Beyond Markowitz: Can Network Theory Improve Portfolio Diversification?

A Correlation-Network and Mean-Variance Portfolio Experiment

Joel Gonzalez Gomez

Independent working paper

July 2026

Abstract

This paper studies whether a simple financial-network layer can complement traditional mean-variance portfolio optimization. Modern Portfolio Theory formalizes diversification through the covariance matrix: a portfolio is attractive when it improves expected return per unit of variance. Yet covariance matrices estimated from financial returns are noisy, unstable, and difficult to interpret economically. Network methods offer an alternative representation of the same correlation structure. By transforming pairwise correlations into distances and extracting a minimum spanning tree, the investor can see which assets are central, peripheral, or redundant in the portfolio's dependence structure. Using daily U.S. equity prices from an open public dataset covering Apple, Microsoft, IBM, Starbucks, and the S&P 500 benchmark from 2007 to 2016, I compare equal-weight, inverse-volatility, minimum-variance, maximum-Sharpe, and network-peripheral portfolios. The empirical exercise is deliberately framed as a methodological prototype rather than a final investment recommendation. In the out-of-sample backtest, the maximum-Sharpe portfolio delivers the highest annualized return and Sharpe ratio, while the network-peripheral portfolio produces competitive returns and a higher Sharpe ratio than the equal-weight benchmark, but with higher volatility and drawdown. The results suggest that network methods are most useful not as a replacement for Markowitz optimization, but as a diagnostic and selection layer: they reveal dependence concentration, identify structurally central assets, and create an interpretable bridge between diversification as a mathematical concept and diversification as an economic intuition.

Keywords: portfolio optimization, network theory, minimum spanning tree, Markowitz,

efficient frontier, covariance matrix, Sharpe ratio, financial econometrics.
JEL-style classification: G11, G12, C58, C61.

Contents

1 Introduction

Portfolio diversification is one of the most familiar ideas in finance, but it is also one of the easiest to misunderstand. A portfolio with many securities can still be highly concentrated if those securities move together. Conversely, a smaller set of assets can provide meaningful diversification if their returns are weakly connected. In practice, diversification is not simply a count of assets. It is a property of the covariance structure.

The classical answer to the diversification problem is the mean-variance framework introduced by Markowitz (1952). In that framework, the investor selects portfolio weights by trading off expected return and variance. The efficient frontier summarizes the set of portfolios that minimize variance for a given target return, or equivalently maximize return for a given risk level. The tangency portfolio is then the risky portfolio with the highest Sharpe ratio when a risk-free asset is available.

The mathematical elegance of the Markowitz model also creates its practical weakness. The model requires estimates of expected returns and the covariance matrix. Expected returns are notoriously difficult to estimate, and covariance matrices can be unstable when the number of assets is large relative to the length of the sample. Small estimation errors can produce large changes in optimized weights. This issue is central in the covariance-noise literature, including work by Pafka and Kondor (2003), who emphasize that empirical correlation matrices can contain substantial noise and that portfolio optimization may be highly sensitive to the inputs.

This paper asks whether network theory can provide a complementary way to think about portfolio diversification. The key idea is simple. Instead of viewing the covariance matrix only as an input to an optimizer, I transform the correlation matrix into a network. Each asset becomes a node. Each pairwise relationship becomes an edge weighted by a distance transformation of correlation. A minimum spanning tree then filters the network to the smallest connected structure that preserves the strongest dependence channels. This filtered representation identifies which assets are central in the dependence structure and which assets are more peripheral.

The empirical question is:

> *Can network-based asset selection improve the diversification profile of a portfolio relative to traditional mean-variance and naive allocation rules?*

To answer this question, I build a small portfolio experiment using daily U.S. equity prices from an open public stock dataset. The sample includes Apple, Microsoft, IBM, Starbucks, and the S&P 500 benchmark from January 2007 through March 2016. The four individual assets create a compact but instructive universe: technology, consumer discretionary/consumer-facing retail, and a broad market benchmark. I estimate log returns, correlations, correlation distances, a minimum spanning tree, full-sample portfolio weights, and rolling out-of-sample performance.

The analysis compares five portfolio rules. The first is equal weight. The second is inverse volatility. The third is the minimum-variance portfolio. The fourth is the maximum-Sharpe or tangency-style portfolio. The fifth is a network-peripheral portfolio that selects the most peripheral assets in the minimum spanning tree and holds them equally. This last portfolio is intentionally simple. The objective is not to produce a trading strategy, but to test whether a network-based selection rule produces a different risk-return profile from traditional optimization.

The results are mixed and informative. In the out-of-sample period, the maximum-Sharpe portfolio performs best by annualized Sharpe ratio. The network-peripheral portfolio performs competitively and outperforms the equal-weight portfolio by return and Sharpe ratio, but at the cost of higher volatility and drawdown. The minimum-variance portfolio reduces volatility and drawdown, but also sacrifices return. These findings suggest that network methods should not be interpreted as a mechanical replacement for optimization. Instead, network methods are useful as a diversification diagnostic, an asset-selection screen, and a visual representation of hidden concentration.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature. Section 3 describes the data. Section 4 explains the mean-variance and network methodology. Section 5 presents descriptive evidence. Section 6 reports portfolio-construction results. Section 7 presents out-of-sample backtests. Section 8 discusses market-adjusted regressions. Section 9 evaluates robustness and limitations. Section 10 concludes.

2 Literature Review

2.1 Mean-variance portfolio theory

The foundation of modern portfolio construction is Markowitz (1952). The core insight is that investors should not evaluate assets in isolation. An asset's contribution to portfolio risk depends on how it co-moves with other assets. This is why covariance is central to the optimization problem. A high-volatility security can still improve a portfolio if it has low or negative covariance with the existing holdings.

Let r_t be the vector of asset returns at time t, $\mu = E[r_t]$ the expected return vector, and $\Sigma = Var(r_t)$ the covariance matrix. For a vector of portfolio weights w, portfolio expected return and variance are:

$$\mu_p = w'\mu, \tag{1}$$

$$\sigma_p^2 = w'\Sigma w. \tag{2}$$

The efficient frontier is obtained by solving:

$$\min_w w'\Sigma w \tag{3}$$

subject to:

$$w'\mathbf{1} = 1, \quad w'\mu = \mu^*, \tag{4}$$

where μ^* is a target return.

The tangency portfolio maximizes the Sharpe ratio. When there is a risk-free rate r_f, the unconstrained tangency portfolio can be written as:

$$w^* = \frac{\Sigma^{-1}(\mu - r_f\mathbf{1})}{\mathbf{1}'\Sigma^{-1}(\mu - r_f\mathbf{1})}. \tag{5}$$

In practice, constraints such as no shorting and maximum position sizes often require numerical optimization.

2.2 The covariance-estimation problem

The Markowitz framework is straightforward once μ and Σ are known. In empirical finance, however, they are estimated from finite samples. This is not a minor detail. Covariance matrices are high-dimensional objects, and sampling error can distort both optimization and risk measurement. Pafka and Kondor (2003) emphasize that financial correlations play a central role in portfolio optimization and risk management, but estimated correlation matrices can be noisy. When an optimizer is given noisy inputs, it may interpret noise as signal and assign extreme weights.

This problem is especially severe for expected returns. A small change in estimated mean return can materially alter the maximum-Sharpe portfolio. For this reason, practitioners often prefer minimum-variance, risk-parity, shrinkage, or constrained optimization frameworks. These approaches reduce reliance on expected-return estimates and place more emphasis on covariance structure.

2.3 Financial networks and correlation filtering

Financial-network methods approach the same covariance information from a different perspective. Instead of feeding the full correlation matrix directly into an optimizer, the researcher treats the correlation matrix as a graph. Each asset is a node. Pairwise correlations become connections between nodes. The challenge is to filter the graph so that the most important dependence relationships remain visible.

Tumminello et al. (2010) discuss how correlation matrices can be converted into hierarchical trees and correlation-based networks. The minimum spanning tree is one of the most common filters. It preserves $N-1$ edges among N assets and connects all nodes while minimizing total distance. In financial applications, the standard distance transformation is:

$$d_{ij} = \sqrt{2(1-\rho_{ij})}, \tag{6}$$

where ρ_{ij} is the correlation between assets i and j. A higher correlation implies a lower distance. The minimum spanning tree therefore connects assets through the strongest dependence channels.

Network methods do not eliminate the need for statistical caution. Millington and Niranjan (2021) show that minimum spanning trees can depend on how correlations are estimated and compare Pearson, Spearman, and Kendall correlation methods. Their work highlights that network topology can be informative, but also subject to sampling instability. This is why the present paper treats the network layer as a complementary tool rather than a standalone optimizer.

2.4 Network-based portfolio construction

Recent applied work explores whether minimum spanning trees can help asset selection and allocation. For example, Berouaga (2023) explicitly studies portfolio-optimization stages using MST-based methods. The motivation is intuitive: if two assets are strongly connected in the network, they may provide less diversification benefit when held together. Conversely, peripheral assets may capture more distinct sources of risk.

This paper follows that intuition in a simple way. It defines a network-peripheral portfolio that selects the assets with the lowest centrality score in the MST. The goal is to test whether a peripheral selection rule produces a portfolio with competitive risk-adjusted performance relative to traditional allocation rules.

3 Data

3.1 Raw source and sample

The empirical dataset is Plotly's public `stockdata.csv`, available through the Plotly datasets repository. The dataset contains daily price observations for Apple, Microsoft, IBM, Starbucks, and the S&P 500 benchmark variable `GSPC`. The sample used in this paper runs from January 3, 2007 to March 1, 2016.

The individual assets are:

- Apple (AAPL),
- Microsoft (MSFT),
- IBM (IBM),
- Starbucks (SBUX).

The benchmark is:

- S&P 500 proxy (GSPC).

This sample is not intended to represent the full U.S. equity market. It is a compact public dataset used to demonstrate the full workflow: data cleaning, return construction, covariance estimation, network filtering, optimization, backtesting, and market-adjusted regression. A larger version of the analysis would extend the same framework to the S&P 100, S&P 500, or a sector-balanced universe.

3.2 Return construction

Daily log returns are calculated as:

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}), \tag{7}$$

where $P_{i,t}$ is the price of asset i on day t. Annualized sample mean returns and covariance matrices are computed using 252 trading days:

$$\hat{\mu}_i = 252 \cdot \frac{1}{T} \sum_{t=1}^{T} r_{i,t}, \tag{8}$$

$$\hat{\Sigma} = 252 \cdot \widehat{Var}(r_t). \tag{9}$$

Table 1: Asset-level annualized summary statistics

Asset	Annual return	Annual volatility	Sharpe	Max drawdown
AAPL	27.3%	33.9%	0.803	-60.9%
MSFT	9.0%	28.9%	0.311	-57.9%
IBM	5.8%	22.9%	0.251	-44.3%
SBUX	15.4%	33.6%	0.459	-80.2%
GSPC	3.7%	21.7%	0.172	-56.8%

Table 1 presents asset-level annualized statistics. Apple has the highest annualized return in the full sample, while IBM has the lowest volatility among the individual stocks. The S&P 500 benchmark has lower return and lower volatility than most individual assets.

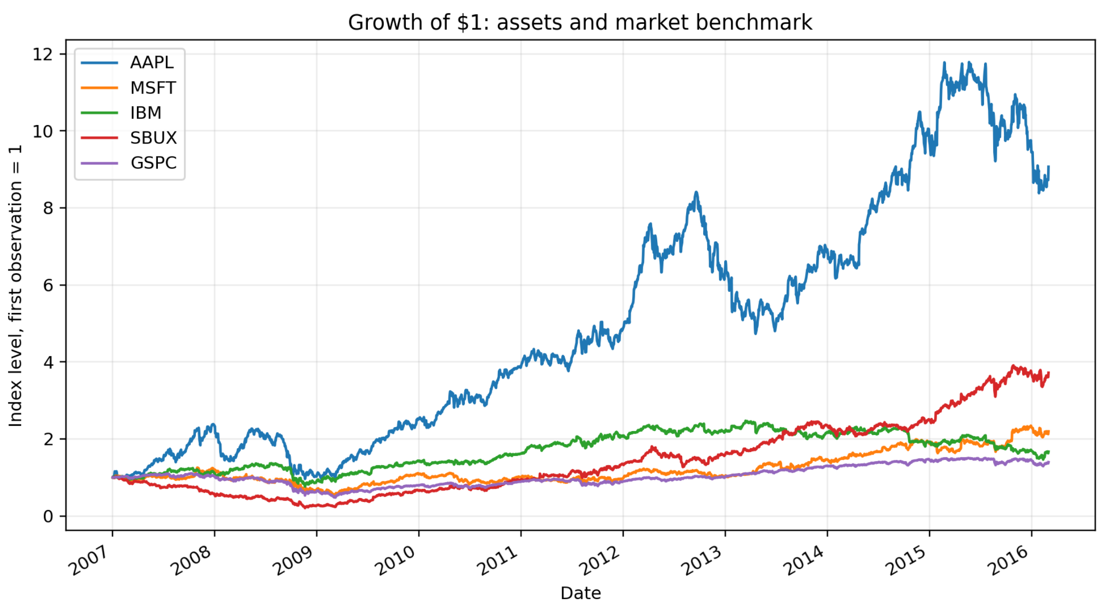

Figure 1: Growth of one dollar invested in the assets and market benchmark.

4 Methodology

4.1 Mean-variance optimization

The first portfolio family is based on traditional mean-variance optimization. The minimum-variance portfolio solves:

$$\min_{w} w'\hat{\Sigma}w \tag{10}$$

subject to:

$$\sum_i w_i = 1, \quad w_i \geq 0. \tag{11}$$

The maximum-Sharpe portfolio solves:

$$\max_w \frac{w'\hat{\mu}}{\sqrt{w'\hat{\Sigma}w}} \tag{12}$$

subject to the same long-only and fully-invested constraints. Because the analysis uses a public historical price dataset and does not merge risk-free rates, the Sharpe ratio is computed relative to zero in the baseline. This is acceptable for comparing portfolios over the same period, but it should not be interpreted as a fully risk-free-adjusted performance measure.

4.2 Naive and risk-based benchmarks

The equal-weight portfolio is:

$$w_i^{EW} = \frac{1}{N}. \tag{13}$$

The inverse-volatility portfolio is:

$$w_i^{IV} = \frac{1/\hat{\sigma}_i}{\sum_{j=1}^{N} 1/\hat{\sigma}_j}, \tag{14}$$

where $\hat{\sigma}_i$ is the annualized volatility of asset i. These benchmarks are useful because they avoid expected-return estimation. Equal weight asks whether complexity adds value. Inverse volatility asks whether simple risk scaling improves performance.

4.3 Network construction

The network layer starts with the return correlation matrix:

$$\hat{\rho}_{ij} = Corr(r_{i,t}, r_{j,t}). \tag{15}$$

Correlations are transformed into distances using:

$$d_{ij} = \sqrt{2(1 - \hat{\rho}_{ij})}. \tag{16}$$

This distance satisfies the intuition that highly correlated assets are close and weakly correlated assets are far apart. A complete graph is constructed with assets as nodes and distances as edge weights. The minimum spanning tree T is the tree connecting all assets with minimum total distance:

$$T^* = \arg\min_{T \in \mathcal{T}} \sum_{(i,j) \in T} d_{ij}, \tag{17}$$

where $\mathcal{T}$ is the set of all spanning trees.

4.4 Centrality and peripheral selection

For each asset, I compute degree, betweenness centrality, and closeness centrality in the MST. The centrality score is defined as the sum of percentile ranks:

$$CS_i = Rank(Degree_i) + Rank(Betweenness_i) + Rank(Closeness_i). \tag{18}$$

Assets with lower CS_i are more peripheral. The network-peripheral portfolio selects the two most peripheral assets and assigns equal weights:

$$w_i^{NET} = \begin{cases} 1/K, & \text{if asset } i \text{ is among the } K \text{ most peripheral assets,} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

In this empirical application, $K = 2$ because the investable universe contains four individual assets.

4.5 Out-of-sample backtest

The backtest uses annual rebalancing. For each calendar year y, parameters are estimated using the previous two years of daily returns. Portfolio weights are then held during year y. This produces a rolling out-of-sample daily return series for each strategy from 2009 through March 2016.

Let $\mathcal{T}_y^{train}$ be the two-year training window before year y and $\mathcal{T}_y^{test}$ be the test period in year y. The rolling design is:

$$\hat{w}_{p,y} = f_p\left(\{r_t : t \in \mathcal{T}_y^{train}\}\right), \tag{20}$$

$$r_{p,t}^{OOS} = \hat{w}_{p,y}' r_t, \quad t \in \mathcal{T}_y^{test}. \tag{21}$$

This structure reduces look-ahead bias because only information available before the test year is used to form the portfolio.

4.6 Performance metrics

Annualized return is calculated as:

$$R_{ann} = \exp(252\bar{r}) - 1. \tag{22}$$

Annualized volatility is:

$$\sigma_{ann} = \sqrt{252} \cdot sd(r_t). \tag{23}$$

The Sharpe ratio is:

$$SR = \frac{R_{ann}}{\sigma_{ann}}. \tag{24}$$

Maximum drawdown is:

$$MDD = \min_t \left(\frac{V_t}{\max_{s \le t} V_s} - 1 \right), \tag{25}$$

where V_t is the cumulative wealth index.

5 Descriptive Analysis

5.1 Correlation structure

Figure 2 shows the full-sample correlation matrix. Correlations are positive across all pairs, but not identical. This matters because mean-variance optimization and network construction both depend on the relative strength of co-movement. The most correlated pair in the MST is Microsoft and IBM, while Apple and Starbucks are more peripheral under the selected network rule.

Table 2: Minimum spanning tree edges

Asset 1	Asset 2	Distance	Correlation
AAPL	IBM	1.015	0.485
MSFT	IBM	0.945	0.553
IBM	SBUX	1.006	0.494

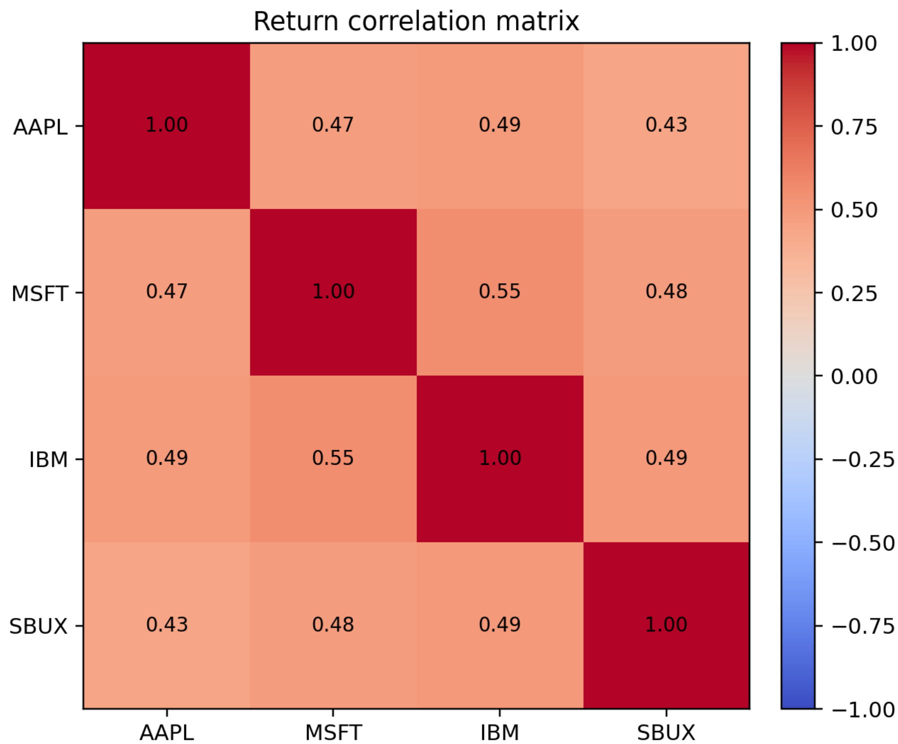

Figure 2: Daily return correlation matrix.

5.2 Minimum spanning tree

Figure 3 presents the minimum spanning tree. IBM occupies the most central position in the filtered network: it connects Apple, Microsoft, and Starbucks. Apple and Starbucks are peripheral nodes. In economic terms, a central node in the MST is not necessarily the best asset. Rather, it is an asset whose return behavior acts as a bridge in the dependence structure. A peripheral node may offer a more distinct source of return variation.

Table 3: Network centrality and peripheral ranking

Asset	Degree	Betweenness	Closeness	Centrality score	Peripheral rank
AAPL	1	0.000	0.601	1.250	1
SBUX	1	0.000	0.602	1.500	2
MSFT	1	0.000	0.618	1.750	3
IBM	3	1.000	1.011	3.000	4

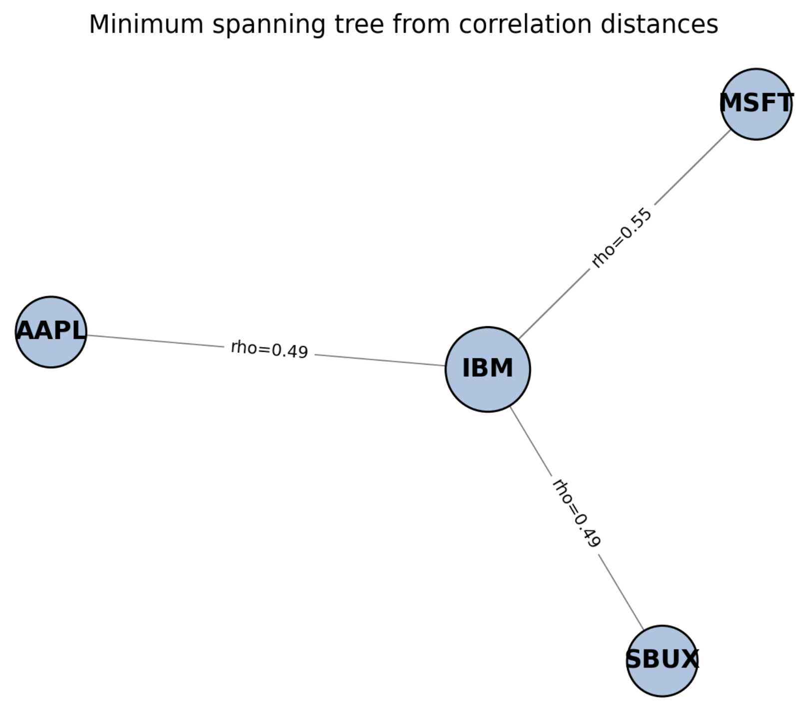

Figure 3: Minimum spanning tree from correlation distances.

6 Portfolio Construction Results

6.1 Efficient frontier

Figure 4 shows the efficient frontier and model portfolios. Random long-only portfolios form a risk-return cloud. The frontier traces the portfolios with the lowest volatility for each target return. The maximum-Sharpe portfolio concentrates in Apple and Starbucks in the full-sample estimation, reflecting their higher return estimates over the sample. The minimum-variance portfolio tilts heavily toward IBM because IBM has lower volatility and a central covariance profile. The network-peripheral portfolio holds Apple and Starbucks because they are the most peripheral assets in the MST.

Table 4: Full-sample optimized portfolio weights

Portfolio	AAPL	MSFT	IBM	SBUX	Return	Volatility	Sharpe
Equal weight	25.0%	25.0%	25.0%	25.0%	13.2%	23.4%	0.563
Minimum variance	8.7%	18.7%	64.2%	8.5%	8.5%	21.7%	0.393
Tangency max Sharpe	81.4%	0.0%	0.0%	18.6%	22.3%	30.8%	0.723
Inverse volatility	21.4%	25.2%	31.7%	21.6%	12.2%	22.8%	0.536
Network peripheral	50.0%	0.0%	0.0%	50.0%	19.2%	28.6%	0.673

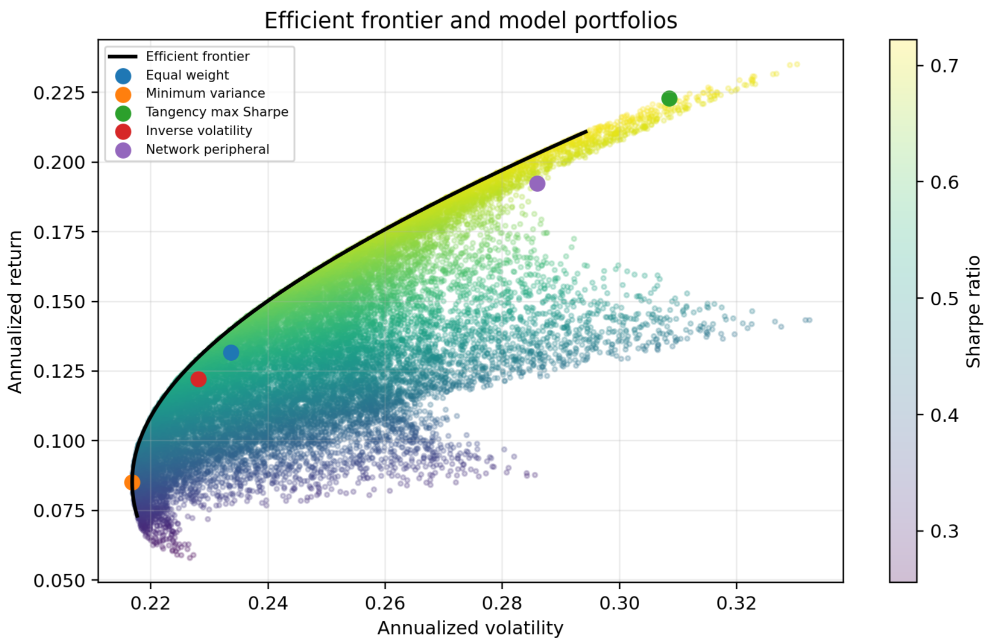

Figure 4: Efficient frontier and full-sample model portfolios.

6.2 Interpretation of optimized weights

The full-sample weights show how differently the methods interpret diversification. Equal weight allocates 25 percent to each asset. Inverse volatility gives the largest weight to IBM because of its lower volatility. Minimum variance assigns more than 60 percent to IBM. Maximum Sharpe concentrates heavily in Apple, with a smaller allocation to Starbucks. The network-peripheral portfolio holds Apple and Starbucks equally.

This contrast is useful. Mean-variance optimization tends to reward assets that appear attractive in the return-covariance tradeoff. Network selection rewards assets that are less central in the dependence structure. These are different definitions of attractiveness. A stock can be attractive because it has a high estimated return; because it reduces portfolio variance; or because it provides a distinct node in the network.

7 Out-of-Sample Backtest

7.1 Cumulative performance

Figure 5 shows out-of-sample cumulative performance. The backtest begins in 2009 because the first two years are used for training. All portfolio strategies outperform the S&P 500 benchmark over this sample. This is partly a function of the chosen assets and period: Apple and Starbucks performed strongly during the post-crisis recovery.

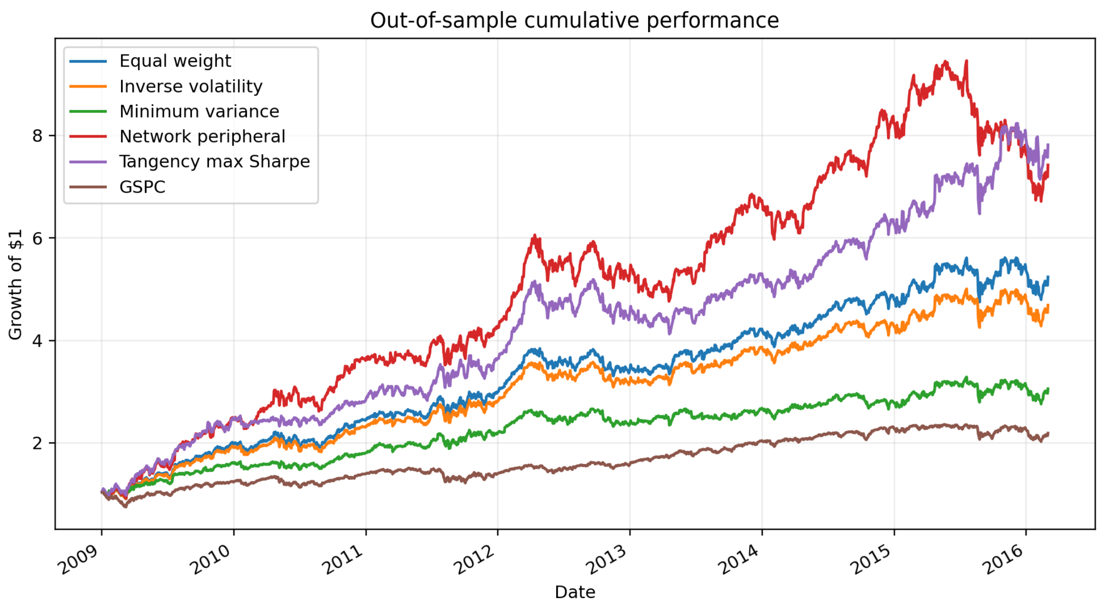

Figure 5: Out-of-sample cumulative portfolio performance.

7.2 Risk-return summary

Table 5 reports annualized backtest performance. The maximum-Sharpe portfolio has the highest annualized return and Sharpe ratio. The network-peripheral portfolio has the second-highest annualized return and a Sharpe ratio above equal weight. However, it also has a larger maximum drawdown than the equal-weight, inverse-volatility, and minimum-variance portfolios. This means that the network rule improves return in this sample, but does not automatically reduce downside risk.

Table 5: Out-of-sample portfolio performance

Portfolio	Annual return	Annual volatility	Sharpe	Sortino	Max drawdown
Tangency max Sharpe	33.3%	22.6%	1.475	2.178	-20.5%
Network peripheral	32.4%	24.3%	1.334	1.956	-29.1%
Equal weight	26.1%	19.9%	1.310	1.858	-19.3%
Inverse volatility	24.1%	19.4%	1.240	1.760	-18.8%
Minimum variance	16.9%	18.8%	0.899	1.296	-16.1%
GSPC	11.6%	18.1%	0.641	0.826	-27.6%

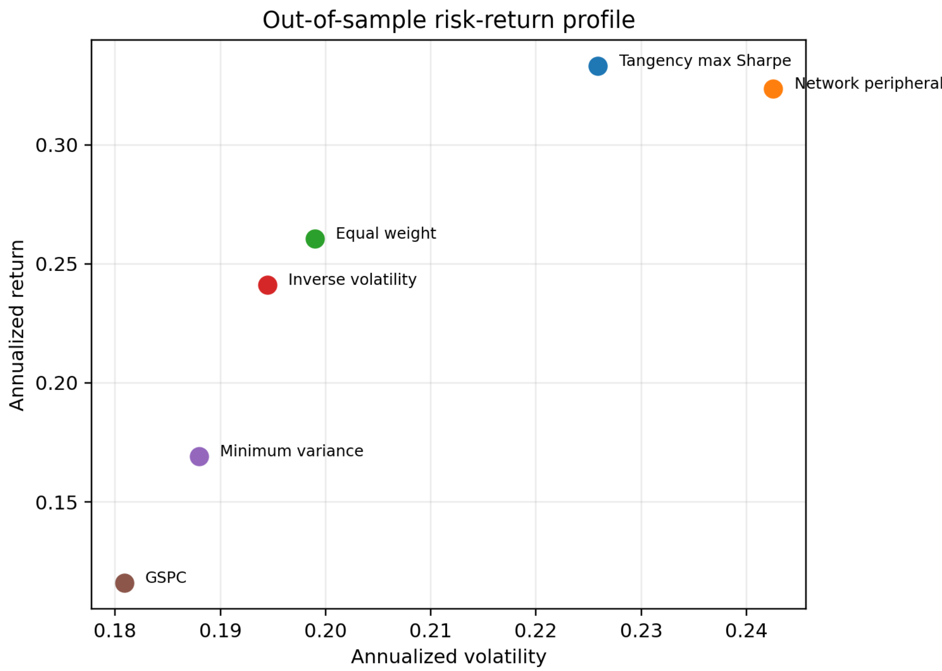

Figure 6: Out-of-sample annualized risk-return profile.

8 Market-Adjusted Regression Analysis

8.1 CAPM-style specification

To evaluate whether the portfolio returns are simply compensation for market exposure, I estimate a simple market-adjusted regression:

$$r_{p,t} = \alpha_p + \beta_p r_{m,t} + \epsilon_{p,t}, \tag{26}$$

Table 6: CAPM-style market-adjusted regressions with HAC standard errors

Portfolio	Annualized alpha	Market beta	Alpha t-stat	Alpha p-value	R-squared
Tangency max Sharpe	20.8%	0.896	3.284	0.001	0.515
Network peripheral	18.7%	0.997	2.881	0.004	0.552
Equal weight	13.9%	0.926	3.315	0.001	0.709
Inverse volatility	12.3%	0.909	3.070	0.002	0.714
Minimum variance	6.9%	0.818	1.620	0.105	0.619

where $r_{p,t}$ is the portfolio return and $r_{m,t}$ is the S&P 500 benchmark return. Standard errors are estimated using a heteroskedasticity and autocorrelation consistent covariance estimator with five lags.

The intercept α_p is interpreted as average daily return not explained by contemporaneous market returns. Since this is not a full Fama-French regression, the alpha should be interpreted cautiously. It is a market-adjusted performance measure, not a proof of abnormal risk-adjusted returns.

8.2 Regression results

Table 6 reports the market-adjusted regressions. The maximum-Sharpe and network-peripheral portfolios have positive annualized alpha estimates and market betas close to one. Equal weight and inverse volatility also have positive alpha estimates. The minimum-variance portfolio has a lower beta and a lower alpha estimate, consistent with its more defensive profile.

The regression results reinforce the backtest interpretation. The strongest return outcomes are not simply the result of holding the market. However, because the asset universe is small and selected ex post from an open dataset, the alpha estimates should be treated as descriptive rather than as evidence of a tradable anomaly.

9 Robustness and Limitations

9.1 Mathematical robustness

The main mathematical risk in this analysis is estimation error. The maximum-Sharpe portfolio is especially sensitive to expected returns. In the full-sample results, the maximum-Sharpe portfolio assigns a large weight to Apple because Apple has a high estimated historical return. If the sample window changes, this allocation may change sharply.

The minimum-variance portfolio is less sensitive to expected-return estimates because it only uses the covariance matrix. However, it can still concentrate in one asset if that asset has low volatility and favorable covariance properties. This occurs in the full-sample minimum-variance allocation, where IBM receives the largest weight.

The network-peripheral portfolio is robust in a different sense. It does not use expected returns. It uses the topology of the correlation matrix. This makes it less sensitive to mean estimates, but still sensitive to correlations and to the centrality definition. A different network filter, such as a planar maximally filtered graph, or a different correlation estimator, such as Spearman rank correlation, could produce a different peripheral ranking.

9.2 Economic robustness

The paper's main economic conclusion is not that network methods dominate Markowitz optimization. The out-of-sample performance does not support such a strong claim. Instead, the evidence supports a narrower claim: network methods produce an interpretable diversification layer that can change asset selection and produce competitive performance.

A larger implementation should test the following robustness checks:

1. A larger equity universe, such as the S&P 100 or S&P 500.
2. Sector-balanced samples to avoid overrepresenting technology stocks.
3. Rolling windows of one, three, and five years.
4. Transaction costs and turnover penalties.
5. Alternative network filters and centrality measures.
6. Factor regressions using Fama-French factors.
7. Stress-period subsamples, including the global financial crisis and COVID period.

9.3 Data limitations

The empirical sample is intentionally compact. It uses four individual assets and a market benchmark from an open public dataset. This is sufficient to demonstrate the methodology, but not sufficient to claim general superiority of one portfolio rule. The small number of assets also makes the MST simple. In a larger universe, network structure would be richer and sector clusters would be more visible.

The analysis is therefore best interpreted as a research prototype. Its value is in the complete workflow: it shows how to transform stock returns into a covariance matrix, how to extract a network, how to construct portfolios, how to backtest, and how to interpret the results using both portfolio theory and financial networks.

10 Conclusion

This paper asked whether network theory can improve portfolio diversification beyond the traditional Markowitz framework. The answer is nuanced. Network theory does not replace mean-variance optimization. It complements it.

The Markowitz framework remains essential because it directly links expected return, risk, and portfolio weights. The efficient frontier and tangency portfolio are still the core mathematical objects of portfolio choice. However, the network representation adds something that the optimizer alone does not provide: interpretability. By converting correlations into distances and extracting a minimum spanning tree, the investor can see which assets are central in the dependence structure and which assets are more peripheral.

Empirically, the maximum-Sharpe portfolio delivered the strongest out-of-sample performance in this sample. The network-peripheral portfolio was competitive and outperformed equal weight by return and Sharpe ratio, but it did not minimize drawdown. Minimum variance produced the most defensive profile but sacrificed return. These results imply that the best use of network theory may be as a diagnostic and selection layer rather than as a mechanical allocation rule.

The broader lesson is that diversification is not just a matter of holding more assets. It is a matter of holding assets that are not redundant. Network theory provides a clean way to detect redundancy. For investors, researchers, and students of quantitative finance, that makes networks a valuable complement to the efficient frontier: Markowitz tells us how to optimize a portfolio; network theory helps us understand what kind of diversification the portfolio actually contains.

A Variable Definitions

Variable	Definition
$P_{i,t}$	Price of asset i on date t.
$r_{i,t}$	Daily log return of asset i, $\ln(P_{i,t}) - \ln(P_{i,t-1})$.
μ_i	Annualized mean return of asset i.
Σ	Annualized covariance matrix of asset returns.
ρ_{ij}	Pairwise return correlation between assets i and j.
d_{ij}	Correlation distance, $\sqrt{2(1-\rho_{ij})}$.
T^*	Minimum spanning tree extracted from the complete correlation-distance graph.
CS_i	Network centrality score for asset i.
w_i	Portfolio weight assigned to asset i.
R_{ann}	Annualized portfolio return.
σ_{ann}	Annualized volatility.
SR	Sharpe ratio.
MDD	Maximum drawdown.

B Additional Notes on Interpretation

The network-peripheral portfolio uses the two assets with the lowest centrality score in the minimum spanning tree. In this application, those assets are Apple and Starbucks. This does not mean that these assets are always "better." It means that, in the estimated dependence network, they occupy less central positions. A peripheral asset may provide distinct exposure, but it may also have higher volatility, weaker expected returns, or worse downside risk. Network position must therefore be interpreted together with return and risk statistics.

C Reproducibility

All raw data, cleaned data, generated figures, tables, and the Python analysis script are included in the working paper pack. The core pipeline is contained in `src/run_analysis.py`. The empirical dataset is the public Plotly `stockdata.csv` file. The script converts prices into returns, estimates the correlation network, constructs portfolios, runs the out-of-sample backtest, and writes the tables used in this paper.

References

Berouaga, Y. (2023). Portfolio Optimization Using Minimum Spanning Tree Method. *Journal of Risk and Financial Management*, 11(2), 53.

Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91.

Millington, T., and Niranjan, M. (2021). Construction of Minimum Spanning Trees from Financial Returns using Rank Correlation. *Physica A*.

Pafka, S., and Kondor, I. (2003). Estimated Correlation Matrices and Portfolio Optimization. *Physica A*.

Plotly Datasets. (2026). `stockdata.csv`. Public GitHub dataset used for demonstration and visualization examples.

Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1), 40–58.