Network models to improve robot advisory-portfolio management

Paolo Giudici

University of Pavia, Pavia, Italy

Gloria Polinesi

Universitá Politecnica delle Marche, Ancona, Italy

Alessandro Spelta

Center for Analysis Decision and Society, Milan, Italy

University of Pavia, Pavia, Italy

Abstract

Robot advisory services are rapidly expanding responding to a growing interest people have in direct managing their savings. Robot advisors may reduce costs and improve the quality of the service, making user involvement more transparent. In particular, they let customers' independently specify the preferred investment risk profile. Against this background however, there exists the possibility that robot advisors underestimate market risks, especially when highly correlated assets are being considered. This may lead to a lack of compliance between investors' expected and actual risk. The aim of the paper is to enhance robot advisory portfolio allocation, taking users' preference into account. In particular, we demonstrate how Random Matrix Theory and network metrics can be combined to construct investment portfolios that provide lower risks with respect to standard Markovitz portfolios. To demonstrate the advantages of this approach we employ the observed returns of a large set of ETFs, which is representative of the financial products at the ground of the activity of robot advisors.

Keywords: Correlation networks, Network centrality, Portfolio optimization, Random Matrix theory.

---

1. Introduction and background

Financial Technologies (FinTech) can be broadly defined as technologically enabled financial innovations that could result in new business models, applications, processes or products with an associated material effect on financial markets, financial institutions, and on the provision of financial services (Carney, 2017). From the last few years, FinTech innovations are increasing exponentially, ranging from payments to lending, from insurance to asset management. The Financial Stability Board (FSB) in its two recent reports (Board (2017a), Board (2017b)) identifies three common drivers for FinTech: shifting consumer preferences on the demand side; evolving technology and changing financial regulation on the supply side.

Within this background, robo-advisor services of automated investment advice are growing fast to address the need of directly managing savings. They are accessible via online platforms and, therefore, allow to act quickly and in the first person. According to Statista, in 2019, the masses managed by the automatic consultancy are estimated around 980 billion dollars reaching in 2023 over 2,552 billion[1].

However, the rapid growth of FinTech activities, and of robo-advisors in particular, has determined the emergence of new financial risks. In this perspective, robo-advisors that build personalised asset portfolios on the basis of automated algorithms have been suspected of underestimating investors' risk profile in their actual portfolio allocation. This is especially because, on one side, the user could not understand the mechanisms underlying the portfolios construction and, on the other side, because computational models that are employed to build portfolios are often simplified, and do not take multivariate dependencies between asset returns properly into account.

In this paper we propose to exploit topological information embedded into similarity networks to cope with this issues. Understanding the structure of a similarity network (see Mantegna and Stanley (1999)) is indeed instrumental for understanding the origin and the distribution of asset returns and to build portfolios more robust against adverse shocks hitting the economy. Similarity patterns between asset returns can be extracted from a distance matrix and they can reveal how asset performances are related to the topology of the network. To account for such topological information we rely on centrality

[1]For more information please see: https://www.statista.com/outlook/337/100/robo-advisors/worldwide

measures (see Newman (2018)). More specifically we employ a global centrality measure that provides information on the position of each node relative to all other nodes; namely the eigenvector centrality (see Bonacich (2007)). This measure assigns relative scores to all nodes in the network, based on the principle that connections to few high scoring nodes contribute more to the score of the node in question than equal connections to low scoring nodes. We show that the inclusion of centrality measures into portfolio generative models improve their performances in terms of risk/returns.

Moreover, we propose a portfolio allocation methodology that correctly takes multivariate dependencies and risk contagion into account and, consequently, improves the matching between the expected and the actual risk profile of investors. The information contained in a correlation matrix is a crucial aspect for portfolio allocation. Our aim is to show that topological methods based on the correlation matrix, not yet employed by robo-advisory platforms, can generate new portfolio allocation strategies able to systematically over-perform the benchmark index and other simpler investment strategies in terms of the derived risk/returns.

The literature on stock and financial networks based on correlation matrices stems from the seminal paper of Mantegna (1999). The author employs correlation matrices to detect the hierarchical organization of stock markets; a distance metric based on correlation matrices is used to determine a "minimal spanning tree" (MST), which connects stock returns on a considered in a time horizon. Tumminello et al. (2005) extend Mantegna (1999) with a topological generalisation of the MST, the Planar Maximally Filtered Graph (PMFG), which retains the same hierarchical properties of the MST but adds more complex structures, such as loops and cliques. Tola et al. (2008) show how, using a range of clustering techniques on correlation matrices, in combination with filtering approaches based on the random matrix theory (RMT), can improve the reliability of portfolios in terms of the ratio between expected and realized risk. They then construct portfolios using Markowitz's optimisation model. Other contributions that follow Tola et al. (2008): León et al. (2017), Raffinot (2017), Ren et al. (2017) and Zhan et al. (2015). This literature also looks for a hierarchical structure of stock returns in financial markets and, on the basis of it, construct investment portfolios based on the Markowitz approach.

We follow the previous stream of literature, with two main original contributions. From an applied viewpoint, we extend the methodology from stock returns to Exchange Traded Fund returns (ETFs). The term ETF identifies a particular type of investment fund with two main features: it is traded on the stock exchange like a stock and it aims to replicate the index to which it refers (benchmark) through totally passive management. An ETF summarizes the characteristics of a fund as a stock, allowing investors to exploit the strengths of both instruments: diversification and risk reduction of the funds; flexibility and information transparency of real-time trading of shares. From a methodological viewpoint, we propose a portfolio optimisation approach different from what proposed by Tola et al. (2008), taking network centrality explicitly into account in Markowitz model. In this work, we do not rely only on a typical indicator of diversification, such as the covariance between assets returns, but we apply a topological measure which embeds also higher order information of asset performances to built suitable portfolios.

The empirical findings obtained from the application of our proposed methods confirm the validity of the proposed approach and, thus, the proposed methodology can constitute a new instrument in robo-advisor toolboxes.

The structure of the paper is as follows: Section 2 presents the Random Matrix Theory, a technique to purge data from noise components, and the minimal spanning tree approach able to build, in a parsimonious way, a correlation network among assets. The Section also presents a new portfolio optimisation strategy that embeds topological information extracted from the network through a centrality measure, namely the eigenvector centrality. Section 3 presents the results of the application of our models to ETFs data managed by a leading European Robot Advisory platform. Section 4 ends with some concluding remarks.

2. Methodology

2.1. The Random Matrix approach

Since the beginning of the 20th century, Random Matrix Theory (RMT) has been used in various applications, ranging from quantum mechanics (Beenakker, 1997), condensed matter physics (Guhr et al., 1998), wireless communications (Tulino et al., 2004), to economics and finance (Potters et al., 2005). In a nutshell, RMT aims at separating the "sistematic" part of a signal embedded into a correlation matrix from the "noise" component.

The basic idea of RMT is to test the subsequent empirical eigenvalues of a correlation matrix: $\lambda_k < \lambda_{k+1}; k = 1, \ldots, n$, against the null hypothesis that they are equal to the eigenvalues of a random Wishart matrix $\mathbf{R} = \frac{1}{T}\mathbf{A}\mathbf{A}^\mathbf{T}$ of the same size, with $\mathbf{A}$ a $N \times T$ matrix containing N time series of length T whose elements are independent and identically distributed random variables, with zero mean and unit variance.

It can be shown (see Marchenko and Pastur (1967)) that, as $N \to \infty$ and $T \to \infty$, with the ratio $Q = \frac{T}{N} \geq 1$ fixed, the density of the sample eigenvalues converges to:

$$f(\lambda) = \frac{T}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{1}$$

where $\lambda \in (\lambda_-, \lambda_+)$, $\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}$.

It follows that, when $\lambda_k > \lambda_+$ the null hypotheses is rejected, from the k-th eigenvalue on-wards. Then, RMT "reconstruct" the correlation matrix applying a singular value decomposition based only on the eigenvectors that corresponds to the eigenvalues that are greater than the k-th.

In other words, RMT eigendecomposes a correlation matrix of time series by returning a filtered correlation matrix (see Eom et al. (2009)). The observed correlation matrix can be used to extract the observed eigenvalues and, then, through Equation 1 to check whether to reject the null hypotheses, for any given eigenvalue. Plerou et al. (2002) show that the characteristic directions of the signal correspond to the eigenvalues that are clearly different from those obtained from the random Wishart matrix. They define a subspace which contains the systemic information related the market structure. This corresponds to the identification of empirically constructed variables that drive the EFT system and, in this framework, the number of surviving eigenvalues is the effective characteristic dimension of this economic space.

More formally, let r_i, for $i = 1, \ldots, n$, be a time series of asset returns, computed, for any given time point t, as the difference between the logarithms of daily asset prices:

$$r_i(t) = logP_i(t) - logP_i(t-1). \tag{2}$$

Given a set of N asset return series, a correlation coefficient between any two pairs can be defined as:

$$c_{ij} = \frac{E(r_i r_j) - E(r_i)E(r_j)}{\sigma_i \sigma_j}, \tag{3}$$

where $E(\circ)$ and $\sigma(\circ)$ indicate, respectively, the mean and the standard deviation operators. Let $\mathbf{C}$ be the matrix that contains all pairwise correlations, the correlation matrix.

According to the RTM the filtered correlation matrix given by:

$$\mathbf{C}' = \mathbf{V \Lambda V^T}, \tag{4}$$

where

$$\mathbf{\Lambda} = \begin{cases} 0 & \lambda_i < \lambda_+ \\ \lambda_i & \lambda_i \geq \lambda_+ \end{cases}$$

and $\mathbf{V}$ represents the matrix of the deviating eigenvectors associated to the eigenvalues greater than λ_+.

2.2. The Minimal Spanning Tree

With the filtered correlation matrix $\mathbf{C}'$ obtained from RMT the next step is to find out the sparse representation of the relationships derived by such matrix. To accomplish this purpose we derive the Minimal Spanning Tree (MST) representation of EFT returns similarities (see Mantegna and Stanley (1999), Bonanno et al. (2003), Spelta and Araújo (2012)). In particular, Mantegna (1999) demonstrates that this particular type of graph, derived from the correlations between the stock returns, reveals a topological arrangement of financial markets that has an important meaning from an economic point of view. Each pairwise correlation obtained through RTM can be converted in an Euclidean distance by the function:

$$d_{ij} = \sqrt{2 - 2c'_{ij}}, \tag{5}$$

and the pairwise distances can be organised in a distance matrix $\mathbf{D} = \{d_{ij}\}$, which can be used to draw the MST[2]. The MST is derived from a single linkage clustering algorithm which, based on the distance matrix, associates each asset node to its closest neighbour, avoiding loops. The term "minimal" refers to the fact that MST allows to reduce the number of links between

[2]Moreover Raffinot (2017) extends the MST considering some clustering variants, such as complete linkage (CL), average linkage (AL) and Ward's Method (WM). He shows however that different algorithms differ in terms of grouping structure, but not in terms of performance.

assets from $\frac{N(N-1)}{2}$ to $N-1$ and the sum of those links provide the minimum weight of the graph. At the initial step, we consider N clusters corresponding to the N EFTs. Then, at each subsequent step, two clusters l_i and l_j are merged into a single cluster if:

$$d\left(l_i, l_j\right) = \min\left\{d\left(l_i, l_j\right)\right\}$$

with the distance between clusters being defined as:

$$\hat{d}\left(l_i, l_j\right) = \min\left\{d_{pq}\right\}$$

with $p \in l_i$ and $q \in l_j$. These operations are repeated until a single cluster emerges[3]. Understanding the structure of the similarity network, and in particular determining which nodes act as hubs in the network, is key for understand how EFT returns behave in a multidimensional space and thus to construct portfolios that suitably take into account risk/return trade-off. Moreover, the investigation of the network topology is meaningless while evaluating the dynamics of the system through time; in order to detect how financial relationships evolve over time Spelta and Araújo (2012) employ a measure called *residuality* coefficient that compares the relative strengths of the connections above and below a threshold distance value, in folmula:

$$R = \frac{\sum\limits_{d_{i,j} > L} d_{i,j}^{-1}}{\sum\limits_{d_{i,j} \leq L} d_{i,j}^{-1}} \tag{6}$$

where L is the highest threshold distance value that ensures connectivity of the MST.

Indeed, it has been empirically observed during financial crisis the structure of the MST is reinforced in the topological sense, as revealed by the *residuality* coefficient. The number of redundant elements that characterize the distinct time periods provides evidence on the structural changes taking place on the network structure. They are due to the emergence of high correlated positions (synchronization) in the network. Overall, R decreases when a network becomes less sparse (the number of links increases), and vice-versa.

[3] We use the symbols $\hat{d}$ and $\hat{\mathbf{D}}$ to denote the distances representing the MST derived from the fully connected network $\mathbf{D}$.

2.3. The eigenvector centrality

The research in network theory has dedicated a huge effort to developing measures of interconnectedness, related to the detection of the most important player in a network. The idea of centrality was initially proposed in the context of social systems, where a relation between the location of a subject in the social network and its influence on group processes was assumed. Various measures of centrality have been proposed in network theory such as the count of neighbors of a node has, i.e. the degree centrality, which is a local centrality measure, or measures based on the spectral properties of the graph (see Perra and Fortunato (2008)). Spectral centrality measures include the eigenvector centrality (Bonacich (2007)), Katzs centrality (Katz (1953)), PageRank (Brin and Page (1998)), hub and authority centralities (Kleinberg (1999)). These measures are feedback, also know as global, centrality measures and provide information on the position of each node relative to all other nodes.

The eigenvector centrality measures the importance of a node in a network by assigning relative scores to all nodes in the network, based on the principle that connections to few high scoring nodes contribute more to the score of the node in question than equal connections to low scoring nodes. More formally, for the i-th node, the centrality score is proportional to the sum of the scores of all nodes which are connected to it, as in the following equation:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} \hat{d_{i,j}} x_j \tag{7}$$

where x_j is the score of a node j, $\hat{d_{i,j}}$ is the $(i;j)$ element of the adjacency matrix of the network, λ is a constant and N is the number of nodes of the network.

The previous equation can be rewritten for all nodes, more compactly, as:

$$\hat{\mathbf{D}}\mathbf{x} = \lambda \mathbf{x} \tag{8}$$

where $\hat{\mathbf{D}}$ is the adjacency matrix, λ is the eigenvalue of the matrix $\hat{\mathbf{D}}$, with associated eigenvector x, an N-vector of scores (one for each node). Note that, in general, there will be many different eigenvalues λ for which a solution to the previous equation exists. However, the additional requirement that all the elements of the eigenvector be positive (a natural request in our

context) implies (by the Perron-Frobenius theorem) that only the eigenvector corresponding to the largest eigenvalue provides the desired centrality measures. Therefore, once an estimate of $\hat{\mathbf{D}}$ is provided, network centrality scores can be obtained from the previous equation, as elements of the eigenvector associated to the largest eigenvalue.
Notice that, in our networks that are based on distances between objects, the higher the centrality measures associated to a node, the more the node is dissimilar with respect to its peers (or with respect to all other nodes in the network).

2.4. Portfolio construction

In this Section, we show how correlation networks and topological measures can be used in a portfolio optimization framework and how they can contribute to improve portfolio performances. Correlation between stocks play a central role in investment theory and risk management being key elements for optimization problems as in the Markowitz portfolio theory Markowitz (1952). Thus, correlation based graphs could be very useful for analyzing the interactions between financial markets and building optimal investment strategy. Onnela et al. (2003) show that the assets of classic Markowitz portfolio (Markowitz (1952)) are always located on the outer nodes of a MST, i.e. the portfolio is mainly composed by asset that lay in the periphery of the network and not in its core. Moreover, Pozzi et al. (2013) show how an efficient portfolio strategy benefits from the knowledge of the hierarchical structure of the market: portfolios including central assets are characterized by greater risk and worse performance with respect to portfolios including peripheral assets. Vỳrost et al. (2018) suggest that network-based asset allocation strategies may improve risk-return profiles. Their work is based on the study of Peralta and Zareei (2016) which find a negative relationships between the centrality of assets within a complex financial network and the optimal weights under the Markowitz model. Other authors have gone beyond the above remarks and proposed novel portfolio optimisation strategies. For example, Plerou et al. (2002) and Conlon et al. (2007) use the correlation matrix, filtered with the random matrix approach, in the Markowitz model. They show that for the obtained portfolios the realized risk is closer to the expected one while Tola et al. (2008), combining the MST with the RMT filtering, provide improvement with respect to Markowitz portfolios. Finally, Tumminello et al. (2010) demonstrate that the risk of the optimized portfolio obtained using a "filtered" correlation matrix is more stable than the non filtered matrix. As the above authors, we intend to exploit topological measures to improve portfolio performances with respect to the standard Markowitz approach. However, differently from previous works which employ RMT and MST as an alternative measure of diversification risk, we extend Markowitz' approach using RMT and MST in the optimisation function itself, rather than applying Markowitz to the filtered and simplified correlation matrix. In our case we minimize the following constrained objective function:

$$\min_{\mathbf{w}} \mathbf{w^T COV' w} + \gamma \sum_{i=1}^{n} x_i w_i \tag{9}$$

subject to

$$\begin{cases} \sum_{i=1}^{n} w_i = 1 \\ \mu_P \geq \frac{\sum_{i=1}^{n} \mu_i}{n} \\ w_i \geq 0 \end{cases}$$

where μ_P indicates the returns of the portfolio, the parameter γ represents a risk aversion coefficient, x_i is the eigenvector centrality associated with EFT i while the i, j element of the $\mathbf{COV'}$ is given by $\sigma_i \sigma_j c'_{i,j}$. The basic principle that governs Markowitz's theory is that, in order to build an efficient portfolio, it is necessary to identify a combination of securities such as to minimize risk and maximize total return by offsetting the asynchronous trends of the individual securities. In a nutshell, securities that make up the portfolio must be uncorrelated or, rather, not perfectly correlated. Within the minimization problem we are face, the component derived from the MST structure is related to additional risk that eventually, an investor want to minimize. In other words, by increasing the value of γ we are asking whether this topological measure is a meaningful measure of risk and whether its inclusion in the minimization problem produces superior performance with respect to the standard Markowitz portfolio. We remark that, in our formulation, the risk aversion coefficient γ expresses aversion towards systemic risk and not, as in classical Markowitz', towards risk in general. Therefore, since a high centrality is inversely related to the distance the asset has with respect to all other ETFs in the network, a high risk propensity (represented by a high value of γ) translates in a portfolio composed by a more systemically risky assets that lay in the central body of the network thus avoiding peripheral ETFS.

3. Empirical findings

The data we consider to apply our model contains 92 time series of returns referred to ETFs traded over the period January 2006 - February 2018 (3173 daily observations).

Table 1 shows the classification of the 92 ETFs in 11 asset classes, according to the classification provided by the Exchange where they are traded.

	ETF class	**Number of ETFs**
1	Aggregate Bonds	4
2	Commodity	8
3	Corporate-euro	11
4	Corporate-not euro	3
5	Corporate-high yield	2
6	Corporate-world	1
7	Emerging Equity-Asia	30
8	Emerging Equity-America	10
9	Emerging Equity-East Europe	4
10	Emerging Equity-world	17
11	Equity-Europe	1

Table 1: **ETFs by Asset classes.** Number of Exchange Traded Funds for each class.

From Table 1 note that the emerging market asset classes are the most frequent, followed by corporate ETFs.

Table 2 describes some summary statistics for the considered asset classes and, specifically, the mean, variance and kurtosis of the returns distribution, to describe their location and variability.

From Table 2 note that the mean value of the returns is around 0 for each asset class, consistently with the efficient market hypothesis suggested by Malkiel and Fama (1970). Differently, the standard deviation depends on the considered asset class: Emerging equity and Commodity classes are more volatile with respect to the corporate classes. Moreover, the high values of the kurtosis confirm known stylized facts: the distribution of most ETF asset returns tends to be non-Gaussian and heavy tailed.

To compare crisis phases with expansive market phases, the dataset has been divided in two chronologically successive batches, from 2006 to 2012

	ETF class	mean	std.	kurtosis
1	Aggregate Bond	0.00014	0.00265	6.66
2	Commodity	-0.00007	0.01052	3.64
3	Corporate-euro	0.00014	0.00155	3.35
4	Corporate-not euro	0.00021	0.00454	5.36
5	Corporate-high yield	0.00040	0.00602	24.42
6	Corporate-world	0.00017	0.00320	4.32
7	Emerging Equity-Asia	0.00036	0.01541	11.43
8	Emerging Equity-America	0.00024	0.01928	8.99
9	Emerging Equity-East Europe	0.00011	0.02380	18.19
10	Emerging Equity-world	0.00026	0.01235	9.10
11	Equity-Europe	0.00018	0.01213	6.96

Table 2: **ETF classes summary statistics.** The table shows summary statistics for Exchange Traded Funds classes compositions: mean, standard deviation and kurtosis computed for the whole dataset.

(crisis) and from 2013 to 2018 (post-crisis). Figure 1 provides temporal boxplots for ETF returns, grouped by their asset class, as in Table 1.

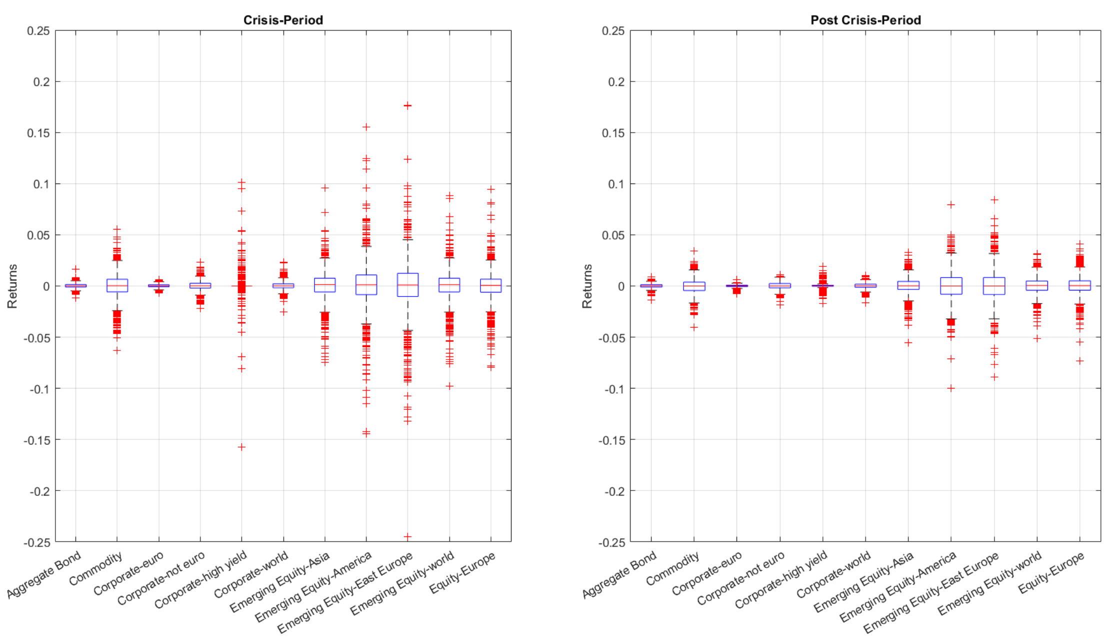

Figure 1: **Summary plots for ETF class returns**. Two different periods are compared: crisis (2006-2012) and post crisis (2012-2013).

Figure 1 shows that the volatility of the ETFs belonging to the Emerging equity classes, regardless of the geographical area considered, as well as that of the Commodity asset class, is larger during the crisis period. This particularly explains why their overall standard deviation, reported in 2, is higher.

3.1. Random Matrix Theory and network topology

To apply Random Matrix Theory filtering, we first need to divide the data into consecutive overlapping time windows. The width of such windows has been set equal to $T = 250$ (12 trading months), with a window step length of one month ($\cong$ 21 trading days) for a total of 140 monthly one-year windows. For each time window, we use 11 months of daily observations to build our model and the remaining month to validate it. This means, in particular, that we calculate 140 correlation matrices between all 92 ETF returns, based on 11 months of data and, then,"filter" them applying the Random Matrix Theory approach.

Figure 2 shows the ordered eigenvalue distribution, for both the actual and the random correlation matrices, for the last time window of the dataset (March 2017- January 2018), employed to predict February 2018.

Figure 2 shows that the most part of the eigenvalues distribution lies between λ_{min} and λ_{max}, which are respectively equal to 0.16 and 2.71. This "bulk" may be considered as being generated by random fluctuations while the six deviating eigenvalues that are greater than λ_{max} represent the effective characteristic dimension of the economic space described by the correlation matrix. Similar considerations occurr for the other considered time windows.

As described in the methodological section, if, for each time window, we reconstruct the correlation matrix using only the eigenvectors that correspond to the deviating eigenvectors, we obtain a sequence of "filtered" correlation matrices which can be used to improve the Minimum Spanning Tree representation. Figure 3 plots, for each time window, the most central node, defined by the ETF with the highest degree (with the largest number of connected nodes), in the MST representation, using both the filtered and the unfiltered correlation matrices.

From Figure 3 note that the RMT approach allows to achieve a more diversified Minimal Spanning Tree over time: central vertices according to the highest degree criterion are different and belong to different ETF classes. On the other hand, the MSTs based the empirical correlation matrices seem to almost always repeat the same structure: the ETF EIMI-IM, belonging to the Asia Emerging Market class, is for most of the time the central node.

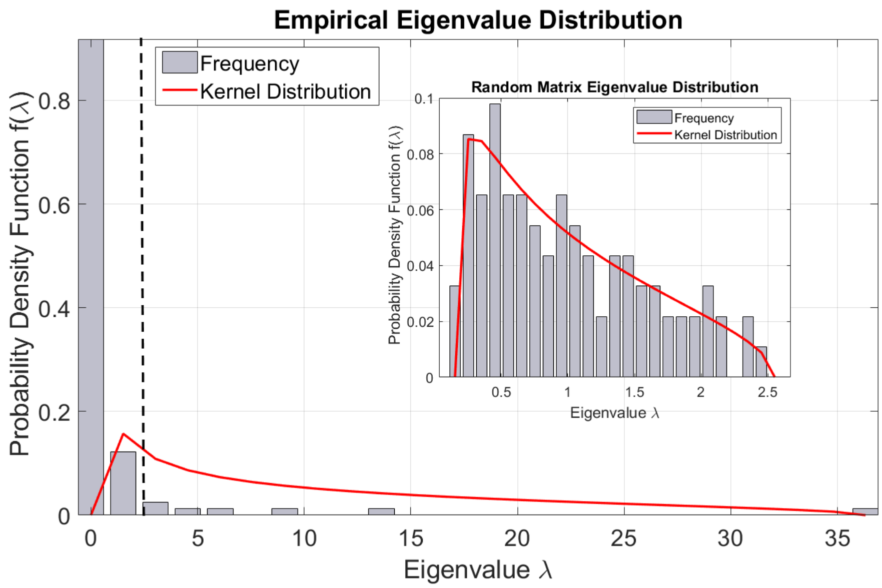

Figure 2: **Eigenvalue distribution for the last time window.** In the figure, the red line shows the kernel density of the eigenvalues associated to the empirical correlation matrix C, in the main graph, and of the random correlation matrix R, in the smaller graph. The dashed vertical line indicates the threshold value λ_{max} which separates the "signal" eigenvalues from the "noise" ones.

Before moving to portfolio optimisation, we evaluate how the MSTs dynamically change over time. To this aim, we employ, as a summary measure of each MST, the Max link, the maximum distance value between two pairs of nodes used in the construction of the tree, and the *residuality* coefficient (introduced in Spelta and Araújo (2012)), which measures the ratio between links eliminated and maintained by the MST. Figure 4 shows how the evolution of these two summaries, over the considered time windows.

From Figure 4 note that, during the 2008 financial crisis, the Max link sharply decreases, due to the decrease of most distances between ETF returns. This can be explained by the increased correlations between all asset returns, which synchronise during crisis, consistently with the literature findings. While the Max link bounces back after the crisis, the residuality coefficient continues its decline until the "normal times" of 2014. This may indicate the persistence of a set of strong connections in the market, that determine the relevance of a limited number of links.

To better understand the previous findings, Figure 5 compares an MST

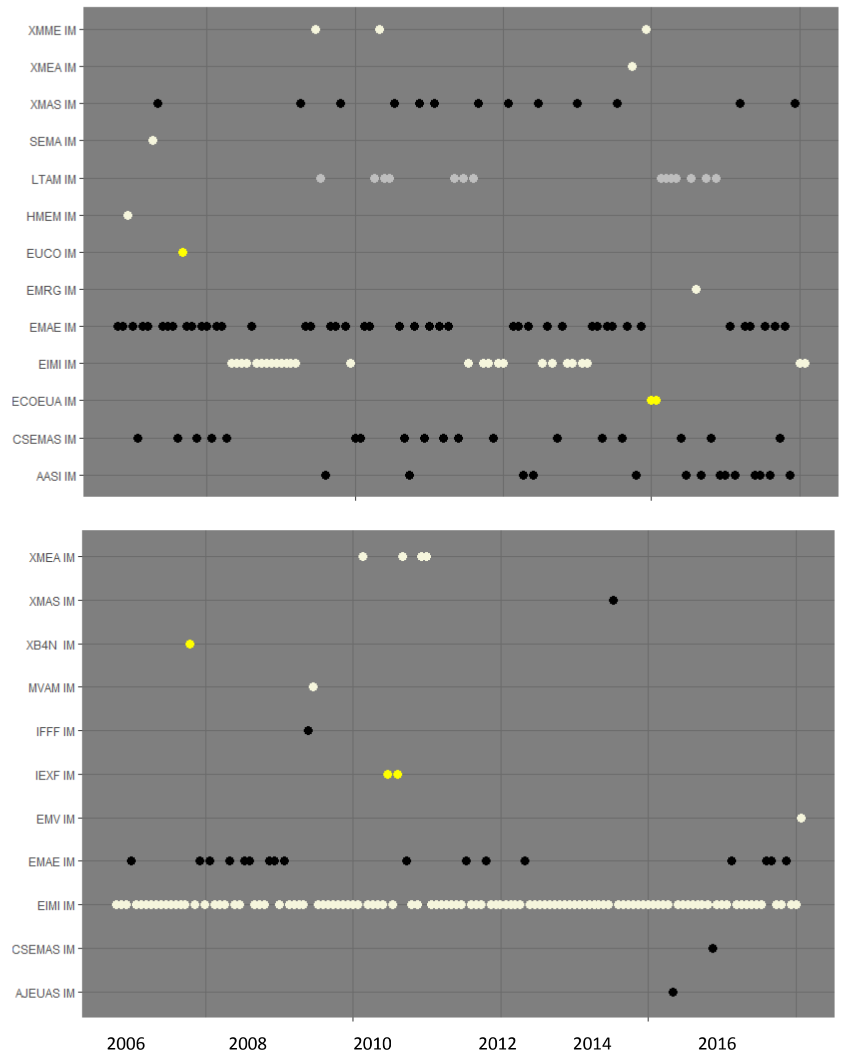

Figure 3: **Central nodes of MSTs, along time windows.** The figure reports the ETF node that has the highest degree in the MST representation, in each of the 140 time windows, considering the filtered correlation matrix (top) and the empirical correlation matrix (bottom). Node colors represent the belonging class of ETFs: Corporate (yellow), Emerging Market Asia (black), America (grey), World (beige).

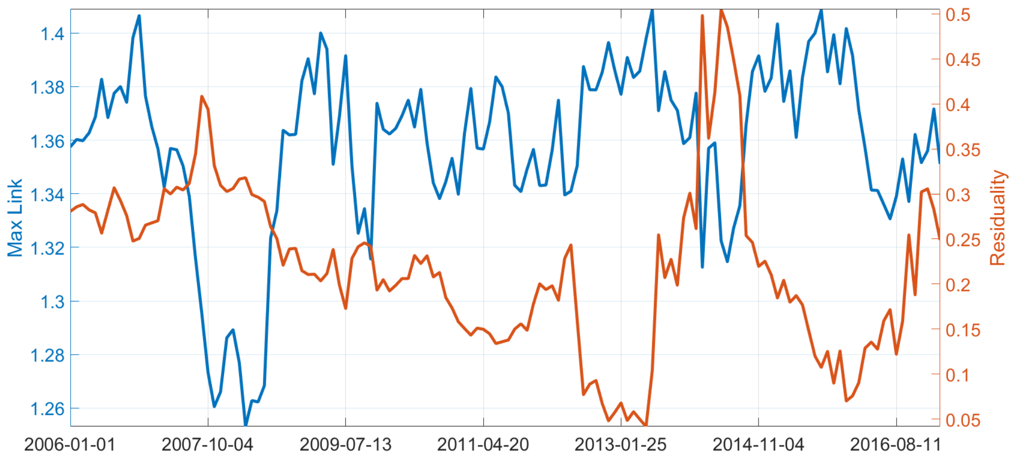

Figure 4: **MST thresholds and residuality coefficients**. The blue line shows the Max link distance, while the red line shows the *residuality* coefficient, whose values are reported respectively on the left and right-y axis.

obtained during the only 2008 time window, representative of crisis times, with respect to the MST of the last time window considered, taken as a reference for a "business as usual" market phase.

Figure 5 reflects how correlations increase during crisis phases, leading to more links in the MST representation. In addition, the crisis period MST indicates the importance of the Asian, American and World Emerging Market classes, which have the highest centralities. The importance of the American emerging market node declines post crisis, but the Asian one remains high. This may explain the persistence of low values for the residuality coefficient, after crisis times.

3.2. Portfolio construction

We now present the application of our proposed portfolio strategy, in which the eigenvector centrality computed on the MST derived from RMT is inserted as an additional diversification measure of risk in the objective function of the Markowitz optimization problem.

The optimal weights are obtained by minimizing the constrained objective function (see Equation 9), where the value of γ is set a priori on the basis of the level of risk aversion of the investors. A high value of γ indicates that,

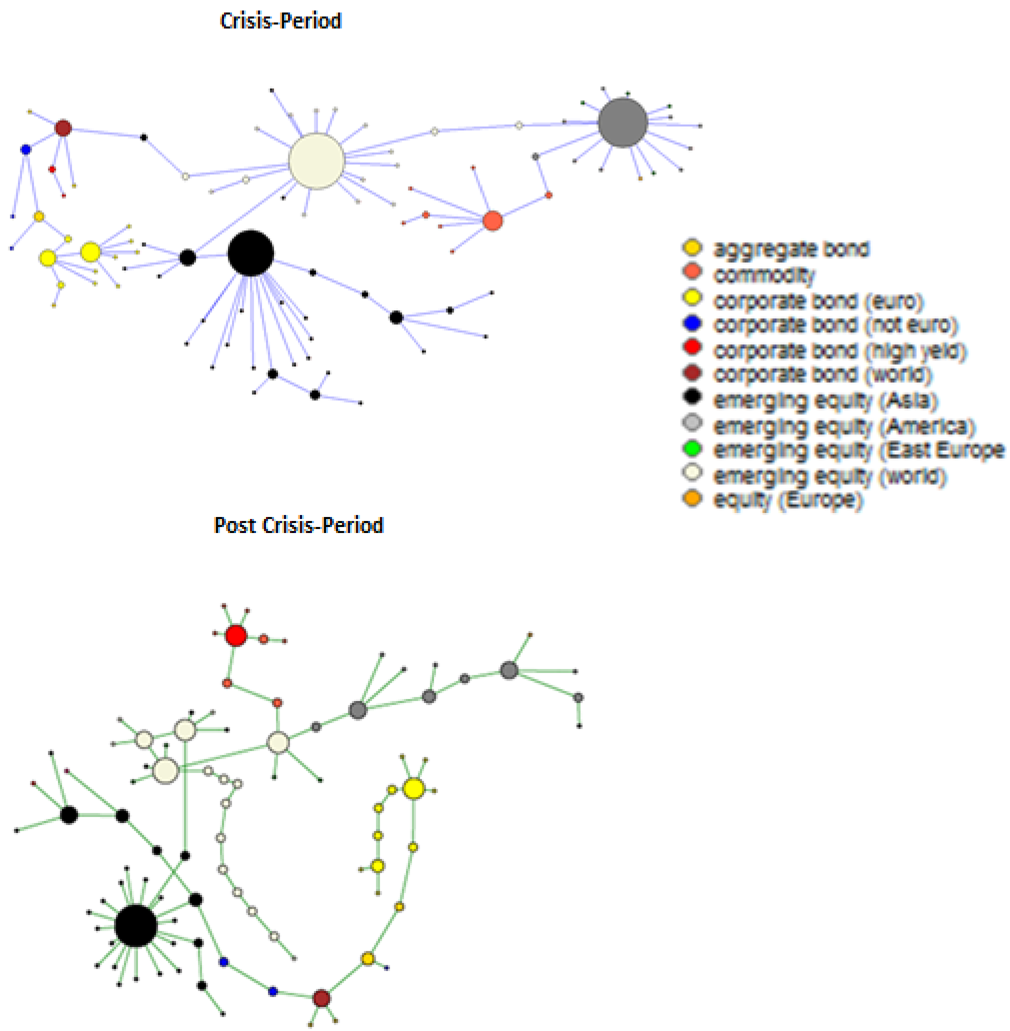

Figure 5: **Minimal Spanning Tree drawn from the RMT filtered correlation matrix for crisis and post-crisis periods**. The nodes in the figure indicate ETFs, the size of the node represents their degree centrality. Colors indicate different asset classes, as reported in the legend.

in the desired allocation of financial assets, more central ETFs (such as the Emerging markets ones) weight more.

Portfolio returns are computed using the last month of each time window, in an out-of-sample manner. The resulting Profit & Losses, occurred in the last month of each time window can then be calculated.

More precisely, we use eleven months of observations as a build-up period, computing asset centralities and the consequent portfolio weights. We then calculate the return of each portfolio over the next month, weighting each ETF with the obtained weights. Finally, we can cumulate each monthly portfolio returns, from December 2006 to February 2018.

Figure 6 represents the cumulative returns reached by performing investment strategies based on different values of γ, using the model in Equation 9, compared with the "naive" strategy portfolio, in which assets are equally weighted, as well as a benchmark portofolio: the MSCI Index.

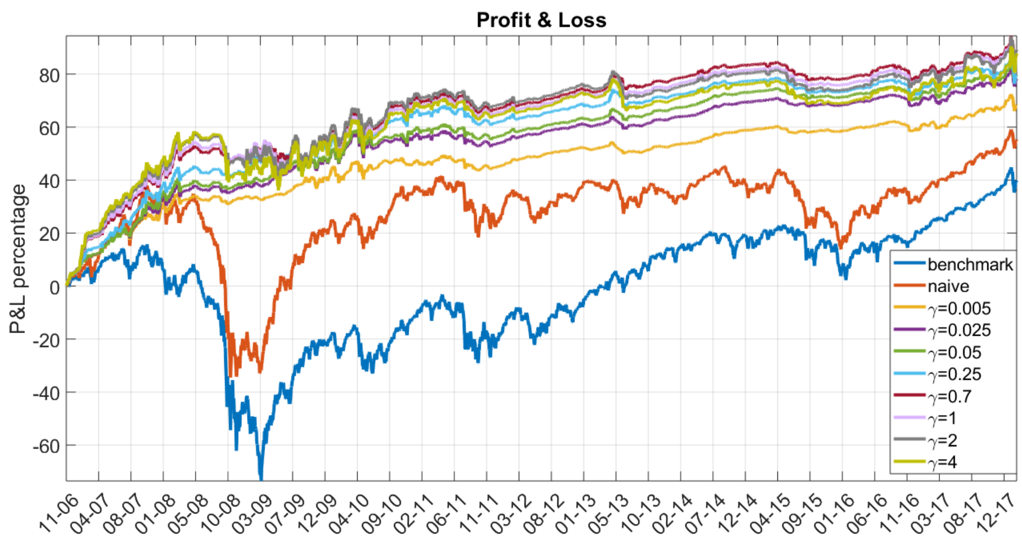

Figure 6: **Cumulative returns for different portfolio strategies**.The plot reports the cumulative Profits & Losses obtained by using our model based on different values of γ, the "naive" strategy portfolio (orange line) and the MSCI benchmark index (blue line).

Figure 6 highlights that our proposed model performs better than the benchmark index and the "naive" portfolio strategy. All of our strategies win in terms of total cumulative returns, regardless the coefficient of the individual risk propension. Increasing values of γ corresponds to higher cumulative returns, as suggested by modern portfolio theory.

Looking in more detail, during the crisis period $(2007 - 2009)$ our strategies are much better with respect to the benchamrk and the naive portfolio, since they all suffer less from financial draw-down. However, they are not able to reproduce the growing rebound at the end of 2009. More generally, during non crisis times our strategies, despite producing positive returns, can not reach the performance of the others.

To provide further insights of how our portfolios look like, we report in Figure 7 the dynamic of the ETF class weights considering $\gamma = 0.7$. For sake of brevity we report only this, since the results with other γ coefficients are qualitatively the same.

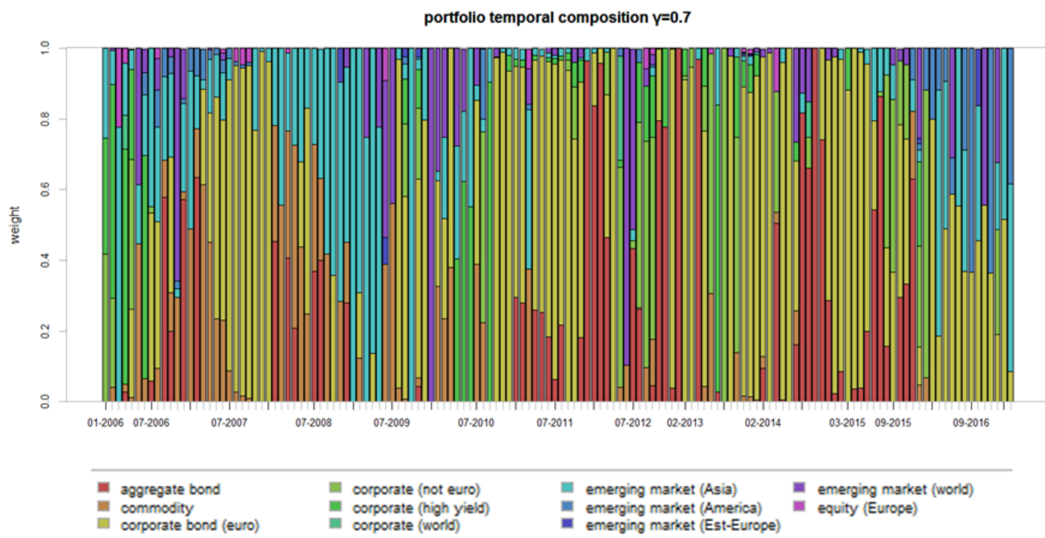

Figure 7: **Portfolio weights along time**. The figure reports the portfolio weights associated with a risk aversion coefficient equal to $\gamma = 0.7$

From Figure 7 it is clear that during crisis times the weight of the ETFs belonging to Emerging Equity classes is the highest relatively to the whole period considered. Differently, during non crisis times, emerging ETFs are substituted, in particular with corporate ones.

To have deeper insights about how portfolio performances change as market conditions mutate, the following tables report, as performance measures that take both risk and returns into consideration, the Sharpe Ratio (Sharpe, 1994) and the α of the Capital Asset Pricing Model (CAPM).

Table 3 specifically refers to the yearly Sharpe Ratio, defined as the ratio between the mean value of the excess returns and its standard deviation.

year	naive	$\gamma = \mathbf{0.005}$	$\gamma = \mathbf{0.025}$	$\gamma = \mathbf{0.05}$	$\gamma = \mathbf{0.25}$	$\gamma = \mathbf{0.7}$	$\gamma = \mathbf{1}$	$\gamma = \mathbf{2}$	$\gamma = \mathbf{4}$
2006	0.18	0.29	0.29	0.28	0.23	0.24	0.24	0.23	0.22
2007	0.10	0.14	0.15	0.15	0.18	0.20	0.20	0.21	0.21
2008	-0.00	0.65	0.61	0.59	0.57	0.52	0.47	0.34	0.21
2009	0.07	-0.18	-0.14	-0.12	-0.08	-0.07	-0.07	-0.04	-0.03
2010	0.03	-0.04	-0.01	0.01	0.02	0.03	0.03	0.02	0.02
2011	-0.02	0.09	0.07	0.07	0.05	0.05	0.05	0.04	0.04
2012	0.01	-0.22	-0.19	-0.19	-0.15	-0.13	-0.12	-0.11	-0.09
2013	-0.16	-0.57	-0.45	-0.39	-0.29	-0.26	-0.27	-0.29	-0.28
2014	-0.02	0.14	0.14	0.13	0.09	0.08	0.08	0.09	0.08
2015	-0.06	0.07	0.04	0.02	-0.03	-0.05	-0.05	-0.06	-0.06
2016	0.02	-0.09	-0.09	-0.10	-0.09	-0.06	-0.04	-0.01	0.01
2017	0.01	-0.15	-0.13	-0.13	-0.10	-0.08	-0.07	-0.07	-0.08
2018	0.03	-0.06	-0.05	-0.06	-0.10	-0.07	-0.04	0.04	0.09

Table 3: **Annual Sharpe Ratio.** The table shows the Sharpe Ratio of portfolios under different strategies. All the measures are computed relatively to the benchmark strategy.

Table 3 highlights how, during market crisis condition (as in 2008), the Sharpe Ratio of our portfolio strategy is higher with respect to the "naive", due to the higher returns reached in this specific phase. Following rebounds (as in 2009) are not captured by our strategy and the low Sharpe Ratio reflects this feature.

The value of α measures the ability to choose potentially profitable assets, reflecting the expertise of asset managers in exploiting market signals and investing accordingly, thus generating positive extra-performances. This, in turn, calls for the identification of systematic patterns in the way investors produce these extra-performances, and whether they are persistent in time.

Table 4 describes the α coefficient of the CAPM model, which reflects portfolio extra/under performances with respect to the benchmark.

Table 4 shows how our portfolios over performs the benchmark strategy reporting values greater then 0, and is generally better than the "naive" portfolio, as the differences in the α coefficients of the two strategies are mostly positive. Only during the growing rebound phase our strategies clearly under performs the "naive" portfolio.

4. Conclusions

In this paper we have shown how correlation networks can be used to improve robot advisory service, providing better risk/return performances.

In particular, we have demonstrated how Random Matrix Theory together with Minimal Spanning Tree and centrality measures can be used to

year	naive	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.25$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	0.12	0.18	0.18	0.18	0.16	0.17	0.17	0.17	0.16
2007	0.09	0.08	0.09	0.10	0.11	0.13	0.13	0.14	0.15
2008	-0.09	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.02
2009	0.12	0.03	0.04	0.03	0.02	0.01	0.00	0.02	0.02
2010	0.03	0.01	0.02	0.03	0.03	0.04	0.04	0.04	0.04
2011	-0.03	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
2012	0.02	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.03
2013	-0.06	-0.00	0.00	0.00	-0.00	-0.00	-0.01	-0.03	-0.03
2014	-0.00	0.02	0.03	0.03	0.02	0.03	0.03	0.03	0.03
2015	-0.04	-0.00	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.03
2016	0.02	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.02
2017	0.04	0.01	0.01	0.00	0.00	0.00	0.00	-0.00	-0.01
2018	0.03	-0.01	-0.01	-0.02	-0.07	-0.05	-0.03	0.05	0.14

Table 4: **Annual α.** The table shows α of the CAPM model of portfolios under different strategies. All the measures are computed in relation to the benchmark strategy and all the reported values are multiplied by a scale factor of 100.

construct investment portfolios that take risk aversion and return correlations into account.

We have applied our proposal to the observed returns of a set of Exchange Traded Funds (ETFs), typically highly correlated, which are representative of the activity of robot advisors. Our empirical findings show that, when the Random Matrix Theory approach is used to filter the correlation matrix, we obtain a network representation of ETFs which is clear, and leads to useful insights. In addition, the Random Matrix Theory approach applied to modern portfolio theory improves the reliability of portfolios in terms of the ratio between predicted and realized risk, as suggested by (Tola et al., 2008).

In addition, when the network centrality parameters are included in the Markowitz optimization function, a further diversification of portfolio risk can be reached, for a given value of returns. In fact, the model takes into account not only the individual and general risk of financial asset but it also incorporates aversion towards systemic risk by managing the coefficient γ.

For this reason, we believe that our proposal could be relevant, especially for regulators aimed a measuring and preventing under estimation of risks and compliance risks coming from the adoption of robot advisory financial consulting.

Possible extensions of this work could investigate the relationship between the risk propensity deriving from the robo-advisor online questionnaires and the parameter γ, and the inclusion of other measures of network centrality in the Markowitz objective function.

C. W. Beenakker. Random-matrix theory of quantum transport. *Reviews of modern physics*, 69(3):731, 1997.

F. S. Board. Financial stability implications from fintech: Supervisory and regulatory issues that merit authorities attention. *June, Basel*, 2017a.

F. S. Board. Fintech credit. *Financial Stability Board Report (27 June, 2017)*, 2017b.

P. Bonacich. Some unique properties of eigenvector centrality. *Social networks*, 29(4):555–564, 2007.

G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130, 2003.

S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

M. Carney. The promise of fintech–something new under the sun. In *Speech at Deutsche Bundesbank G20 Conference, by Bank of England Governor Mark Carney, January 25th*, 2017.

T. Conlon, H. J. Ruskin, and M. Crane. Random matrix theory and fund of funds portfolio optimisation. *Physica A: Statistical Mechanics and its applications*, 382(2):565–576, 2007.

C. Eom, G. Oh, W.-S. Jung, H. Jeong, and S. Kim. Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series. *Physica A: Statistical Mechanics and its Applications*, 388(6):900–906, 2009.

T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4-6):189–425, 1998.

L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

D. León, A. Aragón, J. Sandoval, G. Hernández, A. Arévalo, and J. Nino. Clustering algorithms for risk-adjusted portfolio construction. *Procedia Computer Science*, 108:1334–1343, 2017.

B. G. Malkiel and E. F. Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

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

R. N. Mantegna and H. E. Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967.

H. Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

M. Newman. *Networks*. OUP Oxford, 2018. ISBN 9780192527493. URL `https://books.google.it/books?id=YdZjDwAAQBAJ`.

J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, and A. Kanto. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5):056110, 2003.

G. Peralta and A. Zareei. A network approach to portfolio selection. *Journal of Empirical Finance*, 38:157–180, 2016.

N. Perra and S. Fortunato. Spectral centrality measures in complex networks. *Physical Review E*, 78(3):036107, 2008.

V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126, 2002.

M. Potters, J.-P. Bouchaud, and L. Laloux. Financial applications of random matrix theory: Old laces and new pieces. *arXiv preprint physics/0507111*, 2005.

F. Pozzi, T. Di Matteo, and T. Aste. Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, 3:1665, 2013.

T. Raffinot. Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, 44(2):89–99, 2017.

F. Ren, Y.-N. Lu, S.-P. Li, X.-F. Jiang, L.-X. Zhong, and T. Qiu. Dynamic portfolio strategy using clustering approach. *PloS one*, 12(1):e0169299, 2017.

W. F. Sharpe. The sharpe ratio. *Journal of portfolio management*, 21(1): 49–58, 1994.

A. Spelta and T. Araújo. The topology of cross-border exposures: beyond the minimal spanning tree approach. *Physica A: Statistical Mechanics and its Applications*, 391(22):5572–5583, 2012.

V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32 (1):235–258, 2008.

A. M. Tulino, S. Verdú, et al. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 1(1):1–182, 2004.

M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10421–10426, 2005.

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

T. Vỳrost, Š. Lyócsa, and E. Baumöhl. Network-based asset allocation strategies. *The North American Journal of Economics and Finance*, 2018.

H. C. J. Zhan, W. Rea, and A. Rea. An application of correlation clustering to portfolio diversification. *arXiv preprint arXiv:1511.07945*, 2015.