IAJ
Indonesian Actuarial Journal
Persatuan Aktuaris Indonesia (PAI)

Vol. 01, No. 02, December (2025)
e-ISSN 3110-6463
https://10.65689/iajvol01no2pp128-136

Identifying Weakly Correlated Dominating Stocks using Maximum Independent Set

Yeremia Bertolla Gimanjar[1], Diah Prastiwi[2], Fendy Septyanto[3*]

[1,3] Division of Economic, Financial, and Actuarial Mathematics, School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia
[2] Department of Mathematics, Faculty of Military Mathematics and Natural Scicences, Indonesia Defence University, Bogor, Indonesia

*Corresponding email: fendy-se@apps.ipb.ac.id

Abstract

In this paper we consider the problem of efficiently finding a (small) set of stocks taken from an index that can replicate the index performance. Furthermore, we add the requirement that the set's returns have weak correlation with each other. Such a selection of stocks may be useful for investors who want to simplify their analysis of the stock index, trying to capture market movement with reduced risk. To solve this problem, we use maximum independent set, a concept from graph theory. As a case study we consider IDX80 in the year 2024.

Article History:
Received: 14 May 2025
First Revised: 17 July 2025
Accepted:21 July 2025
Published: 31 December 2025

Keywords:
Stocks, Investements, Diversification, Portfolio Optimization

1. INTRODUCTION

In Indonesia, the popularity of stocks as an investment vehicle has risen substantially among investors [1], evidenced by a consistent increase in the number of stock investors over recent years. According to the Indonesian Capital Market Statistics published by PT Kustodian Sentral Efek Indonesia (KSEI) in December 2023, the investor base in stocks and other securities expanded by 210% from 2020 to 2023, with an 18.87% increase occurring within 2023 alone. IDX80 is one of several stock indices maintained by PT Bursa Efek Indonesia (BEI). IDX80 measures the stock price performance of 80 stocks listed in IDX Composite with relatively large market capitalization, high liquidity, and good fundamentals, weighted by free float adjusted market capitalization capped at 9% [2].

The large number of different stocks from a number of different sectors may be overwhelming for investors to analyze. There is an incentive to identify a small number of "key stocks" that can capture market movement. The following is an intuitive method to achieve this goal:

1. Start with a stock index, say IDX80.
2. Choose a smaller subset $S = \{s_1, \dots, s_n\}$ of the constituent stocks.
3. Measure how well can S capture market movement. Technically, this can be done by choosing weights $w_1, \dots, w_n$ so that the porfolio return $R_P = w_1R_1 + \cdots + w_nR_n$ approximates the index's performance as close as possible (similar to a linearly regressing index return to the assets' return). Suppose that the closeness is measured by sum of squared error (SSE).

Ideally, we want to find S such that the SSE is as small as possible (or at least small enough). In principle, this can be done by brute force, trying out all possible subsets of the index. However, this is not practical: if the index has m constituent stocks, there are $2^m - 2$ possibility for a non-empty subset that is smaller than the entire index. This number is prohibitively large. For example, IDX80 lists $m = 80$ stocks, and the number of subsets to be checked is $2^{80} - 2 = 1{,}208{,}925{,}819{,}614{,}629{,}174{,}706{,}174$. This number can be reduced by screening the stocks with respect to certain measures of performance (e.g. mean daily return). But the total number of possible subsets can still get very large. The methods we use in this paper will help to significantly reduce the number of subsets/combinations to check.

Instead of brute force, we will use graph theory. Price fluctuations among stocks have complicated relationships [3], considering the stock market as a complex system [4]. Graph theory provides an approach to complex systems with many interacting units [5]. Recent studies have applied graph theory to the stock market by constructing "market graphs", representing a financial network [6]. Peralta [7] argued that optimal portfolio weights are negatively correlated to "centrality" in the financial network. Various notions of "centrality" can be leveraged to create new optimization models. Rafsanjani and Rahimnezhad [8] used closeness centrality, betweenness centrality, and eigenvector centrality to optimize a modified Sharpe ratio. Berouaga et al. [9] used minimum spanning tree (MST) to identify key relationships in the financial network. These studies indicate that many aspects of graph theory are increasingly viewed as useful tools in the investigation of financial markets.

Rational investors pursue the maximization of returns while simultaneously minimizing risks [10]. A viable strategy to reduce risk entails the formation of a diversified stock portfolio, which involves allocating capital across multiple equities to mitigate potential losses in one stock through gains in others [11]. As indicated in [12], stock correlations critically influence the efficacy of diversification in risk management. A high correlation between stocks can cause portfolio risk to remain high despite the number of stocks being held, while a low correlation can reduce portfolio risk. Thus it is important that our selection of stocks is weakly correlated with each other. Prastiwi and Septyanto [13] used graph theory and vertex coloring to produce several sets at once with that property.

We will use the concept of Maximum Independent Set (MIS). Previous research by Hidaka et al. [14] used a quantum-inspired parallel algorithm to find MIS in order to produce a correlation-diversified portfolio. In this paper, we will use a classical recursive algorithm to list all the MIS. While [14] used equal weights and inverse volatility weights, in this paper we use SSE minimizing weights and Sharpe ratio maximizing weights.

The main purpose of this research is to described a method to select dominant stocks in the stock market and to allocate capital to maximize investment efficiency without short position.

2. METHODS

2.1 Data and Processing

For case study we use the stock price data of IDX80 for the year 2024, collected from *Yahoo Finance*. The index is revised quarterly, and we choose only the stocks that are consistently listed in IDX80 throughout the four quarters of 2024. Further screening is done to remove stocks whose mean daily return is lower than the risk free rate.

2.2 Assumptions and Notations

The return of each stock is assumed to be a random variable. The observed return of a stock is computed from adjusted close price of a stock, as a simple return

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \tag{1}$$

where $P_{i,t}$ is the adjusted closing price of stock i at time t, and $R_{i,t}$ is its return.

Given n assets (e.g. stocks), a portfolio is a collection of weights $\boldsymbol{w}^T = (w_1, \dots, w_n)$ with the assumption that $w_1, \dots, w_n \geq 0$ signifying long-only position, and $w_1 + \cdots + w_n = 1$ signifying that all capital is invested in the risky assets. Portfolio return is a random variable $R_P = w_1 R_1 + \cdots + w_n R_n$ with expected return

$$\mu_P = E(R_P) = w_1 E(R_1) + \cdots + w_n E(R_n) = \boldsymbol{w}^T \boldsymbol{\mu} \tag{2}$$

where $\boldsymbol{\mu} = \left(E(R_1), \dots, E(R_n)\right)^T$ is the vector of expected returns of individual stocks. Portfolio variance is

$$Var(R_P) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j Cov(R_i, R_j) = \boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w} \tag{3}$$

where $\boldsymbol{Q} = \left[Cov(R_i, R_j)\right]_{i,j=1}^{n}$ is the covariance matrix. We take the standard deviation $\sigma_P = \sqrt{Var(R_P)}$ as a measure of the stock's volatility or risk.

In portfolio optimization, asset returns are usually assumed normal, so that portfolio return is also normal and hence completely described by just two parameters: mean and variance. In this paper we do not assume normality. Thus, variance is just a partial measure of risk and our assertions about risk shall be understood to refer to only the standard deviation.

Sharpe ratio measures the amount of (excess) return gained for every unit of risk taken. Mathematically,

$$SR = \frac{E(R_P) - r_f}{\sigma_P} = \frac{\boldsymbol{\mu}^T \boldsymbol{w} - r_f}{\sqrt{\boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w}}} \tag{4}$$

where r_f is the risk-free rate. The risk-free rate is derived from the average BI 7 Days Repo Rate during the observation period, taken as nominal and converted to daily effective return. A positive Sharpe ratio implies that on average the portfolio performs better than a risk-free obligation.

2.3 Graphs

We use standard definition and notation of graph theory, such as contained in [15]. A graph is often used to represent a relationship between objects. The objects are represented by vertices/nodes, and a pair of related objects are represented by drawing an edge/line between the pair. For example, **Figure 1** below may represent the relationship between 9 stocks. Supposedly, an edge/line between two stocks indicate that their returns are "highly" correlated. The meaning of "high" is of course relative, exceeding some predetermined treshold. In **Figure 1**, the central stock's return is highly correlated to the other 8 stocks, but the 8 stocks are weakly correlated among themselves (as there are no edges between the 8 stocks).

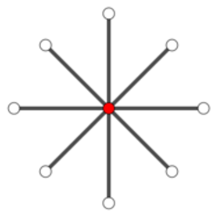

Figure 1. A simple example of graph

2.4 Correlation Graph

We are going to construct a graph from stock return data (only capital gain, ignoring dividend) as follows:

1. Create one vertex for every stock.
2. Compute the correlation matrix, then delete (turn into 0) the diagonal entries.
3. Delete (turn into 0) all correlations below a certain treshold. In this paper, the treshold is the average of pairwise correlation between the stocks' returns.
4. If the (i, j)'th entry in the matrix is non-zero, draw an edge between the i'th and j'th stocks.

Thus, an edge in the correlation graph represents similarity or high correlation.

2.5 Independent Domination

If two vertices are connected by an edge, we say that they are adjacent or neighbours. The degree of a vertex is the number of its neighbours. A set of vertices is ***independent*** if its members are pairwise non-adjacent. A set of vertices is ***dominating*** if every vertex not in the set is adjacent to at least one member of the set. A set of vertices is ***independent dominating*** if it is both an independent and dominating set. For example, the set consisting of 8 "white" vertices in Figure 1 is independent dominating. In the same figure, the set consisting of only 1 "red" vertex is also independent dominating. These concepts were studied in depth in [16].

Proposition A: If an independent set is ***maximal*** (we cannot add any new member without destroying independence) then that set is also dominating.

Proof: Suppose that S is a maximal independent set. If S is not dominating, then by definition there is a vertex v outside of S that does not have any neighbour in S. Then $S \cup \{v\}$ is also independent, contradicting the maximality assumption on S. Therefore, S must be dominating. QED

Finding a maximal independent set can be assisted by a related concept. An independent set is ***maximum*** if its size (the number of its members) is as large as possible. Note that maximal and maximum independent sets are different concepts. For example, the set consisting of 1 "red" vertex in Figure 1 is maximal because if we add any of the white vertices then the set will have adjacent vertices. That set with one member is clearly not maximum, because there is a larger independent set (in terms of size), namely the set consisting of 8 "white" vertices. So, a maximal independent set is not necessarily maximum. However, the converse is true.

Proposition B: If an independent set is maximum (we cannot find any other independent set with larger size) then that set is maximal.

Proof: Suppose that S is a maximum independent set. If S is not maximal, then by definition we can add a vertex v outside of S such that $S \cup \{v\}$ is independent. However, $S \cup \{v\}$ has one more member compared to S, contradicting the assumption that S is maximum. Therefore, S must be maximal. QED

The point of Propositions A and B is that we can find an independent dominating set (our main goal) by finding a maximum independent set. The latter is easier to find, and we shall describe a recursive algorithm for that task. The algorithm is based on the following.

Proposition C: If a vertex is a member of an independent set, then all neighbours of that vertex are excluded from the set.

Proof: Suppose that S is an independent set and v is a member of S. Let w be any neighbour of v. If w is also a member of S, then S will have two adjacent member, contradicting independence. Therefore, w cannot be a member of S. QED

Recursive Algorithm to Find Maximum Independent Set (MIS):

1. If a graph does not have any vertex, define its MIS as the empty set.
2. If a graph does not have any edge, define its MIS as the entire vertex-set.
3. Suppose that a graph $G = (V, E)$ has at least one edge.
   a. Choose a vertex v with the largest degree.
   b. Create a new graph G_1 from G by removing v.
   c. Compute S_1 = MIS(G_1).
   d. Create a new graph G_2 from G by removing v and all of its neighbours.
   e. Compute $S_2 = \{v\} \cup$ MIS(G_2).
   f. Choose the largest among S_1 and S_2 as MIS(G). Output the result as MIS(G).

In general the maximum independent set is not unique, in fact the number of maximum independent sets can be very large, so the above algorithm may give different results each time. However, the algorithm can be slightly modified to produce the complete list of all maximum independent sets. This algorithm is implemented in R using the `ivs` command in the `igraph` library.

2.6 Weight Allocation: Minimizing Sum of Squared Error

Given several candidates (MIS) we choose the best one by measuring how good they can replicate index return. Suppose that we are considering an independent dominating set of stocks $S = \{s_1, \dots, s_n\}$ (in the correlation graph) from a stock index (in this case, IDX80). Let stock s_i have weight w_i and return $R_{i,t}$ at time t. The portfolio return at time t is $R_{P,t} = w_1 R_{1,t} + \cdots + w_n R_{n,t}$. Suppose that the index return at time t is r_t. We seek to minimize the sum of squared error (SSE):

$$SSE(w_1, \dots, w_n) = \sum_{t=1}^{\tau} \left(r_t - R_{P,t}\right)^2 \tag{5}$$

where τ is the number of days of observation, with $w_1, \dots, w_n \geq 0$ and $w_1 + \cdots + w_n = 1$. We transform this to a standard quadratic form, so that numerical solution can be found with `solve.QP` command in `quadprog` library in R.

Proposition D: The SSE is minimized by a weight vector $\boldsymbol{w} = (w_1 \ \cdots \ w_n)^T$ satisfying the following optimization problem

$$\min\left(-\boldsymbol{r}^T \boldsymbol{R} \boldsymbol{w} + \frac{1}{2} \boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{R} \boldsymbol{w}\right) \tag{6}$$

where $\boldsymbol{r} = (r_1 \ \cdots \ r_\tau)^T$ is the vector of the index's daily return and

$$\boldsymbol{R} = \begin{bmatrix} R_{1,1} & R_{2,1} & \cdots & R_{n,1} \\ R_{1,2} & R_{2,2} & \cdots & R_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1,\tau} & R_{2,\tau} & \cdots & R_{n,\tau} \end{bmatrix} \tag{7}$$

is the matrix where every column consists of a stock's daily return, subject to the constraints $\boldsymbol{1}^T \boldsymbol{w} = 1$ (where $\boldsymbol{1}$ is the all 1's column vector) and $\boldsymbol{w} \geq 0$.

Proof: Note that the error function (4) can be written as a matrix product SSE $= \boldsymbol{e}^T \boldsymbol{e}$ where

$$\boldsymbol{e} = \begin{bmatrix} r_1 - R_{P,1} \\ r_2 - R_{P,2} \\ \vdots \\ r_\tau - R_{P,\tau} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_\tau \end{bmatrix} - \begin{bmatrix} R_{P,1} \\ R_{P,2} \\ \vdots \\ R_{P,\tau} \end{bmatrix} = \boldsymbol{r} - \begin{bmatrix} w_1 R_{1,1} + \cdots + w_n R_{n,1} \\ w_1 R_{1,2} + \cdots + w_n R_{n,2} \\ \vdots \\ w_1 R_{1,\tau} + \cdots w_n R_{n,\tau} \end{bmatrix} = \boldsymbol{r} - \boldsymbol{R}\boldsymbol{w}$$

Therefore,

$$\text{SSE} = \boldsymbol{e}^T \boldsymbol{e} = (\boldsymbol{r} - \boldsymbol{R}\boldsymbol{w})^T (\boldsymbol{r} - \boldsymbol{R}\boldsymbol{w}) = (\boldsymbol{r}^T - \boldsymbol{w}^T \boldsymbol{R}^T)(\boldsymbol{r} - \boldsymbol{R}\boldsymbol{w}) = \boldsymbol{r}^T \boldsymbol{r} - \boldsymbol{r}^T \boldsymbol{R}\boldsymbol{w} - \boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{r} + \boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{R}\boldsymbol{w}$$

Note that $\boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{r}$ is a number, so it is equal to its own transpose $\boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{r} = (\boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{r})^T = \boldsymbol{r}^T \boldsymbol{R}\boldsymbol{w}$. Thus,

$$\text{SSE} = \boldsymbol{r}^T \boldsymbol{r} - 2\boldsymbol{r}^T \boldsymbol{R}\boldsymbol{w} + \boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{R}\boldsymbol{w} = \boldsymbol{r}^T \boldsymbol{r} + 2\left(-\boldsymbol{r}^T \boldsymbol{R}\boldsymbol{w} + \frac{1}{2} \boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{R}\boldsymbol{w}\right)$$

Since $\boldsymbol{r}^T \boldsymbol{r}$ is constant, minimizing SSE is equivalent to minimizing $-\boldsymbol{r}^T \boldsymbol{R}\boldsymbol{w} + \frac{1}{2} \boldsymbol{w}^T \boldsymbol{R}^T \boldsymbol{R}\boldsymbol{w}$. QED

2.7 Weight Allocation: Maximizing Sharpe Ratio

After selecting the best MIS by the previous allocation method, we optimize the portfolio performance by maximizing Sharpe ratio. Given a vector of stocks' expected return $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{Q}$, we seek a weight vector $\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{n}}$ satisfying the following constrained optimization problem

$$\max SR(\boldsymbol{w}) = \max \frac{\boldsymbol{\mu}^T \boldsymbol{w} - r_f}{\sqrt{\boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w}}} \tag{8}$$

$$\boldsymbol{1}^T \boldsymbol{w} = 1, \qquad \boldsymbol{w} \geq \boldsymbol{0}$$

where $\boldsymbol{1} = (1,1,\dots,1)^T$ and $\boldsymbol{0} = (0,0,\dots,0)^T$ are $n \times 1$ vectors.

Exact solution to (8) is difficult to obtain analytically due to the non-negativity constraint. Similar to the previous section, we transform this to a quadratic optimization.

Proposition E: Consider the problem of finding $(\boldsymbol{x}, k) \in \mathbb{R}^n \times \mathbb{R}$ (so $\boldsymbol{x} \in \mathbb{R}^n$ is a vector and $k \in \mathbb{R}$ is a number) satisfying the following optimization problem

$$\min \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \tag{9}$$

$$\left(\boldsymbol{\mu} - r_f \boldsymbol{1}\right)^T \boldsymbol{x} = 1, \qquad \boldsymbol{1}^T \boldsymbol{x} - k = 0, \qquad \boldsymbol{x} \geq 0, \qquad k \geq 0.$$

If $(\boldsymbol{x}_0, k_0)$ is an optimal solution of (9), then $\boldsymbol{w}_0 = \frac{x_0}{k_0}$ is a optimal solution of (8).

Proof. Suppose that $(\boldsymbol{x}_0, k_0)$ is an optimal solution to (9). It can be checked that $\boldsymbol{w}_0 = \frac{x_0}{k_0}$ belongs to the feasible set of (8), namely $\mathbf{1}^T\boldsymbol{w}_0 = 1$ and $\boldsymbol{w}_0 \geq 0$. Moreover,

$$SR(\boldsymbol{w}_0) = \frac{\boldsymbol{\mu}^T\boldsymbol{w_0} - r_f}{\sqrt{\boldsymbol{w}_0^T\boldsymbol{Q}\boldsymbol{w}_0}} = \frac{\frac{\boldsymbol{\mu}^T\boldsymbol{x_0}}{k_0} - r_f}{\sqrt{\frac{\boldsymbol{x}_0^T}{k_0}\boldsymbol{Q}\frac{\boldsymbol{x}_0}{k_0}}} = \frac{\boldsymbol{\mu}^T\boldsymbol{x}_0 - r_f k_0}{\sqrt{\boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{x}_0}} = \frac{\boldsymbol{\mu}^T\boldsymbol{x}_0 - r_f\mathbf{1}^T\boldsymbol{x}_0}{\sqrt{\boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{x}_0}} = \frac{\left(\boldsymbol{\mu} - r_f\mathbf{1}\right)^T\boldsymbol{x}_0}{\sqrt{\boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{x}_0}} = \frac{1}{\sqrt{\boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{x}_0}}$$

Let $\boldsymbol{w}$ be any vector in the feasible set for (8), so $\mathbf{1}^T\boldsymbol{w} = 1$ and $\boldsymbol{w} \geq 0$. If $SR(\boldsymbol{w}) \leq 0$ then $SR(\boldsymbol{w}_0) > 0 \geq SR(\boldsymbol{w})$. Now assume $SR(\boldsymbol{w}) > 0$, so that $\boldsymbol{\mu}^T\boldsymbol{w} - r_f > 0$. Define

$$k \coloneqq \frac{1}{\left(\boldsymbol{\mu} - r_f\mathbf{1}\right)^T\boldsymbol{w}} = \frac{1}{\boldsymbol{\mu}^T\boldsymbol{w} - r_f\mathbf{1}^T\boldsymbol{w}} = \frac{1}{\boldsymbol{\mu}^T\boldsymbol{w} - r_f} > 0$$

and $\boldsymbol{x} \coloneqq k\boldsymbol{w}$. It can be checked that $(\boldsymbol{x}, k)$ is in the feasible set of (9):

- $\left(\boldsymbol{\mu} - r_f\mathbf{1}\right)^T\boldsymbol{x} = \left(\boldsymbol{\mu} - r_f\mathbf{1}\right)^T k\boldsymbol{w} = k\left(\boldsymbol{\mu} - r_f\mathbf{1}\right)^T\boldsymbol{w} = 1,$
- $\mathbf{1}^T\boldsymbol{x} = \mathbf{1}^T k\boldsymbol{w} = k\mathbf{1}^T\boldsymbol{w} = k.$

Similar to $SR(\boldsymbol{w}_0) = \frac{1}{\sqrt{\boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{x}_0}}$, we also have $SR(\boldsymbol{w}) = \frac{1}{\sqrt{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}}}$. Since $(\boldsymbol{x}_0, k_0)$ is an optimal solution to (9), we have $\boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{x}_0 \leq \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$ so $SR(\boldsymbol{w}_0) \geq SR(\boldsymbol{w})$, proving that $SR(\boldsymbol{w}_0)$ is maximum. QED

In general, the constraints of (9) are not always satisfied. For example, when $\boldsymbol{\mu} - r_f\mathbf{1}$ has some negative entries (i.e. when some stocks have negative Sharpe ratios), the first constraint $\left(\boldsymbol{\mu} - r_f\mathbf{1}\right)^T\boldsymbol{x} = 1$ may necessitate some x_i to be negative. However, when $\boldsymbol{\mu} - r_f\mathbf{1} > 0$ i.e. when all entries of $\boldsymbol{\mu} - r_f\mathbf{1}$ are positive, this problem is avoided. This provides another reason for removing stocks with negative Sharpe ratio at the screening process.

3. RESULT AND DISCUSSION

There are 71 stocks consistently listed in IDX80 in all four quarters of 2024. Of these 71 stocks, 28 stocks have positive Sharpe ratio. **Figure 2** shows the correlation graph of the 28 stocks. An edge indicates "high" (above average) correlation. From this figure we can already glean some insights. For example, HEAL is directly connected to very few other stocks (only 3), so HEAL is highly correlated to only a small number of other stocks; therefore, HEAL can safely be included in most portfolio without significant risk. On the other hand, PGAS is directly connected to many stocks (20 out of 27 other stocks), so inclusion of PGAS in a portfolio will likely increase the portfolio's risk.

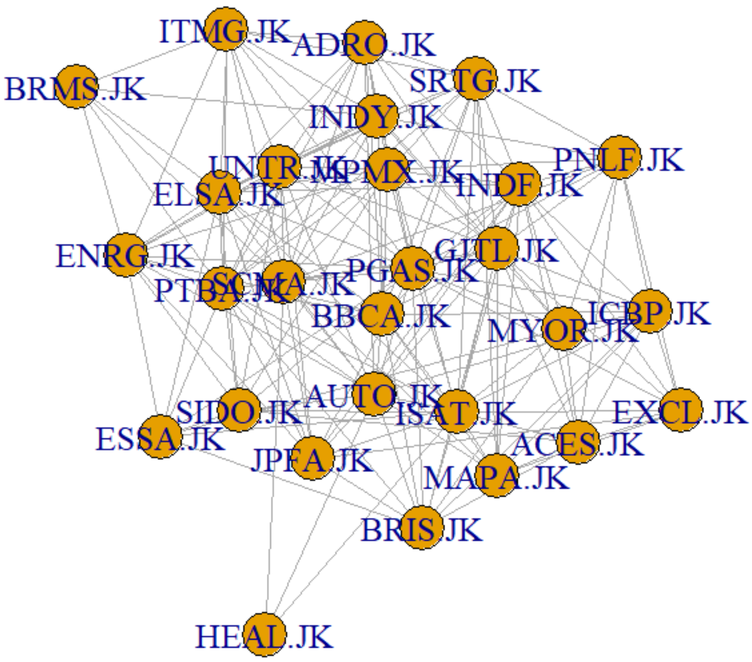

Figure 2. Correlation graph of 28 stocks in IDX80 2024 with positive Sharpe ratio

After running the modified recursive algorithm in Section 2.5, we discovered 4 different maximum independent sets, each consisting of 7 members:

Set 1: ADRO, BRIS, BRMS, HEAL, MAPA, PNLF, SIDO

Set 2: ADRO, BRMS, ESSA, HEAL, JPFA, MAPA, PNLF

Set 3: ADRO, BRMS, HEAL, JPFA, MAPA, MYOR, SIDO

Set 4: ADRO, BRMS, HEAL, JPFA, MAPA, PNLF, SIDO,

The following table shows the minimum SSE for each set, together with the corresponding weights.

Table 1. Minimizing SSE for each set

Weight Allocation								Minimum SSE
Set 1	ADRO 11.31%	BRIS 18.92%	BRMS 9.37%	HEAL 13.39%	MAPA 12.21%	PNLF 12.69%	SIDO 22.11%	0.0205
Set 2	ADRO 12.68%	BRMS 10.45%	ESSA 11.64%	HEAL 16.46%	JPFA 17.90%	MAPA 16.98%	PNLF 13.90%	0.0252
Set 3	ADRO 8.32%	BRMS 7.80%	HEAL 11.42%	JPFA 14.19%	MAPA 9.96%	MYOR 27.29%	SIDO 21.02%	0.0202
Set 4	ADRO 10.35%	BRMS 8.56%	HEAL 13.64%	JPFA 16.71%	MAPA 13.06%	PNLF 13.22%	SIDO 24.45%	0.0229

Among the four sets, Set 3 has the smallest minimum SSE. Therefore, we choose Set 3 as the constituents of our portfolio. Its members have weak correlation among themselves, and any stock outside the set is highly correlated with at least one member. Set 3 serves as a well diversified group of stocks that can capture market movement. The following figure compares Set 3's return (with the SSE-minimizing weights) with index return.

The two returns have a positive correlation of 0.5423. The SSE-minimum portfolio has a mean daily return of 0.15%, a cumulative 1-year return of 42.50%, standard deviation of 0.98%, and Sharpe ratio of 0.1319.

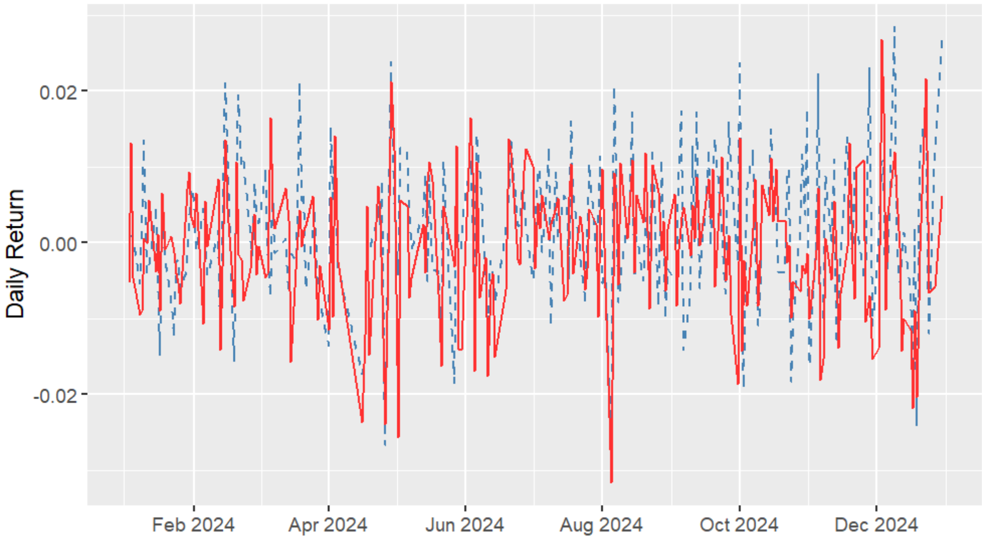

Figure 3. Comparing IDX80 return (solid red line) with Set 3 return (dashed blue line)

Finally, we recalculate the weights to maximize Sharpe ratio. The maximum Sharpe ratio portfolio has a mean daily return of 0.24%, a cumulative 1-year return of 72.56%, standard deviation of 1.30%, and Sharpe ratio of 0.1648.

Table 2. Maximum Sharpe ratio weights of Set 3

Weight Allocation							
Set 3	ADRO 19.80%	BRMS 20.36%	HEAL 4.12%	JPFA 28.02%	MAPA 14.79%	MYOR 7.75%	SIDO 5.16%

4. CONCLUSIONS

In this paper, we have considered the stocks from IDX80, filtered for consistency in the inclusion during all quarters of 2024, and positive Sharpe ratio. We obtained an independent dominating set of 7 stocks, namely ADRO, BRMS, HEAL, JPFA, MAPA, MYOR, SIDO. These stocks have weak correlation among themselves, but every other stock is highly correlated with at least one of them. We calculated the weights to minimize SSE, and we have observed graphically and numerically that the 7 stocks can replicate the index performance quite well. Then we recalculated the weights to maximize Sharpe ratio instead.

The methods of this paper can be generalized to other assets, not just stocks. The graph theoretic technique of maximum independent set (MIS) can be applied to any other "similarity" network, not just based on correlation.

5. REFERENCES

[1] V. Verina, F. Cuandra. (2022). Analisis tingkat kesehatan dan tingkat pengembalian saham perbankan. *SEIKO: Journal of Management & Business*. 5(1):188-197.

[2] PT. Bursa Efek Indonesia. (2021). *IDX Handbooks*.

[3] W. Q. Huang, X. T. Zhuang, S. Yao. (2009). A network analysis of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications*. 388(14):2956-2964.

[4] R. N. Mantegna. (1999). Hierarchical structure in financial markets. *The European Physical Journal B*. 11:193–197.

[5] J.P. Onnela, K. Kaski, J. Kertész. (2004). Clustering and information in correlation-based financial networks. *The European Physical Journal B*. 38:353-362.

[6] V. Kalyagin, A. Koldanov, P. Koldanov, V. Zamaraev. (2014). Market graph and Markowitz model. In: T. M. Rassias, C. A. Floudas, S. Butenko, editors. (2014). *Optimization in Science and Engineering: In Honor of the 60th Birthday of Panos M. Pardalos*. New York (NY): Springer. 293-306.

[7] G. Peralta. (2016). A Network Approach to Portfolio Selection. [S.l.] : SSRN.

[8] Z. H. Rafsanjani, F. Rahimnezhad. Portfolio Selection by Optimizing Risk and Return Based on Complex Network Analysis (Case Study: Tehran Stock Exchange). *International Journal of Industrial Engineering & Management Science*. 8(1):1-15.

[9] Y. Berouaga, C. El Msiyah, J. Madkour. (2023). Portfolio Optimization Using Minimum Spanning Tree Model in the Moroccan Stock Exchange Market. *Int. J. Financial Stud*. 2023, 11, 53.

[10] M. Ramadhan, T. Suharti, I. Nurhayati. (2020). Diversifikasi saham dalam pembentukan portofolio untuk meminimumkan risiko. *Manager: Jurnal Ilmu Manajemen*. 3(4):450-458.

[11] Z. Bodie, A. Kane, A. J. Marcus. (2011). *Investments 9th Ed*. New York (NY): McGraw Hill.

[12] N. Safelia. (2012). Konsep dasar keputusan investasi dan portfolio. *Jurnal Manajemen Terapan dan Keuangan*. 1(3):217-226.

[13] D. Prastiwi, F. Septyanto. (2024). Pewarnaan simpul untuk seleksi aset pada pembentukan portofolio investasi saham. MILANG. 20(2):111-121.

[14] R. Hidaka, Y. Hamakawa, J. Nakayama, K. Tatsumura. (2023). Correlation diversified portfolio construction by finding maximum independent set in large scale market graph. IEEE Access. 11: 142979–142991.

[15] G. Chartrand, P. Zhang. (2012). *A First Course in Graph Theory*. Mineola, New York: Dover Publications.

[16] E. J. Cockayne and S. T. Hedetniemi. Towards a theory of domination in graphs. *Networks*, 7:247–261, 1977.