Asset Allocation based on a Partitional Clustering Algorithm

Flávio Gabriel Duarte
Natural Computing and Machine Learning Laboratory
Graduate Program in Electrical Engineering and Computing
Mackenzie Presbyterian University
São Paulo, SP 01302-000
Email: flaviogabriel@gmail.com

Leandro Nunes de Castro
Natural Computing and Machine Learning Laboratory
Graduate Program in Electrical Engineering and Computing
Mackenzie Presbyterian University
São Paulo, SP 01302-000
Email: lnunes@mackenzie.br

Abstract—**This paper proposes a method for asset allocation based on partitional clustering. This method is different from the approaches already proposed in the literature, which essentially use either an optimization-based approach or a hierarchical clustering algorithm to allocate resources in assets. After finding the clusters, the method uniformly allocates the resources over the clusters and then within the clusters, thus guaranteeing that all assets are allocated. The method was tested using data from the Brazilian Stock Exchange (B3) and the assets eligible to enter the allocation were those that were part of the Ibovespa Index at the time of portfolio rebalancing. The results were compared with the Ibovespa index for different metrics, such as volatility, return, sharpe ratio, turnover and drawdown. The proposed approach illustrates the potential of machine learning techniques in portfolio allocation.**

I. INTRODUCTION

With the advances in computing, the evolution of internet and the popularization of data-driven decision-making, the use of algorithms in asset management by professional managers has become common. Five of the six largest investment funds in the world use computers for most or all of their decision-making processes [1]. The rapid development and absorption of *Artificial Intelligence* (AI) techniques has created increasingly efficient mechanisms for *portfolio selection* or *portfolio optimization*, in order to optimize (maximize or minimize) one or more objectives [2], [3], such as the *expected return*, which must be maximized, and the *risk*, which must be minimized.

The work of [2] is pioneer in the introduction of a mean-variance model for the optimization of portfolios, assuming that the diversification of assets allows to increase the portfolio return whilst reducing its risk. Over time, several techniques have been introduced to improve the portfolio allocation process, such as the models proposed by [4] and [5], both inspired by the optimization proposal of [2].

Portfolio optimization algorithms usually start from the premise that it is necessary to incorporate the expected returns and risks in the optimization process. A wrong estimate of future return and risk leads to a non optimal portfolio as a result of the optimization [5]. The existence of more than one objective to be optimized simultaneously, as well as the a priori ignorance of the problem's decision surface, makes the application of AI techniques very convenient for this task.

Among these, *multi-objective evolutionary algorithms* have received a lot of attention from the community as efficient tools for portfolio optimization [3].

The above methods have limitations, however, as they depend on the future values of return and risk, which can compromise the result of the optimization if the estimates are wrong. There is also an algebraic constraint in which increasing the size of the correlation matrix implies an increase in the matrix conditioning number, causing numerical errors that make the matrix unstable [6]. Given the problems mentioned above, portfolio allocation techniques known as *risk parity* have become popular, in which allocations are made based on the expected risk without the need to incorporate expected future returns [7]. In [6] the authors noted that, in these methods, the inversion of the positive-definite covariance matrix that can lead to numerical errors and instability is still necessary. Thus, he proposed a technique called "Hierarchical Risk Parity" (HRP) for asset allocation, which uses a *hierarchical clustering* algorithm to perform the allocation.

While a hierarchical clustering performs a hierarchical decomposition of the dataset, a partitional clustering method performs a division of data into partitions. Based on López de Prado's proposal [6] for the application of hierarchical clustering for asset allocation, [8] proposed and tested the use of other hierarchical clustering algorithms for asset allocation.

Inspired by the HRP approach, this paper presents a method based on a partitional clustering algorithm, known as k-medoids, for asset allocation [9]. The allocation using partitional clustering allows flexibility in the portfolio construction, as both the intragroup and intergroup information allow the allocation in different ways, according to the portfolio that is sought.

Therefore, the objective of this work is to propose a portfolio allocation approach based on the k-medoids partitional clustering algorithm, testing different values for k (number of groups). The assets used will be stocks traded in the Brazilian stock market (B3), and the assets eligible to form the clusters are the same as those that comprise the theoretical portfolio of the Ibovespa index in the respective rebalancing date. All assets that pass through the clustering algorithm will be allocated forming long-only portfolios. The allocation is

tested between 2009-2017 with monthly rebalancing.

The paper is organized as follows. Section 2 introduces clustering and discourses about the most recent and most important applications of clustering for asset allocation. Section 3 introduces the asset allocation method proposed based on partitional clustering. Section 4 presents the experimental results obtained and discusses its performance. The paper is concluded in Section 5 with a general discussion and trends for future research.

II. CLUSTERING AND ASSET ALLOCATION

Clustering is a term used to designate numerical methods of multivariate data analysis for the purpose of discovering homogeneous groups of data [10]. It can be defined as the organization of a set of objects into groups based on their similarity. Put another way, clustering is the process of partitioning a dataset into subsets so that the objects in each group share common characteristics, usually proximity, based on some measure of similarity or distance. A group can be defined according to the internal cohesion and external isolation of its objects [10].

The concept of *natural cluster* was introduced by [11] and defines that natural clusters are those that share continuous, relatively dense and object-populated regions of the space, surrounded by relatively empty regions.

Most clustering algorithms focus on obtaining k groups of n similar objects according to some pre-established criteria. Finding an optimal clustering solution is considered an NP-hard problem, since the number of clusters in a dataset is not known a priori and not all datasets have natural cluster separations [12]. Furthermore, different initializations of the clustering algorithms may result in different performances.

Most clustering methods assume, as a starting point for grouping, an array X of size $n \times M$, $X \in R^{n \times m}$, which represents n objects with m attributes each. Another data structure widely used to perform clustering is a distance matrix, D, with dimension $n \times n$, $D \in R^{n \times n}$, where each element of D corresponds to a quantitative measure of distance between all objects.

Clustering algorithms can be roughly divided into: partitional, hierarchical, density-based, and graph-based [13]. The most well-known hierarchical algorithms are single-link and complete-link, and both can be implemented in various forms. The minimum spanning tree is an implementation used for graph-based clustering solutions, and can be seen in [10].

Recently, some contributions were made in the field of asset allocation and clustering. In [6] the authors proposed the use of machine learning for portfolio allocation using a technique called *hierarchical risk parity*, which uses a hierarchical clustering algorithm to perform the allocation based on the clustering result.

In [14] the authors noted that, although portfolio optimization is conceptually simple, involving the creation of efficient frontiers, the solutions are complex. Usually, this problem is solved with quadratic parametric programming, which uses the inverse matrices for its solution [15], [14].

In portfolio optimizers, as assets are added to this matrix used for the optimization of a portfolio, the matrix conditioning number increases, making the inverse calculation unstable [6]. The method proposed by López de Prado does not need an inverse matrix, since the allocation process is performed by the clustering algorithm based on the covariance matrix. Inspired by the work of López de Prado, [8] explored asset allocation using the main hierarchical clustering algorithms.

The most commonly used algorithms for partitional clustering are *k-means*, *k-medoids* and variations of both [10]. This paper presents the use of a *k*-medoids partitional algorithm, in its variation known as *Partitioning Around Medoids* (PAM) [16]. The application of *k*-medoids is due to the fact that the distance matrix is obtained from the correlation matrix. Thus, the distance between the objects (assets) and a prototype that is not another data object does not exist and, therefore, the prototype (medoid) must be a data object (asset). Since the *k*-medoids algorithm operates by receiving an already calculated distance matrix as input, this algorithm is suitable to be applied to this task.

III. ASSET ALLOCATION BASED ON PARTITIONAL CLUSTERING

The method proposed here contains five steps, as follows: A) calculate the correlation matrix between assets and, then, determine their distance matrix; B) apply the *k*-medoids partitional clustering algorithm; C) perform intragroup allocation; D) perform intergroup allocation; and E) evaluate performance. Each of these steps will be detailed in the following sections.

A. Calculating the Correlation and Distance Matrix

To perform clustering in financial time series, it is necessary to perform some transformations in the series. According to [17], a technique for time series clustering was defined by [18] for hierarchical clustering and has been widely adopted ever since. The technique can be defined as follows:

- Let N be the number of assets to be clustered.
- Let $P_i(t)$ be the price of asset i at time t, where $1 \leq i \leq N$.
- Let $r_i(t)$ be the return in time t of asset i, where:

$$r_i(t) = \log P_i(t) - \log P_i(t-1). \tag{1}$$

- For each pair of assets i, j, the correlation $\rho_{i,j}$ is calculated:

$$\rho_{i,j} = \frac{\sum (r_i - \overline{r_i})(r_j - \overline{r_j})}{\sqrt{\sum (r_i - \overline{r_i})}\sqrt{\sum (r_j - \overline{r_j})}} \tag{2}$$

- The correlation coefficients $\rho_{i,j}$ are converted into distance $d_{i,j}$:

$$d_{i,j} = \sqrt{2(1 - \rho_{i,j})} \tag{3}$$

B. Partitional Clustering

After the distance matrix is calculated, clustering is performed using a partitional clustering algorithm, more specifically the k-medoids in its PAM variation, as will be briefly reviewed in the following.

The k-medoids algorithm is a partitional clustering method similar to k-means, but whose *prototype*, that is the cluster center, is always an object from the dataset, and not a hypothetical object like in k-means. While k-means is sensitive to noisy and discrepant data objects, because of the average measure that the algorithm uses to define and update the cluster centroids (prototypes), k-medoids is more robust to both problems, because instead of using the average of the group to represent its center, k-medoids uses an object from the cluster itself to represent the center. The prototypes in k-medoids are named *medoids* and represent the most central objects in each cluster based on the minimum sum of distances for the other objects in the cluster [19].

A common implementation found in the literature for k-medoids is called *partitioning around medoids* (PAM), proposed by [16]. PAM consists of selecting k objects to be the initial medoids. Then, each object is associated with the cluster whose medoid is the closest. Repositioning of the medoids is performed by minimizing the objective function, which is the sum of the distances of objects from the cluster to the medoid. Following the repositioning of the medoids, the process of assigning the objects to the clusters and recalculating the medoid placement is performed iteratively until the algorithm has converged and no further medoid and cluster updates happens [19]. The k-medoids algorithm pseudocode is presented in Algorithm 1. PAM operates by minimizing the intracluster distance [20] and most partitioning clustering algorithms aim to minimize the quadratic error between objects in a cluster and their prototype (intracluster distance).

In the k-medoids algorithm there is no need to recalculate the distance between objects and the prototype at each iteration, since distances are already calculated in the distance matrix, significantly reducing the computational cost when compared to k-means [21].

Algorithm 1 k-medoids PAM Algorithm Pseudocode

```
def kmedoids(k, data):
    medoids = select_medoids(k, data)
    medoids_copy = medoids
    while not medoids == medoids_copy:
        groups = calc_groups(data, medoids, k)
        medoids_copy = update_medoids(
            groups, medoids)
        medoids = medoids_copy
    return groups
```

In general, the k-medoids algorithm is sensitive to the initial medoids, i.e., different initializations result in different data partitions [22]. By always selecting the same initial medoids,

we eliminate the stochastic factor of the algorithm and all runs will have the same result. In spite of using a random or a single initialization of the medoids, it is still necessary to choose the data objects that will serve as initial prototypes. To do so, we propose the following heuristic: assets with high or low volatility will be used as the initial prototypes for the cluster search. It is expected that the clusters found have medoids whose assets have high or low volatility, or that the medoids have a low dissimilarity with the initial search medoids, thus creating clusters whose medoids represent high or low volatility assets.

After the clusters are determined by the partitional clustering algorithm, it is necessary to define the weight of each asset, allocating resources within each group (intragroup allocation) and over the groups (intergroup allocation). The approach to be proposed here is based on the work of [23], who showed that optimization-based approaches are not consistently better than a naive $1/N$ (uniform) allocation.

C. Intragroup Allocation

The intragroup allocation is responsible for calculating the weight for each asset in a group and here we use the naive diversification, as explored by [23]. When using this form of allocation, each asset will have the same weight:

$$w_{i,j} = \frac{1}{N_j}, \tag{4}$$

where $w_{i,j}$ is the weight to each asset i in the group j, and N_j is the number of objects in cluster j.

D. Intergroup Allocation

After the weight is defined with intragroup allocation, it is necessary to calculate the final weight considering the number of clusters. To do so, we use the same naive approach of [23]:

$$w_{fi,j} = \frac{w_{i,j}}{k}, \tag{5}$$

where $w_{i,j}$ is the weight defined during intragroup allocation, k is the number of clusters and $w_{fi,j}$ is the final weight.

E. Performance Measures

To assess the proposed method, the following performance measures will be used:

- **Annualized Return**: This is the rate of return scaled to one year.

$$R = \frac{p(t)}{p(i)}^{(1/n)} - 1 \tag{6}$$

where $p(t)$ is the accumulated value of the portfolio at the observed time, $p(i)$ is the initial value and n is the number of years.

- **Annualized Volatility**: This is a statistical measure of the dispersion of returns.

$$\sigma_p = \sigma_{daily} * \sqrt{252} \tag{7}$$

where σ_p is the volatility of the series in the period, σ_{daily} is the standard deviation of the series of daily returns in

the period. This measure informs the realized portfolio volatility.

- **Sharpe Ratio**: This is a measure that indicates the risk-adjusted return on investment [24]:

$$SR = \frac{\mu - r_f}{\sigma},\qquad(8)$$

where μ is the return of the portfolio, r_f is the risk-free rate and σ is the portfolio volatility.

- **Turnover Ratio**: This is the measure that indicates the portfolio change with each rebalancing.

$$TR = \sum_{i=1}^{N} w_i(t) - w_i(t-1)\qquad(9)$$

where $w_i(t)$ is the weight of the asset in the current month and $w_i(t-1)$ is the weight of the asset in the previous month.

- **Maximum Drawdown**: This is the number that indicates in percent the greatest fall from peak to valley in the return series.
- **Sum of Squared Portfolio Weights**: This is a measure that indicates the level of diversification of an equity portfolio [25]:

$$SSPW = \sum_{i=1}^{N} w_i^2\qquad(10)$$

This measure indicates how concentrated a portfolio is, from 0 (less) concentrated to 1 (most) concentrated.

IV. Performance Assessment

The partitional clustering approach proposed here has five sequential phases, namely:

1) Calculate the correlation and distance matrix between assets;
2) Apply the partitional clustering algorithm;
3) Perform intragroup allocation;
4) Perform intergroup allocation;
5) Evaluate performance.

To make the asset allocation, it will be used assets traded in the Brazilian stock exchange (B3) and which are part of the Ibovespa index in the respective rebalancing. The method was tested between 2009-2017. The correlation coefficients will be determined according to Equation 2, and the distance will be calculated according to Equation 3. The partitional clustering algorithm with the initilization heuristic based on the volatility of assets is then applied. Intragroup and intergroup allocation are performed using the naive approach of [23] and the performance considers the measures presented in the Performance Measures subsection.

A. Experimental Methodology

The portfolio allocation process using partitional clustering is executed monthly, with rebalancing in the first business day of each month. To calculate the correlation matrix, it will be considered the closing price information in the previous

12 months. In each rebalancing, a set of assets is needed to calculate the correlation between the assets and then to perform the clusterings.

Two test scenarios were made. In the first scenario the experiments were performed using low volatility stocks as the initial medoids for the partitional clustering algorithm, and in the second scenario high volatility stocks were selected as the initial medoids.

As it is not possible to know a priori the ideal number of groups in partitional clustering, as previously mentioned, portfolio allocations will be performed using values of k between 3 and 20. The assets that will be part of each rebalancing are those that compose the theoretical portfolio of the Ibovespa index in the respective month of the rebalancing assessment.

B. Results for Scenario 1: Initialization with Low Volatility Assets

Table I summarizes the results obtained by the proposed approach with the partitional clustering method initialized with low volatility stocks and compares it with the Ibovespa index for different values of k ranging from 3 to 20. It is possible to observe that for all values of k the proposed method outperforms the index in terms of return, volatility, drawdown and sharpe. For some values of k, for instance k = 12, 15 and 16, the sharpe ratio provided by the method introduced is better than that of the Ibovespa Index.

The k-medoids algorithm is sensitive to the initial medoids, and tends to find clusters with low dissimilarity in relation to the initial medoids. It is possible to observe, from the results of Table I, where the initial medoids were the low volatility assets, that this happens because the resulting portfolios have lower volatility and smaller drawdowns when compared to the Ibovespa index.

C. Results for Scenario 2: Initialization with High Volatility Assets

Table II summarizes the results obtained by the proposed approach with the partitional clustering algorithm initialized with high volatility stocks and compares it with the Ibovespa index for different values of k ranging from 3 to 20.

It is possible to note that the algorithm also found clusters with low dissimilarity with the initial clusters, as they resulted in more volatile portfolios, with larger drawdowns and a worse risk-return ratio.

V. Conclusions and Future Research

This paper proposed the use of a partitional clustering method, more specifically, the k-medoids algorithm, with an initialization heuristic based on volatility, to allocate resources to a set of assets. The clustering method is responsible for finding clusters of assets based on their distance obtained from their price correlations. After segmenting the assets into clusters of similar assets, the resources are uniformly divided within the clusters and then among clusters.

The method was tested with stocks traded in the Brazilian Stock Exchange (B3) and which are part of the Ibovespa index

k	Return	Volatility	Turnover	Drawdown	Sharpe	SSPW
3	3.3%	20.9%	48.0%	-48.1%	-0.4	0.04
4	4.0%	20.3%	55.0%	-43.3%	-0.3	0.04
5	4.5%	19.4%	59.7%	-40.1%	-0.3	0.05
6	6.0%	19.3%	56.7%	-40.9%	-0.3	0.05
7	6.3%	19.0%	60.4%	-39.4%	-0.2	0.05
8	7.8%	18.4%	57.7%	-31.2%	-0.2	0.05
9	8.1%	18.2%	55.2%	-33.8%	-0.2	0.05
10	9.4%	18.1%	56.7%	-33.5%	-0.1	0.05
11	10.5%	17.7%	56.4%	-30.6%	-0.0	0.05
12	11.3%	17.3%	48.4%	-31.8%	0.0	0.05
13	8.9%	17.3%	44.9%	-35.6%	-0.1	0.04
14	9.9%	17.4%	43.9%	-34.3%	-0.1	0.04
15	11.2%	17.4%	42.3%	-32.6%	0.0	0.04
16	11.2%	17.3%	40.3%	-31.6%	0.0	0.04
17	10.5%	17.3%	38.3%	-29.9%	-0.0	0.04
18	9.8%	17.4%	39.1%	-29.8%	-0.1	0.04
19	9.4%	17.4%	37.6%	-30.1%	-0.1	0.03
20	8.6%	17.4%	37.5%	-31.1%	-0.1	0.03
Ibovespa	1.4%	22.7%	–	-48.6%	-0.4	0.05

k	Return	Volatility	Turnover	Drawdown	Sharpe	SSPW
3	4.1%	23.1%	57.7%	-54.3%	-0.3	0.06
4	-1.2%	23.0%	57.5%	-63.8%	-0.5	0.05
5	2.5%	24.0%	61.9%	-63.0%	-0.4	0.05
6	3.0%	24.7%	59.5%	-58.8%	-0.3	0.05
7	-0.3%	25.4%	59.6%	-66.4%	-0.4	0.05
8	1.7%	25.8%	59.2%	-63.6%	-0.4	0.05
9	-0.5%	25.8%	59.0%	-65.5%	-0.4	0.05
10	-1.9%	25.5%	56.6%	-70.4%	-0.5	0.04
11	-1.9%	25.8%	55.3%	-70.2%	-0.5	0.04
12	-0.0%	26.1%	54.9%	-66.1%	-0.4	0.04
13	2.4%	26.5%	54.0%	-64.0%	-0.3	0.04
14	1.4%	26.5%	49.9%	-66.2%	-0.4	0.04
15	2.2%	26.3%	48.9%	-66.4%	-0.3	0.04
16	2.4%	26.1%	48.1%	-65.0%	-0.3	0.04
17	4.2%	26.0%	46.3%	-60.6%	-0.3	0.03
18	3.7%	26.0%	46.4%	-61.7%	-0.3	0.03
19	4.9%	26.1%	45.1%	-59.3%	-0.2	0.03
20	5.2%	25.9%	42.5%	-58.0%	-0.2	0.03
Ibovespa	1.4%	22.7%	-	-48.6%	-0.4	0.05

over the period from 2009 to 2017. The performance was measured in terms of return, volatility, sharpe ratio, turnover, drawdown and portfolio diversification.

From the results observed, it is possible to note that the portfolios found presented the expected results, that is, when the medoids were initialized with low volatility assets, the resulting portfolio presented low volatility, and vice-versa. This behavior is similar to the effect discussed by [26].

The results obtained presented in Table I showed that our approach is superior to the Ibovespa index in all measures. The poor sharpe ratio can be explained by the poor performance of the Brazilian stock market during the period and by the high brazilian risk-free rate, which was 10.9% annualized during the period.

As future work, it is possible to explore the use of evaluation metrics for machine learning algorithms, both in assessing the quality of the clusters created in relation to the portfolio performance, and in the application of other techniques to realize intragroup and intergroup allocation. Another possibility for future work is the application of the statistical framework proposed by [27] to analyze which is the most suitable time series period for the application of the partitional clustering.

REFERENCES

[1] Institutional Investor, The Hedge Fund 100: Quant Funds Excel — at Asset Gathering, 2018. https://www.institutionalinvestor.com/article/b19dxvr2pvgjq2/The-Hedge-Fund-100-Quant-Funds-Excel-at-Asset-Gathering

[2]   Markowitz, Harry. "Portfolio selection." The journal of finance 7, no. 1, 77–91, 1952.

[3]   Ponsich, Antonin, Antonio Lopez Jaimes, and Carlos A Coello Coello. "A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem andother finance and economics applications." IEEE Transactions on Evolutionary Computation 17, no. 3, 321-344, 2013.

[4]   Black, Fischer, and Robert Litterman. "Global portfolio optimization." Financial Analysts Journal 48, no. 5, 28–43, 1992.

[5]   Michaud, Richard O. "The Markowitz optimization enigma: Is 'optimized'optimal?" Financial Analysts Journal 45, no. 1, 31–42, 1989.

[6]   López de Prado, Marcos. "Building diversified portfolios that outperform out-of-sample," 2016.

[7]   Clarke, Roger, Harindra De Silva, and Steven Thorley. "Risk parity, maximum diversification, and minimum variance: An analytic perspective." The Journal of Portfolio Management 39, no. 3, 39-53, 2013.

[8]   Raffinot, Thomas. "Hierarchical clustering based asset allocation." Available at SSRN 2840729, 2017.

[9]   Duarte, Flavio Gabriel, and de Castro, Leandro Nunes. "A Framework to Perform Asset Allocation Based on Partitional Clustering," IEEE Access, vol. 8, pp. 110775-110788, 2020, doi: 10.1109/ACCESS.2020.3001944.

[10] de Castro, L. N., and D. G. Ferrari. Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações. 116–119. 2016.

[11] Carmichael, J. W., and R. S. Julius. "Finding Natural Clusters." Systematic Biology 17, no. 2, 144-150, 1968.

[12] Aloise, D., A. Deshpande, P. Hansen, and P. Popat. "NP-hardness of Euclidean sum-of-squares clustering." Machine Learning75, no. 2, 245-249, 2009.

[13] Jain, A. K., M. N. Murty, and P. J. Flynn. "Data Clustering: A Review." ACM Computing Surveys 31, no. 3, 264-323, 1999.

[14] Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions ofrisk." The journal of finance 19, no. 3, 425-442, 1964.

[15] Markowitz, Harry. "The optimization of a quadratic function subject to linear constraints." Naval research logistics Quarterly 3, nos. 1-2, 111–133, 1956.

[16] Kaufman, Leonard, and Peter Rousseeuw. Clustering by means of medoids. North-Holland, 1987.

[17] Marti, Gautier, Frank Nielsen, Mikołaj Binkowski, and Philippe Donnat. "A review of two decades of correlations, hierarchies, networks and clustering in financial markets." arXiv preprint arXiv: 1703.00485, 2017.

[18] Mantegna, Rosario N. "Hierarchical structure in financial markets."The European PhysicalJournal B-Condensed Matter and Complex Systems 11, no. 1, 193-197, 1999.

[19] Jin, Xin, and JW Han. K-Medoids clustering, Encyclopedia of machine learning, 564–565, 2010.

[20] Krishnapuram, Raghu, Anupam Joshi, and Liyu Yi. "A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering." Fuzzy Systems Confer-ence Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International, 3: 1281–1286. IEEE, 1999.

[21] Reynolds, Alan P, Graeme Richards, and Vic J Rayward-Smith. "The application of k-medoids and pam to the clustering of rules." In International Conference on Intelligent Data Engineering and Automated Learning, 173–178. Springer, 2004

[22] Meng, Ying, Ke Luo, and Jian Hua Liu. "An improved k-medoids clustering algorithm." In Advanced Materials Research, 562: 2106–2110. Trans Tech Publ, 2012

[23] DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification:How inefficient is the 1/N portfolio strategy?" The review of Financial studies 22, no. 5, 1915–1953, 2009.

[24] Sharpe, William F. "Mutual fund performance." The Journal of business 39, no. 1, 119–138, 1966.

[25] Goetzmann, William N, and Alok Kumar. "Equity portfolio diversification." Review of Finance12, no. 3, 433–463, 2008.

[26] Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler. "Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly."Financial Analysts Journal67, no. 1, 40–54, 2011.

[27] Marti, Gautier, Sébastien Andler, Frank Nielsen, and Philippe Donnat. "Clustering financialtime series: How long is enough?" arXiv preprint arXiv: 1603.04017, 2016.