UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

JAKOB PIRŠ

UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

PORTFOLIO OPTIMIZATION WITH GRAPH THEORY BASED ALGORITHMS

Ljubljana, July 2023

JAKOB PIRŠ

AUTHORSHIP STATEMENT

The undersigned Jakob Pirš, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title Portfolio optimization with graph theory based algorithms, prepared under supervision of prof. dr. Igor Lončarski PhD.

DECLARE

1. this written final work of studies to be based on the results of my own research;

2. the printed form of this written final work of studies to be identical to its electronic form;

3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;

4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;

5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;

6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;

7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;

8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;

9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;

10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

11. that I have verified the authenticity of the information derived from the records using artificial intelligence tools.

Ljubljana, July 11th, 2023 Author's signature: ___________________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CAPM	Capital Asset Pricing Model
CMC	Cross Maximal Clique Centrality
EMH	Efficient Market Hypothesis
EW	Equal Weight
MPT	Modern Portfolio Theory
MV	Mean-Variance
n. d.	No Date
NP	Non-deterministic Polynomial
P	Polynomial
PO	Portfolio Optimization
DC	Degree Centrality
CC	Closeness Centrality
BC	Betweenness Centrality
EC	Eigenvector Centrality

INTRODUCTION

Robert Arnott pointed out that in the realm of investing, what is comfortable is rarely profitable. This perspective highlights the importance of ongoing innovation and persistent exploration of more efficient strategies in finance. One of the pioneering models in this field is Markowitz's mean-variance optimization model. Proposed in 1952, this model sets the cornerstone of modern portfolio theory, suggesting that asset allocation within a portfolio should be influenced by the individual risk and return profiles of each asset, as well as their intercorrelations (Markowitz, 1952). The objective is to strike a balance between risk and return, aiming to maximize expected returns while minimizing risk, especially correlation risk (Buraschi, Porchia & Trojani, 2010).

The poor performance of Markowitz's rule stems from the large estimation errors on the vector of expected returns (Merton, 1980) and the covariance matrices (Jobson & Korkie, 1980) leading to the well-documented error-maximizing property discussed by Michaud and Michaud (2007). The magnitude of this problem is clear when we acknowledge the modest improvements achieved by those models specifically designed to tackle the estimation risk (DeMiguel, Garlappi & Uppal, 2009). Moreover, the evidence indicates that the simple yet effective equally-weighted portfolio rule has not been consistently outperformed by more sophisticated alternatives (Bloomfield, Leftwich, & Long, 1977; DeMiguel, Garlappi & Uppal, 2009). But since the end of the 20th century, numerous papers from various fields have characterized financial markets as networks, in which securities correspond to nodes and the links relate to the correlation of returns (Mantegna, 1999; Peralta & Zareei, 2016). In particular, the minimum spanning tree has been used by Onnela, Chakraborti, Kaski, Kertész and Kanto (2003). The authors Pozzi, Di Matteo and Aste (2013) use Planar Maximally Filtered Graphs, while in Zhan, Rea & Rea (2015) hierarchical clustering trees and neighbour-nets have been applied to reduce the complexity of the network, characterizing the heterogeneous spreading of risk across a financial market.

The work of Peralta and Zareei (2016) establishes a bridge between Markowitz's framework and the network theory, showing a negative relationship between optimal portfolio weights and the centrality of assets in the financial market network. As a result, the centrality measures of constructed networks can be used to facilitate the portfolio selection. In most of these papers only betweenness and eigenvector centralities are used (Výrost, Lyócsab & Baumöhl, 2019). Moreover, the algorithms incorporated are mostly variations of minimum spanning trees (Onnela, Chakraborti, Kaski, Kertész & Kanto, 2003), network clusters (Boginski, Butenko, Shirokikh, Trukhanov & Lafuente, 2013) and its subsequent variations (Clemente, Grassi & Hitaj 2022).

This thesis contributes by incorporating two additional centrality measures, namely closeness and degree centrality. Furthermore, it uses two less commonly used algorithms, namely a variation of the *k*-core smallest vertex with degeneracy ordering and the BK algorithm for determining the optimal cross maximal clique centrality. Both being used to create portfolios, the first one to maximize diversification, whilst the other to maximize return. The goals are to investigate the use of graph theory algorithms for finding the optimal weights for diversification. To critically assess and compare the performance of portfolios constructed using graph theory algorithms against an equally weighted index and mean-variance portfolios and consider the possibility of a combined approach using four different centrality measures. To the potential challenges and limitations of employing graph theory in portfolio optimization, as well as identifying its unique contributions to portfolio management. To provide a clear direction for this thesis, the following research questions have been formulated:

- RQ1: How does the performance of portfolios constructed using graph theory algorithms compare to those derived from a traditional MPT (Modern Portfolio Theory) approach, specifically during bull and bear markets?
- RQ2: Can graph-theory algorithms be seamlessly integrated into the traditional Markowitz framework to enhance portfolio optimization? Alternatively, could these algorithms replace conventional methods entirely, offering a wholly independent and superior strategy for portfolio construction?
- RQ3: In the application of graph theory in portfolio optimization, which centrality measure results in the highest Sharpe ratio of the strategy?

To answer these questions, the thesis is organized into two main parts. The first part focuses on building a solid theoretical foundation, combining the principles of the Markowitz model with graph theory centrality measures, and later detailing the application of the *k*-core and BK algorithms. In this context, using the *k*-core algorithm to shape a portfolio with an emphasis on minimizing risk, while the BK algorithm is used to form a portfolio to maximize returns. The second part adopts an empirical approach and is divided into three sections. It begins by explaining the data and methodology used in the study, then outlines the performance metrics and discusses the construction of a network of stocks based on optimal correlation thresholds. Finally, the performance of various portfolio strategies during the in-sample (2017-2019) and out-of-sample (2020-2022) periods are compared.

1 PORTFOLIO OPTIMIZATION

Portfolio optimization, the core of modern finance, seeks to curate a mix of financial assets that maximize returns while mitigating risk. This critical endeavour has long been a focus in financial research, to pinpoint the ideal distribution of assets within a portfolio to achieve targeted outcomes, such as balancing risk and returns (Qu, Zhou, Xiao, Liang & Suganthan, 2017). It is one of the problems most frequently encountered by financial practitioners and appears in various forms in the context of trading, risk management, and capital allocation (De Prado, 2013). A typical optimization model focuses on allocating limited resources among various alternatives to maximize an objective function, such as total profit. Three crucial components in any optimization problem are decision variables, the objective function, and constraints. Optimization problems without constraints are known as unconstrained optimization problems, while those with constraints are referred to as constrained optimization problems (Beasley, 2013). Problems that lack an objective function are called feasibility problems. In some cases, problems may have multiple objective functions, which can be tackled by transforming them into a single-objective optimization problem or a series of such problems (Cornuéjols, Peña & Tütüncü, 2018).

Portfolio optimization originated from Harry Markowitz's influential 1952 paper, "Portfolio Selection," where he proposed the mean-variance optimization model, a foundation of modern portfolio theory. The goal of MPT is to maximize expected returns while minimizing portfolio variance, guiding asset allocation based on each asset's unique risk-return profiles and their intercorrelations. Correlation risk, the chance that assets might move in a synchronized manner, potentially leading to larger than expected losses, can be mitigated through diversification into low-correlation assets, as described by Buraschi, Porchia & Trojani (2010).

1.1 Markowitz optimization

In the early 1960s, the concept of risk was widely discussed within the investment community, but no specific metric existed for quantifying it. To construct a portfolio model, investors needed a way to quantify this risk variable. Harry Markowitz (1952) pioneered the basic portfolio model by deriving the expected rate of return for a portfolio of assets and an associated risk measure. He demonstrated that, under a reasonable set of assumptions, the variance of the rate of return served as a meaningful measure of portfolio risk. Importantly, he established the formula for calculating a portfolio's variance. This formula not only highlighted the importance of investment diversification to reduce overall portfolio risk but also provided guidance on effective diversification strategies.

The Markowitz model is based on several assumptions about investor behavior:

- Investors perceive each investment option as characterized by a probability distribution of expected returns spanning a specific holding period.
- Investors aim to maximize one-period expected utility, and their utility functions display diminishing marginal utility of wealth.
- Investors assess portfolio risk based on the variability of expected returns.
- Investors make decisions only based on expected return and risk, so, their utility functions depend on expected return and the expected variance of returns.
- Given a specific risk level, investors favour higher returns over lower ones. Similarly, for a particular level of expected return, investors opt for lower risk rather than higher risk.

- Under these assumptions, an asset or a portfolio of assets is deemed efficient if no other asset or portfolio of assets provides a higher expected return with equal or lower risk, or a lower risk with equal or higher expected return.

To understand the fundamental Markowitz mean-variance portfolio optimization model, let's first define some notation. Let

– N represent the number of available assets (e.g., stocks or vertices),
– μ_i denote the expected return (average or mean return per time period) of asset i,
– ρ_{ij} symbolize the correlation between returns for assets i and j ($-1 \leq \rho_{ij} \leq +1$),
– σ_i stand for the standard deviation in return for asset i, and
– R signify the desired expected return from the chosen portfolio.

The expected return E[x] can be expressed as the sum of the product of the proportions and expected returns of the individual assets (Cornuéjols, Peña & Tütüncü, 2018):

$$\mathrm{E}[x] = x_1\mu_1 + x_2\mu_2 + \cdots + x_n\mu_n = \sum_{i=1}^{n} x_i\,\mu_i = \mu^T x \tag{1}$$

Var[x], can be calculated as the sum of the product of correlation coefficients, standard deviations, and proportions for each pair of assets i and j (Cornuéjols, Peña & Tütüncü, 2018):

$$\mathrm{Var}[x] = \sum_{i=1}^{n}\sum_{i=1}^{n} \rho_{ij}\,\sigma_i\sigma_j\,x_i x_j = x^T \Sigma\, x \tag{2}$$

As variance σ_p^2 is always nonnegative, it follows that:

$$x^T \Sigma\, x \;\geq\; 0\;\forall\, x \tag{3}$$

implying that Σ is positive semidefinite.[1]

Employing the standard Markowitz mean-variance approach, we can formulate the portfolio optimization problem as follows:

$$\text{minimize} \sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{ij}\;\sigma_i\sigma_j\;x_i x_j \tag{4}$$

[1]A positive definite matrix essentially implies that there are no redundant assets in the portfolio. Redundant assets would mean that there is a perfect linear relationship between some of the assets, which would result in a determinant of zero for the covariance matrix, making it not positive definite. Hence, if the covariance matrix is positive definite, it implies that no asset is a perfect linear combination of other assets (Cornuejols & Tütüncü, 2006).

$$\text{subject to} \sum_{i=1}^{N} x_i \mu_i = R \tag{5}$$

$$\sum_{i=1}^{N} x_i = 1 \tag{6}$$

$$0 \leq x_i \leq 1,\ i = 1,..,N$$

In Equation (4), our goal is to minimize the total variance associated with the portfolio. Within the Markowitz framework, risk is represented by the variance in portfolio return. Equation (5) ensures that the portfolio achieves an expected return of R. Equation (6) guarantees that the proportions of the portfolio allocations sum up to one, indicating that all available funds must be invested in assets without any remaining uninvested money. This formulation represents a straightforward nonlinear programming problem (Beasley, 2013). Frequently, instead of employing standard deviations and correlations, we express the objective in terms of covariance. Let σ_{ij} denote the covariance between the returns for assets i and j; then (given that $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$), we can rewrite the objective (Equation (4)) as:

$$\text{minimize} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \sigma_{ij} \tag{7}$$

The decision variables are represented by x_i, which indicate the proportion of the total investment associated with (invested in) asset i ($0 \leq x_i \leq 1$). These proportion variables x_i are referred to as weights. In this case, we impose non-negativity ($x_i \geq 0$). Allowing negative weights (x_i can be positive or negative) would permit shorting (Beasley, 2013). But, given that portfolio problems often involve holding the decided portfolio for an extended period, making shorting impractical, we exclude shorting here. The cvxopt.solvers.qp function is used to solve the quadratic programming problem using the interior point method (Andersen & Vandenberghe, 2023).

$$\text{minimize } \lambda x^T \sum x \tag{8}$$

$$\text{s. t.} \quad x \geq 0 \qquad \text{no shorting}$$

$$I^T x = 1 \qquad \text{budget constraint}$$

Where: x is a column vector of dimension $n \times 1$, n is the number of assets and it represents the portfolio weights, μ is a column vector of dimension $n \times 1$, representing the expected returns of the assets, λ denotes the risk aversion parameter (for simplicity it is set equal to 1), Σ is a symmetric matrix of dimension $n \times n$, representing the covariance matrix of asset returns and finally I denotes the unit vector. Alternatively we could maximize the Sharpe ratio, which involves formulating a convex quadratic programming problem equivalent to the original problem. For this, we need to make two assumptions: Firstly, the sum of all portfolio weights, x_i, equals 1 for any feasible portfolio x. This is a logical assumption as the x's represent the proportion of the portfolio invested in different asset classes. Secondly, we assume that there exists a feasible portfolio, denoted as x, such that the expected return of this portfolio, $x^T\mu$, is greater than the risk-free rate. This assumption is necessary because if all feasible portfolios have expected returns that are at most equal to the risk-free rate, then there is no benefit in optimizing - the risk-free investment would dominate all other portfolios (Cornuéjols, Peña & Tütüncü, 2018).

$$\text{maximize } \frac{x^T\mu - rf}{(x^T\Sigma\, x)^{1/2}} \tag{9}$$

$$\text{s. t.} \quad x \geq 0 \qquad \text{no shorting}$$

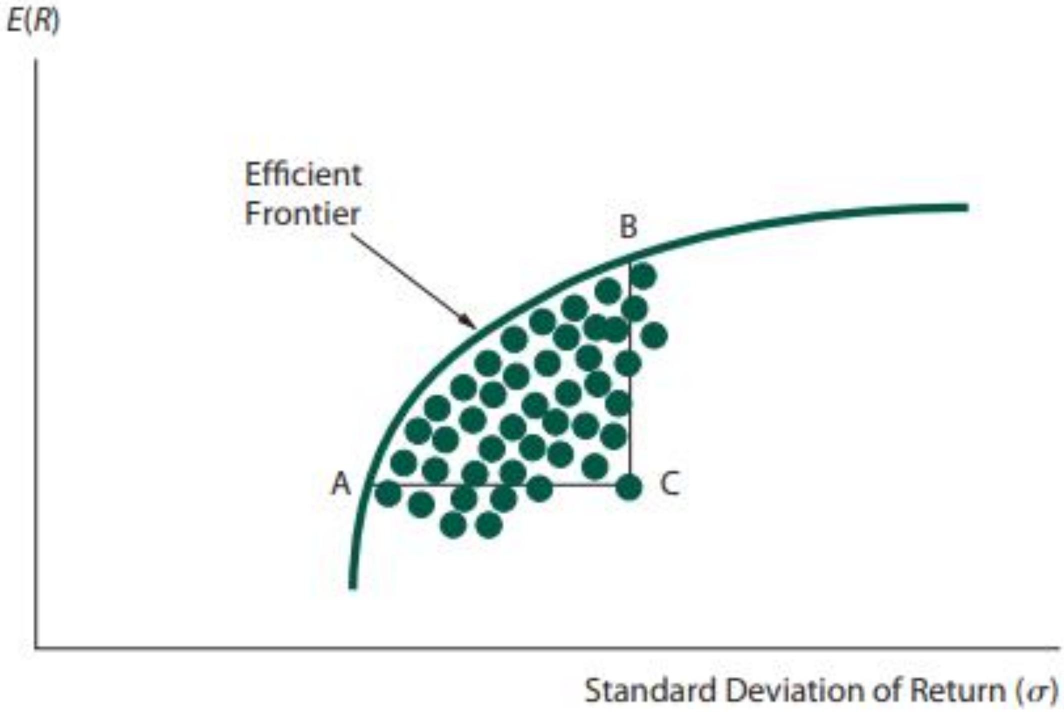
$$I^T x = 1 \qquad \text{budget constraint}$$

In this context, the quadratic objective and the positive semidefinite characteristic of σ_{ij} play important roles in enabling us to derive the optimal solution for any specific dataset in practice (Beasley, 2013). This optimization problem's outcome leads us to an important concept in portfolio theory, known as the efficient frontier. The efficient frontier embodies an array of optimal portfolios that yield the highest expected return for a given level of risk or the least risk for a specified expected return. It can be visualized as a curved line on a graph, where the x-axis stands for risk (typically represented as the standard deviation of the portfolio's returns), and the y-axis denotes the expected return.

This graphical representation elegantly encapsulates the trade-off between risk and return, forming the cornerstone of MPT. To construct an efficient frontier, we initially select a group of securities for our portfolio. For every possible mix of these securities, we estimate the portfolio's expected return (Equation 1) and risk (Equations 2-7). When these portfolios are charted on a graph, they create a cloud of points, with each point symbolizing a distinct portfolio. The efficient frontier represents the upper boundary of this cloud, illustrating the portfolios that yield the most return for their respective risk level. The portfolios under this curve are suboptimal since for the same risk level, there exist alternative portfolios on the frontier that provide higher returns.

Figure 1 provides a visual representation of the efficient frontier. Portfolios located on the efficient frontier offer either higher returns for the same level of risk or lower risk for the same level of returns compared to portfolios below the frontier. For example, in Figure 1, portfolio *A* dominates portfolio *C* since it provides the same return rate but incurs significantly less risk. Similarly, portfolio *B* dominates portfolio *C* as it offers the same level of risk but anticipates a higher return rate. Diversification among assets with imperfect correlations allows us to achieve these advantages, and thus, the efficient frontier typically consists of combinations of investments rather than individual securities.

Figure 1: Efficient Frontier for Alternative Portfolios



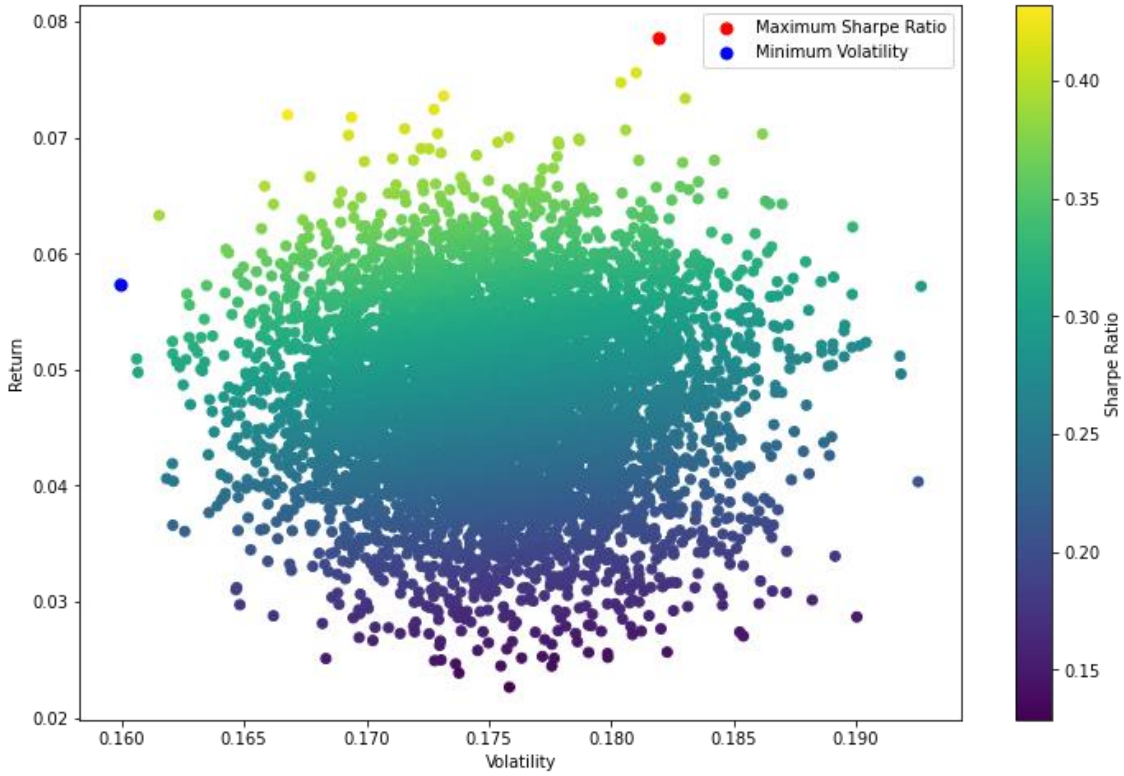
Source: Elton, Gruber, Brown & Goetzmann (2014).

Figure 2 depicts a visual representation of simulated portfolio allocations using the Euro Stoxx 50 dataset in the empirical analysis. Each point or dot on the graph represents an individual portfolio resulting from various combinations of asset allocations within the Euro Stoxx 50 constituents. The color of the dots indicates the corresponding portfolios' Sharpe ratios.

The Sharpe ratio is a measure of risk-adjusted returns, providing insights into the potential return of a portfolio relative to its assumed risk. Portfolios with higher Sharpe ratios are represented by warmer shades, while those with lower Sharpe ratios are depicted with cooler tones. By drawing a boundary line through the portfolios with the highest Sharpe ratios on the outskirts of the scatter plot, we can effectively outline the efficient frontier.

The portfolio with the highest Sharpe ratio is denoted by a red dot on the graph. This particular allocation represents the desired outcome when aiming to maximize returns while considering the level of risk. On the other hand, the blue dot represents the portfolio with the lowest volatility, signifying a more risk-averse allocation. In other words, it represents the allocation that prioritizes minimizing risk over maximizing returns.

Figure 2: Simulated portfolio allocations (Euro Stoxx 50)



Source: own work.

1.1.1 Constraints

The complexity of the portfolio optimization problem largely depends on the constraints involved (Maringer, 2008). Real-world financial constraints significantly increase the level of complexity. For example, the cardinality constraint, which requires a limited number of assets to be included in the portfolio, turns the problem into a non-convex one (Jin, Qu & Atkin, 2016). As a result, it is not always suitable to use exact methods to find optimal solutions. So, most of the current literature focuses on heuristic approaches for solving the constrained PO problem.

For the MV (Mean-Variance) model, specialized methods like the simplex method (Wolfe, 1959) and branch and bound methods can be used to find solutions efficiently. These techniques can also adjust arbitrary linear constraints, such as quantity constraints (Borchers & Mitchell, 1997). However, as the number of assets grows and additional constraints are introduced, the problem becomes increasingly more complex. For example, when the cardinality constraint is added, the problem transforms into a mixed-integer nonlinear programming problem, which is NP-hard. Bienstock (1996) introduced a branch and cut algorithm for the cardinality-constrained portfolio optimization problem, involving up to 3,897 assets with varying cardinality values. At the time of publication, it was suggested that solving larger problems to proven optimality within a reasonable timeframe might be infeasible. Some studies necessitate the inclusion of exactly K assets in a portfolio such as Chang, Meade, Beasley and Sharaiha (2000), Xu, Zhang, Liu and Huang (2010) and Jin, Qu and Atkin (2014), while others use a more relaxed version (Schaerf, 2002; Ruiz-Torrubiano & Suarez, 2010). Behr, Guettler and Miebs (2011) showed that the constrained minimum-variance portfolio with lower and upper portfolio weight constraints achieves substantial out-of-sample variance reductions in comparison to various minimum-variance portfolios.

Other common constraints used (Gilli, Maringer & Schumann, 2019):

- leverage $\|w\|_1 \leq \gamma$,
- turnover $\|w - w_0\|_1 \leq \tau$,
- max position $\|w\|_\infty \leq u$,
- sparsity $\|w\|_1 \leq K$

Where:

– $\gamma \geq 1$ controls the amount of shorting and leveraging
– $\tau > 0$ controls the turnover (to control the transaction costs in the rebalancing)
– u limits the position in each stock
– K controls the cardinality of the portfolio (to select a small set of stocks from the universe).

1.1.2 Limitations

Even though Markowitz's mean-variance optimization model is very popular, it does have its flaws. It works on the assumption that the returns you get from assets will follow a certain pattern, it heavily relies on past data to predict risk and returns, and it can be quite sensitive to changes in input data (Elton, Gruber, Brown & Goetzmann, 2014).

a) Return Distributions

The assumption about how asset returns are distributed is particularly important. If returns deviate from the pattern we expect, using variance (the measure of how spread out numbers are) as a measure of risk can skew our results (Boasson, Boasson & Zhou, 2017). In real-world scenarios, investment returns usually don't follow the normal distribution but are often skewed, following something closer to a lognormal distribution (Wang, 2023). This departure from symmetry means that variance isn't a very effective measure of risk because it doesn't differentiate between gains and losses (Boasson, Boasson & Zhou, 2017). So, investors who only rely on variance or standard deviation (how spread out numbers are) may not make the best investment decisions. Moreover, the skewness and peakedness in the distribution of returns can lead us to underestimate risk (Wang, 2023).

b) Risk Aversion and Investor Utility

The model also overlooks how risk-averse different investors can be. Basically, using variance as a measure of risk gives us a general idea of how returns can vary but doesn't consider individual investors' preferences towards risk (Boasson, Boasson & Zhou, 2017). Additionally, relying only on the mean and variance of returns might not give us a complete picture of what an investor wants. This point of view has been discussed both in academia and in by industry practitioners, highlighting the limitations of mean-variance efficiency in accurately representing what an investor wants (Michaud & Ma, 2001).

c) The Impact of Estimation Error and Unpredictability

One key downside of the Markowitz model is how sensitive it is to mistakes in estimations. Often, when creating portfolios, we treat input estimates as precise, unchanging values. But in reality, these estimates can often be wrong (Wang, 2023). This can lead to errors in the optimization process, resulting in portfolios that don't perform as expected when tested in real-world scenarios (Michaud & Ma, 2001). This sensitivity, often known as "instability" or "ambiguity," shows the model's heavy reliance on input parameters, meaning that even small changes can result in significant shifts in portfolio allocations.

d) Limited Relevance for Long-term Investment

The framework is focused on a single period, making it less suitable for investors with long-term investment goals like pension plans and endowment funds. It's another shortcoming of the mean-variance approach, as it doesn't effectively address long-term investment objectives (Michaud & Ma, 2001). The model can often underperform when tested on future data, sometimes even falling behind the simple *1/N* naive portfolio strategy (De Prado, 2016). Considering these limitations, it's clear that more advanced developments, like machine learning or in this case graph theory algorithms, might offer better strategies for portfolio optimization. Graph theory, with its ability to manage complex networks and relationships, could be an interesting solution to the issues with the Markowitz model.

1.1.3 Subsequent developments

These limitations have opened the door for new and improved portfolio optimization methods. These include models like the Black-Litterman and Konno-Yamazaki (Cornuéjols, Peña & Tütüncü, 2018) as well as methods using machine learning algorithms (Kalayci, Ertenlice & Akbay, 2019), genetic algorithms (Chen, Peng & Zhang, 2016), and models based on network theory (Gilli, Maringer & Schumann, 2019). Techniques like robust optimization, fuzzy logic, and prediction have been adopted to reduce the risk of estimation errors, with significant implications for both theory and practice. At the same time, the Black-Litterman and Konno-Yamazaki models, machine learning, and genetic algorithms have emerged to counter the limitations of the mean-variance model (Cornuéjols, Peña & Tütüncü, 2018; Kalayci, Ertenlice & Akbay, 2019; Chen, Peng, Zhang & Rosyida, 2016).

1.2 Bridging the gap between graph theory and MPT

In contrast to the limitations of MPT, graph theory offers a different perspective on portfolio optimization. It considers assets as nodes in a network, with links between nodes representing correlations between assets (Rankin & Robinson, 2013). This approach allows for a more in-depth understanding of the complex interconnections among various investment choices. It helps identify clusters of similar assets and exposes the structure of the overall network. How we treat correlations is a big difference between MPT and graph theory. MPT uses a correlation matrix, whereas graph theory provides an intuitive visual representation that simplifies the relationships between assets (Boginski, Butenko & Pardalos, 2005). For example, a concept borrowed from graph theory and applied to financial models, Minimum Spanning Trees, uses correlations to uncover the hierarchical structure in financial markets (Mantegna, 1999). Recent studies by Haluszczynski, Lau and Modest (2017) have employed centrality measures in their networks to provide insights into the behavior of assets. Tse, Liu, & Lau (2010) have successfully utilized graph theory to visualize a network of US stock prices, illustrating the complex interdependence between them.

When it comes to application, there hasn't been much written about the use of graph theory in the portfolio selection process (Peralta & Zareei, 2016). However, the existing literature suggests that networK-based asset allocation strategies can enhance risk-return profiles (Výrost, Lyócsa & Baumöhl, 2019). This research promotes combination portfolios, which evenly distribute investments between a benchmark portfolio and a set of networK-based portfolios. Another important area where graph theory has been applied is the study of interconnectedness risk (Billio, Getmansky, Lo & Pelizzon, 2012). By identifying highly interconnected assets, it can assist in portfolio diversification (Baitinger & Papenbrock, 2017). Additionally, graph theory can provide measurements and visualizations of systemic risk and can be combined with machine learning to address inequality-constrained portfolio optimization problems (López de Prado, 2016; Kaya, 2015).

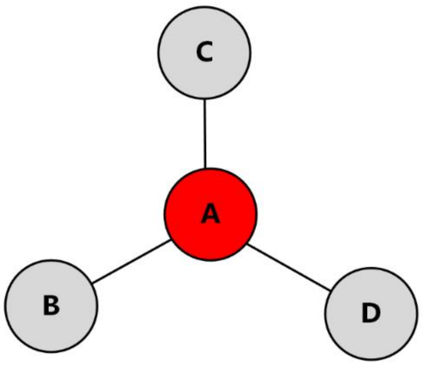
The use of graph and network theory in finance has seen significant attention and development in the past decade. For example, Mantegna (1999) proposed constructing a network where edge weights are inversely proportional to correlation, reducing network complexity. Building on this, Peralta & Zareei (2016) established a connection between Markowitz's framework and network theory, showing a negative relationship between optimal portfolio weights and asset centrality. Similarly, this thesis also adopts a graph theory-based approach to optimal asset allocation as an alternative to MPT. However, the findings suggest that asset allocation strategies based on centrality measures did not improve risk-return profiles compared to the benchmark portfolio, which differs from the research by

Výrost, Lyócsa & Baumöhl (2019). It's important to consider transaction costs and other constraints as well. Out of the four centrality measures used in the optimization problem, only one of them (eigenvector centrality) showed relatively stable performance, performing well during the in-sample period but remaining flat during the out-of-sample period. This indicates that centrality measures need to be carefully incorporated into the portfolio selection problem to yield meaningful results. In this thesis, the approach is kept simple by directly incorporating the vector of centrality measures into the objective function. Although the results may not have been outstanding, the centrality measures provide insights into the importance, influence, and "centrality" of nodes within a network. These metrics are crucial in understanding social networks, traffic systems, the internet, and more recently, financial markets. In the next section, I describe the most common centrality measures and explain how they are applied in the optimization problem.

1.2.1 Degree Centrality

Among the centrality measures used, degree centrality stands out due to its simplicity and intuitive understanding (Coppola & Elgazzar, 2020). In a financial network, a stock with high degree centrality represents an asset that has a large number of correlations with other assets. This could be a well-established company like Microsoft or Apple, whose stock price movements are closely monitored and can have a significant impact on a wide range of other assets. The high degree centrality of these assets reflects their importance and influence within the network.

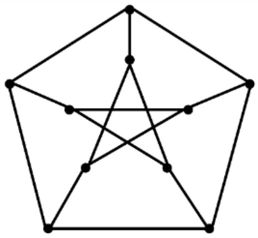
Figure 3: Small financial network of 4 stocks



Source: own work.

In figure 3, the stock with the highest degree is stock A. It is connected to 3 other stocks, thereby having a degree of 3. The degree of a vertex x represents the number of its neighbours, expressed as $d(x) = |N(x)|$. A graph G is called regular if all vertices have the same degree. The Petersen graph is regular with degree 3.[2] We the notation is used $N[x] = N(x) \cup x$ to include both the neighborhood of x and the vertex x itself.

Figure 4: The Petersen graph



Source: own work.

The concept of centrality is useful for identifying important points in a graph, which can prove crucial when the objective is to trace the origins of a virus infection or to analyze a social network (Metcalf & Casey, 2016). In the domain of financial markets, if a graph is developed to represent the connections between different asset prices, a centrality measure can be applied to highlight a central asset. This central asset could potentially be a pivotal player that drives market trends, an asset that aids others, or an asset that triggers substantial changes in the market. Reader should note, that centrality doesn't have a universal measure, it primarily involves comparing different points. Also, the choice of a centrality measure is contingent based on the graph's configuration. For example, a measure that focuses on cliques won't be effective in a graph deprived of cliques. However, for this discussion, only connected graphs are considered. Determining the significance of a point becomes challenging if it doesn't connect to some other points.

The degree centrality measure, derived from Freeman's general formula, underlines the significance of a point by considering its degree. It embodies the concept of "He with the most toys, wins," signifying that the number of neighbors a point has is significant. If the network is expansive, it should exhibit low centralization. If the centralization is high, then points with larger degrees should be seen easily in the graph (Metcalf & Casey, 2016). For the calculation of degree centralization, the point with the highest degree is first identified and referred to as $v*$. Subsequently, H is defined as $(|V| - 1)(|V| - 2)$. Degree centrality is then calculated using the provided equation (equation 10), where the top part of the sum takes into account the degree variations between the point with the highest degree and the given point. The bottom part of the sum is selected to normalize the results. If the degree of a point is identical to the highest degree, the sum remains unchanged. The closer the degree is to the highest degree, the less it contributes to the total sum.

[2] Julius Petersen. There is a graph encountered by Petersen during his research that has become famous and is named for him. The Petersen graph is a 3-regular graph of order 10 (Benjamin, Chartrand & Zhang, 2017).

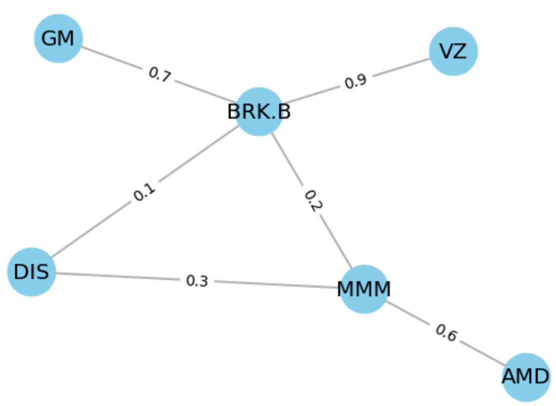
$$C_D(G) = \sum_{v \in G} \frac{[deg(v^*) - deg(v)]}{|H|} \tag{10}$$

By incorporating degree centrality into the portfolio optimization problem, the objective is to identify the optimal portfolio that not only yields the highest average return but also includes assets with a high degree of connectivity within the financial network. This approach should improve risk management as a stock with numerous connections may display behavior that is more predictable based on market trends. Here we should note that a high degree centrality could mean that a stock is influenced by many other factors and stocks, which could also make it riskier. So, we need to find a balance, and whether maximizing degree centrality is beneficial or not depends on what the financial network and model look like.

1.2.2 Closeness Centrality

Closeness centrality is a measure that indicates the nearness of a node to all other nodes in the graph (Coppola & Elgazzar, 2020). For example, a participant that has a high closeness centrality score can swiftly interact with others. The areas of interaction may span from information exchange, transaction speed, to proximity to centralized trading hubs where stocks are traded. Faster access to price information is a significant advantage, offered to those closer to the trading center, potentially giving them an edge in their trading activities. Assets marked by high closeness centrality might act as indicators of broader market trends and could be seen as indicators for the overall health of the market. Attaching assets with high closeness centrality into a portfolio can help ensure that the portfolio mirrors the performance of the market at large. For instance, in Figure 5, assets like BRK.B and MMM would have the highest closeness centrality as they are the closest to all other nodes in the graph. This indicates their robust interconnectedness and potential influence within the financial network.

Figure 5: A weighted graph representing the relationship between several stocks



Source: own work.

In the same graph, the weight of each edge, denoted by the number alongside it, symbolizes the correlation between two stocks. As an example, the edge linking stocks MMM and AMD carries a weight of 0.6, signifying a relatively strong correlation. In contrast, the edge joining stocks BRK.B and DIS has a lower weight of 0.1, suggesting a weaker correlation. This provides an initial understanding of how the Clique and Degeneracy indexes are built in the practical simulation part of this thesis.

Closeness centrality and farness centrality are closely related. The closeness is defined so that if a vertex is close to every other vertex, then the value is larger than if the vertex is not close to everything else. The closeness of a vertex is defined as:

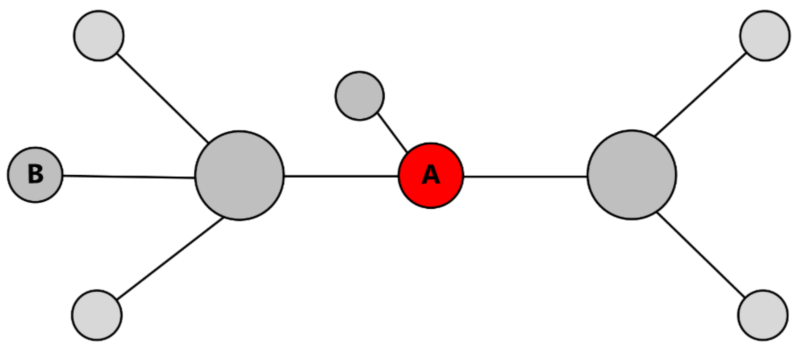
$$C(v) = \sum_{w \in G} \frac{1}{d(v,w)} \tag{11}$$

The equation implies that if the sum of the distances is large, the closeness is small, and vice versa. A stock with high closeness centrality indicates that it has close relationships with many stocks. Farness centrality is the inverse of closeness centrality, so if the closeness is small, the farness is large, and vice versa. In other words, it is the sum of all distances from the vertex v to every other vertex in the graph. If it is substantial, the vertex does not have close relationships with many vertices. That vertex is connected to the graph but is distant from most other vertices in the graph.

1.2.3 Betweenness Centrality

An asset with high betweenness centrality often serves as a kind of “bridge” within the financial market. This could be a commodity like oil, which has a direct effect on the energy sector and indirectly on other sectors such as manufacturing and transport. Assets with high betweenness centrality could be valuable for spreading out investment risks because their performance is tied to many different parts of the market.

Figure 6: Illustration of a stock with highest BC in a financial network



Source: own work.

In Figure 6, stock A has a higher betweenness centrality (BC) than stock B. This is because a change in the price of stock A will influence more stocks along a shorter path compared to stock B. Let's say the left and right sides of the stock network are not related at all. If so, picking stock A as an investment would be a good choice if we wanted to spread out our investments as much as possible. Metcalf and Casey (2016) provided an intuitive analogy to explain the concept of betweenness. They used the saying "All roads lead to Rome" to illustrate that if we consider Rome as a central point, any two cities would have to pass through Rome to connect with each other. In the context of a graph, we aim to find the equivalent of Rome, so a point that lies between all other points when considering the shortest paths between them. If there is no exact equivalent, we seek the closest approximation. This concept forms the basis of the definition of betweenness.

Let $\sigma_{(u,v)}$ be the number of shortest paths between points u and v. Then let $\sigma_{(u,v)}(w)$ stand for the number of shortest paths between u and v that go through the point w. The betweenness of a point can be found using equation 12:

$$g(w) = \sum_{u,v \in G} \frac{\sigma_{(u,v)}(w)}{\sigma_{(u,v)}} \tag{12}$$

So, betweenness is the sum of the ratio of the number of paths that pass-through a given point divided by the total number of paths that exist. If the ratio is 1, then all the shortest paths between u and v go through w. If it is 0, then none of them go through w. If the sum of all the fractions is $|V| - 1$ where $|V|$ is the number of points, then every shortest path between two points has to pass through w. In other words, we have found our 'Rome'.

1.2.4 Eigenvector Centrality

Different from degree centrality, which counts the number of connections a vertex has, eigenvector centrality considers both the number and quality of connections. This implies that an asset with high eigenvector centrality may not necessarily be linked to many other assets, but the ones it is linked to are important within the market. For instance, influential companies in the market are often leaders in areas such as AI or other technology solutions. An important feature of a network or graph is its adjacency matrix, represented as A, which is an n×n symmetric matrix in the simplest form, with n denoting the number of vertices in the network.

$$A_{ij} = \begin{cases} 1, & \text{if there is an edge between vertices } i \text{ and } j, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

If there is an edge between vertices i and j, A_{ij} equals 1, and 0 otherwise.[3] The adjacency matrix captures the connections between different nodes or vertices, and can accommodate

[3] The matrix is symmetric since if there is an edge between i and j then clearly there is also an edge

networks with weighted edges, directed networks, self-edges, and hyperedges (Newman, 2010). For centrality measures, the eigenvector centrality of a vertex, denoted by x_i, is proportional to the average of the centralities of the vertex's network neighbors. It can be mathematically represented as:

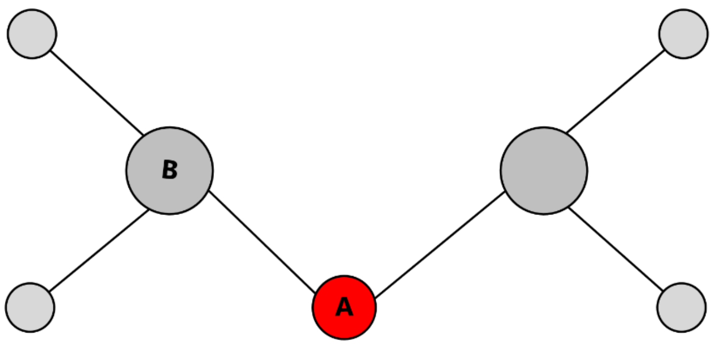
$$x_i = \frac{1}{\lambda}\sum_{j=1}^{n} A_{ij}x_j \tag{14}$$

Where, λ is a constant, and the centrality vector $x = (x_1, x_2, \ldots x_i)$. Rewriting this equation in matrix form gives:

$$\lambda x \ = \ Ax \tag{15}$$

This implies that x is an eigenvector of the adjacency matrix with eigenvalue λ. To ensure that the centralities are non-negative, λ must be the largest eigenvalue of the adjacency matrix and x must be the corresponding eigenvector, as stated by the Perron–Frobenius theorem (Knill, 2011). This form of eigenvector centrality gives each vertex a centrality based on the number and quality of its connections. A vertex with a larger number of high-quality contacts may have higher centrality than one with a greater number of lower-quality contacts. A well-known application of eigenvector centrality is Google's PageRank algorithm for web page ranking (Bryan & Leise, 2006).

Figure 7:: Illustration of a stock with highest EC in a financial network



Source: own work.

In Figure 7, it can be seen that two stocks with the same degree centrality can have different eigenvector centralities. Although both stocks *A* and *B* are linked with two other stocks, the importance of the stocks they are connected to differs. Stock *A* is linked with stocks that have greater influence in the market, as shown by their size and the number of other connections they have. In the context of portfolio optimization, eigenvector centrality can be used to identify important assets in the market network, considering not just the number

between *j* and *i*. So $A_{ij} = A_{ij}$.

but also the quality of connections. This measure provides insights into how closely an asset is connected to influential or well-connected assets, and can aid in constructing a well-diversified portfolio that accounts for both the asset's performance and its interconnectedness within the broader market.

1.2.5 Using centrality measures in the optimization problem

The use of centrality measures in portfolio optimization allows investors to align their investment strategy with their market outlook and risk tolerance. The choice of the centrality measure can affect the portfolio's performance based on the specific characteristics of each centrality measure. Investors who anticipate that well-connected, influential companies are key drivers of market trends might prioritize degree centrality and eigenvector centrality (Kaya, 2015). These measures could potentially maximize returns during periods of market growth, given that they focus on nodes with many connections, such as degree centrality or nodes connected to other well-connected nodes in the case of eigenvector centrality. But, their performance might be worse during periods of market instability or downturns because these influential assets are also likely to be at the epicenter of market corrections. On the other hand, investors seeking diversification or aiming to mirror the overall market trend might focus on closeness centrality and betweenness centrality (Peralta & Zareei, 2016). Closeness centrality emphasizes assets that are close to all other assets, providing a diversified exposure to the market. Betweenness centrality focuses on assets that usually act as bridges between other assets, potentially making them less correlated with specific market segments and providing diversification benefits. But still, these measures might not provide the highest returns in strong bull markets, where a concentrated portfolio in influential companies might outperform (i.e., Clique or even Eigenvector Index)

Incorporating these centrality measures can be done in various ways (Peralta & Zareei, 2016), and one straightforward method is to include them directly into the objective function of the portfolio optimization problem. By extending the traditional approach of minimizing portfolio variance to also controlling for the centrality measures, investors can potentially balance their portfolio's risk and return while factoring in their view on market structure and connectivity. This can be written as:

$$\text{minimize } \frac{1}{2} x^T \sum x - c^T x \tag{16}$$

$$\text{s. t.} \quad x \geq 0 \quad \text{no shorting}$$

$$I^T x = 1 \quad \text{budget constraint}$$

Where c is a column vector of dimension $n \times 1$, representing the combined return (realized return and centrality measure). The term $- c^T x$ means we are subtracting the combined return, so in essence, we are aiming to minimize the negative of the combined return, which equates to maximizing the combined return. By solving this optimization problem, we're looking for the portfolio with the best trade-off between risk and return, adjusted for centrality.

Among the challenges of applying centrality measures to portfolio optimization is the need for significant computational power. This need arises because complex optimization problems might not be solvable in linear or polynomial time (Skiena, 2012). However, using approaches such as the maximum clique problem, which seeks the largest complete subgraph in a graph, can help identify highly interconnected asset clusters (Conte, 2021). To solve this, algorithms like the Bron-Kerbosch algorithm have been developed. They are described in the next chapter, specifically the *k*-core with degeneracy ordering and the BK algorithm (Matula & Beck, 1983; Chin, Chen, Wu, Ho, Ko & Lin, 2014). These algorithms help identify an optimal independent vertex set and a highly concentrated group of assets, respectively. For the in-sample (2017-2019) and out-of-sample (2020-2022) performance, it can be expected that different centrality measures will achieve returns based on the market sentiment and volatility. For example, in stable or growing market conditions, degree centrality and eigenvector centrality might provide higher returns due to their focus on well-connected, influential assets. Conversely, in volatile or declining markets, closeness centrality and betweenness centrality might offer better portfolio performance due to their diversification benefits. The results are shown in the Practical simulation part of the thesis.

2 GRAPH THEORY ALGORITHMS

In this part, the two graph theory algorithms used for constructing the indexes are described. There are many choices available, and even more variations. The ones selected are the *k*-core algorithm using degeneracy ordering and the BK algorithm, based on the study and research of GitHub user Je-Suis-Tm (2023). Both algorithms are part of graph theory and utilize similar concepts but leverage them to obtain completely different results. The main purpose of the *k*-core algorithm is to locate the vertices that are the least connected to all other nodes in the graph.

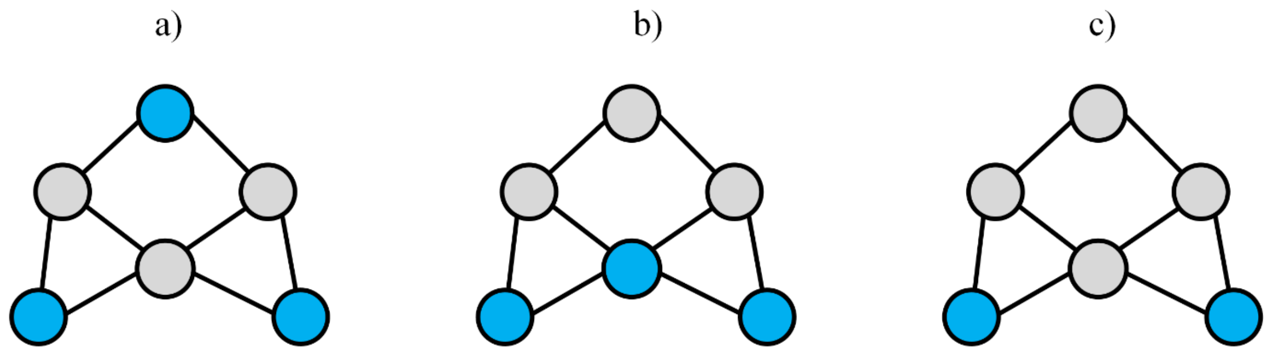
In portfolio optimization, this helps us achieve a well-diversified portfolio based on how we define the correlations between assets. For instance, suppose we prefer stocks in our portfolio that share at least a 40% correlation with others. The algorithm would then eliminate all stocks that do not meet this 40% correlation requirement. As a result, all stocks in the finalized portfolio will not exceed a 40% correlation with one another. On the other hand, the BK algorithm uses correlations to determine the most correlated stocks in the network. This strategy is better suited for investors with a higher risk tolerance who prefer a highly concentrated asset combination. Again, this is achieved by defining a set threshold.

2.1 Degeneracy Index

An independent set is the opposite of a clique in a graph. It is a collection of vertices where no vertex is adjacent to any other vertex. In the context of a social network, an independent set represents a group of people who are all strangers to one another. Researchers in graph theory are often interested in finding the largest clique and the largest independent set in a graph, which can be challenging tasks in large graphs (Rankin & Robinson, 2013).

In a graph G, a subset X of $V(G)$ is considered independent if the subgraph of G induced by X is null.[4] The size of the largest independent subset of $V(G)$ is denoted by $\alpha(G)$. A graph H is considered a subgraph of another graph G if the vertices of H form a subset of the vertices of G, and the edges of H constitute a subset of the edges of G.

Figure 8: Example of what is considered an independent set



Source: own work.

Figure 8 illustrates the idea of an independent set. Readers can notice that in graph a), the blue dots are not connected, thereby forming an independent set. The same cannot be said about example b) as the blue dots are connected, so b) is not an independent set. To be considered an independent set, the middle blue dot would have to be removed. Example c) is an independent set, however, it's not a maximal independent set. A maximal independent set is shown in example a) as there are no more available dots that could be added that are independent of one another. How to achieve the maximal independent set in linear time is shown in this section with the use of the k-core algorithm.

2.1.1 k-core

In simple terms, a k-degenerate graph is a type of undirected graph where every subgraph has at least one vertex connected to k or fewer edges in that subgraph (Bader & Hogue,

[4] A subgraph is said to be induced by a set of vertices if it includes all the vertices in the set and includes an edge between any two vertices if the edge is present in the original graph (Kloks & Xiao, 2022).

2003). The degeneracy of a graph refers to the smallest value of k that makes it a k-degenerate graph.

Degeneracy has been given various names, such as k-core number (Bader & Hogue, 2003), width (Freuder, 1982), and linkage (Kirousis & Thilikos, 1996). It is also closely related to the coloring number and the Szekeres-Wilf number (Szekeres & Wilf, 1968). k −degenerate graphs have sometimes been called k −inductive graphs as well (Irani, 1994). Lick and White introduced the term k −degenerate in 1970. The concept has been presented under different names before and since then. Notably, 'k −dense tree' has been used for 'k −degenerate graph', and 'k −arch graph' has been used for 'maximal k −degenerate graph'.

The idea of a k-core was initially introduced to investigate the clustering structure of social networks (Seidman, 1983) and to describe the evolution of random graphs (Bollobás, 1984). A comprehensive overview of the subject, including key concepts,[5] essential algorithmic methods, and some application areas, can be found in Malliaros, Giatsidis, Papadopoulos & Vazirgiannis (2019). The fundamental properties of maximal k-degenerate graphs were initially established by Lick and White (1970) and Mitchem (1977). An early overview of the findings can be found in Pereira's work titled: "A survey of K-degenerate graphs" (1976).

Figure 9 illustrates the iterative process of determining the k-core of graph G.[6] The 3-core of graph G is $2K_4$ (represented by two K_4 subgraphs). Note, that the final step is not depicted since the 4 −core requires each vertex to have a degree of at least 4. In our case, all vertices have a degree less than four, leading to their removal. Consequently, the 4-core of graph G either corresponds to the null graph, containing no vertices or edges, or it simply does not exist.

This process illustrates how the degeneracy index is constructed later on. For example, the first step is determining the highest k-core number. In this example, that is 4. The next step is to apply degeneracy ordering in linear time, so the nodes are sorted in ascending order from the smallest degree to the largest. The algorithm then picks the lowest degree stock and appends it to the independent set. It then checks whether the next stock is correlated to any other stock already in the independent set. If it's not, then it's added and repeated until no stocks are left.[7] In Figure 9 (step 5), this would mean that for a 4-core ordering, we would have picked 1 stock from each subset.

---

[5] Graph coloring is very similar for identifying k-cores in a graph. This is because a graph can be k-colored if and only if all its vertices have degree less than k. The k-core of a graph is the largest subgraph where every vertex has a degree of at least k. Therefore, a k-core cannot be k-colored, because all of its vertices have degree k or more. By removing vertices from a graph one by one in order of their colors, the remaining graph after removing the first i colors will be a subgraph where every vertex has degree at least i. Therefore, the k-core is the remaining graph after removing the first (k-1) colors (Saoub, 2021).

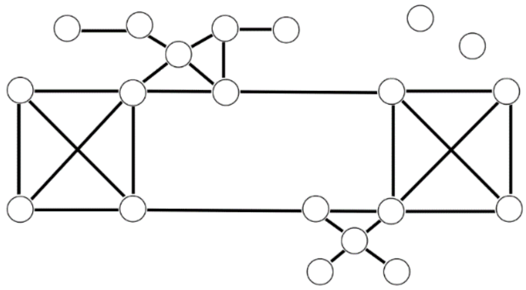
[6] Repeated execution of a block of code as long as a certain condition is met. This is usually achieved through structures like for-loops and while-loops. Each execution of the loop is called an iteration.

[7] Some definitions of k-core require the k-core to be connected. If that is the case, k-cores lose their

Figure 9: Successive cores of a particular graph G

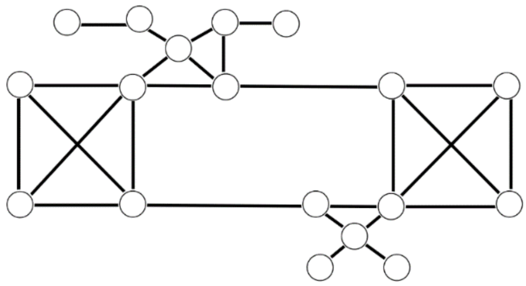
Step 1: Determining the $0-$ core.

The $0-$core of a graph is the maximal subgraph where every vertex in the subgraph has a degree of at least zero. Since every vertex in every graph has a degree of at least zero, the $0-$ core of a graph is the entire graph itself.



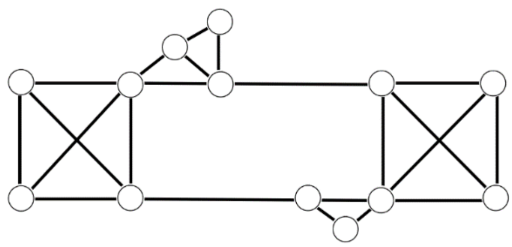
Step 2: Determining the 1 – core

The $1-$core of a graph is the maximal subgraph where every vertex has a degree of at least one. Therefore, we only need to remove the isolated vertices (vertices with degree zero) to get the $1-$core of the graph.



Step 3: Determining the 2 – core

The $2-$core of a graph requires every vertex to have a degree of at least two. Therefore, vertices that couldn't be in the $1-$core cannot be in the $2-$core either. We systematically delete vertices that have a degree less than two until no more such vertices are left.



Step 4: Determining the 3 – core

Similar to the $2-$core, the $3-$core requires every vertex to have a degree of at least three. We systematically delete vertices that have a degree less than three until no more such vertices are left. Therefore the $3-$core of G is $2K_4$.

---

uniqueness, and every individual connected subgraph would be a k-core.

Adapted from Wrath of Math (2019).

The Smallest-Last Vertex Ordering algorithm, developed by David W. Matula and Leland L. Beck in 1983, introduced a linear-time approach to determining the degeneracy ordering of a graph. Though innovative and efficient, we have decided against utilizing this algorithm in our present work given that it is not accessible in the Python package we are using. Instead, we will be using a potentially more effective version developed by Vladimir Batagelj and Matjaž Zaveršnik from the University of Ljubljana. Their algorithm significantly improves the process of network core decomposition. The main premise is that if all vertices with a degree less than *k* are recursively removed from a given graph, the remaining graph will be the *k*-core. The algorithm's execution time is directly proportional to the number of edges, denoted by '*m*', making it an *O(m)* algorithm. In the experimental section of this study, each stock within the index is treated as a vertex in a graph, with their correlations represented as edges. Degeneracy ordering is used to systematically isolate the least correlated stocks. Conceptualize this as a continuous procedure where we persistently identify and remove the stock with the smallest degree, essentially the one with the least correlation with others in the portfolio. The degeneracy of the graph, and consequently our portfolio, is then established by the stock that had the highest degree or most correlations at the time it was removed.

The procedure to identify the optimal set of least correlated stocks, represented as the maximal independent vertex set in our graph, is as follows:

1. Obtain the degeneracy ordering in linear time (Matula & Beck, 1983).
2. Select the vertex with the lowest order.[8]
3. If the selected vertex is not connected to any other vertices currently in the output set, it is added to our output set.
4. Steps 2 and 3 are repeated until all vertices have been considered, ensuring that every stock in the portfolio has been evaluated.

The end goal is to systemically select the least correlated set of stocks within the portfolio.

2.2 Clique index

Kloks & Xiao (2022) define a clique as a subgraph that is a complete graph, meaning every vertex in the subgraph is adjacent to every other vertex in the subgraph. The clique-size of a graph, $\omega(G)$, is the largest integer *n* such that K_n is a subgraph of G but K_{n+1} is not (Saoub, 2021). We can think of cliques in the context of a financial network, where vertices represent

[8] Stock with the least correlations.

stocks and edges are their correlations. In this case, a clique would be a selection of stocks which are all correlated.

In a graph G, a clique is a nonempty set $C \subseteq V$ such that any two vertices of C are adjacent. A clique in G is an independent set in the complement graph $\bar{G}$, and vice versa. The maximal cardinality of a clique in G is denoted by $\omega(G)$, then

$$\omega(G) = \alpha(\bar{G})$$

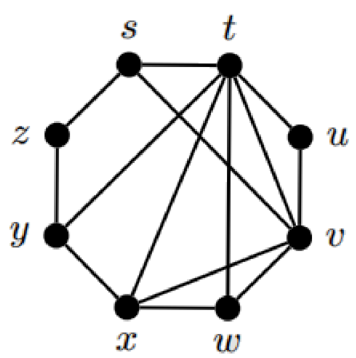
A clique with three vertices is called a triangle. Bipartite[9] graphs have no triangles, as they are odd cycles.[10] Consequently, $\omega(G) \leq 2$ when G is bipartite. A graph is called a clique if every pair of its vertices are adjacent. In this case

$$\omega(G) = |V(G)| = \chi(G) \quad and \quad \alpha(G) = 1\,.$$

On the other hand, a graph is an independent set if it has no edges, resulting in

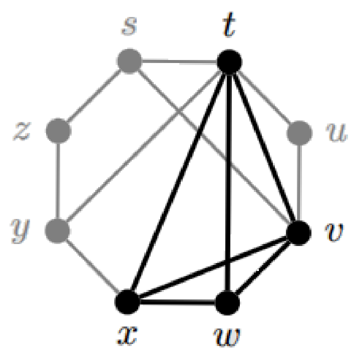
$$\omega(G) = \chi(G) = 1 \quad and \quad \alpha(G) = |V(G)|\,.$$

Figure 10: G_1 with removal



Source: own work.

Figure 11: G_1 with removal



Source: own work.

In graph G_1, vertices *t,v,w,* and *x* are all connected, as illustrated. However, it's impossible to find a set of 5 vertices that are all interconnected, due to insufficient vertices with a degree of at least 4. Therefore $\omega(G_1) = 4$. This serves as the basis for constructing the clique index, which will be a portfolio of maximally connected stocks. But determining a clique of that size is a computationally difficult task, so an optimal algorithm has to be used, in this case, the BK algorithm.

[9] Bipartite graphs are a type of graph where the vertices can be divided into two distinct sets, and every edge connects a vertex in one set to a vertex in the other set. There are no edges within a set (Benjamin, Chartrand & Zhang, 2017).
[10] A cycle is a non-empty path in which the first node is also the last one, meaning, it forms a loop. Each edge and vertex is distinct in a cycle, except for the first and last vertex, which are the same. A triangle, which has three vertices, is the simplest example of an odd cycle (Kloks & Xiao, 2022).

2.2.1 Cross-Maximal Clique Centrality

Cross-maximal clique centrality (CMCC) is a unique centrality measure used for understanding the connectivity of a node to different cliques within a network. This is the only centrality measure that is not directly used in portfolio optimization but is used to form the Clique index (Kloks & Xiao, 2022). Given the complexity and computational inefficiency associated with calculating CMCC due to the NP-hard nature of the problem, incorporating it directly into the portfolio optimization framework would not be feasible.

This measure was initially proposed by Borgatti, Jones and Everett (1998) as clique-overlap centrality and was later used by Faghani and Nguyen (2013). It is important in identifying key stocks in our index. A high degree of centrality, or in the specific case of CMCC, a high degree of connectivity to multiple cliques, signifies the asset's importance in the market network. Unlike degree centrality, which focuses on the number of connections a node has, CMCC underscores the number of strongly connected groups, meaning cliques that include that node.

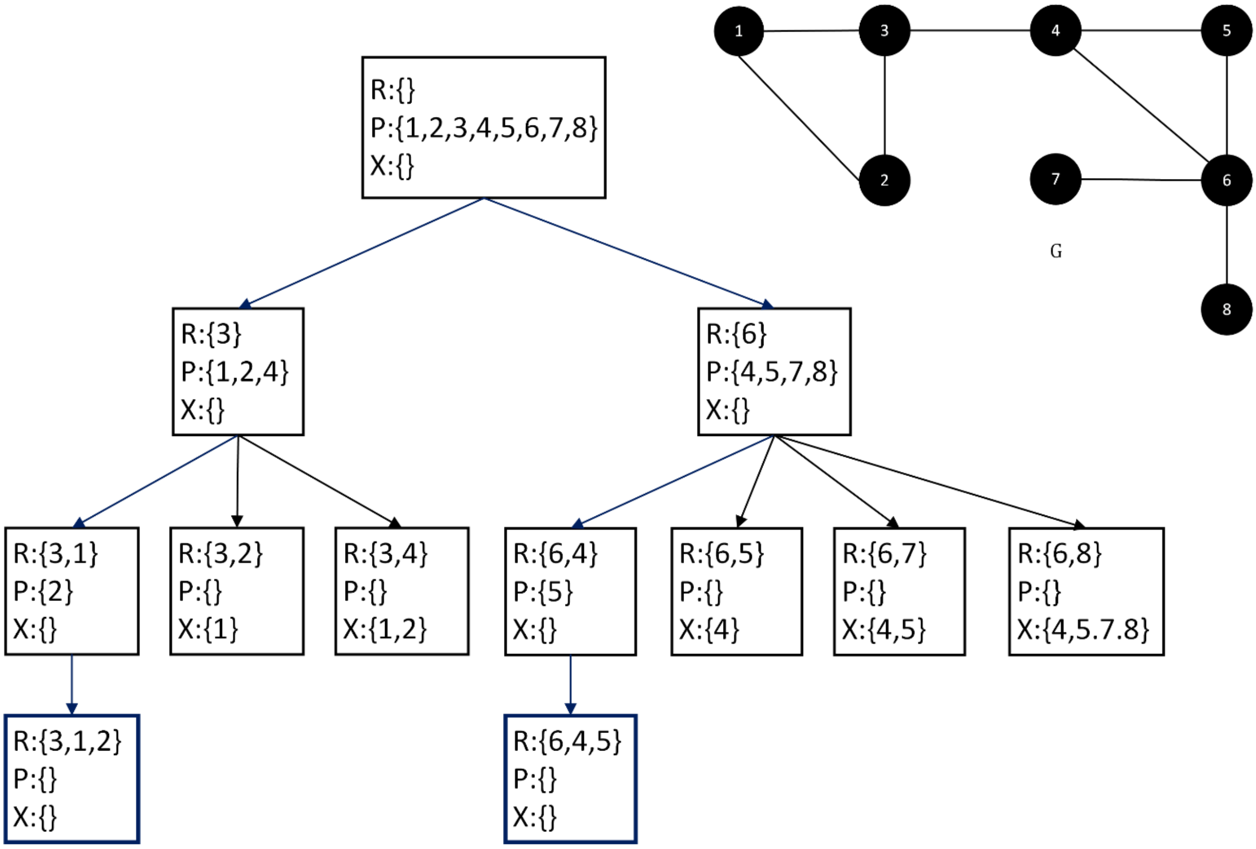
It is interesting and important to mention that an asset may have a high degree centrality but lower CMCC if it isn't a part of many maximal cliques. Conversely, an asset with fewer connections might have a higher CMCC if it's part of many maximal cliques (Coppola & Elgazzar, 2020). The Bron-Kerbosch algorithm, which works via recursive backtracking, is the most commonly used method to locate the maximal cliques in a graph. But since this algorithm is computationally expensive, it limits its practical application in portfolio optimization, especially when the number of securities in the portfolio is large.

2.2.2 Bron-Kerbosch algorithm

The Bron-Kerbosch enumeration algorithm is the most efficient method for finding maximal cliques in an undirected graph. It lists all subsets of vertices that have the properties of being completely connected, and no listed subset can have any additional vertices added to it while preserving its complete connectivity.

The algorithm was designed by Dutch scientists Coenraad Bron and Joep Kerbosch, who published its description in 1973 (Akkoyunlu, 1973). There are essentially three variations of the algorithm: the original without pivoting, a version with pivoting introduced by Tomita, Tanaka, & Takahashi (2006), and a version with vertex ordering proposed by Eppstein, Löffler and Strash (2010). For the construction of the Degeneracy index, the variation with pivoting is used.

Figure 12: BK Algorithm with pivoting

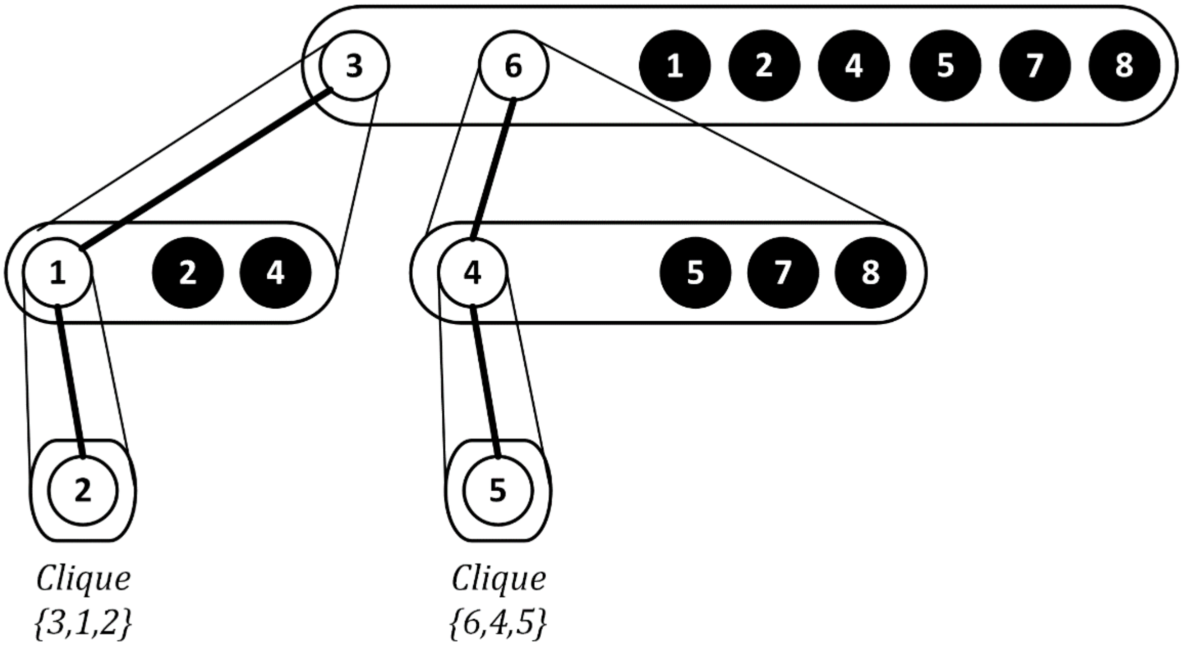


Source: own work.

Figure 12 depicts two maximal cliques: {3,2,1} and {6,4,5}. Once the first clique is identified, the algorithm proceeds to assess neighboring vertices that have not yet been evaluated. In this case, these are vertices 2 and 4. As vertex 2 has already been considered, it gets moved to set *X*. Similarly, vertex 4 follows the same procedure. The crucial part is the selection of the pivot once all neighboring vertices are evaluated. The chosen pivot is the one with the maximum intersection with set *P*, effectively minimizing the search area as vertex 3's neighbors have already been processed.

As a result, vertex 6 is chosen as the new pivot. The BK algorithm repeats this procedure, eventually finding the second maximal clique. Figure 13 illustrates this process using a different search forest, where the black vertices indicate those already processed or those not considered potential candidates for forming larger cliques.

Figure 13: BK Algorithm with pivoting



Source: own work.

2.3 P, NP, NP-Hard, NP-complete

In computer science, problems are classified based on their relative computational complexity into several categories: P, NP, NP-Complete, and NP-Hard. These classes provide valuable insight into the computational resources needed to solve a problem and significantly influence algorithm selection and design (Pokharel, 2020). Understanding these categories is particularly pertinent in the context of portfolio optimization, where both the volume and complexity of the data involved can be considerable.

- $O(1)$ – constant-time
- $O(log2(n))$ – logarithmic-time
- $O(n)$ – linear-time
- $O(n^2)$ – quadratic-time
- $O(n^k)$ – polynomial-time
- $O(n^n)$ – exponential-time
- $O(n!)$ – factorial-time

Where k is a constant and n is the size of the input (Pokharel, 2020). To classify a problem into any one of these classes, it must be computable, meaning there must exist some algorithm that can solve the problem. Problems that can be computed are often referred to as "solvable", "decidable", or "recursive" problems (Sharma, 2020). On the other hand, noncomputable problems are those for which no algorithm exists to solve them. In general, we categorize problems into four distinct classes:

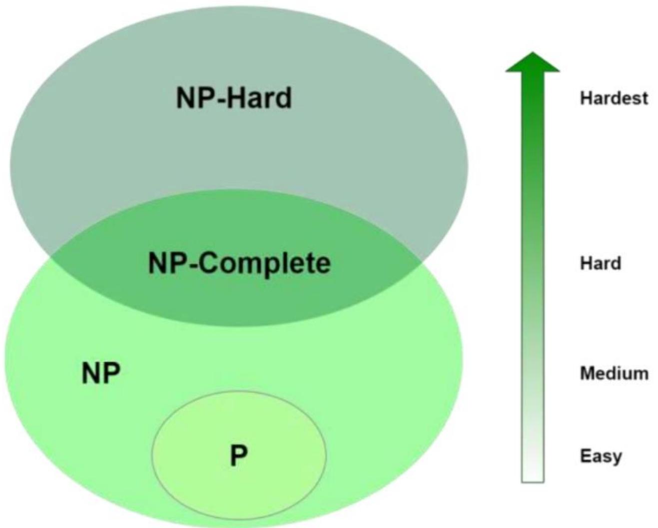
1. Class P: Class P includes problems that can be solved by deterministic algorithms within polynomial time, denoted as *O(p(n))* where *p(n)* is a polynomial function of *n*. This class is of particular importance for portfolio optimization problems. Convex optimization problems, including quadratic optimization, which is a key component of Markowitz optimization, generally fall within this class as there exist polynomial-time algorithms that efficiently solve them (Erciyes, 2021).

2. Class NP: The class NP contains decision problems that can be verified in polynomial time, even if finding the solution might not be as efficient (Erciyes, 2021).

3. Class NP-Complete: An important subset of NP, NP-Complete problems are as hard as the hardest problems in NP (Erciyes, 2021).

4. Class NP-Hard: The NP-Hard class includes problems that are at least as hard as the hardest problems in NP. The task of identifying a clique with the highest group betweenness centrality value in a graph, for example, is a problem of this class and is often solved using the Bron-Kerbosch algorithm in our context (Rysz, Mahdavi, & Pasiliao, 2018).[11]

Figure 14: Easy-to-Hard scale



Source: Pokharel (2020).

The distinction between these complexity classes and understanding where a problem lies is vital to comprehend the computational effort required and the trade-off between the precision of the solution and the computational resources. This is particularly important in

[11] The Clay Mathematics Institute emphasises the importance of this problem by offering a million-dollar reward for anyone who can definitively prove or disprove that P = NP (Dickson, 2020).

portfolio optimization where trade-offs between risk and return are fundamental. The Bron-Kerbosch algorithm is known to be NP-Hard and it is used in the calculation of the Clique index. Its worst-case running time of $O(3^{\frac{n}{3}})$ aligns perfectly with the upper limit of maximal cliques as established by Moon and Moser (1965), thus making it an optimal algorithm for the task despite its high complexity class (Segundo, Artieda & Strash, 2018). On the other hand, the problem of detecting the maximum independent set via degeneracy ordering, which we tackle using the Smallest-Last Vertex Ordering algorithm or its variant developed by Vladimir Batagelj and Matjaž Zaveršnik, falls within the P category, specifically when implemented on certain types of graphs, such as trees (Heinold, 2019). This makes the Degeneracy index computationally more tractable, even for large portfolios.

3 DATA

3.1 Data and benchmark selection

To conduct the empirical portion of the thesis, an index with a moderate number of components was required. Consequently, the Euro Stoxx 50 index was chosen. The Euro Stoxx 50 is a stock index of Eurozone stocks, designed by STOXX, an index provider owned by Deutsche Börse Group. This index comprises 50 stocks from 11 Eurozone countries (Barone & Barone, 2022). The index is calculated using the Laspeyres formula, which measures price changes against a fixed base quantity weight (STOXX, 2023).

Given the long-term superior performance of equal-weighted indices compared to market capitalization-weighted indices, we have decided to include the former as an additional benchmark. The choice of an equal weight index is backed by a body of research that showcases its advantages in certain market conditions. One of the appealing characteristics of equal-weighted portfolios is their inherent mean-reversion tendencies (Maillard, Roncalli, & Teïletche, 2010). Since De Bondt and Thaler (1985), numerous authors have argued that mean-reversion occurs in stock markets, implying that buying past underperformers and selling past outperformers could lead to higher returns. This strategy could be incorporated into the selection criterion, constructing a portfolio that is maximally oversold, subject to constraints. However, persistence in the optimization criterion is not necessarily desirable, implying the objective function might contradict what we intend. Despite these complexities, it is worth noting that mean-reversion often occurs and could be a profitable investment strategy under the right conditions. Ernst, Thompson, and Miao (2017) empirically demonstrated the superiority of an equally weighted S&P 500 portfolio over Sharpe's market capitalization weighted S&P 500 portfolio from 1958 to 2016. Interestingly, there have been periods where the equal-weighted portfolios underperformed the market capitalization-weighted portfolios, as shown by Taljaard and Maré (2021). The underperformance in these instances was attributed to increased concentration in the market capitalization-weighted portfolio and a significantly lower level of diversification benefits.

Despite these short-term periods of underperformance, the evidence leans towards equal-weighted portfolios tending to outperform their market capitalization-weighted counterparts in the long term. Taljaard and Maré (2021) also suggested an approach to improve portfolio performance by dynamically selecting a market cap or an equal weighting using a basic linear regression model. This provides investors with a strategy to potentially enhance their returns depending on market conditions. Moreover, Plyakha, Uppal and Vilkov (2012) further proved the superior performance of equal-weighted portfolios over value- and price-weighted portfolios over the last four decades. They discovered that, despite the higher portfolio risk, the equal-weighted portfolio outperforms in terms of total mean return, four-factor alpha, Sharpe ratio, and certainty-equivalent return. This outperformance is attributed to the higher return for bearing systematic risk and a higher alpha measured using the four-factor model of the equal-weighted portfolio. Intriguingly, the higher alpha of the equal-weighted portfolio is due to the monthly rebalancing needed to maintain equal weights. This mechanism effectively acts as a contrarian strategy that exploits the reversal and idiosyncratic volatility of stock returns.

3.2 Data preparation and cleaning

Necessary libraries such as Pandas and "yfinance" were imported for data manipulation and retrieval. The Euro Stoxx 50 index components were defined in a list, and the data-fetching date range was set from January 1, 2017, to May 1, 2023. A loop iterated through each ticker, downloading historical stock data via the Yahoo Finance API. Missing values were replaced with the average of the previous three values.

3.3 Environment

The practical simulation was executed using various open-source Python packages. The following packages were utilized for specific purposes:

- *Os*: This package was used for changing the working directory to the appropriate folder containing the data and other relevant files.
- *Pandas*: Used for data manipulation, handling, and processing tasks.
- *Networkx*: For creating, analyzing, and visualizing complex networks and graph structures.
- *Matplotlib.pyplot*: This library was employed for generating visualizations and plotting graphs to better understand the data.
- *Matplotlib.lines*: Customized line appearances in visualizations for clarity.
- *Numpy*: Used for numerical operations and linear algebra calculations.
- *Seaborn*: For creating more advanced and aesthetic visualizations, such as heatmaps.
- *Cvxopt*: Employed for solving convex optimization problems, such as the optimization of the portfolio in the simulation.

4 PRACTICAL SIMULATION

4.1 Performance metrics

To evaluate the performance of the constructed portfolios, several widely accepted metrics were used.

4.1.1 Average Annual Rate of Returns (AARR)

This measure provides an insight into the annualized profitability of the index. It is calculated by first converting the daily returns (r) into daily growth rates ($1+r$), and then taking the product of these growth rates over the given period. The cumulative product is then raised to the power of the ratio of the number of trading days in a year (n) to the total number of trading days (N), before subtracting 1 to convert back to a return and multiplying by 100 to convert to a percentage. This approach effectively calculates the geometric mean return, which is then annualized (Reilly & Brown, 2012).

$$AARR = \left(\left(\prod_{i=1}^{N} (1 + r_i) \right)^{\frac{n}{N}} - 1 \right) \times 100 \tag{17}$$

While the metric provides a good sense of the portfolio's profitability, it does not account for the risk associated with the portfolio.

4.1.2 Annual Volatility

Volatility is in this case interpreted as a measure of risk, providing a view of the portfolio's inconsistency or the standard deviation of its returns. It is calculated by multiplying the standard deviation of daily returns σ_r by the square root of the number of trading days in a year ($\sqrt{N}$) (Reilly & Brown, 2012).

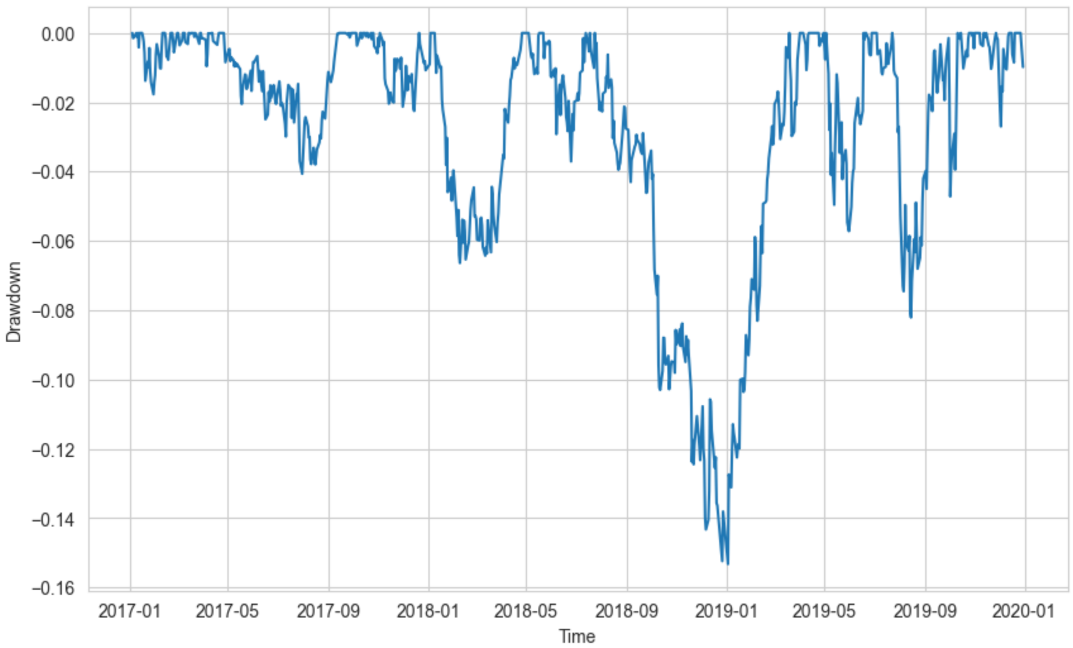
$$Annual\ Volatility = \sigma_r\ *\ \sqrt{N} \tag{18}$$

Despite its widespread usage, the reader should note that annual volatility is a simplistic measure of risk. It assumes that returns are normally distributed and do not capture severe losses, often associated with tail risks.

4.1.3 Maximum Drawdown (MDD)

MDD is a purely empirical and backwards-looking measure of risk, not associated with any probabilistic distribution. Originating from the work of Zhou and Grossmann (1993), this measure is widely used by institutional investors. Drawdown refers to a drop in the portfolio's value compared to its previous peak, whereas Maximum Drawdown specifically identifies the most significant drop, expressed as a percentage. This is a critical distinction from metrics like VaR and Expected Shortfall, where they estimate potential future adverse scenarios, whereas MDD measures the worst loss that has already occurred in the past (Basile & Ferrari, 2016). An illustration of the Maximum Drawdown can be seen in Figure 15, where the drawdown over the in-sample period is plotted for the EW index. In this example, a maximum drawdown of approximately -15.5% means that at its lowest point during the in-sample period, the index experienced a decline of 15.5% from its previous peak.

Figure 15: Drawdown over time – EW index



Source: own work.

The maximum drawdown can be calculated as:

$$MDD = min(DD_t) \tag{19}$$

$$DD_t = min(\frac{VM_{P,t}}{max(VM_{P,0\ldots t})} - 1;\ 0) \tag{20}$$

Where t represents a given point in time, ranging from 0 to T. DD_t is the drawdown at time t. $VM_{P,t}$ is the portfolio value at time t and $max(VM_{P,0\ldots t})$ is the maximum portfolio value between time 0 and time t. In essence, MDD captures the largest peak-to-trough decline during a specific period, representing the portfolio's worst historical loss.

4.1.4 Sharpe Ratio

The Sharpe ratio is a widely used metric for evaluating risk-adjusted returns. It is defined as the difference between the portfolio's return (R_p) and the risk-free rate (R_f), divided by the portfolio's standard deviation of returns (σ_p) (Reilly & Brown, 2012).[12]

$$\frac{(R_p - R_f)}{\sigma_p} \tag{21}$$

Sharpe ratio is dimensionless, and thus it does not have units of measurement. Its interpretation is straightforward: a higher Sharpe ratio indicates a higher excess return for the amount of risk taken. The limitation of it is that it assumes returns are normally distributed and the ratio does not handle well the situations with large deviations (Reilly & Brown, 2012).

4.1.5 Returns Over Maximum Drawdown (RoMaD)

RoMaD is a performance metric that compares the return of an investment relative to its worst-case scenario loss (maximum drawdown). It's calculated by taking the annual return of a portfolio and dividing it by the absolute value of its maximum drawdown (Pfaff, 2016).

$$RoMaD = \frac{AARR}{|MDD|} \tag{22}$$

For example, if a portfolio has an AARR of 10% and an MDD of -20%, the RoMaD would be 0.5. This means that for every 1% of maximum drawdown risk, the portfolio has historically earned 0.5% annual return. A higher RoMaD indicates better risk-adjusted performance because it signifies more return per unit of drawdown risk. Where it becomes useful is when comparing two portfolios with many different values. So if another portfolio has an AARR of 15% but an MDD of -30%, its RoMaD would also be 0.5. Even though the second portfolio has a higher return, it also has a larger maximum drawdown, leading to the same RoMaD. This is useful when comparing an index that relies on a riskier strategy than the other, for instance, the Clique Index and the Variance Index. So it accounts for both the upside potential and downside risk of a portfolio. The limitation of this measure is that it does not account for the frequency or duration of drawdowns, only the magnitude.

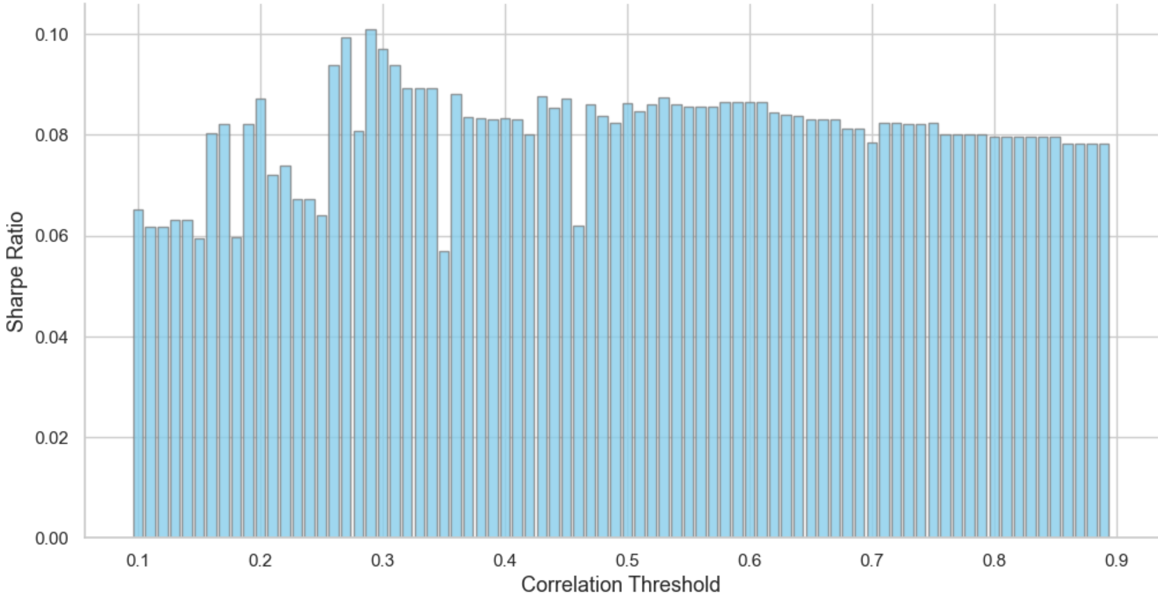
4.2 Graph creation

The dataset was split 50/50, with optimal correlation and centrality thresholds determined using the in-sample data [2017-2019] and tested on the out-of-sample data [2020-2022]. To determine the optimal correlation threshold between the stocks, an iteration was done over

[12] Assuming a 3.5 % risk free rate.

different values and compared their Sharpe ratios. Note that this was done for the in-sample period, where the market was in an uptrend. Since the out-of-sample period is far more volatile, this could greatly affect the performance. Figure 16 shows that the highest Sharpe ratio based on the in-sample period is achieved when the correlation threshold between all stocks in our graph is set at no less than 29%. So any two stocks whose mean return is correlated by at least 29% will have to be added to the graph.

Figure 16: Optimal empirical correlation



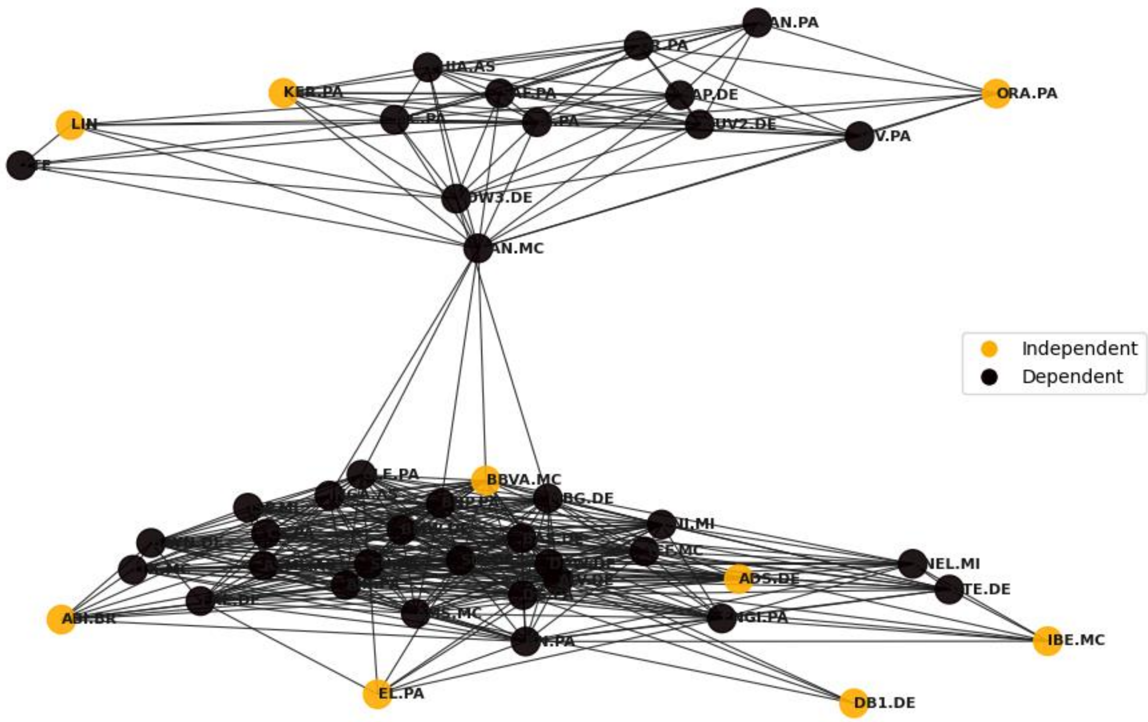
Source: own work.

4.3 Constructing Degeneracy Index

After constructing the graph, the next step is to create the Degeneracy Index. This index is a blend of the Independent Set and the Outliers. The main objective behind constructing this index is to maximize diversification, which means we are aiming to include the least correlated stocks within the graph while also integrating the most influential stocks.

As described in Chapter 2.1, we begin this process by identifying the largest *k*-core within our network, which will provide us with the most interconnected assets. We then proceed to apply Degeneracy Ordering to these assets. This involves arranging the chosen stocks in ascending order of their degree, allowing us to select the least correlated stocks from our densely interconnected subset. Consequently, we generate the Independent Set. Figure 17 visually represents the stocks that form the foundation of the Independent Set. You'll notice they all sit on the periphery of the graph, indicating their lower level of connectivity to other stocks.

Figure 17: Independent vertex set



Source: own work.

The other part is to determine the outliers, i.e. stocks which were by default less or even negatively correlated with all other assets in our graph and thus not added to the graph. This is beneficial since our objective is to minimize risk and maximize diversification. Outliers with zero correlation to other strongly connected components represent independent stocks that are not subject to the spillover effect.[13] If the outliers have a negative correlation with other strongly connected components, it is even more advantageous, as we will have negatively correlated assets to hedge against market turmoil. Our graph contains three such outliers:

- AD.AS - Ahold Delhaize (Retail industry)
- CRH.L - CRH plc (Building Materials industry)
- NOKIA.HE - Nokia Corporation (Telecommunications Equipment industry)
- UNA.AS - Unilever NV (Consumer Goods industry)

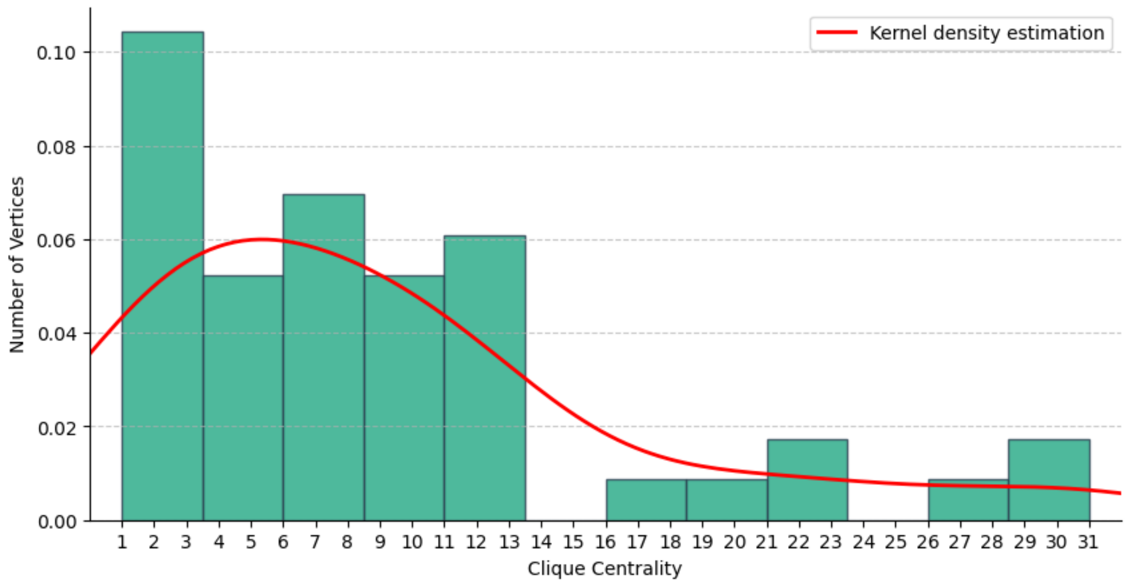
[13] The spillover effect refers to the impact that apparently unrelated events in one context can have on outcomes in another context. For example, a recession in a large economy like the United States can have a spillover effect on global economic conditions (Satchell, 2016).

4.4 Constructing the Clique Index

Our second portfolio, referred to as the Clique Index, adopts an approach contrasting that of the Degeneracy Ordering. Instead of aiming for assets with minimal correlation to maximize diversification, the Clique Index focuses on identifying a tightly combined group of assets that share strong connections.

The logic behind this strategy is to focus on a specific set of stocks that have the potential to enhance overall returns by eliminating non-performing or redundant assets. The objective here is to discover stocks with cross-maximal Clique Centrality exceeding a predefined threshold. Simply put, these stocks hold significant influence.

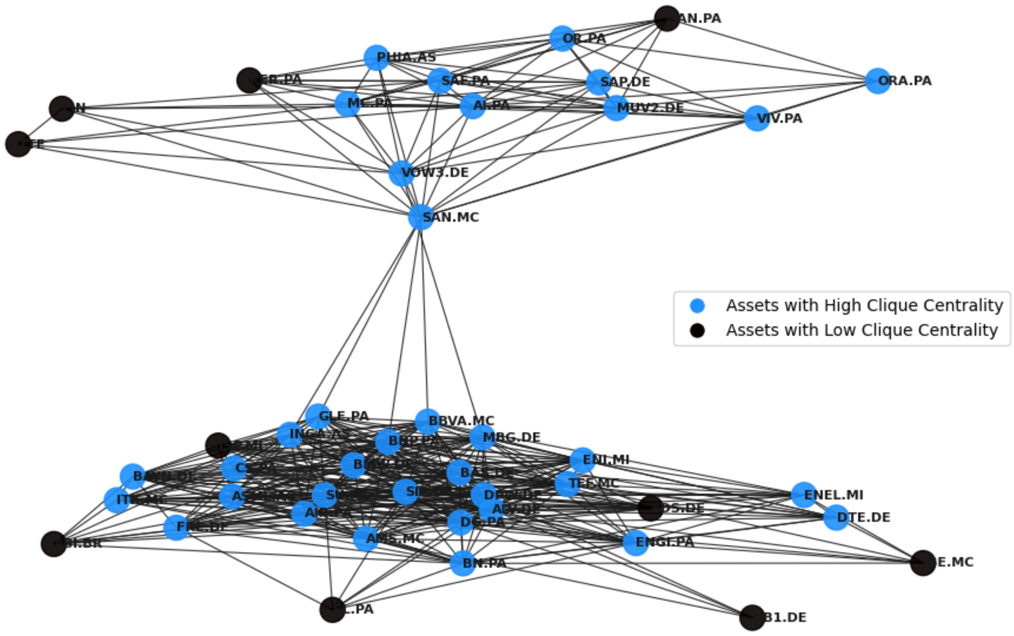
Figure 18: Clique centrality distribution (CCD)



Source: own work.

As illustrated in Figure 18, a cross-maximal clique centrality of 5 means that a stock is part of 5 different maximal cliques. This suggests that the stock has a significant influence on the network, as it is strongly interconnected with several different groups of stocks. These stocks might be considered key components of the portfolio, as their performance could potentially affect the performance of several different groups of stocks. On the other hand, they might also carry more risk, as negative performance can also spread through their multiple connections.

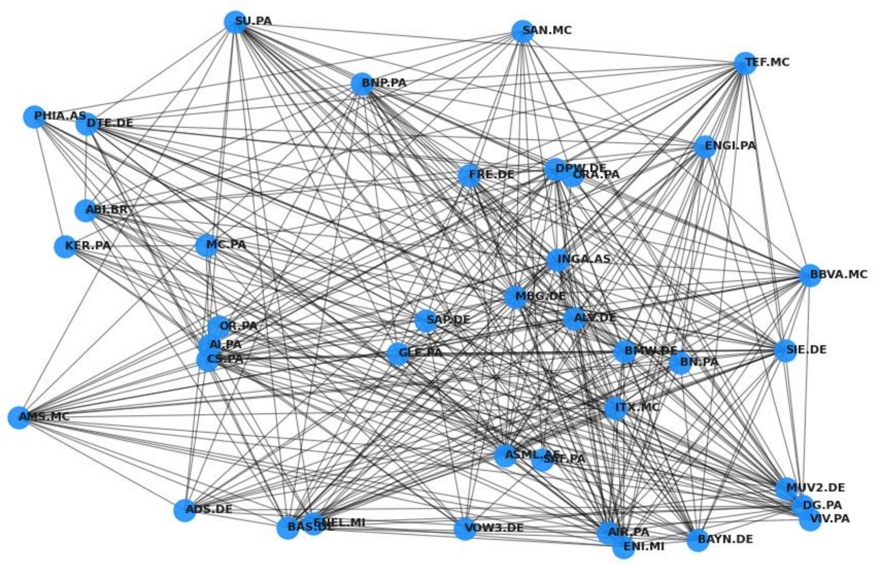
Figure 19: Highly concentrated asset combination



Source: own work.

In contrast to the Independent Set, Figure 19 depicts how the Clique Index is situated at the heart of the graph, indicating that these stocks share more connections than those on the periphery.

Figure 20: Clique Index



Source: own work.

Figure 20 displays all the stocks included in the Clique Index. Here, one can see these stocks are tightly linked. It's hard to find a stock that doesn't have many connections to other assets in the index. This shows how closely the market moves together. In the face of a black swan event, it's normal to see the whole market go down. Stocks that don't fall much when the market is going down are not common. These special stocks can be seen in the Degeneracy Index, shown in Figure 17. The Degeneracy Index has fewer assets than the Clique Index. This is because there aren't many stocks that don't move together with the other stocks in the portfolio.

4.5 Systemic Risk Analysis

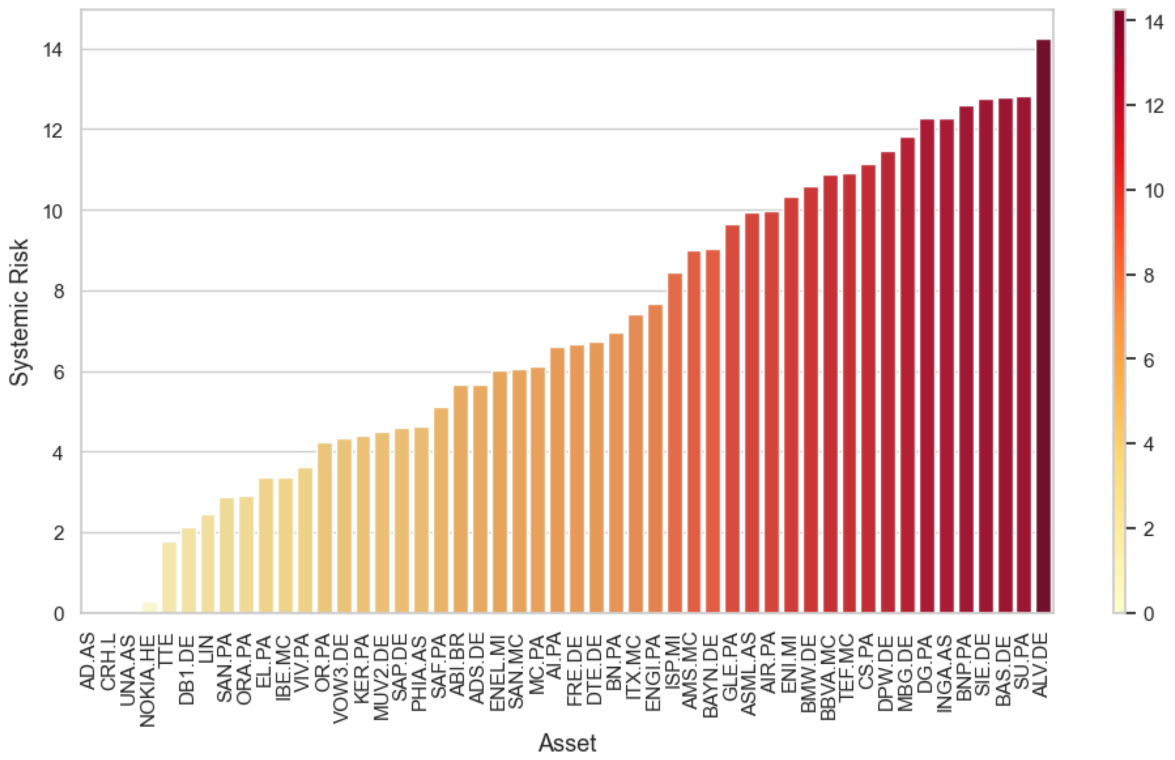
After creating two graph theory portfolios, a key question is: how can we measure the risk of the individual stocks in these portfolios? In theory, stocks in the Degeneracy Index should be less risky, and those in the Clique Index should be riskier. To check this, we can use an idea from Network Theory called "Network Robustness".

Network Robustness looks at how removing a stock, or "node", from a portfolio affects the whole network. A stock with high systemic risk has a bigger effect on the portfolio when removed than a stock with lower systemic risk. In essence, this tests how vulnerable a network is to the removal of nodes. The problem is in identifying and eliminating crucial nodes from these networks, a struggle encountered in many real-world applications.

This idea is important in many real-world situations. For instance, law enforcement often tries to break down criminal networks into smaller parts, or to increase the distance between nodes. After the events of September 11, 2001, there was a lot more research into the effects of removing nodes from complex networks (Krebs, 2001). In this study, the main focus when creating the index is on how much the stocks are correlated. Jahanpour and Chen (2013) found that using degree and betweenness centralities to decide which nodes to remove works better than other methods. Because of this, the study uses degree centrality for node removal.

Figure 21 shows stocks from least to most risky or connected. For example, removing ALV.DE, would affect 14 other stocks in the portfolio. But removing AD.AS, the portfolio would not be affected. This suggests that the least risky or connected stocks are probably in the Degeneracy Index, and the riskiest or connected stocks are likely in the Clique Index. Looking at the stocks in the Degeneracy Index: ['NOKIA.HE', 'TTE', 'DB1.DE', 'OR.PA', 'IBE.MC', 'EL.PA', 'ASML.AS', 'AD.AS', 'CRH.L', 'UNA.AS'], all but ASML.AS are on the left side of the chart, which means they are less risky. The first eight stocks listed from left to right in the Clique Index’s components are not included, showing that the portfolios have been created correctly.

Figure 21: Systemic Risks of Assets



Source: own work.

4.6 MPT Indexes

4.6.1 Variance Index

Following the creation of the Degeneracy Index and Clique Index, we move forward with the Variance Index. The objective is to achieve optimal diversification where the overall variance of the portfolio is minimized, thereby minimizing risk. In practical terms, the process of forming the Variance Index begins with computing the covariance matrix for the selected assets. This matrix reflects the behavior and relationships between the assets. A high covariance value between two assets indicates that they generally move in sync, either rising or falling together. On the other hand, a low or negative covariance suggests that the assets do not closely follow each other and may even move in opposite directions.

The next step is to determine the optimal weight for each asset in the portfolio using quadratic programming. The constraints ensure that all weights are non-negative and sum up to one, allowing us to utilize our entire investment without shorting any stocks. It's important to note that this process focuses solely on minimizing risk and does not consider the expected returns of the stocks.

In a bull market, the performance of the Variance Index might not be as spectacular as that of an index focusing on returns, given its emphasis on risk reduction. A strong market rally might leave it trailing as it would not necessarily be heavily invested in the best-performing stocks. It is essentially designed to safeguard against downturns rather than to capitalize on upturns. Contrarily, in bear markets, the index could show its true value. As it is configured to minimize risk, it is better equipped to endure market downturns. The weights assigned by the Variance optimization process tend to favour fewer volatile stocks, which often demonstrate resilience in down markets.

One interesting characteristic of the Variance Index is its potential "dislike" for high volatility. High volatility implies greater uncertainty and increased risk, which the Variance Index strives to steer clear of. By design, it would assign lower weights to such volatile assets, muting their impact on the overall portfolio. The Variance Index's greatest weakness might be its disregard for returns. It exclusively focuses on minimizing risk, potentially leading to a sub-optimal return profile in a bullish market. When compared to graph theory portfolios, the Variance Index takes a more traditional finance approach. While the Degeneracy and Clique Indexes consider asset correlations from a network perspective, the Variance Index considers them from a covariance matrix. Its advantage lies in its simplicity and robustness, but it may lack the sophisticated insight offered by the graphical approach.

4.6.2 Sharpe Index

The Sharpe Index operates on the premise of obtaining the best return per unit of risk, based on the Sharpe ratio. It seeks to identify the portfolio that offers the most favorable return relative to the level of risk undertaken. The process of forming the Sharpe Index commences similarly to the Variance Index, starting with the computation of the covariance matrix. Then, as before, we solve an optimization problem. But this time, instead of just minimizing risk, the goal is to maximize the Sharpe ratio.

The quadratic programming gives a set of weights that yield the highest ratio of expected return to portfolio risk. Readers should note that the Sharpe Index might take on more risk than the Variance Index, as it factors in returns along with risk. Each is based on a different goal, i.e., the Variance Index might appeal to the more risk-averse investor, while the Sharpe Index could be a better option for those willing to accept a higher level of risk for potentially higher returns.

In a bull market, the Sharpe Index would more than likely perform better. It doesn't just consider risk, but returns as well. If a high-performing stock carries with it a level of risk the index finds acceptable, it won't hesitate to assign significant weight to it. This could make the Sharpe Index a better choice in a market rally, specifically in the in-sample period. In contrast, in a bear market, the Sharpe Index might perform suboptimally. Despite its risk-considerate nature, it could experience more significant drops compared to a purely risk-minimizing portfolio. Its dual consideration of risk and return could expose it to downturns when the market breaks. Unlike the Variance Index, it doesn't "dislike" high volatility. It takes into account whether the potential rewards are worth the risk. The Sharpe Index's major drawback could be its reliance on expected returns. These expectations might not always materialize, leading to sub-optimal portfolio performance.

When comparing both the Variance and Sharpe Indexes to an equally weighted portfolio, the latter often tends to outperform in the long run. This is attributed to the periodic rebalancing that accompanies equal-weight strategies, effectively capturing a 'rebalancing bonus'. The rebalancing allows buying low and selling high, capitalizing on the mean-reverting nature of stock returns (Dubikovsky & Susinno, 2015).

4.7 In-sample results

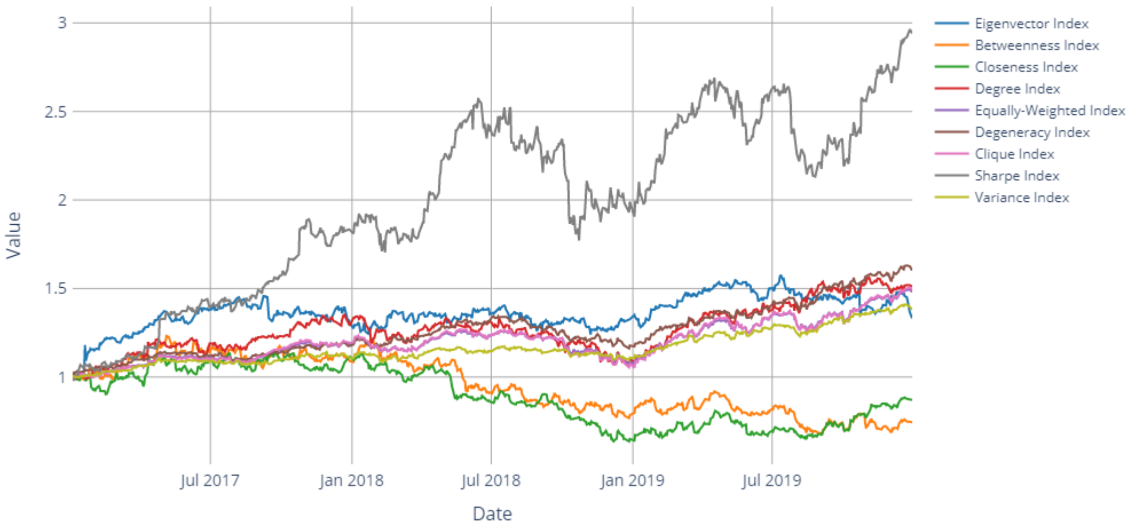
After establishing the optimal asset allocations for our portfolio collection, we took a closer look at their performance during the in-sample period. Predictably, the Sharpe Index topped the charts, as it's designed to chase the highest risk-adjusted returns. However, the performance of the various portfolios wasn't uniform. A good number of indexes showed strong performance, which isn't surprising given the strong upward trend in the market during this period. This was the perfect environment for market growth with low-interest rates, minimal political risks, and a thriving economy.

The purely graph-theory-based portfolios, namely the Degeneracy Index and the Clique Index, showed only average performance. Even more surprising, the Closeness Index and the Betweenness Index, which merge graph theory's centrality measure into the portfolio problem, performed rather poorly. This was unexpected as these models typically provide diversification benefits, which should help them perform at least on par with the market average.

Interestingly, the Clique Index and the Equally Weighted Index showed similar performance as depicted in Figure 22. This suggests that even though some models might be more complex to calculate, they don't necessarily yield better results. In terms of computational complexity, the EW Index is the simplest as it assigns an equal share of the total investment to each asset. Other models are more complex because they involve more advanced mathematical problems.

However, the *k*-core algorithm, which is used for the Degeneracy Index, is quite efficient as its complexity grows linearly with the number of edges and vertices in the graph. Despite the high complexity involved in the calculation of the Clique Index, it outperformed most hybrid models, reinforcing the idea that complexity does not always equate to better performance.

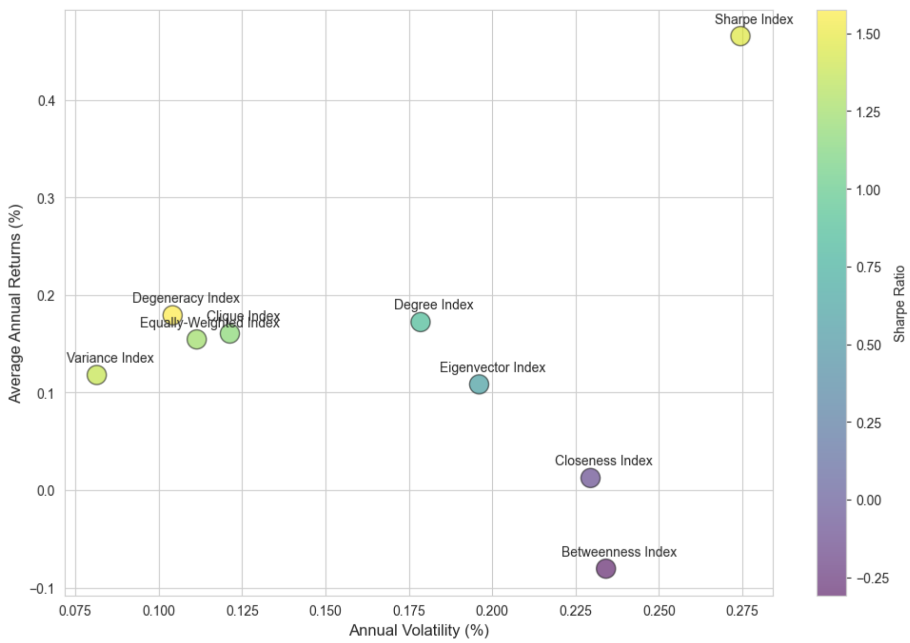
Figure 22: In-sample normalized comparison



Source: own work.

As shown in Figure 23, from a risk diversification standpoint, the Variance Index and Degeneracy Index provided the best allocations, though not necessarily in terms of returns. In terms of performance metrics, the Sharpe Index and Degeneracy Index outperformed the rest.

Figure 23: In-sample mean variance comparison



Source: own work.

The annualized Sharpe ratio was high for the Sharpe Index, as expected, because this ratio measures performance after adjusting for risk. Interestingly, this doesn't mean that the

Sharpe Index had the highest Sharpe ratio. The highest scoring index among this performance metric was the Degeneracy Index, as it had an Annualized Sharpe ratio of 1.72, beating the Sharpe Index by just over 0.02 points.

Both the Variance Index and the EW Index also had respectable Sharpe ratios, indicating that they offered good returns per unit of risk. RoMaD, which measures an investment's return relative to its worst-case loss, was notably high for the Degeneracy Index. Even though it didn't provide the highest returns, it managed downside risk effectively.

As mentioned in the metric description, even though the Sharpe Index had multiple short but frequent drawdowns, it still achieved a high RoMaD, despite the volatility. This is because the metric doesn't take into account the frequency or duration of drawdowns, only the magnitude. This is why the Variance Index scored the highest. Its combination of low volatility and average returns allowed for optimal performance.

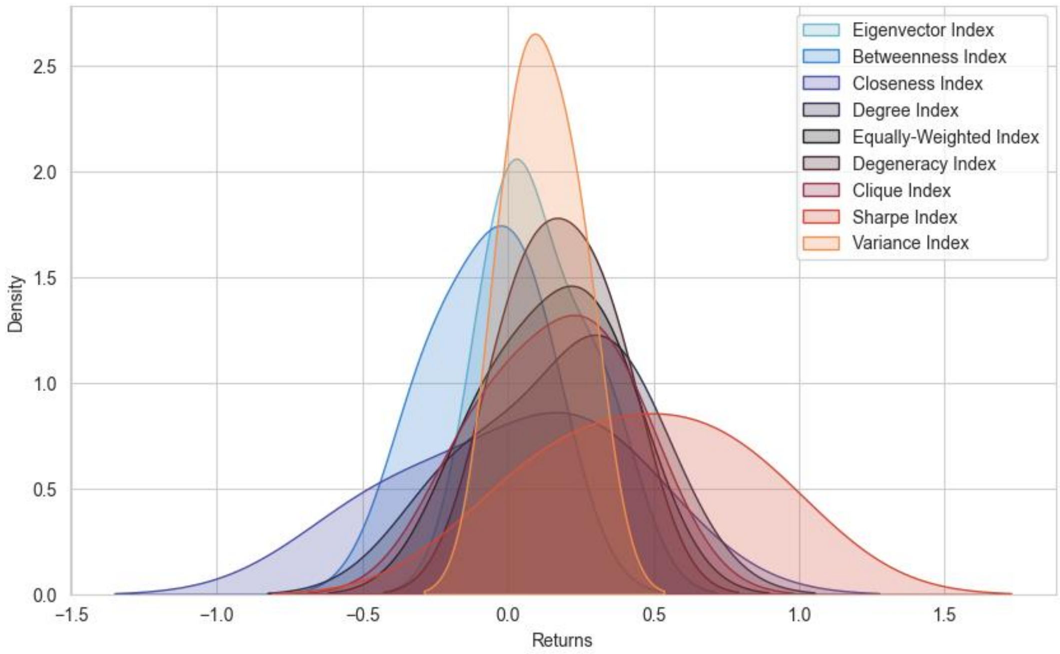
Table 1: In-sample performance metrics (annualized)

	Avg Annual RoR	**Annual Volatility**	**Maximum Drawdown**	**Annualized Sharpe Ratio**	**Returns Over Max Drawdown**
Eigenvector Index	11%	20%	-16%	0.55	70%
Betweenness Index	-8%	23%	-45%	-0.34	18%
Closeness Index	1%	23%	-44%	0.05	3%
Degree Index	17%	18%	-20%	0.96	85%
EW Index	15%	11%	-15%	1.39	101%
Degeneracy Index	18%	10%	-14%	1.72	132%
Clique Index	16%	12%	-17%	1.32	94%
Sharpe Index	47%	27%	-31%	1.70	149%
Variance Index	12%	8%	-7%	1.45	180%

Source: own work.

Examining the distribution statistics in Figure 24 gives us some interesting insights. A tighter distribution for the Variance Index suggests lower risk and more consistent performance, similar to the Degeneracy Index. Conversely, the broader and right-leaning distribution of the Sharpe Index indicates a higher likelihood of larger gains, although with higher volatility and risk. This is consistent with the results obtained from Table 1.

Figure 24: Distribution of returns (in-sample)



Source: own work.

The fact that the Sharpe Index, despite recording the highest average return, also displayed the highest volatility and maximum drawdown, underscores the principle that higher returns often involve higher risks.

4.8 Out-of-sample results

In a change from the in-sample results, the Sharpe Index didn't perform quite as well during the out-of-sample period. This particular index tends to excel when the market is doing well, but when the market is volatile, the performance tends to drop substantially. The market turbulence following the COVID-19 pandemic in 2020 is a clear example of this. Both portfolios initially fell but then quickly recovered to all-time highs. However, as the market took a turn for the worse, the portfolios followed suit.

An interesting point to note is the resilience of the Variance Index. This index consistently performed well across different market conditions, especially during challenging times marked by high volatility or economic downturns. By focusing on minimizing risk through diversification, it was able to weather the storm more effectively.

Figure 25: Out-of-sample normalized comparison

Source: own work.

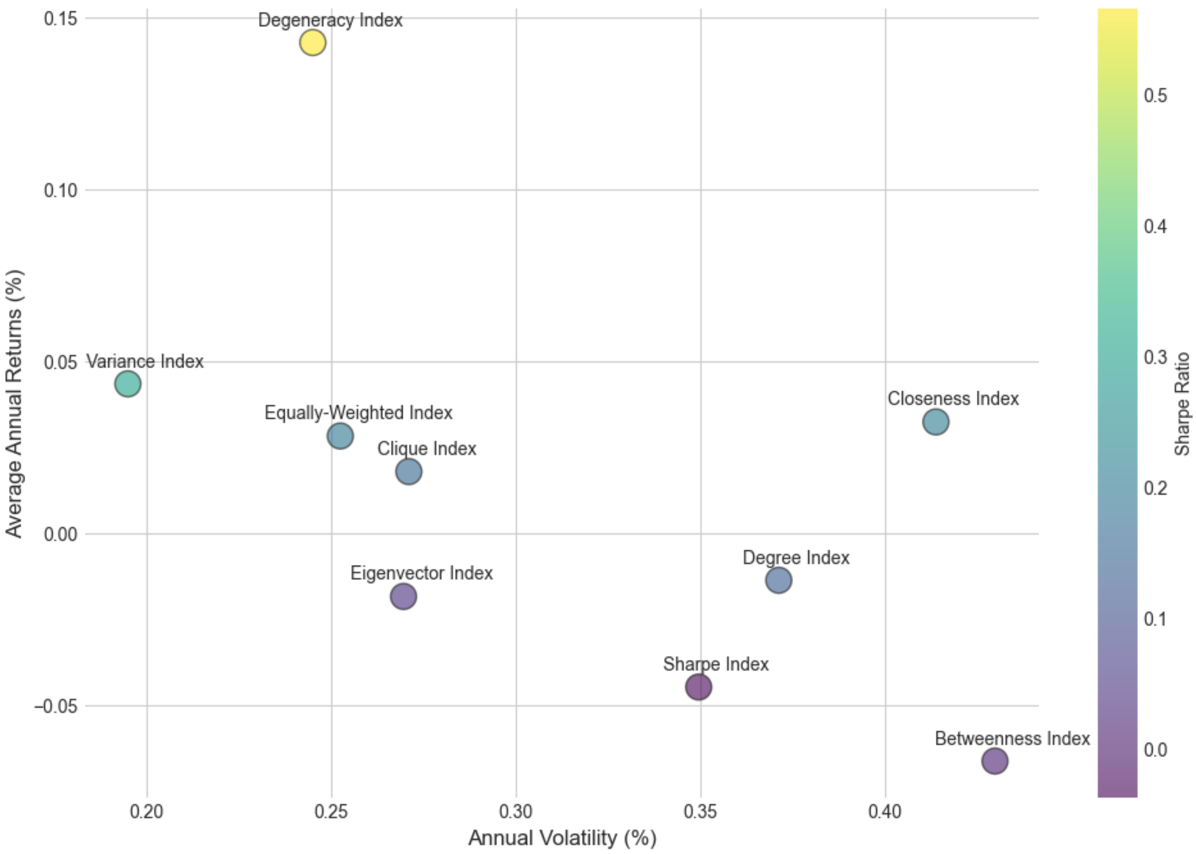
Another portfolio that showed consistent performance was the EW Index. This is an index where equal weights are assigned to all constituent assets, providing a balanced exposure to the market. The EW Index's performance underlines the benefits of this balanced approach, as it helps mitigate the impact of individual stock fluctuations and offers more stability during uncertain market conditions.

The Degeneracy Index also stood out with its impressive performance during the market downturn following the COVID-19 pandemic. It rose to almost 1.8 times its initial value at its peak, reinforcing the benefits of diversification and optimal asset allocation.

This demonstrates that indexes focusing on risk minimization and diversification can perform well over the long term, given a mix of bearish and bullish phases. On the other hand, the Betweenness Index didn't perform well.

Looking at the results, we could assume that it's designed to minimize returns, given its subpar performance. The Degree and Closeness Indexes also ranked lower due to their higher risk profiles.

Figure 26: Out-of-sample mean variance comparison



Source: own work.

A surprising finding was the weak performance of the Clique Index during the bear market. This index's approach of concentrating on correlated assets left it vulnerable to market downturns. Furthermore, the complex calculations required to determine optimal weights can become challenging when dealing with a larger set of assets. This suggests that indexes that maintain a more balanced and diversified asset allocation, such as the Degeneracy Index, Variance Index, and EW Index, may be better equipped to handle volatile market environments. On the other hand, portfolios that lean towards correlated assets, like the Clique Index, could face more difficulties during market turbulence.

When we look at performance measures, we see that a diversified index is crucial for successful management. While the Sharpe Index initially looked promising in a stable market, it did not perform well under more unpredictable market conditions. Not only that, looking at the results from Table 2, it can be seen that the Sharpe Index had a RoMaD of 10%, which would be much worse if we were to account for the frequency of those drawdowns, which is also important to consider. The indexes based on graph theory algorithms had different outcomes, reflecting their contrasting goals of diversification versus concentration. It is interesting that despite the Degeneracy Index's goal to minimize risk, it had almost the same annual volatility as the Clique Index, which was not designed to account for volatility.

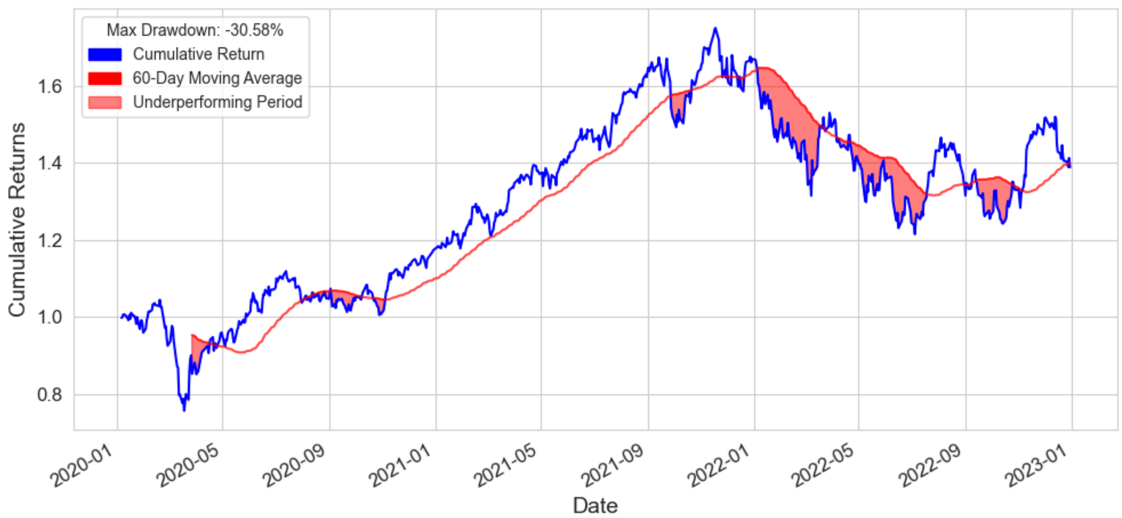
Table 2: Out-of-sample performance metrics (annualized)

	Avg Annual RoR	**Annual Volatility**	**Maximum Drawdown**	**Annualized Sharpe Ratio**	**Returns Over Max Drawdown**
Eigenvector Index	-2%	27%	-32%	-0.07	6%
Betweenness Index	-7%	43%	-61%	-0.15	11%
Closeness Index	3%	41%	-54%	0.08	6%
Degree Index	-1%	37%	-47%	-0.04	3%
EW Index	3%	25%	-38%	0.11	7%
Degeneracy Index	14%	25%	-31%	0.58	47%
Clique Index	2%	27%	-40%	0.07	4%
Sharpe Index	-4%	35%	-44%	-0.13	10%
Variance Index	4%	20%	-32%	0.22	13%

Source: own work.

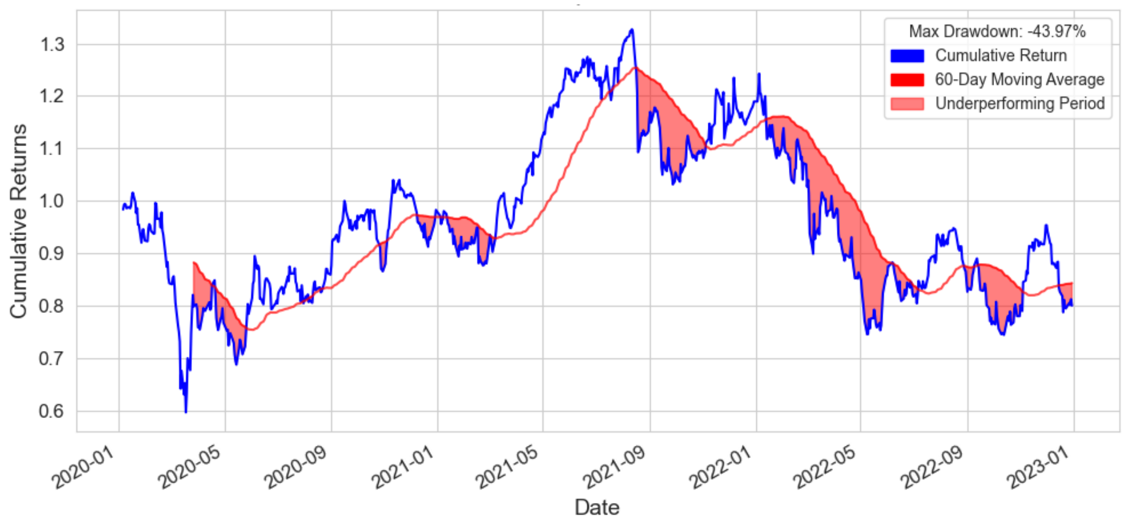
Comparing the Sharpe and Degeneracy indexes side by side in Figures 27 and 28 also paints a clear picture. The Degeneracy Index has a smooth, almost linear 60-day moving average, while the Sharpe Index's moving average shows much more fluctuation. This suggests that the Degeneracy Index provides more stable returns over time, while the Sharpe Index has more dramatic ups and downs.

Figure 27: Cumulative returns over time – Degeneracy index



Source: own work.

Figure 28: Cumulative returns over time – Sharpe index



Source: own work.

This result underscores the challenges associated with having a highly concentrated index of correlated assets, especially during periods of market turbulence. The so-called Hybrid Models did not perform well in either period, providing little value in terms of diversification or return. This leads us to conclude that the graph theory algorithms cannot act as complements to the traditional Modern Portfolio Theory approach, at least not when it comes to centrality measures. However, they could potentially be more effectively used as standalone tools or as substitutes for the Markowitz portfolio optimization approach.

CONCLUSION

This research aimed to investigate the application of graph theory-based algorithms for portfolio optimization and contrast their performance with the traditional Modern Portfolio Theory (MPT). Additionally, it aimed to evaluate the feasibility of integrating graph theory centrality measures with MPT to potentially enhance portfolio performance.

The results of the study indicate that portfolio diversification using graph theory concepts, specifically through the Degeneracy Index, can potentially enhance performance, particularly during market turbulence. This index effectively navigated the market volatility during the COVID-19 pandemic and subsequent inflationary pressures, emphasizing the importance of minimized correlations and diversified portfolio construction. This observation highlights the criticisms of MPT that its reliance on past performance does not necessarily predict optimal or even satisfactory future performance. Consequently, these methods usually fall short during unprecedented historical events such as black swan or tail events. Hence, our research provides a favorable response to our first research question (RQ1) by showcasing the superior diversification benefits of graph theory portfolios over a traditional MPT approach during both bullish and bearish markets.

To address the second research question (RQ2), we incorporated four commonly used centrality measures into the optimization problem, anticipating that at least one of them would outperform the benchmark. However, the combined approach of graph theory and Markowitz optimization delivered underwhelming results. The poor performance of the hybrid models like the Closeness Index and Betweenness Index was unexpected as they were expected to perform well due to the benefits of diversification. Also, the out-of-sample results demonstrated that the hybrid models didn't perform well in either period, providing little value in terms of diversification or return. This could be due to the improper formulation of the problem or an inherent incompatibility between the two concepts. This outcome suggests that graph theory, specifically centrality measures, might better serve as an alternative to MPT rather than a complement to it.

Our third research question (RQ3) examined which centrality measure resulted in the highest Sharpe ratio of the strategy. For the in-sample data, the highest Sharpe ratio was achieved by the Degree Index, which also performed well in the out-of-sample period when considering only centrality measure-based portfolios. For the out-of-sample period, the index with the highest Sharpe ratio was the Closeness Index. However, this index had a negative Sharpe ratio in the in-sample period, rendering any definitive conclusions unreliable. The practical simulations also showcased some key observations regarding the Variance Index, EW Index, and the Clique Index. The Variance Index's robust performance across different market conditions highlighted its resilience, owing to its focus on minimizing risk through diversification. The EW Index maintained stability during uncertain market conditions thanks to the benefits of a balanced market exposure. The Clique Index, despite its complex calculation methods and concentration on correlated assets, showed weak performance during the bear market, suggesting the advantages of maintaining a balanced and diversified asset allocation.

Despite the promising findings, this research has certain limitations. It heavily relies on historical data, which may not always be a reliable predictor of future market behavior. The research does not account for potential transaction costs or taxes that could significantly impact portfolio performance. The limited number of assets examined facilitated the construction of graph theory portfolios, despite the NP-completeness. However, if more assets were involved, the performance could potentially be less promising. Future research could explore the integration of other asset classes such as bonds, commodities, or cryptocurrencies with graph theory for more diverse results. The impact of different rebalancing strategies on portfolio performance is another interesting area to explore, as the timing and frequency of portfolio rebalancing can significantly influence performance. In conclusion, this thesis showcases that graph theory can offer unique ways to rethink portfolio construction. While the traditional MPT approach remains a significant pillar of portfolio optimization, these alternative models can provide valuable alternative strategies, particularly when there is market turbulence or in the case of unpredictable future events.

REFERENCES

1. Akkoyunlu, E. A. (1973). The Enumeration of Maximal Cliques of Large Graphs. *SIAM Journal on Computing*, *2*, 1–6. doi: 10.1137/020200
2. Andersen, M., & Vandenberghe, L. (2023). Cone Programming — CVXOPT User's Guide. Retrieved March 28, 2023, from cvxopt.org website: http://cvxopt.org/userguide/coneprog.html#quadratic-programming.
3. Bader, G. D., & Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, *4*(1), 2. doi: 10.1186/1471-2105-4-2
4. Baitinger, E., & Papenbrock, J. (2017). Interconnectedness risk and active portfolio management. *The Journal of Investment Strategies*, *6*(2), 63–90. doi: 10.21314/JOIS.2017.081
5. Barone, E., & Barone, G. (2022). Tracking The euro stoxx 50. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4203380
6. Basile, I., & Ferrari, P. (2016). *Asset Management and Institutional Investors* (1st ed.). Switzerland: Springer. doi: 10.1007/978-3-319-32796-9
7. Batagelj, V. (2003, October 25). An O(m) Algorithm for Cores Decomposition of Networks. Retrieved from arXiv.org website: https://arxiv.org/abs/cs/0310049
8. Beasley, J. E. (2013). Portfolio Optimisation: Models and Solution Approaches. In *Theory Driven by Influential Applications* (pp. 201–221). doi: 10.1287/educ.2013.0114
9. Behr, P., Löffler, G., & Miebs, F. (2011). On Portfolio Optimization: Imposing the Right Constraints. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1618756
10. Benjamin, A., Chartrand, G., & Zhang, P. (2017). *The Fascinating World of Graph Theory*. United Kingdom of Great Britain and Northern Ireland: Princeton University Press.
11. Benjelloun, H. (2010). Evans and Archer - forty years later. *Investment Management and Financial Innovations*, *7*(1), 98–104.
12. Bickle, A. (2010). *The K-cores of a Graph* (PhD Dissertation). Western Michigan University.
13. Bienstock, D. (1996). Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, *74*(2), 121–140. doi: 10.1007/bf02592208
14. Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, *104*(3), 535–559. doi: 10.1016/j.jfineco.2011.12.010
15. Bloomfield, T., Leftwich, R., & Long, J. A. (1977). Portfolio strategies and performance. *Journal of Financial Economics*, *5*(2), 201–218. doi: 10.1016/0304-

405x(77)90018-6

16. Boasson, V. W., Boasson, E., & Zhou, Z. (2017). Portfolio optimization in a mean-semivariance framework. *Investment Management & Financial Innovations*, *8*(3), 58–68. Retrieved from https://businessperspectives.org/journals/investment-management-and-financial-innovations/issue-84/portfolio-optimization-in-a-mean-semivariance-framework
17. Bollobás, B. (1984). The evolution of random graphs. *Transactions of the American Mathematical Society*, *286*(1), 257–274. doi: 10.1090/s0002-9947-1984-0756039-5
18. Borchers, B., & Mitchell, J. C. (1997). A computational comparison of branch and bound and outer approximation algorithms for 0–1 mixed integer nonlinear programs. *Computers & Operations Research*, *24*(8), 699–701. doi: 10.1016/s0305-0548(97)00002-6
19. Borgatti, S. P., Jones, C., & Everett, M. G. (1998). Network Measures of Social Capital. *Connections*. Retrieved from http://fs.nashaucheba.ru/tw_files2/urls_1/6221/d-6220657/7z-docs/3.pdf
20. Bratanic, T. (2022). Introduction to social network analysis. In *Graph Algorithms for Data Science* (2nd ed., pp. 127–160). New York, United States of America: Manning.
21. Bryan, K., & Leise, T. L. (2006). The $25,000,000,000 Eigenvector: The Linear Algebra behind Google. *Siam Review*, *48*(3), 569–581. doi: 10.1137/050623280
22. Buraschi, A., Porchia, P., & Trojani, F. (2010). Correlation Risk and Optimal Portfolio Choice. *Journal of Finance*, *65*(1), 393–420. doi: 10.1111/j.1540-6261.2009.01533.x
23. Cazals, F., & Karande, C. (2008). A note on the problem of reporting maximal cliques. *Theoretical Computer Science*, *407*(1–3), 564–568. doi: 10.1016/j.tcs.2008.05.010
24. Chang, T.-C., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, *27*(13), 1271–1302. doi: 10.1016/s0305-0548(99)00074-x
25. Chen, L., Peng, J., Zhang, B., & Rosyida, I. (2016). Diversified models for portfolio selection based on uncertain semivariance. *International Journal of Systems Science*, *48*(3), 637–648. doi: 10.1080/00207721.2016.1206985
26. Chin, C. Y., Chen, S., Wu, H.-H., Ho, C.-W., Ko, M.-T., & Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8*(S4). doi: 10.1186/1752-0509-8-s4-s11
27. Clemente, G. P., Grassi, R., & Hitaj, A. (2022). Smart network based portfolios. *Annals of Operations Research*, *316*(2), 1519–1541. doi: 10.1007/s10479-022-04675-7
28. Conte, A., & Tomita, E. (2021). Overall and Delay Complexity of the cliques and Bron-Kerbosch Algorithms. *Lecture Notes in Computer Science*. doi: 10.1007/978-3-030-68211-8_16
29. Coppola, C., & Elgazzar, H. (2020). Novel Machine Learning Algorithms for Centrality

and Cliques Detection in Youtube Social Networks. *ArXiv (Cornell University)*. doi: 10.48550/arxiv.2002.03893

30. Cornuéjols, G., Peña, J., & Tütüncü, R. (2018). Optimization Methods in Finance. *Cambridge University Press*. doi: 10.1017/9781107297340
31. De Bondt, W. F. M., & Thaler, R. H. (1985). Does the Stock Market Overreact? *Journal of Finance*, *40*(3), 793–805. doi: 10.1111/j.1540-6261.1985.tb05004.x
32. De Prado, M. L. (2013). A Journey Through the "Mathematical Underworld" of Portfolio Optimization. *Advanced Risk & Portfolio Management*. doi: 10.2139/SSRN.2214771
33. De Prado, M. L. (2016). Building Diversified Portfolios That Outperform Out-of-Sample (Presentation Slides). *Social Science Research Network*. doi: 10.2139/ssrn.2713516
34. DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/*N* Portfolio Strategy? *Review of Financial Studies*, *22*(5), 1915–1953. doi: 10.1093/rfs/hhm075
35. Dickson, D. (2000). Mathematicians chase the seven million-dollar proofs. *Nature*, *405*(6785), 383. doi: 10.1038/35013216
36. Dubikovsky, V., & Susinno, G. (2015). Demystifying Rebalancing Premium and Extending Portfolio Theory in the Process. *Social Science Research Network*. doi: 10.2139/ssrn.2927791
37. Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). *Modern Portfolio Theory and Investment Analysis* (9th ed.). New Jersey, United States of America: John Wiley & Sons.
38. Eppstein, D., Löffler, M., & Strash, D. (2010). Listing All Maximal Cliques in Sparse Graphs in Near-Optimal Time. In *Lecture Notes in Computer Science* (pp. 403–414). Berlin, Germany: Springer. doi: 10.1007/978-3-642-17517-6_36
39. Erciyes, K. (2021). *Discrete Mathematics and Graph Theory: A Concise Study Companion and Guide*. Springer Nature.
40. Ernst, P., Thompson, J. C., & Miao, Y. (2017). Tukey's transformational ladder for portfolio management. *Financial Markets and Portfolio Management*, *31*(5), 317–355. doi: 10.1007/s11408-017-0292-1
41. Evans, J. S. O., & Archer, S. L. (1968). Diversification and the Reduction of Dispersion: An Empirical Analysis. *Journal of Finance*, *23*(5), 761. doi: 10.2307/2325905
42. Faghani, M. R., & Nguyen, U. T. (2013). A Study of XSS Worm Propagation and Detection Mechanisms in Online Social Networks. *IEEE Transactions on Information Forensics and Security*, *8*(11), 1815–1826. doi: 10.1109/tifs.2013.2280884
43. Freuder, E. C. (1982). A Sufficient Condition for BacktracK-Free Search. *Journal of the ACM*, *29*(1), 24–32. doi: 10.1145/322290.322292

44. Gilli, M., Maringer, D., & Schumann, E. (2019). Optimization problems in finance. In *Elsevier eBooks* (2nd ed., pp. 219–228). San Diego, United States of America: Elsevier. doi: 10.1016/b978-0-12-815065-8.00022-4

45. Grossman, S. J., & Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, *3*(3), 241–276. doi: 10.1111/j.1467-9965.1993.tb00044.x

46. Haluszczynski, A., Laut, I., Modest, H., & Räth, C. (2017). Linear and nonlinear market correlations: Characterizing financial crises and portfolio optimization. *Physical Review*, *96*(6). doi: 10.1103/physreve.96.062315

47. Heinold, B. (2019). A Simple Introduction to Graph Theory. In *Bipartite Graphs and Trees* (1st ed., Vol. 1, pp. 1–66). Emmitsburg, United States of America: Mount St. Mary's University.

48. Himmel, A.-S., Molter, H., Niedermeier, R., & Sorge, M. (2017). Adapting the Bron–Kerbosch algorithm for enumerating maximal cliques in temporal graphs. *Social Network Analysis and Mining*, *7*(1). doi: 10.1007/s13278-017-0455-0

49. Irani, S. (1994). Coloring inductive graphs on-line. *Algorithmica*, *11*(1), 53–72. doi: 10.1007/bf01294263

50. Jahanpour, E., & Chen, X. (2013). Analysis of complex network performance and heuristic node removal strategies. *Communications in Nonlinear Science and Numerical Simulation*, *18*(12), 3458–3468. doi: 10.1016/j.cnsns.2013.04.030

51. GitHub. (2018). *Portfolio optimization*. Retrieved March 14, 2023 from https://github.com/je-suis-tm/graph-theory/tree/master/Portfolio%20Optimization%20project#intro

52. Jin, Y., Qu, R., & Atkin, J. a. D. (2014). A Population-Based Incremental Learning Method for Constrained Portfolio Optimisation. In *2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. doi: 10.1109/synasc.2014.36

53. Jin, Y., Qu, R., & Atkin, J. a. D. (2016). Constrained Portfolio Optimisation: The State-of-the-Art Markowitz Models. In *5th International Conference on Operations Research and Enterprise Systems*. doi: 10.5220/0005758303880395

54. Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, *75*(371), 544. doi: 10.2307/2287643

55. Kalayci, C. B., Ertenlice, O., & Akbay, M. A. (2019). A comprehensive review of deterministic models and applications for mean-variance portfolio optimization. *Expert Systems With Applications*, *125*, 345–368. doi: 10.1016/j.eswa.2019.02.011

56. Kaya, H. (2015). Eccentricity in asset management. *The Journal of Network Theory in Finance*, *1*(1), 1–32. doi: 10.21314/jntf.2015.003

57. Kirousis, L. M., & Thilikos, D. M. (1996). The Linkage of a Graph. *SIAM Journal on*

Computing, *25*(3), 626–647. doi: 10.1137/s0097539793255709

58. Kloks, T., & Xiao, M. (2022). *A Guide to Graph Algorithms*. Singapore: Springer.

59. Knill, O. (2011). *Linear Algebra with Probability, Lecture 34: Perron Frobenius theorem*. Retrieved April 4, 2023, from https://people.math.harvard.edu/~knill/teaching/math19b_2011/handouts/lecture34.pdf

60. Krebs, V. (2001). Mapping Networks of Terrorist Cells. *Connections*. Retrieved from https://ecsocman.hse.ru/data/517/132/1231/mappingterroristnetworks.pdf

61. Lick, D. R., & White, A. (1970). *K*-Degenerate Graphs. *Canadian Journal of Mathematics*, *22*(5), 1082–1096. doi: 10.4153/cjm-1970-125-1

62. Maillard, S., Roncalli, T., & Teiletche, J. (2010). The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, *36*(4), 60–70. doi: 10.3905/jpm.2010.36.4.060

63. Malliaros, F. D., Giatsidis, C., Papadopoulos, A. N., & Vazirgiannis, M. (2019). The core decomposition of networks: theory, algorithms and applications. *The Vldb Journal*, *29*(1), 61–92. doi: 10.1007/s00778-019-00587-4

64. Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems*, *11*(1), 193–197. doi: 10.1007/s100510050929

65. Maringer, D. (2008). Heuristic Optimization for Portfolio Management [Application Notes]. *IEEE Computational Intelligence Magazine*, *3*(4), 31–34. doi: 10.1109/mci.2008.929847

66. Markowitz, H. M. (1952). Portfolio selection*. *Journal of Finance*, *7*(1), 77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x

67. Matula, D. W., & Beck, L. L. (1983). Smallest-last ordering and clustering and graph coloring algorithms. *Journal of the ACM*, *30*(3), 417–427. doi: 10.1145/2402.322385

68. Merton, R. C. (1980). On estimating the expected return on the market. *Journal of Financial Economics*, *8*(4), 323–361. doi: 10.1016/0304-405x(80)90007-0

69. Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *Journal of Finance*, *42*(3), 483–510. doi: 10.1111/j.1540-6261.1987.tb04565.x

70. Michaud, R. O., & Ma, T. (2001). Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation. *Review of Financial Studies*, *14*(3), 901–904. doi: 10.1093/rfs/14.3.901

71. Michaud, R. O., & Michaud, R. (2007). Estimation Error and Portfolio Optimization: A Resampling Solution. *Social Science Research Network*. doi: 10.2139/ssrn.2658657

72. Mitchem, J. (1977). Maximal K-degenerate graphs. *Utilitas Mathematica*, *11*, 101–106.

73. Moon, J. C., & Moser, L. (1965). On cliques in graphs. *Israel Journal of Mathematics*,

3(1), 23–28. doi: 10.1007/bf02760024

74. Newman, M. (2010). *Networks: An Introduction* (1st ed.). Oxford, United Kingdom of Great Britain and Northern Ireland: Oxford University Press.

75. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. J. (2003a). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review*, *68*(5). doi: 10.1103/physreve.68.056110

76. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. J. (2003b). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review*, *68*(5). doi: 10.1103/physreve.68.056110

77. Pardalos, P. M., & Xue, J. (1994). The maximum clique problem. *Journal of Global Optimization*, *4*(3), 301–328. doi: 10.1007/bf01098364

78. Pasiliao, E. L., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, *48*(2), 431–443. doi: 10.1016/j.csda.2004.02.004

79. Pasiliao, E. L., Butenko, S., Shirokikh, O., Trukhanov, S., & Lafuente, J. G. (2013). A networK-based data mining approach to portfolio selection via weighted clique relaxations. *Annals of Operations Research*, *216*(1), 23–34. doi: 10.1007/s10479-013-1395-3

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

81. Pereira, J. M. S. (1977). A survey of K-degenerate graphs. *Graph Theory Newsletter*, *5*, 1–7.

82. Pfaff, B. (2016). *Financial Risk Modelling and Portfolio Optimization with R* (1st ed.). Chichester, United Kingdom of Great Britain and Northern Ireland: John Wiley & Sons. doi: 10.1002/9781119119692

83. Plyakha, Y., Uppal, R., & Vilkov, G. (2012). Why Does an Equal-Weighted Portfolio Outperform Value- and Price-Weighted Portfolios? *Econometric Modeling: Capital Markets - Portfolio Theory*. doi: 10.2139/ssrn.1787045

84. Pokharel, M. (2018). Computational Complexity Theory (P,NP,NP-Complete and NP-Hard Problems). *Central Department of Computer Science and IT, Tribhuvan University*.

85. Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific Reports*, *3*(1). doi: 10.1038/srep01665

86. Qu, B., Zhou, Q.-L., Xiao, J. Q., Liang, J., & Suganthan, P. N. (2017). Large-Scale Portfolio Optimization Using Multiobjective Evolutionary Algorithms and Preselection Methods. *Mathematical Problems in Engineering*, *2017*, 1–14. doi: 10.1155/2017/4197914

87. Rankin, S. A., & Robinson, I. J. W. (2013). *Discrete Mathematics: Number theory,*

modular arithmetic, and graph theory (2nd ed.). New York, United States of America: Kinson Publishing, Ltd.

88. Rea, A., Zhan, H. C. J., & Rea, W. (2015). A Comparision of Three Network Portfolio Selection Methods - Evidence from the Dow Jones. *Working Papers in Economics, University of Canterbury, Department of Economics and Finance*. doi: 10.13140/2.1.3269.4562

89. Reilly, F. K., & Brown, K. C. (2012). *Investment Analysis and Portfolio Management* (10th ed.). Boston, United States of America: Cengage Learning.

90. Ruiz-Torrubiano, R., & Suárez, A. (2010). Hybrid Approaches and Dimensionality Reduction for Portfolio Selection with Cardinality Constraints. *IEEE Computational Intelligence Magazine*, *5*(2), 92–107. doi: 10.1109/mci.2010.936308

91. Rysz, M., Pajouh, F. M., & Pasiliao, E. L. (2018). Finding clique clusters with the highest betweenness centrality. *European Journal of Operational Research*, *271*(1), 155–164. doi: 10.1016/j.ejor.2018.05.006

92. Saoub, K. R. (2021). *Graph Theory: An Introduction to Proofs, Algorithms, and Applications* (1st ed.). Boca Raton, United States of America: CRC Press.

93. Satchell, S. (2016). *Asset Management: Portfolio Construction, Performance and Returns*. London, United Kingdom of Great Britain and Northern Ireland: Palgrave MacMillan.

94. Schaerf, A. (2002). Local Search Techniques for Constrained Portfolio Selection Problems. *Computational Economics*, *20*(3), 177–190. doi: 10.1023/a:1020920706534

95. Segundo, P. S., Artieda, J., & Strash, D. (2018). Efficiently enumerating all maximal cliques with bit-parallelism. *Computers & Operations Research*, *92*, 37–46. doi: 10.1016/j.cor.2017.12.006

96. Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, *5*(3), 269–287. doi: 10.1016/0378-8733(83)90028-x

97. Skiena, S. (2012). Graph Problems: Hard Problems. In *The Algorithm Design Manual* (pp. 523–561). London, United Kingdom of Great Britain and Northern Ireland: Springer. doi: 10.1007/978-1-84800-070-4_16

98. Stoxx. (2023). *Creating an investment intelligence advantage*. Retrieved March 23, 2023, from https://www.stoxx.com/document/Indices/Common/Indexguide/stoxx_calculation_guide.pdf

99. Szekeres, G., & Wilf, H. S. (1968). An inequality for the chromatic number of a graph. *Journal of Combinatorial Theory*, *4*(1), 1–3. doi: 10.1016/s0021-9800(68)80081-x

100. Taljaard, B., & Maré, E. (2021). Why has the equal weight portfolio underperformed and what can we do about it? *Quantitative Finance*, *21*(11), 1855–1868. doi: 10.1080/14697688.2021.1889020

101.Tomita, E., Tanaka, A., & Takahashi, H. (2006). The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science*, *363*(1), 28–42. doi: 10.1016/j.tcs.2006.06.015

102.Tse, C. K., Liu, J., & Lau, F. C. M. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, *17*(4), 659–667. doi: 10.1016/j.jempfin.2010.04.008

103.Výrost, T., Lyócsa, Š., & Baumöhl, E. (2019). NetworK-based asset allocation strategies. *The North American Journal of Economics and Finance*, *47*, 516–536. doi: 10.1016/j.najef.2018.06.008

104.Wang, Z. (2023). Analysis of the Limitations of Portfolio Theory. *Highlights in Business, Economics and Management*, *5*, 43–47. doi: 10.54097/hbem.v5i.4991

105.Wolfe, P. (1959). The Simplex Method for Quadratic Programming. *Econometrica*, *27*(3), 382. doi: 10.2307/1909468

106.Wrath of Math. (2019). *What is the K-core of a Graph?* [Youtube]. Retrieved from https://www.youtube.com/watch?v=rHVrgbc_3JA

107.Xu, R.-T., Zhang, J., Liu, O. L., & Ruizhang, H. (2010). *An Estimation of Distribution Algorithm Based Portfolio Selection Approach*. doi: 10.1109/taai.2010.57

APPENDICES

Appendix 1: Povzetek (Summary in Slovene language)

Robert Arnott je nekoč poudaril, da je v svetu investiranja tisto, kar je udobno, redko dobičkonosno. Ta misel poudarja pomen konstantne inovacije in nenehnega raziskovanja učinkovitejših strategij na področju upravljanja premoženja. Zavedanje, da tradicionalne metode morda ne prinašajo optimalnih rezultatov, vodi do neprestanega napredka in iskanja novih, obetavnih metod. Klasičnemu modelu optimizacije portelfja tj. Markowitz predlagamo alternativni pristop z uporabo algoritmov teorije grafov ki ciljajo na maksimiranje diverzifikacije in donosa. Dodatno definiramo tudi takoimenovane hibridne modele, ki so kombinacija obeh pristopov. Ideja teh modelov je optimizacija na podlagi klasičnega MPT pristopa, vendar z vpeljavo štirih različnih mer središčnosti oziroma centralnosti.

V tej nalogi torej raziščemo, kako se uporaba teorije grafov lahko uporabi za izboljšanje strategij optimizacije portfelja. Nadalje raziščemo tudi ali se lahko uporabi kot samostojna strategija. Naloga je strukturirana v dva glavna dela. Prvi del se osredotoča na izgradnjo teoretičnih temeljev, drugi del pa predstavlja empirično študijo, ki vključuje razlago uporabljenih podatkov in metodologije ter izvedbo in primerjavo uspešnosti različnih strategij portfelja. Simulacija se izvede v obdobju vzorca (2017-2019) in izven vzorca (2020-2022). Ugotovitve kažejo, da hibridni modeli, ki združujejo Markowitzovo teorijo portfelja in mere centralnosti, niso dosegli pričakovanih rezultatov. V obeh preučevanih časovnih obdobjih so se ti modeli izkazali za manj uspešne v primerjavi z ostalimi strategijami. Nasprotno pa je indeks degeneracije, ki temelji na teoriji grafov, v obeh obdobjih presegel vse ostale. To je mogoče pripisati prednostim diverzifikacije na katerih temelji ta strategija. Dosegla je nadpovprečne rezultate v smislu donosnosti in tveganja v obeh obdobjih.

Appendix 2: Pseudocode for BK algorithm

Algorithm 1

function BronKerbosch(P, R, X)
- R : a clique
- $P \cup X$: set of all vertices v such that $R \cup \{v\}$ is a clique and where
  - vertices in P have not yet been considered as additions to R and
  - vertices in X already have been considered in earlier steps

    if $P \cup X = \emptyset$ **then**
        add R to the solution
    end if
    choose pivot vertex $u \in P \cup X$ with $|P \cap N(u)| = \max_{v \in P \cup X} |P \cap N(v)|$
    for $v \in P$ **do**
        BronKerbosch$(P \cap N(v), R \cup \{v\}, X \cap N(v))$)
        $P \leftarrow P \setminus \{v\}$
        $X \leftarrow X \cup \{v\}$
    end for
end function

Adapted from Tomita, Tanaka, & Takahashi (2006).

The parameters (P, R, X) are defined as follows:

- R is either an empty set ($R = \emptyset$) or a clique in G.
- P is a priority queue containing vertices that are candidates for being added to the clique. They are vertices that are connected with all vertices in R. Initially, this can be set as all vertices in the graph.
- X contains vertices that have already been processed and should not be included in the clique.
- P and X are disjoint sets[14], and $P \cup X = \{y \mid R \subseteq N(y)\}$. This means that the set $P \cup X$ contains those vertices y in $V \setminus R$ such that $\{y\} \cup R$ is a clique. Due to the invariant, the set R is a maximal clique exactly when $P \cup X = \emptyset$ (Cazals & Karande, 2008).

The algorithm starts by checking if both P and X are empty. If they are, it means that the algorithm has found a maximal clique, which is yielded as a result. Next, the algorithm chooses a pivot vertex u from the union of P and X. This pivot vertex is chosen as it has the maximum intersection with P.[15] This strategy is used to delay the addition of the neighbours of the pivot vertex to the clique, which reduces the number of recursive calls (Cazals & Karande, 2008).[16] If there's an error in choosing the pivot (which happens if P and X are both empty), the algorithm defaults to consider all vertices in P as the neighbors of the pivot. The

[14] Two sets are disjoint if they have no elements in common, in the context of the BK algorithm, the sets P and X are disjoint, meaning no vertex belongs to both P and X.
[15] Referring to the largest overlap between two sets, specificaly choosing the vertex which has the largest number of shared neighbours with the current set of candidates (P) in the graph.
[16] A process where a function calls itself, consequently allowing it to be repeated several times.

algorithm then iterates over each vertex v in the set N (which is either the difference of P and the neighbors of u, or just P if an error occurred). For each vertex v, the function recursively calls itself with updated arguments: $R \cup \{v\}$ as the new R, and the intersections of the old P and X with the neighbors of v as the new P and X (Himmel, Molter, Niedermeier & Sorge, 2017). After the recursive call, v is removed from P and added to X, indicating that it has been processed and should not be included in the clique. This process continues until all vertices have been processed and all maximal cliques have been found. The use of a pivot and the careful management of the sets P and X helps to reduce the search space and improve the efficiency of the algorithm. Finally, the remaining set of maximal cliques, which are those that do not contain x, are listed via the B & K call.

Appendix 3: Pseudocode for k-core algorithm

Algorithm 2

```
         INPUT: Graph G = (V, L) [Represented by lists of neighbors]
         OUTPUT: Table core [Contains the core number for each vertex]

1.1      Compute the degrees of all vertices
1.2      Arrange the vertices (V) in increasing order based on their degrees
2        for each v ∈ V in the order do begin
2.1          core[v] := degree[v];
2.2          for each u ∈ Neighbors(v) do
2.2.1            if degree[u] > degree[v] then begin
2.2.1.1              degree[u] := degree[u] − 1;
2.2.1.2              reorder V accordingly
                 end
         end;
```

Adapted from Batagelj & Zaveršnik (2003).

The algorithm begins by computing the degree of all vertices and sorting the vertices based on their degrees in ascending order. Next, the algorithm iterates over each vertex 'v' in the ordered list. It assigns to 'core[v]' the degree of vertex 'v', which essentially signifies the current 'core number' of the vertex. The algorithm then examines each neighboring vertex 'u' of the current vertex 'v'. If the degree of 'u' is greater than the degree of 'v', it decreases the degree of 'u' by 1. This is based on the idea that if 'v' is part of a k-core, then its neighbors should connect to at least 'k' vertices in the same k-core. If a neighbor 'u' has a higher degree, it means 'u' has connections outside of the current k-core and reducing its degree would not affect the current k-core status of 'v'. After adjusting the degree of 'u', the algorithm reorders the vertices in ascending order based on their new degrees. This step ensures that the vertices are processed in the correct order.

Appendix 4: Components

Ticker	Name
OR.PA	L'Oreal SA
DG.PA	Vinci SA
BBVA.MC	Banco Bilbao Vizcaya Argentaria SA
SAN.MC	Banco Santander SA
ASML.AS	ASML Holding NV
PHIA.AS	Koninklijke Philips NV
TEF.MC	Telefonica SA
FP.PA	TOTAL SE
AI.PA	Air Liquide SA
CS.PA	AXA SA
BNP.PA	BNP Paribas SA
BN.PA	Danone SA
VIV.PA	Vivendi SA
EL.PA	EssilorLuxottica SA
MC.PA	LVMH Moet Hennessy Louis Vuitton SE
KER.PA	Kering SA
AMS.MC	Amadeus IT Group SA
SAF.PA	Safran SA
AD.AS	Koninklijke Ahold Delhaize NV
UNA.AS	Unilever NV
IBE.MC	Iberdrola SA

INGA.AS	ING Groep NV
LIN	Linde PLC
ITX.MC	Industria de Diseno Textil SA
ISP.MI	Intesa Sanpaolo SpA
ENI.MI	Eni SpA
ENGI.PA	Engie SA
ORA.PA	Orange SA
ABI.BR	Anheuser-Busch InBev SA/NV
SAN.PA	Sanofi
GLE.PA	Societe Generale SA
ENEL.MI	Enel SpA
NOKIA.HE	Nokia Oyj
SU.PA	Schneider Electric SE
ALV.DE	Allianz SE
AIR.PA	Airbus SE
BAYN.DE	Bayer AG
BMW.DE	Bayerische Motoren Werke AG
CRH.L	CRH PLC
BAS.DE	BASF SE
SIE.DE	Siemens AG
VOW3.DE	Volkswagen AG
MUV2.DE	Munich Re
FRE.DE	Fresenius SE & Co KGaA

SAP.DE	SAP SE
ADS.DE	Adidas AG
DTE.DE	Deutsche Telekom AG
DPW.DE	Deutsche Post AG
DAI.DE	Daimler AG
DB1.DE	Deutsche Boerse AG

Source: own work.