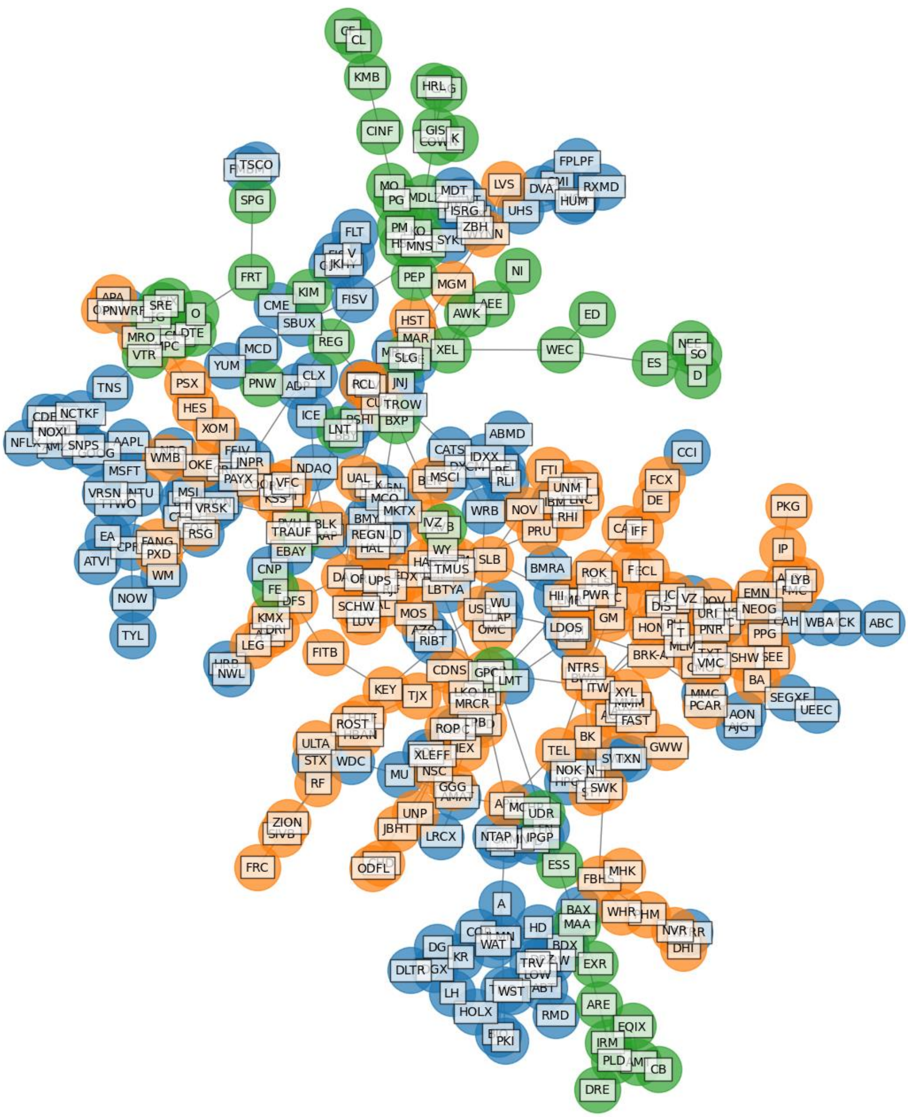

Figure 1. MST-Pearson network of 371 stocks in the S&P 500 index obtained from the Pearson correlation matrix computed by all daily index returns during the period 2013–2022. Source: Research Findings

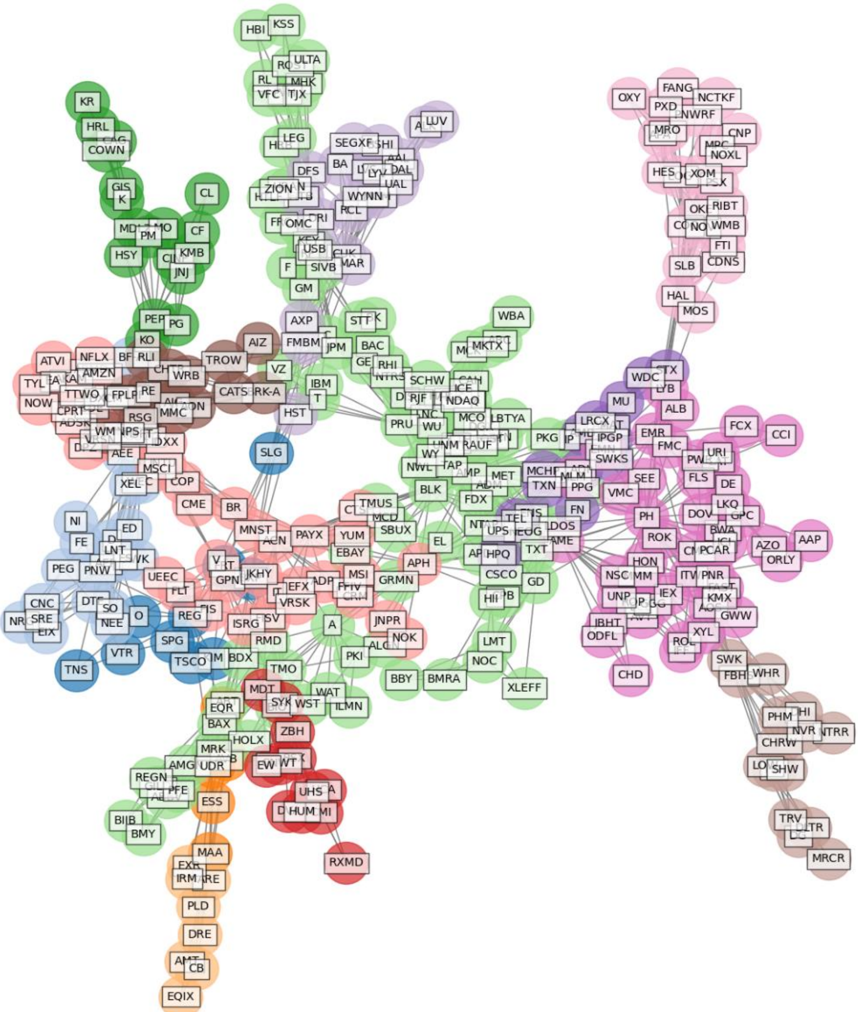

Figure 2. Planar Maximally Filtered Graph (PMFG) (Pearson, DBHT linkage & spring layout) of 371 stocks in the S&P 500 index during the period 2013–2022 for PCC. Source: Research Findings

Figures 2 and 3 show the networks generated from the PMFG-Pearson method using Directed Bubble Hierarchical Tree (DBHT) linkage, employing both the Kamada layout and Spring layout. These networks represent the distance values among 371 stocks in the US stock market, utilizing daily index returns from the period spanning 2013 to 2022. The results indicated that (1) PMFG can maintain the hierarchical structure of MST while featuring more connections, (2) the sectoral clustering effect is more pronounced in PMFG, and (3) the network structures in PMFG are more complex and contain more information. In essence, a network consists of several clusters, with community structure representing one of the most critical frameworks for clustering within the network. Nodes within a community display a high degree of interconnectivity, whereas the connections between distinct communities tend to be relatively sparse. Such community structures are frequently observed in the PMFG networks of financial markets (Wang & Xie, 2015; Song et al., 2011). For instance Buccheri et al. (2013) identified four distinct communities within the PMFG network comprising 49 industrial indices in the U.S. equity markets, spanning the period from 1969 to 2011. Such analysis can aid in investment decision-making and enhances the understanding of cross-stock correlations.

The community structure within financial networks holds substantial practical relevance. For example, a community identified within a stock correlation network comprises a set of stocks whose price movements exhibit similar patterns. Consequently, the price fluctuation of one stock within this community is likely to impact the price movements of other stocks in the same group. This network structure effectively categorizes stocks based on their price changes, providing a valuable framework for market participants in stock selection and portfolio construction. This type of analysis can help identify trends and patterns in the market, allowing investors and analysts to make better decisions and help us understand which stocks are more correlated with each other.

Figure 3. Planar Maximally Filtered Graph (PMFG) (Pearson, DBHT linkage & kamada layout) of 371 stocks in the S&P 500 index during the period 2013–2022. Source: Research Findings

This graph represents a comprehensive analysis of 371 stocks within the US stock market over the period from 2013 to 2022. Figure 4 presents the Planar Maximally Filtered Graph (PMFG), constructed using Pearson correlation, DBHT linkage, and a planar layout. This graph represents a comprehensive analysis of 371 stocks within the US stock market over the period from 2013 to 2022. From Figure 4, we find that: (1) Company RMD is positioned at the top of the PMFG, suggesting it has a central role in this network, which indicates its significant impact in relation to other companies. (2) Companies such as YUM, VRSN, and MNST are clustered together, which indicates a close relationship among them. Additionally, companies like XOM and WMB are also located near each other, suggesting possible collaboration or competition between them. (3) Some companies, like YUM and RMD, have a higher number of connections, indicating they engage more with other companies. Conversely, companies with fewer connections may play a smaller role within this network.

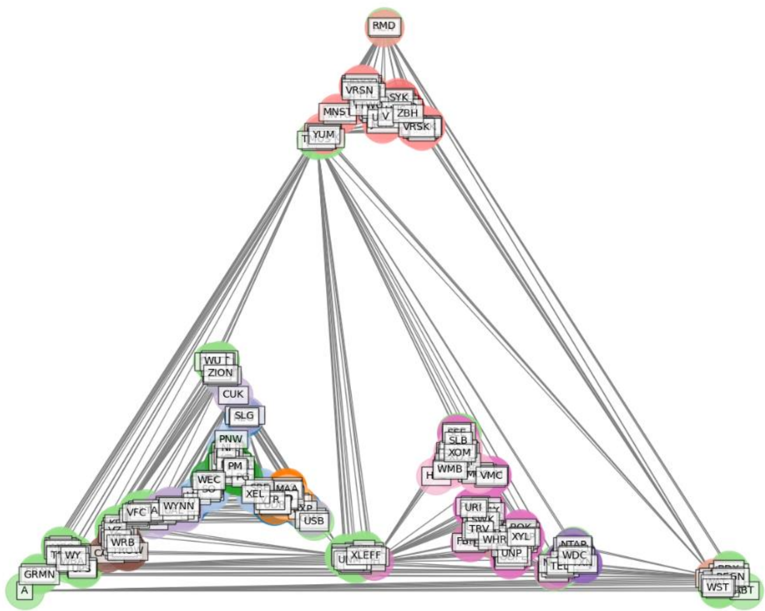

Figure 4. Planar Maximally Filtered Graph (PMFG) (Pearson, DBHT linkage & planar layout) of 371 stocks in the S&P 500 index during the period 2013–2022. Source: Research Findings

This PMFG in Figure 4 represents a complex network of relationships between companies. Company RMD stands out as a central point, while other companies are clustered around it. In general, the applications of MST and PMFG in financial analysis are as follows: (1) Market Structure Analysis: Both MST and PMFG help in understanding the hierarchical structure of markets, revealing how different sectors and stocks are interconnected. (2) Risk Management: By identifying clusters and central nodes, these methods assist in assessing systemic risks and potential contagion effects in times of market stress. (3) Portfolio Optimization: The network structure can guide diversification strategies by selecting stocks from different clusters or based on their centrality measures. (4) Market Dynamics: Tracking changes in MST or PMFG over time can provide insights into evolving market conditions and sector rotations. (5) Anomaly Detection: Unusual changes in the network structure can signal potential market anomalies or upcoming shifts in market behavior.

4.2 Optimization and the Efficient Frontier Analysis

This section presents the results of the stock portfolio optimization method, which determines the efficient frontier for a portfolio of 371 S&P500 stocks from 2013 to 2022. Using the Sharpe ratio for analysis, the portfolio was optimized, leading to the selection of 40 optimal stocks. The Sharpe ratio serves as a vital metric in evaluating investment performance in relation to risk, thereby facilitating the identification of stocks that exhibit the most favorable return-to-risk characteristics. Consequently, these selected stocks are detailed in the accompanying table, along with their corresponding optimal weights. Following the resolution of the optimization equation, the results are presented in Table 2.

Table 2

Expected Return and Risk of the Portfolio with Optimal Weights

Stock	Return	Risk	Optimal weights	Stock	Return	Risk	Optimal weights
ABBV	0.2364	0.2724	0.0170	KR	0.1872	0.2818	0.0393
ABMD	0.4517	0.4621	0.0448	MKTX	0.2682	0.3101	0.0239
AMD	0.1780	0.2520	0.0176	MRCR	0.7998	1.4331	0.0218
ATVI	0.2485	0.3251	0.0040	MSCI	0.3274	0.2889	0.0061
AWK	0.1863	0.2208	0.0171	NCTKF	0.0949	0.3543	0.0377
AZO	0.2293	0.2510	0.0218	NEE	0.2094	0.2229	0.0069
BBY	0.2122	0.3865	0.0739	NFLX	0.4232	0.4775	0.0277
BMRA	0.6743	1.3422	0.0241	NOC	0.2546	0.2375	0.0539
BSHI	0.1557	0.3802	0.0404	NOXL	13.7787	31.6943	0.0008
CNWT	4.2072	5.1133	0.0086	NTRR	1.4441	3.0519	0.0085
DG	0.2171	0.2598	0.0077	ODFL	0.2991	0.2871	0.0036
DPZ	0.2609	0.2857	0.0447	ORLY	0.2617	0.2658	0.0211
DXCM	0.4629	0.4808	0.0295	PNWRF	0.7021	1.2615	0.0199
EA	0.2695	0.3183	0.0247	RLI	0.2125	0.2556	0.0033
EXR	0.2083	0.2455	0.0076	RXMD	2.7076	3.4580	0.0118
FANG	0.3324	0.5148	0.0006	SEGXF	0.1839	0.4058	0.0449
FMBM	0.0993	0.2228	0.1065	TRAUF	0.1137	0.2735	0.0286
FN	0.3135	0.3995	0.0166	UEEC	0.7352	1.0694	0.0324
FPLPF	0.1073	0.6671	0.0131	WM	0.1989	0.1846	0.0536
HSY	0.1592	0.2109	0.0223	XLEFF	1.4965	2.5967	0.0116

Source: Research Findings

Stocks with zero weights were systematically excluded from the portfolio, reflecting a strategic focus on those equities that possess the highest Sharpe ratios. This optimization approach empowers investors to make informed, data-driven decisions, thereby enhancing the overall efficiency of their investment portfolios. Optimal weights indicate the proportion of the portfolio allocated to each stock. Stocks with higher returns might have higher risks, exemplifying the risk-return trade-off. As shown in Table 2, Highest Return: NOXL has a return of 13.7787% with a high risk of 31.6943. Lowest Risk: FMBM presents a return of 0.0933% with a lower risk of 0.2228. Investors can use this information to optimize their portfolios based on their risk tolerance and return. The optimal weights can guide how much capital to allocate to each stock to achieve the desired risk-return profile.

Table 3

Portfolio Optimization, including expected return, expected volatility, and optimal Sharpe ratio

Expected return of the optimal portfolio	Expected volatility of the optimal portfolio	Optimal Sharpe Ratio
0.3761	0.1415	2.6581

Source: Research Findings

In summary, the optimal portfolio is characterized by a favorable expected return of approximately 37.61%, a manageable volatility of about 14.15%, and a very strong Sharpe ratio of 2.6581. This combination of metrics indicates a well-optimized investment strategy that balances risk and return effectively, making it an appealing choice for investors seeking strong performance with a reasonable level of risk. Based on all the obtained results, the efficient frontier curve of the portfolio has been calculated and is presented in Figure 5 as follows:

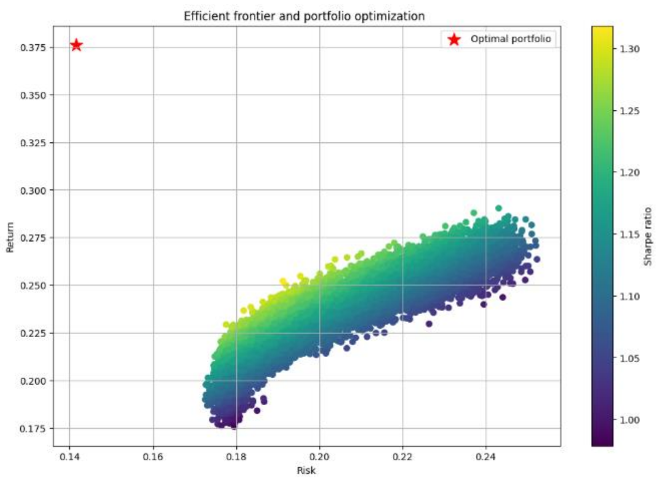

Figure 5. Efficient Frontier of 371 stocks in the S&P 500 index during the period 2013–2022.
Source: Research Findings

5 Conclusion

This study demonstrates the effectiveness of correlation-based network approaches, specifically the MST and PMFG, in capturing the structural organization of the S&P 500 stock market and supporting portfolio optimization under the Markowitz framework using the Sharpe ratio. The empirical analysis, based on daily closing prices of 371 S&P 500 stocks over the period 2013–2022, shows that PMFG preserves more information than MST while maintaining its hierarchical structure. Both methods effectively reveal meaningful market topology, where firms are organized into sectoral clusters and interconnections reflect underlying financial correlations.

The network structure provides several important insights. First, stocks within the same sector tend to exhibit stronger proximity, indicating higher correlation in their price movements. Second, central nodes within the networks highlight systemically important firms that may play a key role in market dynamics and risk transmission. Third, the observed interconnectedness allows for a better understanding of how shocks may propagate across sectors during periods of financial stress.

In addition, the integration of network-based insights with portfolio optimization shows that these structures can improve investment decision-making. By selecting assets across different clusters and considering network centrality, investors can achieve better diversification and more efficient risk-return trade-offs.

The Markowitz-based optimization results confirm the practical value of this framework. The optimized portfolio achieves an expected annual return of 37.61%, a volatility of 14.15%, and a Sharpe ratio of 2.6581, indicating strong risk-adjusted performance. Overall, the findings suggest that MST and PMFG provide a powerful framework for analyzing financial market structure, improving portfolio construction, and enhancing risk management. Future research may extend this approach to multi-scale network analysis or other asset classes such as commodities and emerging markets.

References

Barroso, B. C., Cardoso, R. T., & Melo, M. K. (2021). Performance analysis of the integration between Portfolio Optimization and Technical Analysis strategies in the Brazilian stock market. *Expert Systems with Applications, 186, 115687*.

Birch, J., Pantelous, A. A., & Zuev, K. (2015). The maximum number of 3-and 4-cliques within a planar maximally filtered graph. *Physica A: Statistical Mechanics and its Applications, 417, 221-229*.

Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational statistics & data analysis, 48(2), 431-443*.

Bower, B., & Wentz, P. (2005). Portfolio optimization: MAD vs. Markowitz. *Rose-Hulman Undergraduate Mathematics Journal, 6(2), 3*.

Buccheri, G., Marmi, S., & Mantegna, R. N. (2013). Evolution of correlation structure of industrial indices of US equity markets. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 88(1), 012806*.

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. *Applied soft computing, 100, 106943*.

Chi, K. T., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance, 17(4), 659-667*.

Deng, G. F., Lin, W. T., & Lo, C. C. (2012). Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization. *Expert systems with applications, 39(4), 4558-4566*.

Guo, X., Zhang, H., & Tian, T. (2018). Development of stock correlation networks using mutual information and financial big data. *PloS one, 13(4), e0195941*.

Guo, H., Yu, H., An, Q., & Zhang, X. (2022). Risk analysis of China's stock markets based on topological data structures. *Procedia Computer Science, 202, 203-216*.

Iqbal, J., Sandhu, M. A., Amin, S., & Manzoor, A. (2019). Portfolio selection and optimization through neural networks and markowitz model: A case of pakistan stock exchange listed companies. *Review of Economics and Development Studies, 5(1), 183-196*.

Ivanova, M., & Dospatliev, L. (2017). Application of Markowitz portfolio optimization on Bulgarian stock market from 2013 to 2016. *International Journal of Pure and Applied Mathematics, 117(2), 291-307.*

Lakzaie, F., & Bahiraie, A. (2024). Visualized Portfolio Optimization of stock market: Case of TSE. *Advances in Mathematical Finance and Applications, 2(2), 707.*

Li, Y., Zhuang, X., Wang, J., & Zhang, W. (2020). Analysis of the impact of Sino-US trade friction on China's stock market based on complex networks. *The North American Journal of Economics and Finance, 52, 101185*.

Ma, Z. (2023). Maximizing Returns and Minimizing Risk: A Data-Driven Portfolio Optimization Analysis Using Markowitz's Theory and Sharpe Ratio. *Academic Journal of Business & Management, 5(9),45-53*.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems, 11(1), 193-197*.

Markowitz, H. (1952). Modern portfolio theory. *Journal of Finance, 7(11), 77-91*.

Namaki, A., Shirazi, A. H., Raei, R., & Jafari, G. R. (2011). Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and its Applications, 390(21-22), 3835-3841*.

Ozyesil, M. (2021, January). *Markowitz portfolio optimization model: An application on listed firms on Borsa Istanbul-30 National Stock Index (BIST-30).* In Proceedings of the 3rd International Istanbul Scientific Research Congress (pp. 52–57). Istanbul, Turkey.

Pratama, A. O., Purba, K., Jamhur, J., Prasetyo, T., & Bayu, P. (2020). Pengaruh Faktor Perilaku Investor Saham Terhadap Keputusan Investasi di Bursa Efek Indonesia. Journal Khatulistiwa Informatika, 7(2), 170-179.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance, 19(3), 425-442*.

Song, D. M., Tumminello, M., Zhou, W. X., & Mantegna, R. N. (2011). Evolution of worldwide stock markets, correlation structure, and correlation-based graphs. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 84(2), 026108*.

Stempien, J. P., & Chan, S. H. (2017). Addressing energy trilemma via the modified Markowitz Mean-Variance Portfolio Optimization theory. *Applied Energy, 202, 228-237.*

Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences, 102(30), 10421-10426.*

Tumminello, M., Di Matteo, T., Aste, T., & Mantegna, R. N. (2007). Correlation based networks of equity returns sampled at different time horizons. *The European Physical Journal B, 55(2), 209-217.*

Výrost, T., Lyócsa, Š., & Baumöhl, E. (2019). Network-based asset allocation strategies. *The North American Journal of Economics and Finance, 47, 516-536*.

Wang, G. J., & Xie, C. (2015). Correlation structure and dynamics of international real estate securities markets: A network perspective. *Physica A: Statistical Mechanics and its Applications, 424, 176-193.*

Wang, G. J., Xie, C., & Chen, S. (2017). Multiscale correlation networks analysis of the US stock market: a wavelet analysis. *Journal of Economic Interaction and Coordination, 12(3), 561-594*.

Wang, G. J., Xie, C., & Stanley, H. E. (2018). Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks. *Computational Economics, 51(3), 607-635*.

Yao, H., & Memon, B. A. (2019). Network topology of FTSE 100 Index companies: From the perspective of Brexit. *Physica A: Statistical Mechanics and its Applications, 523, 1248-1262*.

Yusan, B. O., & Riyadi, S. (2024). Portfolio optimization: Application and comparison of markowitz model and single index model on lq45 stocks in indonesia stock exchange. *International Journal of Management Science and Application, 3(1), 57-85.*

Zhao, L., Wang, G. J., Wang, M., Bao, W., Li, W., & Stanley, H. E. (2018). Stock market as temporal network. *Physica A: Statistical Mechanics and its Applications*, *506, 1104-1112*.