

Signed network models for portfolio optimization

Bibhas Adhikari

Fujitsu Research of America, Inc.
badhikari@fujitsu.com
Santa Clara, CA, USA

Abstract. In this work, we consider weighted signed network representations of financial markets derived from raw or denoised correlation matrices, and examine how negative edges can be exploited to reduce portfolio risk. We then propose a discrete optimization scheme that reduces the asset selection problem to a desired size by building a time series of signed networks based on asset returns. To benchmark our approach, we consider two standard allocation strategies: Markowitz’s mean–variance optimization and the $1/N$ equally weighted portfolio. Both methods are applied on the reduced universe as well as on the full universe, using two datasets: (i) the Market Champions dataset, consisting of 21 major S&P500 companies over the 2020–2024 period, and (ii) a dataset of 199 assets comprising all S&P500 constituents with stock prices available and aligned with Google’s data. Empirical results show that portfolios constructed via our signed network selection perform as good as those from the classical Markowitz model and the equal-weight benchmark in most occasions.

Keywords: signed networks, hedge, portfolio

1 Introduction

The framework of combinatorial graphs or networks serves as a powerful mathematical tool across a variety of data analysis techniques. In financial applications, networks play a central role in modeling dependencies among assets via their correlation strengths. By representing assets as vertices and encoding correlations as (weighted) edges, numerous methods have been developed for tasks such as asset-price prediction and risk analysis [11] [34] [3] [10] [23] [40] [19]. Compared to purely statistical approaches, network analysis offers the advantage of capturing both pairwise interactions and higher-order group dynamics among assets. Several surveys and monographs explore the role of networks in finance and economics more broadly [28] [26] [1].

On the other hand, a signed graph augments a standard graph by assigning each edge a sign - positive or negative. When vertices represent random variables, a positive (resp. negative) edge indicates that the corresponding variables are positively (resp. negatively) correlated. For a comprehensive review of signed-graph theory and its applications, see Zaslavsky’s annotated bibliography of

recent developments [43]. Structural balance theory, which hinges on the sign-configuration of triangles, is fundamental in the study of signed social networks [9] [21]. Triangles are classified by the number of negative edges they contain: if T_j denotes a triangle with j negative edges for $j = 0, 1, 2, 3$, then T_0 and T_2 are balanced, whereas T_1 and T_3 are unbalanced (see Figure 1(a)). A signed graph is called *balanced* if its vertex set can be partitioned into two subsets such that every positive edge lies within a subset and every negative edge connects vertices across subsets [21]. Empirical evidence shows that real-world signed networks are typically unbalanced, inspiring various measures to quantify this lack of balance [2] [41] [14].

Harary et al. [22] introduced the notion of balance signed graphs for well-structured equities portfolios that could contain risk in the portfolio. In their model, assets are considered as vertices, and the existence of positive and negative edges in the corresponding signed graph is defined by the correlation between returns of the associated pair of assets. Thus the edges indicate the tendency or manner in which the value of the assets change relative to each other. A positive edge between a pair of assets reflects that the valuation of the assets tend to move in tandem, whereas a negative edge implies that the valuations of the assets move in opposite direction, if one goes up the other goes down. Following the idea of Harary et al., a number of articles considered to investigate financial markets through signed graph models and vice versa, for instance see [25] [2] [15] [17] [44] and the references therein. Recently, in [4], the authors show that the global balance index of financial correlation networks can be used as a systemic risk measure. We note that, even though weighted correlation networks are considered in several context in the literature, weighted signed network models for financial networks are rare to find [37].

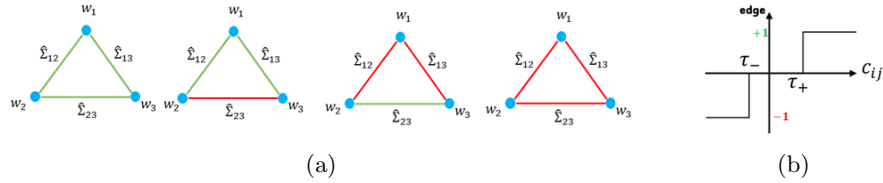


Fig. 1: (a) Triangles in a signed graph, T_0, T_1, T_2, T_3 (from left to right) the green edges are positive and red edges are negative (b) Threshold function [22] for signed network formation. c_{ij} denotes the covariance or correlation strength for the assets i and j

Despite its elegance, Markowitz's portfolio construction is plagued in practice by estimation errors in the covariance matrix and expected return. Consequently, an optimized portfolio based on an estimated covariance matrix will almost surely deviate from the true Markowitz solution [18]. To reduce the dimension of the problem, we propose a discrete optimization problem by incorporating

co-movement statistics of asset returns over all times $t \in \{1, \dots, T\}$, thus generalizing Kendall’s Tau [27]. Our two-step framework for designing a diversified, hedge-protected portfolio is as follows:

1. **Dimensionality reduction via signed-graph models.** We construct a time series of signed graphs on the asset set by comparing each pair’s returns to their own sample means over a rolling window of length T . An edge is assigned a positive sign if both returns lie on the same side of their means, and a negative sign otherwise. Negative edges therefore capture hedge relationships, instances where returns move in opposite directions, directly targeting variance reduction without explicit covariance estimation. We then score assets by the frequency with which they exhibit negative edges against others, and select the top candidates by maximizing these weighted counts.
2. **Final allocation on the reduced universe.** Having selected a smaller asset subset, we apply any standard allocation method such as Markowitz’s mean–variance model and the $1/N$ rule [13] to compute the investment weights.

By filtering the investment universe in the first step, we reduce problem size while retaining hedge-relevant information, and then leverage established optimization techniques on this subset with a desired number of assets. To demonstrate our findings on real financial data, we consider two datasets. The first is the Market Champions dataset from Kaggle , which contains daily stock prices for 21 prominent S&P 500 companies across multiple sectors, covering the period from January 1, 2020 to December 31, 2024. The second dataset is from Kaggle and consists of 199 S&P 500 stocks with price data aligned with Google’s dataset. We evaluate the performance of our proposed method through backtesting and observe that it performs comparably to both the standard Markowitz optimization and the $1/N$ equally weighted strategy applied to the full set of assets in the financial market.

Finally, note that quantum computing constitutes a fundamentally novel paradigm for portfolio optimization. A spectrum of quantum algorithmic frameworks including quantum annealing, variational quantum algorithms such as the Quantum Approximate Optimization Algorithm (QAOA) has been employed to address Markowitz’s mean–variance problem [24] [33] [30] [42]. These approaches are intrinsically designed for large-scale instances, however, their implementation on fault-tolerant hardware remains a future prospect. In contrast, Noisy Intermediate-Scale Quantum (NISQ) platforms enable empirical assessment of both purely quantum and hybrid quantum-classical algorithms on moderately sized portfolios [8]. By integrating our dimension-reduction methodology with the operational capacity of NISQ devices, we show a promising pathway for the practical realization of hybrid quantum-classical portfolio optimization.

2 Financial markets as signed graphs

In this section, we consider weighted signed graph models for representing financial markets using correlation matrices. Since the actual correlation between

the returns is unobserved, the correlation is often estimated by employing several statistical estimators [35]. Denoting the unobserved covariance matrix as Σ for a random vector $\mathbf{R} = (R_1, \dots, R_N)$, we denote an estimator of Σ as $\hat{\Sigma} = [\hat{\Sigma}_{ij}]$, where $\hat{\Sigma}_{ij} = \text{Cov}(R_i, R_j) = E[(R_i - \mu_{R_i})(R_j - \mu_{R_j})]$ denotes the estimated covariance corresponding to the random variables R_i and R_j . Here, $\mu_X = E[R]$, the expected value of the random variable R . In financial time-series data, let $R_i(t)$ denote the random variable corresponding to an index associated with the asset i at time t (for example, a day or month or year). Then a popular unbiased estimator for Σ is the sample covariance matrix, whose entries are defined by $\hat{\Sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^T (r_i^t - \mu_{R_i})(r_j^t - \mu_{R_j})$, where $R_i(t) = r_i^t$ and $R_j(t) = r_j^t$, $\mu_R = \frac{1}{T} \sum_{t=1}^T r^t$, and $t \in \{1, \dots, T\}$ with T is the total time window. The sample correlation coefficient matrix is then defined as $\hat{\rho} = [\hat{\rho}_{ij}]$, with $\rho_{ij} = \text{Cov}(R_i, R_j) / \sqrt{\text{Var}(R_i)\text{Var}(R_j)}$, where $\text{Var}(X) = \frac{1}{T-1} \sum_{t=1}^T (x^t - \mu_X)^2$ is nonzero, and $\hat{\rho}$ estimates the population Pearson correlation matrix. Note that $-1 \leq \hat{\rho}_{ij} \leq 1$ with $\hat{\rho}_{ij} = 1$ if $i = j$. If $\hat{\rho}_{ij} > 0$ then the random variables X_i and X_j are said to be positively correlated and they are negatively correlated if $\hat{\rho}_{ij} < 0$.

For financial time-series data, such as in stock market, let $S_n(t)$ denote the random variable for the price of the n -th stock at time t . Then the random variable $R_n(t)$ which represents return of the n -th stock for a fixed time horizon Δt is defined as: $(S_n(t + \Delta t) - S_n(t)) / S_n(t)$ (Linear return) or $\log S_n(t + \Delta t) - \log S_n(t)$ (log return). Often the value of Δt is considered as 1. For Markowitz's portfolio theory applications, a correlation coefficient estimator matrix must be non-singular, and hence positive definite. We mention here that there are other powerful methods to model the return time-series, such as the GARCH process introduced by Bollerslev [6], a generalization of the ARCH process proposed by Engle in [16].

However, it is demonstrated in literature that for finite time-series data i.e. when $T < \infty$, there is a random offset to every correlation coefficient and these values are dressed up with noise [20], it can be validated by comparing eigenvalue density of a correlation matrix to a random matrix [32]. An important observation from the financial data is that the effect of noise strongly depends on the ratio N/T , where N is the size of the portfolio and T the length of the available time series [39], see also [29][12].

The weighted signed graph $G^s(\hat{\Sigma}_D)$, which represents a model financial market associated with a (denoised) correlation estimator matrix $\hat{\Sigma}_D = [\hat{\Sigma}_{ij}^D]$, is defined as follows.

Definition 1. (*Weighted signed graph models of financial markets*) The vertex set of $G^s(\hat{\Sigma}_D)$ is the set of assets in a portfolio index by $1, 2, \dots, N$. Then the edge set $E \subseteq V \times V$ is defined by the two following ways.

1. Without thresholding: there is an edge between a pair of vertices (i, j) if and only if $\hat{\Sigma}_{ij}^D \neq 0$. The sign of an edge (i, j) is positive if $\hat{\Sigma}_{ij}^D > 0$ and negative if $\hat{\Sigma}_{ij}^D < 0$. The weight of the edge is $\hat{\Sigma}_{ij}^D$.

2. *With thresholding: let $0 < \tau_+ < 1$ and $-1 < \tau_- < 0$. Then There is a positive edge for the vertex pair (i, j) with weight $\hat{\Sigma}_{ij}^D$ if $\hat{\Sigma}_{ij}^D > \tau_+$ and a negative edge for the vertex pair (i, j) with weight $\hat{\Sigma}_{ij}^D$ if $\hat{\Sigma}_{ij}^D < \tau_-$.*

A signed graph representation of a financial market is the underlying signed graph obtained by relaxing the edge weights of a weighted signed portfolio graph. This can be achieved in two ways: directly from the estimated correlation matrix with thresholding and from the denoised correlation matrix. In both cases, the threshold function may or may not be applied. In [22], Harary et al. considered using a threshold function directly from the estimated correlation matrix as described in Figure 1 (b). As they explained, the edges in the normalized market graph represent the tendency of the return values of the associated assets (vertices).

In social signed network systems, structural balance theory plays a pivotal role to investigate the dynamics of the underlying systems and it is believed that social networks evolve toward balance, however it may not be true in all real-world social networks [14]. It is also demonstrated using real-world data that the number of unbalance triangles T_1 and T_3 is significantly lesser than the number of balance triangles T_0 and T_2 . In financial normalized networks, if the edges represent tendencies of going up or down of the return values, then for a triangle of type T_3 of three assets X, Y, Z would mean the following: if the return of X goes up then the returns of Y or Z must go down (due to the negative edges (X, Y) and (X, Z)), however if both the return values of Y and Z go down then they must be positively correlated which contradicts the fact that they are negatively correlated. Thus, the crucial point here is the rates of going up or down of the pairs of return values, which are decided by the correlation values. A similar argument can also be given for the existence of T_1 type triangles. Thus we conclude that structural properties of financial (unweighted) signed networks is strikingly different from social signed networks.

Now we establish from the viewpoint of containing risk that negative edges in a signed graph representation plays an important role to contain portfolio risk than a portfolio with all positive edges (positively correlated assets). We consider the variance of the portfolio as a measure of risk from the perspective of Markowitz's portfolio theory (MPT) [36] [38]. According to MPT, for a diversified portfolio, an investor's goal is to minimize the portfolio variance where the minimum-variance portfolio problem can be written as $\min_w w^\dagger \hat{\Sigma} w$ such that $\mathbf{1}^\dagger w = 1$, where $\hat{\Sigma}$ is the risky assets' (return) estimated covariance matrix, $w = [w_1, \dots, w_N]^T$, $w_j \geq 0$ is the vector of portfolio weights i.e. the proportion of wealth invested in the assets, and $\mathbf{1}$ is the all-one vector of dimension N , the number of total number of assets in the portfolio. The condition $w_j \geq 0$ means that the portfolio does not contain any short positions. Then we have the following theorem.

Theorem 1. *Let $w = [w_1, \dots, w_N]^\dagger$ with $w_i \geq 0$ and $\sum_{i=1}^N w_i = 1$. Suppose $G^s(\hat{\Sigma})$ is the underlying (weighted) signed graph with at least one negative edge. Then $w^\dagger \hat{\Sigma} w \leq w^\dagger |\hat{\Sigma}| w$, where $|\hat{\Sigma}| = [|\hat{\Sigma}_{ij}|]$*

Proof. The proof follows from the fact that

$$w^\dagger \widehat{\Sigma} w = \sum_{i=1}^N w_i^2 \widehat{\Sigma}_{ii} + \sum_{\substack{i \neq j \\ i,j=1}}^N 2 \text{sign}(\widehat{\Sigma}_{ij}) |\widehat{\Sigma}_{ij}| w_i w_j.$$

Triangles constitute a fundamental motif, prevalent both in social networks and in correlation-based financial networks. In the context of portfolio construction, a natural question then arises: which triangle configurations in the weighted signed graph of asset returns most effectively contribute to variance reduction and thus help contain risk? From Figure 1 (a), it is easy to verify that $w^\dagger \widehat{\Sigma} w$ is minimum for the (unbalance) triangle T_3 when $w_i > 0$ for all i . Indeed, when short selling is allowed, it becomes a different story. In the case of short selling, the portfolio weights corresponding to assets which hold short positions are considered negative. Then in Figure 1 (a), considering $w_1 < 0$ and w_2, w_3 to be positive it follows that the unbalance triangle T_1 achieves the minimum portfolio risk. However, changing the assignment of signs of the edges but keeping the balance/unbalance property of the triangles fixed, the minimum risk could be achieved by a different type of triangle.

3 Signed network based hedge-protected portfolio formation

Theorem 1 affirms that negative edges act like hedges in a portfolio, as defined in [5]. Now note that the sample covariance of return values of a pair of assets is given by $\widehat{\Sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^T (R_i^t - \mu_{R_i})(R_j^t - \mu_{R_j})$ for a time period T , where R_k^t denotes the return of asset k at time t , and μ_{R_k} is the mean of the return values of the asset k for the time period T . If $\widehat{\Sigma}_{ij} < 0$, it indicates that one of the assets had a few ‘bad days’ compare to its own mean return value than the other asset in terms of their return values, although for the other days their return values could be at per compare to their own mean return values. Whereas, if $\widehat{\Sigma}_{ij} > 0$ then it would mean that they have the same ‘bad days’ and ‘good days’ i.e. return values of both the assets go up or down together corresponding to their own mean return values in most of the days or the values go up or down quite deep together on a few days compare to the days when pairwise go in opposite directions making a pair (up,down) or (down,up). In an extreme case, one “very good” or “very bad” day of either or both the assets can flip the sign of $\widehat{\Sigma}_{ij}$ from positive to negative or vice-versa. By compressing these finer co-movement patterns into $\widehat{\Sigma}$, the Markowitz mean-variance formulation masks this local return dynamics. This interpretation applies equally to raw and denoised (or thresholded) covariance estimators; henceforth, “covariance matrix” refers to either form.

Recall that the original mean-variance model (OMV) model is formulated as

$$\text{OMV1: } w^* = \arg \min_{w \in \Delta_K} w^\dagger \widehat{\Sigma} w \text{ s.t. } \mu^\dagger w = \epsilon \quad (1)$$

$$\text{or OMV2: } w^* = \arg \min_{w \in \Delta_K} -\mu^\dagger w + \gamma w^\dagger \widehat{\Sigma} w, \quad (2)$$

where μ is the mean vector consists of the means of the asset returns and $\Delta_K = \{w \in \mathbb{R}_{\geq 0}^N : \sum_{i=1}^N w_i = 1\}$ [31]. Thus Markowitz's model recommends formation of portfolio to ensure some level of ϵ (also called target return) of portfolio return $\mu^\dagger w$ and minimizing the portfolio variance given by equation (1), and simultaneously maximizing the return and minimizing the portfolio variance simultaneously with a mixing parameter (also called *risk aversion parameter*) $\gamma \in (0, \infty)$ in equation (2). Thus, both the models urge to gain more return and withstand less risk. From the computational perspective, note that this is a convex optimization problem and efficient methods are available to solve such optimization problems. Indeed, considering $w \in \mathbb{R}^N$ (with short selling), the analytical solution of problem OMV2 is given by [7]

$$w^* = \frac{1}{2\gamma} \widehat{\Sigma}^{-1}(\mu + \nu^* \mathbf{1}), \quad \nu^* = \frac{2\gamma - \mathbf{1}^\dagger \widehat{\Sigma}^{-1} \mu}{\mathbf{1}^\dagger \widehat{\Sigma}^{-1} \mathbf{1}}. \quad (3)$$

In the weighted graph representation of a portfolio, we observe that a negative edge helps to reduce portfolio risk. As proved in Theorem 1, for any invest allocation vector, the risk can be contained more by having negative edges (negatively correlated assets) than positively correlated edges (positively correlated assets) of equal strengths. Observing this, we define *hedge score* of an asset in a portfolio by introducing a time-series of portfolio graphs for a time period T and the negative degrees of the vertices as follows.

We define a normalized market signed graph $G_t^s(\mu, \mathbf{R}_N) = (V, E_t)$ of N assets at a time $t \in \{1, \dots, T\}$ with $V = \{1, \dots, N\}$ as the set of assets, μ is the mean return vector of the assets and $\mathbf{R}_N = (R_1^t, \dots, R_N^t)$ is the observed empirical return values. The edge set is defined as follows. For a pair of assets (i, j) there is a positive edge if $(R_i^t - \mu_{R_i})(R_j^t - \mu_{R_j}) \geq 0$ and a negative edge if $(R_i^t - \mu_{R_i})(R_j^t - \mu_{R_j}) < 0$. Then based on the statistics of negative degree (the number of negative edges at a vertex is adjacent to) of an asset, we have the following definition preserving Markowitz's model through the proposed time-varying normalized market graph representation.

Definition 2. (*Hedge score*) Let $S_n^t : V \setminus \{n\} \rightarrow \{0, 1\}$ be a function $S_n^t(j) = 1$ if $(R_n^t - \mu_{R_n})(R_j^t - \mu_{R_j}) < 0$ and $S_n^t(j) = 0$ otherwise, where $t \in \{1, \dots, T\}$. Then the hedge score of an asset n is defined as

$$H(n, T) = \frac{\sum_{j \in V} \sum_{t=1}^T S_n^t(j)}{T(N-1)}. \quad (4)$$

Note that S_n^t counts the negative degree of the vertex n in the graph $G_t^s(\mu, \mathbf{R}_N)$. Besides, $0 \leq H(n, T) \leq 1$. Then we propose the following optimization problem

for selecting a potential subset of assets for the design of a hedge-protected diversified portfolio as follows.

$$\text{OPT: } \arg \max_{S \subseteq V} \sum_{n \in S} H(n, T) \mu_{R_n}(T) = \arg \max_{S \subseteq V} H_S^\dagger(T) \mu_S(T), \quad (5)$$

where the mean value of the returns of an asset i for a time period T and N is the total number of assets in the market. For a set of assets $S \subseteq V$, denote the $H_S(T) = [H(s_1, T), \dots, H(s_{|S|}, T)]^\dagger$ and $\mu(S, T) = [\mu_{R_{s_1}}(T), \dots, \mu_{R_{s_{|S|}}}(T)]^\dagger$ as the column vectors of hedge scores and mean return values respectively, of the assets $s_k \in S$ within a time period T . Note the theoretical maximum value of equation (5) would be given by the complete graph $G_t^s(\boldsymbol{\mu}, \mathbf{R}_N)$ will all edges are negative for all t , however, from financial data such a graph can never be realized for moderate size value of $|S|$.

We mention that the time complexity of solving the optimization problem is $O(N \log N)$, which follows from the fact that the indices $H(k, T) \mu_k(T)$ can be stored in an array of length N and the optimizer S is then obtained by sorting this array. We could add a constraint $|S| = K \leq N$ to equation (5) and the value of K could be decided by the investor's input and by performing portfolio risk analysis of the potential choices. Overall, this optimization method significantly reduces the dimension of investment allocation problem. Once the set S of assets is determined by solving the optimization problem, the invest allocation vector w can be chosen by employing methods such as $1/|S|$ method or the original Markowitz's mean-variance method.

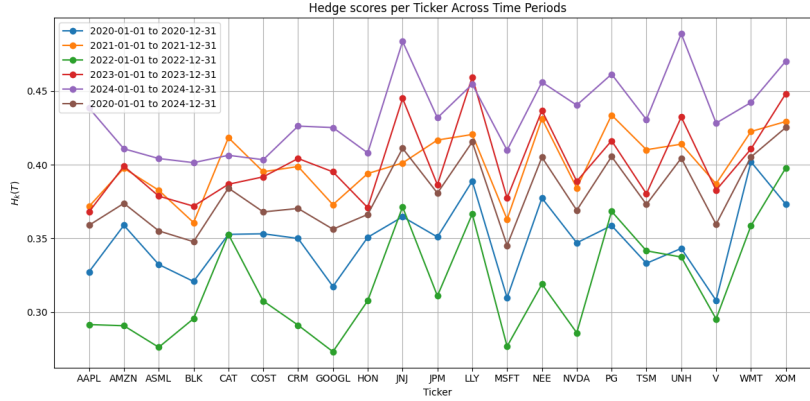


Fig. 2: Hedge scores of all tickers during 2020 to 2024

4 Empirical analysis

To test the proposed methodology for portfolio construction, we consider two datasets. First we consider a data of S&P500 index from January 1, 2020, to December 31, 2024, available online in Kaggle. This data contains stock prices of major companies, called Market Champions: Leading Stocks Dataset, from different sectors including Technology & AI: Apple (AAPL), Microsoft (MSFT), Alphabet (GOOGL), Amazon (AMZN), NVIDIA (NVDA), Taiwan Semiconductor (TSM), Healthcare: Johnson & Johnson (JNJ), UnitedHealth Group (UNH), Eli Lilly (LLY), Energy: ExxonMobil (XOM), NextEra Energy (NEE), Financial: JPMorgan Chase (JPM), Visa (V), BlackRock (BLK). Consumer: Walmart (WMT), Costco (COST), Procter & Gamble (PG), Industrial: Caterpillar (CAT), Honeywell (HON). Software/Cloud: Salesforce (CRM), ASML Holding (ASML).

Table 1: The stock selection based on the proposed optimization method for the Market Champions dataset, setting $K \in \{5, 8, 12, 15\}$. The set $\mathcal{S}_K$ of assets for a year 20XX is obtained by using the data of all the 252 days in the year 20XX.

Year	$\mathcal{S}_5$	$\mathcal{S}_8$	$\mathcal{S}_{12}$	$\mathcal{S}_{15}$
2020	AAPL, AMZN, ASML, NVDA, TSM	$\mathcal{S}_5$ with BLK, CRM, MSFT	$\mathcal{S}_8$ and COST, GOOGL, LLY, NEE	$\mathcal{S}_{12}$ and CAT, UNH, WMT
2021	ASML, GOOGL, LLY, NVDA, XOM	$\mathcal{S}_5$ and COST, MSFT, UNH	$\mathcal{S}_8$ and AAPL, BLK, JPM, NEE	$\mathcal{S}_{12}$ and CAT, CRM, PG
2022	CAT, HON, LLY, UNH, XOM	$\mathcal{S}_5$ and JNJ, V, WMT	$\mathcal{S}_8$ with COST, JPM, NEE, PG	$\mathcal{S}_{12}$ and AAPL, BLK, CAT
2023	AMZN, CRM, GOOGL, LLY, NVDA	$\mathcal{S}_5$ and AAPL, COST, MSFT	$\mathcal{S}_8$ and ASML, JPM, TSM, V	$\mathcal{S}_{12}$ and BLK, CAT, WMT
2024	AMZN, JPM, NVDA, TSM, WMT	$\mathcal{S}_5$ and AAPL, COST, GOOGL	$\mathcal{S}_8$ with BLK, CAT, CRM, LLY	$\mathcal{S}_{12}$ and MSFT, NEE, V

Forming the signed graphs $G_t^s(\mu, \mathbf{R}_N)$, for each day t the hedge-score for each asset is calculated following equation (4) for the time period T , which is considered a year such as 2020, 2021, 2022, 2023, 2024, and for the entire period January 2020 to December 2024 in Figure 2. Based on these hedge score statistics of all the assets and setting $|\mathcal{S}| = K \in \{5, 8, 12, 15\}$, we determine the potential subset $\mathcal{S}_K$ of 21 assets in the market by solving equation (5) for each time period T , described in Table 1.

Next we consider the dataset of the entire collection of assets whose data are aligned with the Google stock in S&P500 index from August 2004 to Dec 2022. This data forms a universe of 199 assets, available in Kaggle. The assets are given by 'A', 'AAP', 'ABMD', 'ABT', 'ACN', 'ADI', 'ADM', 'ADP', 'ADSK', 'AJG', 'AKAM', 'ALB', 'ALGN', 'ALK', 'AMAT', 'AMD', 'AME', 'AMGN',

'AMT', 'AMZN', 'AOS', 'APA', 'APD', 'ARE', 'ATVI', 'AVY', 'BAC', 'BAX', 'BBY', 'BDX', 'BEN', 'BIIB', 'BIO', 'BRK-A', 'BSX', 'BWA', 'BXP', 'CAG', 'CB', 'CCI', 'CDE', 'CHD', 'CHRW', 'CINF', 'CLX', 'CMI', 'CNC', 'COO', 'COP', 'CPB', 'CPRT', 'CRM', 'CSCO', 'CTAS', 'CTSH', 'CUK', 'D', 'DGX', 'DOV', 'DPZ', 'DVA', 'EA', 'EBAY', 'ECL', 'EFX', 'EL', 'EMN', 'ES', 'EW', 'EXR', 'FAST', 'FIS', 'FISV', 'FITB', 'FLS', 'FMC', 'FTI', 'GGG', 'GILD', 'GIS', 'GOOG', 'GPC', 'GPN', 'GWW', 'HAS', 'HBAN', 'HD', 'HES', 'HRB', 'HRL', 'HST', 'HSY', 'HUM', 'IDXX', 'IFF', 'ILMN', 'ISRG', 'ITW', 'IVZ', 'JBHT', 'JCI', 'JKHY', 'JNPR', 'JPM', 'K', 'KIM', 'KMB', 'KSS', 'LEG', 'LH', 'LNC', 'LNT', 'LOW', 'MAA', 'MAR', 'MCHP', 'MCO', 'MDLZ', 'MLM', 'MMC', 'MOS', 'MSFT', 'NEE', 'NEOG', 'NFLX', 'NI', 'NOC', 'NOV', 'NTAP', 'NTRS', 'NVR', 'NWL', 'O', 'ODFL', 'OMC', 'ORLY', 'OXY', 'PAYX', 'PCAR', 'PH', 'PHM', 'PKG', 'PKI', 'PLD', 'PNW', 'PPG', 'PRU', 'PVH', 'RCL', 'REG', 'RF', 'RHI', 'RLI', 'ROK', 'ROL', 'ROP', 'SBUX', 'SCHW', 'SEE', 'SHW', 'SIVB', 'SLB', 'SLG', 'SNPS', 'SO', 'SPG', 'SRE', 'STT', 'SWK', 'SYK', 'T', 'TJX', 'TMO', 'TRV', 'TSCO', 'TSN', 'TTWO', 'TXT', 'TYL', 'UDR', 'URI', 'VFC', 'VMC', 'VRSN', 'VZ', 'WAT', 'WBA', 'WDC', 'WEC', 'WHR', 'WM', 'WMB', 'WRB', 'WST', 'WYNN', 'XEL', 'YUM', 'ZBH', 'ZION'. As in the preceding dataset, we determine potential asset sets $\mathcal{S}_n$ for $n = 20$ and 50 based on solving the optimization problem stated in equation (5). These scores are computed from the signed graphs $G_t^s(\boldsymbol{\mu}, \mathbf{R}_N)$ constructed for each day t over a time period T , where T corresponds to one year for each of the years from 2005 to 2022. We report the reduced universe obtained using the proposed optimization scheme in Table 3.

We employ the backtesting method for analyzing the performance of the proposed method and the results are given in Table 2 and Table 4. We determine the optimized portfolio allocation vector solving the Markowitz's with short selling and no short selling optimization problems as described by OMV1 and OMV2 respectively. Using the data of the previous year we form the portfolios and test the performance of these portfolios using standard statistics for its performance for the next year. For instance, we use the stock data of 2020 to form the portfolio and test the performance by finding the Total return (%), annual return (%), annual volatility (%), and the Sharpe ratio using the data of 2021. Then we employ these methods to construct portfolio from the set of assets $\mathcal{S}_K$ and calculate the above mentioned statistics for these portfolios. We consider the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ as the sample mean and sample variance of the data. For OMV1, we set the target return (ϵ) as the maximum of the mean return values for the Market Champions data of 21 leading stocks, and average of the 75-quartile mean returns for the S&P500 data of 199 assets. We also derive the above mentioned statistics for the portfolio allocation vector using the $1/N$ method, we call the associated portfolio as equally weighted portfolio (EWP). We observe that the proposed dimension reduction technique along with either Markowitz or EWP gives better results compare to employing these methods on all assets of the entire market in several occasions in both the datasets.

Table 2: Comparison of backtesting results for the stock market data. Proposed Method (PM), Markowitz’s Portfolio with short selling (MP), Markowitz’s Portfolio with no short selling (MPNS), Equally Weighted Portfolio (EWP).

Year	Method	Total Return		Annual Return		Annual Volatility		Sharpe Ratio	
		$K = 5$	$K = 8$	$K = 5$	$K = 8$	$K = 5$	$K = 8$	$K = 5$	$K = 8$
2021	PM+MP	-50	298.38	-61	153.56	39.94	53.99	-1.54	2.84
	PM+MPNS	102.97	102.97	81.09	81.09	44.77	44.77	1.81	1.81
	PM+EWP	35.81	33.94	34.31	31.87	26.72	22.47	1.28	1.42
	MP	-7.98		-7.25		14.80		-0.49	
	MPNS	102.97		81.09		44.77		1.81	
	EWP	29		26.55		13.91		1.91	
2022	PM+MP	-58.59	14.39	-74.77	17.19	53	27.02	-1.41	0.64
	PM+MPNS	-60.26	-60.26	-72.90	-72.90	63.23	63.23	-1.15	-1.15
	PM+EWP	-18.49	-17.90	-15.10	-15.69	33.23	28.98	-0.45	-0.54
	MP	-15.05		-14.71		18.57		-0.79	
	MPNS	-60.26		-72.90		63.23		-1.15	
	EWP	-19.76		-19.05		25.07		-0.76	
2023	PM+MP	-17.52	-7.15	-17.13	-6.77	21.73	12.16	-0.79	-0.56
	PM+MPNS	-8.97	-8.97	-6.40	-6.40	24.99	24.99	-0.26	-0.26
	PM+EWP	11.84	9.82	12.38	10.17	14.56	11.81	0.85	0.86
	MP	-14.46		-15.04		12.42		-1.21	
	MPNS	-8.97		-6.40		24.99		-0.26	
	EWP	29.81		27.27		13.09		2.08	
2024	PM+MP	-100	-26.39	-434.13	-22.38	368.12	41.09	-1.18	-0.54
	PM+MPNS	149.38	149.38	105.72	105.72	52.10	52.10	2.03	2.03
	PM+EWP	53.91	44.42	46.40	38.97	24.07	19.43	1.93	2.01
	MP	23.60		21.77		8.99		2.42	
	MPNS	149.38		105.72		52.10		2.03	
	EWP	28.84		26.31		12.35		2.13	

We also observe an interesting phenomena while solving OMV1 for the Market Champions dataset. From Table 2, note that the values of all the statistics obtained for Markowitz’s no short selling method (MPNS) match with the corresponding statistics obtained by employing our proposed method combining hedge scores of the assets and the MPNS. This happened due to the fact that the naive MPNS for all the assets construct the same portfolio as $\mathcal{S}_5$. Thus we see a deep connection of hedge scores with the solution of MPNS that makes our proposed dimension reduction technique using hedge statistics important and opens an avenue for future research.

Table 3: The stock selection based on the proposed optimization method, setting $K \in \{20, 50\}$ from the S&P500 dataset of 199 stocks. The set $\mathcal{S}_K$ of assets for a year 20XX is obtained by using the data of all the 252 days in the year 20XX.

Year	reduced universe
2005	$\mathcal{S}_{20}$: 'ISRG', 'NFLX', 'GOOG', 'CRM', 'NOV', 'HUM', 'WDC', 'CCI', 'HES', 'GPN', 'AKAM', 'ILMN', 'SLB', 'AMT', 'GILD', 'WMB', 'WRB', 'AAP', 'OXY', 'FLS' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'APA', 'MLM', 'FTI', 'AMD', 'MCO', 'COP', 'TSCO', 'DPZ', 'ORLY', 'A', 'PRU', 'BEN', 'EFX', 'IDXX', 'CHRW', 'RHI', 'ROP', 'DVA', 'FAST', 'SLG', 'VMC', 'ATVI', 'CB', 'SCHW', 'IVZ', 'AMGN', 'PHM', 'SRE', 'MCHP', 'URI'
2006	$\mathcal{S}_{20}$: 'ILMN', 'AKAM', 'ALGN', 'WST', 'SLG', 'ALB', 'WYNN', 'TYL', 'CSCO', 'PVH', 'IVZ', 'ABMD', 'VFC', 'BXP', 'TMO', 'CTSH', 'KSS', 'T', 'IFF', 'MOS' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'ES', 'NTAP', 'MAR', 'FMC', 'LH', 'SHW', 'AMT', 'FTI', 'HAS', 'CAG', 'PCAR', 'LNT', 'ADM', 'KIM', 'CPB', 'NEE', 'UDR', 'MLM', 'MDLZ', 'CPRT', 'SNPS', 'VMC', 'CMI', 'WAT', 'SCHW', 'SPG', 'REG', 'VZ', 'HST', 'ACN'
2007	$\mathcal{S}_{20}$: 'MOS', 'ISRG', 'NOV', 'AMZN', 'HES', 'CMI', 'NEOG', 'FTI', 'FLS', 'ATVI', 'CRM', 'JNPR', 'SLB', 'APA', 'WAT', 'OXY', 'IDXX', 'VRSN', 'BWA', 'ILMN' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'WDC', 'GOOG', 'TXT', 'PH', 'ADM', 'GILD', 'AME', 'FMC', 'CPRT', 'WMB', 'HUM', 'BRK-A', 'APD', 'SYK', 'IVZ', 'CCI', 'COP', 'YUM', 'SCHW', 'TMO', 'MLM', 'ALGN', 'BIO', 'ROL', 'BAX', 'PCAR', 'PKG', 'CHD', 'NEE', 'CHRW'
2008	$\mathcal{S}_{20}$: 'AMGN', 'ODFL', 'EW', 'ALK', 'HRB', 'HAS', 'NFLX', 'GILD', 'RLI', 'ABMD', 'AJG', 'GIS', 'WRB', 'CHRW', 'TSCO', 'SHW', 'CHD', 'PHM', 'WM', 'JBHT' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'DGX', 'LOW', 'SO', 'TYL', 'ROL', 'ABT', 'ORLY', 'WST', 'BAX', 'GWW', 'O', 'ADP', 'NEOG', 'VMC', 'ACN', 'LEG', 'MMC', 'FAST', 'AAP', 'HSY', 'HD', 'AOS', 'NVR', 'DVA', 'MAA', 'CLX', 'CB', 'ILMN', 'WEC', 'TRV'
2009	$\mathcal{S}_{20}$: 'AMD', 'WDC', 'AMZN', 'NTAP', 'SBUX', 'CTSH', 'ISRG', 'CRM', 'FTI', 'CCI', 'COO', 'ALGN', 'CDE', 'NFLX', 'SLG', 'PVH', 'GOOG', 'A', 'WHR', 'MOS' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'DPZ', 'TJX', 'TYL', 'EMN', 'WAT', 'FLS', 'EW', 'RCL', 'NOV', 'AKAM', 'ADI', 'GPN', 'NVR', 'SIVB', 'PKG', 'EBAY', 'IVZ', 'PRU', 'MSFT', 'JCI', 'BEN', 'CMI', 'HST', 'WBA', 'ALB', 'SPG', 'APD', 'IDXX', 'MCHP', 'KSS'
2010	$\mathcal{S}_{20}$: 'NFLX', 'URI', 'ILMN', 'CMI', 'BWA', 'EW', 'DPZ', 'AKAM', 'ZION', 'HBAN', 'TSCO', 'CRM', 'RCL', 'NEOG', 'AAP', 'ODFL', 'EL', 'ALK', 'NTAP', 'WYNN' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'ORLY', 'COO', 'PH', 'CTSH', 'CDE', 'PVH', 'PCAR', 'ROL', 'AME', 'FTI', 'FITB', 'ADSK', 'TSN', 'HAS', 'NOV', 'HST', 'MAR', 'ROK', 'EXR', 'UDR', 'ALB', 'ROP', 'FAST', 'HRL', 'AMZN', 'FMC', 'SBUX', 'YUM', 'GWW', 'SHW'
2011	$\mathcal{S}_{20}$: 'DPZ', 'ABMD', 'BIIB', 'ISRG', 'HUM', 'TJX', 'VFC', 'CNC', 'TSCO', 'EL', 'TYL', 'SBUX', 'FAST', 'HSY', 'RLI', 'ORLY', 'NI', 'CHD', 'HRB', 'WMB' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'ALK', 'EXR', 'GWW', 'URI', 'EA', 'TSN', 'COO', 'CPRT', 'D', 'SPG', 'SO', 'MCO', 'ODFL', 'WRB', 'YUM', 'CTAS', 'ABT', 'MDLZ', 'ALGN', 'HD', 'FTI', 'CAG', 'KMB', 'WEC', 'LNT', 'AMGN', 'AMT', 'XEL', 'NEE', 'GIS'
2012	$\mathcal{S}_{20}$: 'PHM', 'WHR', 'BAC', 'ILMN', 'GILD', 'SHW', 'CRM', 'EBAY', 'CCI', 'TYL', 'EMN', 'RF', 'URI', 'PVH', 'PPG', 'PKG', 'EXR', 'AOS', 'PKI', 'HD' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'DVA', 'NEOG', 'WST', 'LOW', 'AMZN', 'MCO', 'FLS', 'WDC', 'EFX', 'NWL', 'AMGN', 'TJX', 'AMT', 'NFLX', 'NVR', 'JBHT', 'TMO', 'COO', 'AME', 'FIS', 'FISV', 'TXT', 'SRE', 'DPZ', 'TSCO', 'FMC', 'BAX', 'RCL', 'BIIB', 'VMC'
2013	$\mathcal{S}_{20}$: 'NFLX', 'BBY', 'ABMD', 'BSX', 'ILMN', 'GILD', 'ALGN', 'TYL', 'BIIB', 'WDC', 'SEE', 'SIVB', 'WST', 'LNC', 'TSN', 'TSCO', 'SCHW', 'ALK', 'VFC', 'ATVI' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'WYNN', 'TMO', 'AOS', 'NOC', 'EA', 'URI', 'PKG', 'AMD', 'HUM', 'PRU', 'GOOG', 'TTWO', 'WBA', 'ADM', 'AAP', 'HRB', 'HES', 'AMZN', 'HAS', 'NEOG', 'TJX', 'FIS', 'DPZ', 'AMAT', 'FLS', 'VRSN', 'CNC', 'ODFL', 'MCO', 'BWA'

2014	$\mathcal{S}_{20}$: 'EW', 'EA', 'CNC', 'RCL', 'ILMN', 'ALK', 'TTWO', 'MAR', 'ODFL', 'ORLY', 'ABMD', 'ARE', 'AAP', 'ISRG', 'SHW', 'IDXX', 'AMAT', 'EXR', 'REG', 'HUM' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'LOW', 'RHI', 'AMGN', 'WBA', 'WEC', 'PNW', 'XEL', 'DPZ', 'LEG', 'UDR', 'LNT', 'WDC', 'ES', 'NI', 'AKAM', 'O', 'COO', 'CHRW', 'URI', 'BXP', 'DGX', 'SLG', 'NVR', 'NEE', 'CTAS', 'NOC', 'SPG', 'SRE', 'APD', 'KIM'
2015	$\mathcal{S}_{20}$: 'ABMD', 'NFLX', 'AMZN', 'ATVI', 'TYL', 'GPN', 'EXR', 'EA', 'HRL', 'GOOG', 'SBUX', 'VRSN', 'VMC', 'BSX', 'TSN', 'ALK', 'EFX', 'AOS', 'NVR', 'CRM' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'ORLY', 'CNC', 'RLI', 'HUM', 'CUK', 'HAS', 'TTWO', 'NOC', 'FISV', 'JKHY', 'EW', 'JNPR', 'HD', 'MLM', 'RCL', 'UDR', 'CLX', 'MDLZ', 'AVY', 'PKI', 'ROL', 'MAA', 'CPB', 'ALGN', 'DPZ', 'ROP', 'NI', 'NEOG', 'CAG', 'MSFT'
2016	$\mathcal{S}_{20}$: 'CDE', 'AMD', 'AMAT', 'IDXX', 'MLM', 'ZION', 'ALB', 'CMI', 'RF', 'DPZ', 'SIVB', 'URI', 'ALGN', 'ODFL', 'TTWO', 'WST', 'FMC', 'APA', 'CPRT', 'BBY' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'PH', 'MCHP', 'WM', 'FITB', 'JBHT', 'NTAP', 'PKG', 'BAC', 'ABMD', 'PCAR', 'VMC', 'SYK', 'BIO', 'JPM', 'LNC', 'COO', 'ROL', 'ITW', 'HES', 'ROK', 'ADI', 'WYNN', 'DGX', 'PRU', 'CINF', 'CTAS', 'AKAM', 'SNPS', 'ADM', 'AJG'
2017	$\mathcal{S}_{20}$: 'ALGN', 'TTWO', 'NVR', 'WYNN', 'PHM', 'CNC', 'ATVI', 'ILMN', 'ISRG', 'ABMD', 'FMC', 'EL', 'BBY', 'MAR', 'AVY', 'AMAT', 'AMZN', 'NTAP', 'GGG', 'PVH' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'URI', 'MCO', 'CPRT', 'NFLX', 'ODFL', 'VRSN', 'BAX', 'SHW', 'ABT', 'CRM', 'SWK', 'AME', 'ALB', 'GPN', 'SNPS', 'A', 'RCL', 'HD', 'WAT', 'VFC', 'PKG', 'ROL', 'ROK', 'PH', 'PKI', 'ROP', 'MCHP', 'AMT', 'ADSK', 'SIVB'
2018	$\mathcal{S}_{20}$: 'AMD', 'ABMD', 'AAP', 'ORLY', 'CHD', 'DPZ', 'BSX', 'NFLX', 'EW', 'VRSN', 'ILMN', 'CRM', 'AMZN', 'ISRG', 'GWW', 'ABT', 'KSS', 'ADSK', 'HRL', 'AJG', $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'TJX', 'IDXX', 'MSFT', 'RLI', 'ROL', 'COO', 'NEE', 'TMO', 'HUM', 'COP', 'AMT', 'ADP', 'O', 'CNC', 'YUM', 'FISV', 'SBUX', 'TSCO', 'CSCO', 'AMGN', 'CPRT', 'MOS', 'ECL', 'JKHY', 'FIS', 'CLX', 'NTAP', 'CTAS', 'WEC', 'EXR'
2019	$\mathcal{S}_{20}$: 'AMD', 'CDE', 'CPRT', 'AMAT', 'TSN', 'GPN', 'MLM', 'NVR', 'WDC', 'TYL', 'CAG', 'BBY', 'MCO', 'CPB', 'FISV', 'HES', 'BIO', 'SNPS', 'EL', 'PLD' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'DOV', 'AMT', 'EW', 'CTAS', 'ODFL', 'SO', 'PHM', 'FMC', 'WST', 'NEE', 'URI', 'VMC', 'MSFT', 'HSY', 'MAA', 'DVA', 'APD', 'SHW', 'SRE', 'ACN', 'ZBH', 'EFX', 'AKAM', 'WEC', 'GIS', 'VFC', 'IDXX', 'ARE', 'TMO', 'AVY'
2020	$\mathcal{S}_{20}$: 'AMD', 'ABMD', 'WST', 'ALB', 'IDXX', 'TTWO', 'NFLX', 'AMZN', 'SNPS', 'ALGN', 'ROL', 'ATVI', 'BIO', 'ADSK', 'TSCO', 'DVA', 'ODFL', 'TYL', 'SIVB', 'PKI' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'TMO', 'CLX', 'EBAY', 'EA', 'DPZ', 'AMAT', 'CDE', 'MSFT', 'A', 'CRM', 'GGG', 'ISRG', 'EFX', 'CPRT', 'URI', 'LOW', 'NEE', 'CHD', 'CTSH', 'FAST', 'GOOG', 'SHW', 'AJG', 'MCHP', 'EL', 'ABT', 'CTAS', 'PH', 'CMI', 'EMN'
2021	$\mathcal{S}_{20}$: 'EXR', 'MAA', 'SPG', 'AMAT', 'APA', 'COP', 'ODFL', 'PLD', 'RHI', 'TSCO', 'GOOG', 'SIVB', 'KIM', 'WST', 'OXY', 'REG', 'JCI', 'MOS', 'UDR', 'AMD' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'LOW', 'TXT', 'DPZ', 'ACN', 'MLM', 'HD', 'SCHW', 'MSFT', 'JNPR', 'LH', 'ALB', 'ORLY', 'FITB', 'EFX', 'AOS', 'HRB', 'AAP', 'MMC', 'WAT', 'EW', 'JBHT', 'PAYX', 'DGX', 'BAC', 'TMO', 'NVR', 'SEE', 'SHW', 'WM', 'SNPS'
2022	$\mathcal{S}_{20}$: 'OXY', 'FTI', 'HES', 'HRB', 'APA', 'COP', 'SLB', 'NOV', 'NOC', 'ADM', 'WRB', 'GIS', 'CPB', 'GPC', 'HSY', 'WMB', 'SRE', 'AMGN', 'GILD', 'MOS' $\mathcal{S}_{50}$: $\mathcal{S}_{20}$ with 'TRV', 'ORLY', 'BIIB', 'K', 'HUM', 'RLI', 'ROL', 'GWW', 'PCAR', 'FMC', 'AJG', 'CAG', 'PNW', 'CB', 'ATVI', 'ALB', 'CMI', 'BSX', 'URI', 'CTAS', 'JKHY', 'APD', 'ADP', 'CNC', 'XEL', 'ABMD', 'TJX', 'SO', 'OMC', 'WM'

Table 4: Comparison of backtesting results for the stock market data. Proposed Method (PM), Markowitz's Portfolio with short selling (MP), Markowitz's Portfolio with no short selling (MPNS), Equally Weighted Portfolio (EWP).

Year	Method	Total Return		Annual Return		Annual Volatility		Sharpe Ratio	
		$n = 20$	$n = 50$	$n = 20$	$n = 50$	$n = 20$	$n = 30$	$n = 20$	$n = 50$
2006	PM+MP	6.60	35.85	8.67	32.70	21.08	18.98	0.41	1.72
	PM+MPNS	4.27	9.55	5.60	10.23	16.65	14.38	0.34	0.71
	PM+EWP	14.67	10.88	15.30	11.66	17.31	15.83	0.88	0.74
	MP	16.70		16.65		14.73		1.13	
	MPNS	11.51		11.73		12.18		0.96	
	EWP	14.11		14.00		11.77		1.19	
2007	PM+MP	-24.37	-98.50	-23.48	-177.54	30.45	208.87	-0.77	-0.85
	PM+MPNS	-14.10	-8.23	-13.34	-7.44	19.88	15.56	-0.67	-0.48
	PM+EWP	2.28	3.21	4.32	4.84	20.22	18.19	0.21	0.27
	MP	6.21		7.02		13.74		0.51	
	MPNS	-3.46		-2.56		14.02		-0.18	
	EWP	1.89		3.19		16.13		0.20	
2008	PM+MP	-59.40	-12.45	-0.09	136.89	132.80	173.87	-0.00	0.79
	PM+MPNS	-42.40	-36.24	-46.43	-39.05	41.50	34.23	-1.12	-1.14
	PM+EWP	-54.55	-47.83	-63.75	-54.43	54.40	45.74	-1.17	-1.19
	MP	-39.26		-41.67		40.48		-1.03	
	MPNS	-34.84		-37.20		33.32		-1.12	
	EWP	-40.62		-42.87		42.72		-1.00	
2009	PM+MP	-5.50	8.57	0.49	10.20	34.42	19.71	0.01	0.52
	PM+MPNS	6.67	5.45	8.23	6.86	18.65	17.46	0.44	0.39
	PM+EWP	0.99	6.87	3.86	9.42	23.96	23.43	0.16	0.40
	MP	17.19		18.63		23.26		0.80	
	MPNS	8.96		9.92		16.11		0.62	
	EWP	20.75		24.07		32.03		0.75	
2010	PM+MP	-370.66	105.98	728.86	140.46	385.55	116.49	1.89	1.21
	PM+MPNS	1.45	22.66	4.74	22.81	25.68	21.40	0.18	1.07
	PM+EWP	26.25	24.22	26.49	24.43	24.81	22.96	1.07	1.06
	MP	8.70		9.97		17.80		0.56	
	MPNS	19.94		19.75		17.24		1.15	
	EWP	19.08		19.44		19.50		1.00	
2011	PM+MP	-55.83	-99.74	-26.59	-280.29	105.30	241.04	-0.25	-1.16
	PM+MPNS	-26.76	-10.64	-27.77	-8.92	26.24	21.73	-1.06	-0.41
	PM+EWP	-17.29	-8.16	-13.96	-4.11	31.81	29.71	0.44	-0.14
	MP	42.73		37.62		19.34		1.94	
	MPNS	-7.27		-5.48		20.38		-0.27	
	EWP	-4.96		-1.75		25.83		-0.07	
2012	PM+MP	55.10	-0.31	62.36	0.94	60.00	15.52	1.04	0.06
	PM+MPNS	19.99	5.70	19.41	6.05	13.81	9.39	1.41	0.64
	PM+EWP	10.72	11.09	11.23	11.35	13.59	11.89	0.83	0.95
	MP	-2.96		-0.02		24.17		-0.00	
	MPNS	5.70		6.05		9.39		0.64	
	EWP	14.67		14.80		13.76		1.08	

Year	Method	Total Return		Annual Return		Annual Volatility		Sharpe Ratio	
		$n = 20$	$n = 50$	$n = 20$	$n = 50$	$n = 20$	$n = 30$	$n = 20$	$n = 50$
2013	PM+MP	33.70	3.84	30.24	6.07	14.58	21.38	2.07	0.28
	PM+MPNS	26.54	27.09	24.75	24.97	14.88	13.36	1.66	1.87
	PM+EWP	35.54	40.19	31.77	34.92	15.60	13.99	2.04	2.50
	MP	-3.38		-2.57		13.32		-0.19	
	MPNS	26.13		24.13		12.71		1.90	
	EWP	29.84		26.94		11.92		2.26	
2014	PM+MP	-58.12	-28.15	-38.51	-25.70	98.85	38.64	-0.39	-0.66
	PM+MPNS	11.23	16.23	11.94	16.12	15.78	14.27	0.76	1.13
	PM+EWP	16.56	15.18	16.88	15.36	17.23	15.20	0.98	1.01
	MP	-0.40		0.57		13.91		0.04	
	MPNS	14.84		14.80		13.45		1.10	
	EWP	13.66		13.56		11.87		1.14	
2015	PM+MP	3.35	-0.18	6.10	11.87	23.56	49.08	0.26	0.24
	PM+MPNS	6.17	7.57	7.48	8.67	17.12	16.36	0.44	0.53
	PM+EWP	12.62	2.74	13.36	3.85	16.85	15.05	0.79	0.26
	MP	-10.18		-9.32		17.06		-0.55	
	MPNS	5.38		6.35		14.76		0.43	
	EWP	-0.26		0.91		15.23		0.06	
2016	PM+MP	-1.95	-7.70	0.73	-6.04	23.24	20.03	0.03	-0.30
	PM+MPNS	-3.57	0.60	-2.51	1.50	15.06	13.46	-0.17	0.11
	PM+EWP	5.42	9.33	6.62	9.99	16.21	14.35	0.41	0.70
	MP	13.90		15.93		23.95		0.67	
	MPNS	3.98		4.72		12.64		0.37	
	EWP	14.92		15.03		14.54		1.03	
2017	PM+MP	5.36	288.53	8.28	228.18	24.52	134.44	0.34	1.70
	PM+MPNS	7.05	17.54	8.14	16.73	15.95	9.28	0.51	1.80
	PM+EWP	30.46	25.86	27.59	23.73	12.42	10.32	2.22	2.30
	MP	49.23		41.62		15.73		2.65	
	MPNS	17.98		17.00		8.16		2.08	
	EWP	18.73		17.57		7.26		2.42	
2018	PM+MP	-23.90	-7.04	-24.79	-5.25	23.30	20.50	-1.06	-0.26
	PM+MPNS	-17.90	-11.80	-18.05	-11.08	19.11	17.69	-0.94	-0.63
	PM+EWP	-12.10	-9.39	-10.38	-7.70	22.86	21.13	-0.45	-0.36
	MP	0.35		1.81		17.06		0.11	
	MPNS	-7.75		-6.97		15.20		-0.46	
	EWP	-10.03		-9.38		15.93		-0.59	

Year	Method	Total Return		Annual Return		Annual Volatility		Sharpe Ratio	
		$n = 20$	$n = 50$	$n = 20$	$n = 50$	$n = 20$	$n = 30$	$n = 20$	$n = 50$
2019	PM+MP	73.80	31.21	64.70	33.11	42.66	34.06	1.52	0.97
	PM+MPNS	17.16	26.85	16.61	24.40	11.89	10.04	1.40	2.43
	PM+EWP	18.21	22.65	18.02	21.30	15.60	12.56	1.15	1.70
	MP	22.74		23.25		23.17		1.00	
	MPNS	27.24		24.68		9.87		2.50	
	EWP	26.18		24.16		12.63		1.91	
2020	PM+MP	-81.08	-26.83	-94.94	14.99	123.30	96.72	-0.77	0.15
	PM+MPNS	4.90	4.60	12.18	11.10	38.24	36.12	0.32	0.31
	PM+EWP	9.00	11.49	16.01	17.27	38.14	35.54	0.42	0.49
	MP	-5.76		3.28		42.30		0.08	
	MPNS	2.80		9.32		35.98		0.26	
	EWP	-1.60		5.61		37.69		0.15	
2021	PM+MP	-35.45	8.14	80.17	11.58	157.42	27.30	0.51	0.42
	PM+MPNS	19.54	11.27	19.81	11.66	19.38	13.68	1.02	0.85
	PM+EWP	23.16	24.38	22.53	23.04	17.85	14.99	1.26	1.54
	MP	-1.97		0.43		22.04		0.02	
	MPNS	13.67		13.60		12.07		1.13	
	EWP	24.31		22.74		13.30		1.71	
2022	PM+MP	-100.00	-99.30	353.52	142.60	843.63	354.21	0.42	0.40
	PM+MPNS	-24.14	-22.28	-25.92	-23.96	26.22	23.76	-0.99	-1.01
	PM+EWP	-23.10	-20.00	-23.78	-20.46	28.70	25.51	-0.83	-0.80
	MP	-8.19		-5.72		25.94		-0.22	
	MPNS	-18.23		-19.11		21.39		-0.89	
	EWP	-17.62		-17.91		23.19		-0.77	

Conclusion: In this paper, we propose a method for dimensionality reduction of Markowitz’s mean–variance portfolio optimization problem by modeling the local dynamics of asset returns through a signed graph framework. Specifically, we define the hedge-score of an asset in terms of the negative degree of the corresponding vertex in the graph representation of the financial market. To evaluate the effectiveness of this approach, we conduct backtesting on two datasets and benchmark the performance of the proposed method on the reduced asset universe against that of Markowitz’s optimization (with and without short selling) as well as the equally weighted portfolio on the full universe.

Our empirical analysis shows that the proposed method outperforms the standard approaches on several occasions, thereby demonstrating its potential efficiency. However, in other cases it fails to achieve comparable performance. Such variability may arise from factors including the choice of K , the number of potential hedge-protected assets to be selected. In future work, we intend to explore the integration of higher-order motifs in the signed graph framework as a means of further enhancing dimensionality reduction.

Acknowledgment. The author thanks Sarvagya Upadhyay and Hannes Leipold for inspiring discussions and valuable suggestions.

References

1. Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608, 2015.
2. Samin Aref and Mark C Wilson. Balance and frustration in signed networks. *Journal of Complex Networks*, 7(2):163–189, 2019.
3. Matteo Barigozzi and Christian Brownlees. Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364, 2019.
4. Paolo Bartesaghi, Fernando Diaz-Diaz, Rosanna Grassi, and Pierpaolo Uberti. Global balance and systemic risk in financial correlation networks. *Physica A: Statistical Mechanics and its Applications*, page 130698, 2025.
5. Dirk G Baur and Brian M Lucey. Is gold a hedge or a safe haven? an analysis of stocks, bonds and gold. *Financial review*, 45(2):217–229, 2010.
6. Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
7. Stephen Boyd, Kasper Johansson, Ronald Kahn, Philipp Schiele, and Thomas Schmelzer. Markowitz portfolio construction at seventy. *arXiv preprint arXiv:2401.05080*, 2024.
8. Giuseppe Buonaiuto, Francesco Gargiulo, Giuseppe De Pietro, Massimo Esposito, and Marco Pota. Best practices for portfolio optimization by quantum computing, experimented on real quantum devices. *Scientific Reports*, 13(1):19434, 2023.
9. Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider’s theory. *Psychological review*, 63(5):277, 1956.
10. Lin Chen, Qian Han, Zhilin Qiao, and H. Eugene Stanley. Correlation analysis and systemic risk measurement of regional, financial and global stock indices. *Physica A: Statistical Mechanics and its Applications* 542:122653, 2020.
11. K Tse Chi, Jing Liu, and Francis CM Lau. A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659–667, 2010.

12. Munki Chung, Yongjae Lee, Jang Ho Kim, Woo Chang Kim, and Frank J Fabozzi. The effects of errors in means, variances, and correlations on the mean-variance framework. *Quantitative Finance*, 22(10):1893–1903, 2022.
13. Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. *The review of Financial studies* 22(5):1915–1953, 2009.
14. Fernando Diaz-Diaz. Mathematical analysis of signed networks: structure and dynamics. PhD thesis, Institute of Cross-Disciplinary Physics and Complex Systems, IFISC, 2025.
15. Maryam Ehsani. The structure of stock markets as signed networks. *Journal of Industrial and Systems Engineering*, 13(1):136–146, 2020.
16. Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.
17. Rosa Figueiredo and Yuri Frota. The maximum balanced subgraph of a signed graph: Applications and solution approaches. *European Journal of Operational Research*, 236(2):473–487, 2014.
18. Lisa R Goldberg, Alex Papanicolaou, and Alex Shkolnik. The dispersion bias. *SIAM Journal on Financial Mathematics*, 13(2):521–550, 2022.
19. Marco Gregnanin, Yanyi Zhang, Johannes De Smedt, Giorgio Gnecco, and Maurizio Parton. Signature-based portfolio allocation: a network approach. *Applied Network Science* 9, no. 1: 54, 2024.
20. Thomas Guhr and Bernd Kälber. A new method to estimate the noise in financial correlation matrices. *Journal of Physics A: Mathematical and General*, 36(12):3009, 2003.
21. Frank Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146, 1953.
22. Frank Harary, Meng-Hiot Lim, and Donald C Wunsch. Signed graphs for portfolio analysis in risk management. *IMA Journal of management mathematics*, 13(3):201–210, 2002.
23. Nikolaus Hautsch, Julia Schaumburg, and Melanie Schienle. Financial network systemic risk contributions. *Review of Finance*, 19(2):685–738, 2015.
24. Narendra N Hegade, Pranav Chandarana, Koushik Paul, Xi Chen, Francisco Albarrán-Arriagada, and E Solano. Portfolio optimization with digitized counterdiabatic quantum algorithms. *Physical Review Research*, 4(4):043204, 2022.
25. Falk Hüffner, Nadja Betzler, and Rolf Niedermeier. Separator-based data reduction for signed graph balancing. *Journal of combinatorial optimization*, 20(4):335–360, 2010.
26. Matthew O Jackson. Networks in the understanding of economic behaviors. *Journal of economic perspectives*, 28(4):3–22, 2014.
27. Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.
28. Dror Y Kenett and Shlomo Havlin. Network science: a useful tool in economics and finance. *Mind & Society*, 14:155–167, 2015.
29. Imre Kondor, Szilárd Pafka, and Gábor Nagy. Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance*, 31(5):1545–1573, 2007.
30. Steven Kordonowy and Hannes Leipold. The lie algebra of xy-mixer topologies and warm starting qaoa for constrained optimization. *arXiv preprint arXiv:2505.18396*, 2025.

31. Zhao-Rong Lai and Haisheng Yang. A survey on gaps between mean-variance approach and exponential growth rate approach for portfolio optimization. *ACM Computing Surveys (CSUR)*, 55(2):1–36, 2022.
32. Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical review letters*, 83(7):1467, 1999.
33. Hannes Leipold and Sarvagya Upadhyay. Train-and-scaling the quantum alternating operator ansatz to solve portfolio diversification. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 2, pages 132–137. IEEE, 2024.
34. Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11:193–197, 1999.
35. Rosario N. Mantegna and H. Eugene Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.
36. Harry M Markowitz. Portfolio selection, the journal of finance. 7 (1). N, 1:71–91, 1952.
37. Naoki Masuda, Zachary M. Boyd, Diego Garlaschelli, and Peter J. Mucha. Introduction to correlation networks: Interdisciplinary approaches beyond thresholding. *Physics Reports* 1136: 1-39, 2025.
38. Giulio Mattera and Raffaele Mattera. Shrinkage estimation with reinforcement learning of large variance matrices for portfolio selection. *Intelligent Systems with Applications*, 17:200181, 2023.
39. Szilárd Pafka and Imre Kondor. Noisy covariance matrices and portfolio optimization ii. *Physica A: Statistical Mechanics and its Applications*, 319:487–494, 2003.
40. Gustavo Peralta and Abalfazl Zareei. A network approach to portfolio selection. *Journal of Empirical Finance*, 38:157–180, 2016.
41. Ranveer Singh and Bibhas Adhikari. Measuring the balance of signed networks and its application to sign prediction. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(6):063302, 2017.
42. Soloviev, Vicente P., Antonio Márquez Romero, Josh Kirsopp, and Michal Krompiec. Scaling Portfolio Diversification with Quantum Circuit Cutting Techniques. *arXiv preprint arXiv:2506.08947*, 2025.
43. Thomas Zaslavsky. A mathematical bibliography of signed and gain graphs and allied areas. *The Electronic Journal of Combinatorics*, pages DS8–Dec, 2012.
44. B. Vasanthi, S. Arumugam, Atulya K. Nagar, and Sovan Mitra. Applications of signed graphs to portfolio turnover analysis. *Procedia-Social and Behavioral Sciences* 211: 1203-1209, 2015.