



# Network models to improve robot advisory portfolios

Paolo Giudici<sup>1</sup> · Gloria Polinesi<sup>2</sup> · Alessandro Spelta<sup>1</sup>

Accepted: 26 July 2021 / Published online: 18 October 2021  
© The Author(s) 2021

## Abstract

Robot advisory services are rapidly expanding, responding to a growing interest people have in directly managing their savings. Robot-advisors may reduce costs and improve the quality of asset allocation services, making user's involvement more transparent. Against this background, there exists the possibility that robot advisors underestimate market risks, especially during crisis times, when high order interconnections arise. This may lead to a mismatch between investors' expected and actual risk. The aim of this paper is to overcome this issue, taking into account not only investors' risk preference but also their attitude towards interconnectdness. To achieve this aim, we combine random matrix theory with correlation networks and extend the Markowitz' optimisation problem to a third dimension. To demonstrate the practical advantage of our proposed approach we employ daily returns of a large set of Exchange Traded Funds, which are representative of the financial products employed by robot-advisors.

**Keywords** Correlation networks · Portfolio optimization · Random matrix theory

## 1 Introduction and literature review

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). In the last few years, FinTech innovations have increased exponentially, delivering new payments and lending methods, and penetrating the insurance and asset management sectors. The Financial Stability Board (FSB) in its two recent reports (FSB 2017a, b) has identified three common drivers for FinTech innovation;

---

✉ Paolo Giudici  
paolo.giudici@unipv.it

Gloria Polinesi  
g.polinesi@pm.univpm.it

<sup>1</sup> University of Pavia, Pavia, Italy

<sup>2</sup> Università Politecnica delle Marche, Piazzale Martelli 8, 60121 Ancona, AN, Italy

namely, the shifting of consumer preferences on the demand side, the evolution of data driven technologies and the changes in the financial regulation on the supply side.

Against this background, robot advisory services for automated investments 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, the masses managed by automatic consultancy are estimated to be over 2552 billion in 2023.<sup>1</sup>

The rapid growth of robot-advisors has determined the emergence of new financial risks. Robot-advisors that build personalised portfolios on the basis of automated algorithms have been suspected of underestimating investors' risk preferences. This may be the effect of the asset allocation models employed by robot advisors which, for the sake of transparency and user's engagement, are often over simplified, and do not take asset returns correlations properly into account.

To improve robot-advisory asset allocation, we propose to embed correlation networks into Markowitz' asset allocation, following the work of Clemente et al. (2019). The authors modify the objective function of the Markowitz minimum variance portfolio taking into account not only the volatility of individual assets but also their interconnectedness, expressed in terms of a clustering coefficient calculated on the correlation network. Their empirical results show that the resulting portfolios are more diversified and show better performances with respect to the classical Markowitz' portfolios.

Similarly to Clemente et al. (2019), our proposal is based on the insertion of correlation network models into Markowitz' objective function. Our additional contribution is a parsimonious correlation network, based on random matrix theory filtering, and an interconnectedness measure that exploits the notion of network centrality. These methodological advancements are particularly useful in the context of robot advisory asset allocation, characterised by many assets, highly correlated with each other.

A related research work is Boginski et al. (2014) who, building on the work of Pattillo et al. (2013) and Boginski et al. (2006) exploit the concept of *clique* relaxations in weighted graphs to find profitable well-diversified portfolios. In their proposal, the weight of each asset corresponds to its return over the considered time period and each pair of assets is connected if the corresponding correlation exceeds a certain threshold value. This mechanism is able to ensure high returns of portfolios that are not guaranteed by the *cliques* themselves.

Another related paper is He and Zhou (2011), who modify Markowitz' problem by exploiting a different utility function, introducing a new measure of loss aversion for large payoffs, called the large-loss aversion degree (LLAD). The measure is applied to portfolio choice under the cumulative prospect theory in Tversky and Kahneman (1992).

A different stream of extensions of the Markowitz portfolio model considers multiobjective evolutionary algorithms (MOEAs), as in Metaxiotis and Liagkouras (2012). Within this framework, Cesarone et al. (2013) propose a heuristic solution based on a reformulation in terms of a Standard Quadratic Program to solve mean-variance portfolio issues related to the introduction of constraints based on cardinality (which limits the number of assets to be held in an efficient portfolios) and allocation shares (which determines the fraction of capital invested in each asset). Whereas Ehrgott et al. (2004) presents a method based on the application of four different heuristic solution techniques to test problems involving up to 1416 assets. Woodside-Oriakhi et al. (2011) consider the application of genetic algorithm, tabu search and simulated annealing meta heuristic approaches to find the cardinality constrained efficient frontier that arises in portfolio optimization. In addition, Doerner et al. (2004) introduces Pareto Ant Colony Optimization as an effective meta-heuristic solution, proposing a two-

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

stage procedure that first identifies the solution space of all efficient portfolios and then locates the best solution within that space.

Further works in the area of employing multidimensional operations research to improve Markowitz' model are Schaerf (2002), Crama and Schyns (2003), Shoaf and Foster (1996), Branke et al. (2009), Bai et al. (2009), El Karoui (2010) and El Karoui (2013).

We contribute to the above literature by proposing a different objective function of Markowitz' portfolio allocation, that extends the classic formulation with an asset centrality term, function of the assets' similarity network.

Similarity networks among asset returns were introduced by Mantegna and Stanley (1999), who expressed the distance between any pair of asset returns as a function of the pairwise correlations among the corresponding time series. They can reveal how assets are related in terms of the topology of a network (Newman 2018), and can allow to calculate centrality measures, which express their "importance" of each asset in the network (see e.g. Avdjiev et al. 2019). Among them, the eigenvector centrality (Bonacich 2007) assigns a relative score 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.

To reduce the computational burden involved with the calculation of the eigenvector centrality on a fully connected matrix, Mantegna (1999) suggested to hierarchically cluster the assets, or groups of assets that are the "closest", leading to a parsimonious "Minimal Spanning Tree" (MST) representation. Tumminello et al. (2005) has extended Mantegna (1999) with a generalisation of the MST, the Planar Maximally Filtered Graph (PMFG), which retains the same hierarchical properties of the MST but adds more complex graph structures, such as loops and cliques.

Another important extension of Mantegna (1999) is Tola et al. (2008) who have shown how a MST calculated on a correlation matrix "filtered" from random noise through the Random Matrix Theory (RMT) approach can improve the performance of the optimal portfolios. Siimilar papers, based on the applications of RMT to asset management, are León et al. (2017), Raffinot (2017), Ren et al. (2017), Zhan et al. (2015), Bun et al. (2017) and Fraham and Jaekel (2005).

Our contribution to the above literature is twofold. From an applied viewpoint, we extend the application of the RMT approach in Tola et al. (2008) to Exchange Traded Fund returns (ETFs), investment funds that aim to replicate the index to which they refer (benchmark) through a totally passive management. From a methodological viewpoint, we propose a portfolio optimisation approach different from what proposed by Tola et al. (2008), as we include network centrality explicitly into the Markowitz objective function. Doing so, we do not rely only on the pairwise covariance between assets returns, but we also consider higher order information on assets' behavior.

In the paper we show that the inclusion of centrality measures into Markowitz model can improve portfolio returns. This because not only bivariate but also multivariate dependencies are taken into account. A further advantage is that investors' risk preferences receive a higher importance and, therefore, the matching between the expected and the actual risk profile improves.

The empirical findings obtained from the application of our proposed method confirm the validity of the proposed approach, which can thus become a new toolbox for robot-advisors. The results demonstrate that centrality measures can generate portfolio allocation strategies able to outperform the benchmark portfolios.

The structure of the paper is as follows: Sect. 2 presents our proposal, in an order that follows the data analysis flow: first we describe the Random Matrix Theory, to filter data

from noise components; then the minimal spanning tree approach, to build a parsimonious correlation network among assets; and, finally, the proposed objective function that adds to Markowitz’ the centrality measures calculated from the obtained correlation network. Section 3 presents the results of the application of the methodology to a database kindly provided by an anonymous robot advisor. Section 4 ends with some concluding remarks.

## 2 Methodology

### 2.1 Random Matrix Theory

Since the mid-nineties, 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), economics and finance (Potters et al. 2005).

The intuition behind RMT is to extract the “systematic” part of a signal embedded in a correlation matrix, separating it from the “noise” component. To achieve this goal, RMT tests the eigenvalues of a correlation matrix:  $\lambda_k < \lambda_{k+1}; k = 1, \dots, N$ , against the null hypothesis that they are equal to the eigenvalues of a random Wishart matrix  $\mathbf{R} = \frac{1}{T}\mathbf{A}\mathbf{A}^T$  of the same size, with  $\mathbf{A}$  being a  $N \times T$  matrix containing  $N$  time series of length  $T$  whose elements are independent and identically distributed standard Gaussian random variables.

Note that RMT does not require any assumption on the distribution of the asset returns. Rather, it compares the eigenvalues of the empirical covariance matrix with those that would be obtained if the returns were drawn from independent Gaussian distributions, leading to a Wishart matrix.

Let  $(x_k = \hat{\lambda}_k) < (x_{k+1} = \hat{\lambda}_{k+1}); k = 1, \dots, N$  be the observed sample eigenvalues. It can be shown (Marchenko and Pastur 1967) that, as  $N \rightarrow \infty$  and  $T \rightarrow \infty$ , with a fixed ratio  $Q = \frac{T}{N} \geq 1$ , the density of each sample eigenvalue converges to:

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

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

When  $x_k > \lambda_+$  the null hypothesis is rejected, for all the eigenvalues greater or equal than  $k$ . This implies that the relevant part of the signal contained in the correlation matrix of the returns can be obtained applying a singular value decomposition based only on the eigenvectors that correspond to eigenvalues that are greater than  $\lambda_+$ . Doing so, RMT simplifies the correlation matrix into a filtered correlation matrix (Plerou et al. 2002; Eom et al. 2009).

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

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

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 the correlation matrix.

According to the RMT the filtered correlation matrix is then given by:

$$\mathbf{C}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad (4)$$

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 eigenvectors associated to the eigenvalues greater than  $\lambda_+$ .

We remark that the fact that the eigenvalues leaking out the “bulk” of the spectral density, the so called “spikes”, contain the important information can also be obtained following Couillet (2015), who assume that the correlation matrix of the stock returns follows a spiked covariance matrix, in which all eigenvalues are equal to one except “r” spikes. In the random matrix theory approach the number of spikes is chosen comparing the sample eigenvalues with the theoretical ones obtained from a Wishart matrix.

## 2.2 Minimal spanning tree

When a large set of asset returns is considered, as in the case of Robot-Advisory, the (filtered) correlation matrix may be difficult to summarise. Its representation in terms of a correlation network (see e.g. Mantegna 1999) can help the task.

A correlation network can be obtained converting pairwise correlations in pairwise distances with the following function:

$$d_{ij} = \sqrt{2 - 2c'_{ij}}, \quad (5)$$

where  $c'_{ij}$  are the elements of the filtered correlation matrix  $\mathbf{C}'$  and  $d_{ij}$  is the Euclidean distance between return  $i$  and  $j$ . The set of all obtained pairwise distances can be organised into a distance matrix  $\mathbf{D} = \{d_{ij}\}$ .

Then, a more parsimonious representation of the correlation matrix can be obtained by means of the Minimal Spanning Tree method (MST, see e.g. (see Mantegna and Stanley 1999; Bonanno et al. 2003; Spelta and Araújo 2012)). The MST is obtained applying to the distance matrix  $\mathbf{D}$  a single linkage clustering algorithm which associates each asset return to its closest neighbour, and avoiding loops. The term “minimal” refers to the fact that MST allows to reduce the number of links between asset returns from  $\frac{N(N-1)}{2}$  to  $N - 1$ , the minimum number of links assuring connectivity of all nodes.

More formally, the MST algorithm proceeds in this way. Initially, it considers  $N$  clusters, corresponding to the  $N$  available asset returns (ETFs in our context). Then, at each subsequent step, two clusters  $l_i$  and  $l_j$  are merged into a single cluster if:

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

with the distance between clusters being defined as:

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

with  $p \in l_i$  and  $q \in l_j$ . The above steps are repeated until a single cluster emerges.<sup>2</sup>

<sup>2</sup> We use the symbols  $\hat{d}$  and  $\hat{\mathbf{D}}$  to denote the distances representing the MST derived from the fully connected network  $\mathbf{D}$ .

We remark that 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 structures, but not in terms of performance.

We also remark that, to detect how financial relationships evolve over time we follow Spelta and Araújo (2012) by employing the *residuality* coefficient measure ( $R$ ) that compares the relative strengths of the connections above and below a threshold value, ias follows:

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

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

We expect that during crisis phases, higher correlation patterns emerge, leading to a lower value of  $R$ , as the number of links increases, and vice-versa.

### 2.3 Eigenvector centrality

Having found a parsimonious representation of the return correlations, we now aim to summarise it, to understand which nodes (asset returns) act as hubs in the network. This is key for understanding how ETF returns behave in a multidimensional space, and to construct optimal portfolios that take the curse of dimensionality into account.

The research in network theory has dedicated much effort to develop measures aimed to detect the most important players in a network. The idea of “centrality” was initially proposed in the context of social systems, where a relationship between the location of a subject in the social network and its influence on the group processes was assumed.

Various measures of network centrality have been proposed, such as the count of the neighbors a node has: the degree centrality, which is a local centrality measure, or measures based on the spectral properties of the adjacency matrix (see Perra and Fortunato 2008), which are global measures. Examples of global centrality measures include the eigenvector centrality (Bonacich 2007), Katz’s centrality (Katz 1953), PageRank (Brin and Page 1998), hub and authority centralities (Kleinberg 1999).

The eigenvector centrality measures the importance of a node 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:

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

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

The previous equation can be rewritten in matrix notation:

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

where  $\hat{\mathbf{D}}$  is the adjacency matrix,  $\lambda$  is the eigenvalue of the matrix  $\hat{\mathbf{D}}$ , with an associated eigenvector  $x$ , an  $N$ -vector of scores (one for each node).

Note that, in general, there will be many different eigenvalues  $\lambda$  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.

We remark that, in the context of correlation networks that we are considering, the higher the centrality score associated to a node, the more the node is dissimilar from all other nodes in the network.

## 2.4 Portfolio construction

We now explain how centrality measures can be embedded in a portfolio optimisation framework, to improve portfolio performances.

Correlations between asset returns play a central role in investment theory and risk management, as key elements for optimisation problems as in Markowitz (1952) portfolio theory. It is natural that correlation networks, based on pairwise correlations, play an important role too.

Indeed, Onnela et al. (2003) have shown that the assets with the highest weights in Markowitz portfolios (Markowitz 1952) are always located in the outer nodes of a Minimal Spanning Tree, so that the optimal portfolios are mainly composed by assets in the periphery of the network, and not in its core. Pozzi et al. (2013) have shown that portfolios which include central assets are characterized by greater risks and lower returns with respect to portfolios which include peripheral assets. Giudici and Polinesi (2021), following Giudici and Abu-Hashish (2019) found a similar behaviour in crypto exchanges. Vřrost et al. (2018) have suggested that network-based asset allocation strategies may improve risk/return trade-offs. Their work is based on the study of Peralta and Zareei (2016) which have found a negative relationships between asset return centralities and the optimal weights obtained under the Markowitz model.

Other authors have built on the above remarks by proposing novel portfolio optimisation strategies. For example, Plerou et al. (2002) and Conlon et al. (2007) have used the correlation matrix, filtered with the random matrix approach in the Markowitz model, and have shown that for the obtained portfolios the realized risk is closer to the expected one. Tola et al. (2008), combining MST with the RMT filtering, have shown performance improvement with respect to Markowitz portfolios. Finally, Tumminello et al. (2010) have demonstrated that the risk of the optimized portfolio obtained using a “filtered” correlation matrix is more stable than that associated with a “non filtered” matrix.

In line with the previous authors, we would like to exploit centrality measures, based on the minimal spanning tree derived from the RMT filtered correlation matrix of asset returns to improve Markowitz portfolios. Differently from the previous authors, we extend Markowitz’ approach using RMT and MST in the optimisation function itself, rather than applying Markowitz to the transformed (filtered and/or simplified) correlation matrix.

More formally, we propose to minimize the constrained objective function:

$$\min_{\mathbf{w}} \mathbf{w}^T \mathbf{COV}' \mathbf{w} + \gamma \sum_{i=1}^N x_i w_i \quad (9)$$

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  $\mu_P$  indicates the mean return of the portfolio, obtained by averaging the mean return of each asset  $\mu_i$ ; the parameter  $\gamma$  represents a risk aversion coefficient, to be specified by investors,  $x_i$  is the eigenvector centrality associated with the ETF return  $i$  and, finally, the  $(i, j)$  element of  $\mathbf{COV}'$  is equal to  $\sigma_i \sigma_j c'_{i,j}$ .

The decision criteria specified by Eq. 9 can be better understood comparing it with the classical Markowitz approach. In the Markowitz model, efficient portfolios are found by solving the optimization problem:

$$\min_{\mathbf{w}} \left( \mathbf{w}^T \mathbf{COV}' \mathbf{w} - \frac{1}{\lambda} \mu * wT \right) \tag{10}$$

where  $\lambda$  is a parameter representing the investor’s risk tolerance. If  $\lambda$  is large,  $\frac{1}{\lambda}$  will be close to zero, meaning that the investor does not have much risk tolerance. Conversely, if  $\lambda$  is small,  $\frac{1}{\lambda}$  will be large, placing more emphasis on returns.

We extend Markowitz model without modifying its underlying quadratic utility function. Specifically, we fix the second term and we minimise a modified version of the first term, which adds to the portfolio variance (first order risk) a new component that measures systemic risk (high order risk). To balance the two risks we introduce a parameter  $\gamma$  that defines their ratio, for a given level of return. When  $\gamma$  increases, investors prefer systemic risks over first order risks, and viceversa when  $\gamma$  decreases. The next Section contains a practical illustration of the implications of our proposal.

As a result, our optimal portfolio allocation formulation includes three dimensions: a first order risk, represented by the variance-covariance matrix, a high order (systemic) risk which depends on the network structure of ETF returns and, finally, the portfolio returns (fixed). We remark that the classical Markowitz efficient frontier is based only on two dimensions: the first order risk and the returns, and is therefore not comparable to our context. However, if we set  $\gamma$  equal to zero, our formulation is two dimensional and can, therefore, be placed on the efficient frontier and compared with Markowitz’ solutions, as will be shown in the next Section.

### 3 Application

The data set we consider to illustrate the application of our proposal is composed by 92 ETFs returns’ time series 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. From Table 1 note that the Emerging Market asset classes are the most frequent, followed by Corporate ETFs.

Table 2 displays 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 value assumed by 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 some known stylized facts: the distribution of most ETFs’ returns tends to be non-Gaussian and heavy tailed.

To compare the behavior of the returns during financial crises and normal times the data set has been divided in two chronologically successive batches, from 2006 to 2012 (crisis)



**Table 1** ETFs by Asset classes

	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

Number of exchange traded funds for each class

**Table 2** ETFs' classes summary statistics

	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

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

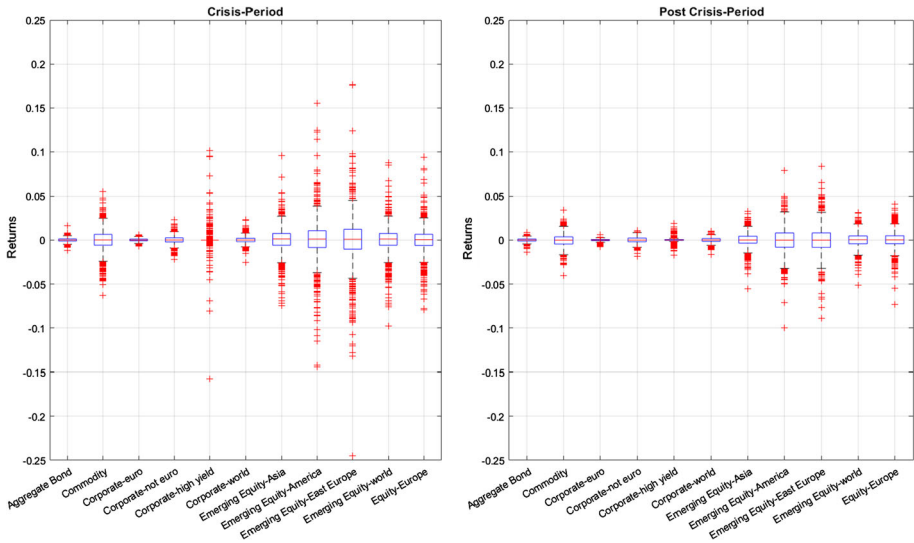
and from 2013 to 2018 (post-crisis). Figure 1 provides temporal boxplots for ETFs' returns, grouped by their asset classes (as described in Table 1).

Figure 1 shows that the volatility of the ETFs belonging to the Emerging Equity classes, as well as that of the Commodity asset class, is larger during the crisis period. This feature explains why their overall standard deviation, reported in 2, is the highest.

### 3.1 Transforming the correlation matrix: RMT and MST

In this subsection we show how the empirical correlation between ETF returns can be filtered and simplified, by means of the application of RMT and MST, described in the methodological section.

We first 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 length of one month ( $\cong 21$  trading days) for a total of 140 overlapping windows. For each time window, we



**Fig. 1** Summary plots for ETFs’ classes. Two different periods are compared: crisis (2006–2012) and post crisis (2013–2018)

use 11 months ( $\cong 229$  trading days) of daily observations to build our model and the remained month to validate it. This means, in particular, that we calculate 140 correlation matrices between all 92 ETFs’ returns, based on 11 months of data, to obtain the filtered correlation matrix applying the RMT approach, to derive the MST and the eigenvector centrality measure, and, finally, to derive the optimal portfolio, which is validated in an out-of-sample fashion using the twelfth month of each window.

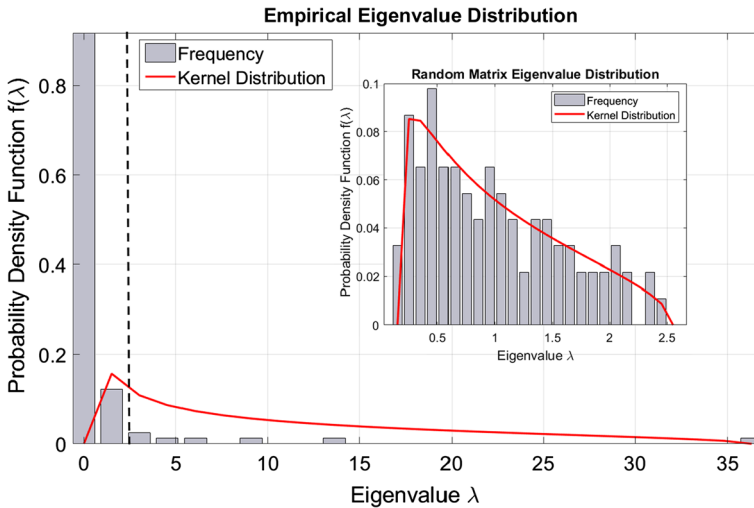
Figure 2 shows the ordered eigenvalue distribution of the empirical correlation matrix, for the last time window of the data set (March 2017–January 2018), compared with the theoretical Wishart correlation matrix that would be observed under random noise.

Figure 2 shows that most of the eigenvalues’ distribution lies between  $\lambda_{min}$  and  $\lambda_{max}$ , 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  $\lambda_{max}$  represent the effective characteristic dimension described by the correlation matrix.

Similar considerations can be made for other time windows, without loss of generality.

As described in the methodological Section, if, for each time window, we reconstruct the correlation matrix using only the eigenvectors that correspond to the largest eigenvalues, we obtain a sequence of “filtered” correlation matrices which can be used to improve the Minimum Spanning Tree representation of the ETFs’ returns. Figure 3 reports for both the filtered and the unfiltered correlation matrices and for each time window, the most central node, defined as the ETF return with the highest degree (the largest number of connected nodes) in the MST representation.

From Fig. 3 note that the application of the RMT filtering approach leads to different Minimal Spanning Tree configurations over time: the most central nodes are different and belong to different ETF classes. On the other hand, the Minimal Spanning Trees based on the unfiltered empirical correlation matrices do not seem to vary: the ETF labelled EIMI-IM, belonging to the Asia Emerging Market class, is for most of the time the most central node.



**Fig. 2** Eigenvalue distribution. In the figure, the red line shows the kernel density of the eigenvalues associated to the empirical correlation matrix  $C$ , in the main box, and of the random correlation matrix  $R$ , in the smaller subplot. The dashed vertical line indicates the threshold value  $\lambda_{max}$  which separates the informative eigenvalues from the “random noise” ones

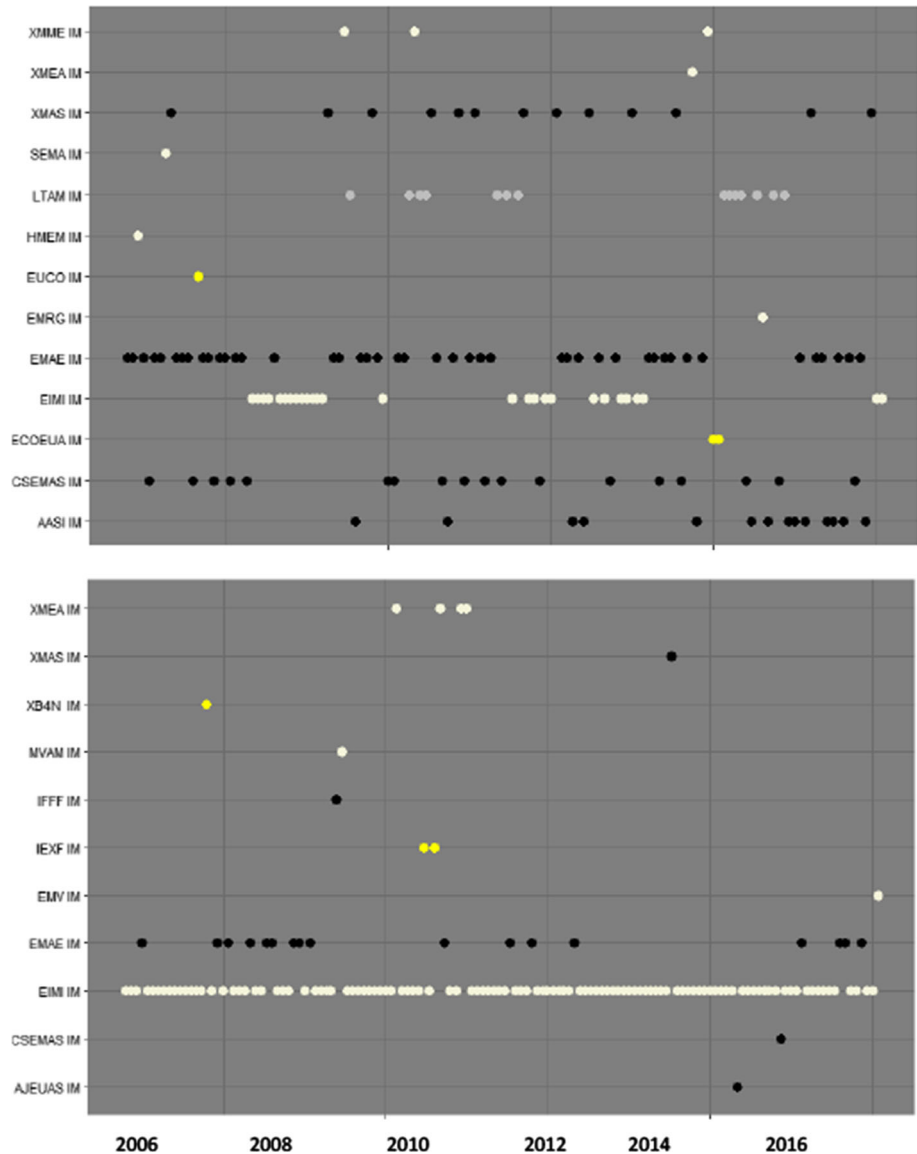
To further evaluate how the MSTs dynamically change over time, we employ, as a summary measure, the Max link: the maximum distance value between two pairs of nodes used in the construction of the tree, and the *residuality* coefficient, which measures the ratio between links eliminated and maintained by the MST construction. Figure 4 shows the evolution of these two quantities over the considered period.

From Fig. 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 returns, which synchronise during the crisis, consistently with the literature findings. While the Max link bounces back after the crisis, the residuality coefficient continues its decline until 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, Fig. 5 shows the MST topology during 2008, as representative of crisis times, and its topology during the last time window, taken as a reference period for a “business as usual” market phase. Figure 5 reflects how correlations increase during the crisis phase, leading to a high number of links in the network. Furthermore, in the crisis period, the MST reveals 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 class centrality remains high. This may explain the persistence of low values in the residuality coefficient, after the crisis phase.

### 3.2 Portfolio construction

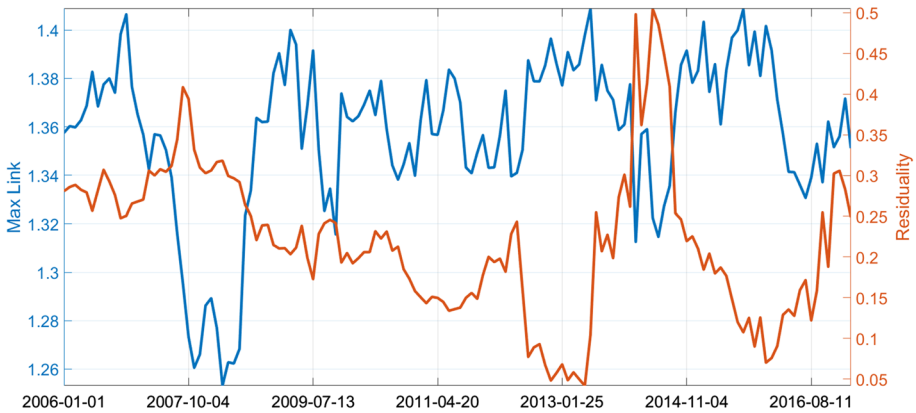
We now present the application of our proposed portfolio strategy, in which the eigenvector centrality computed on the MST derived from the application of RMT to the empirical



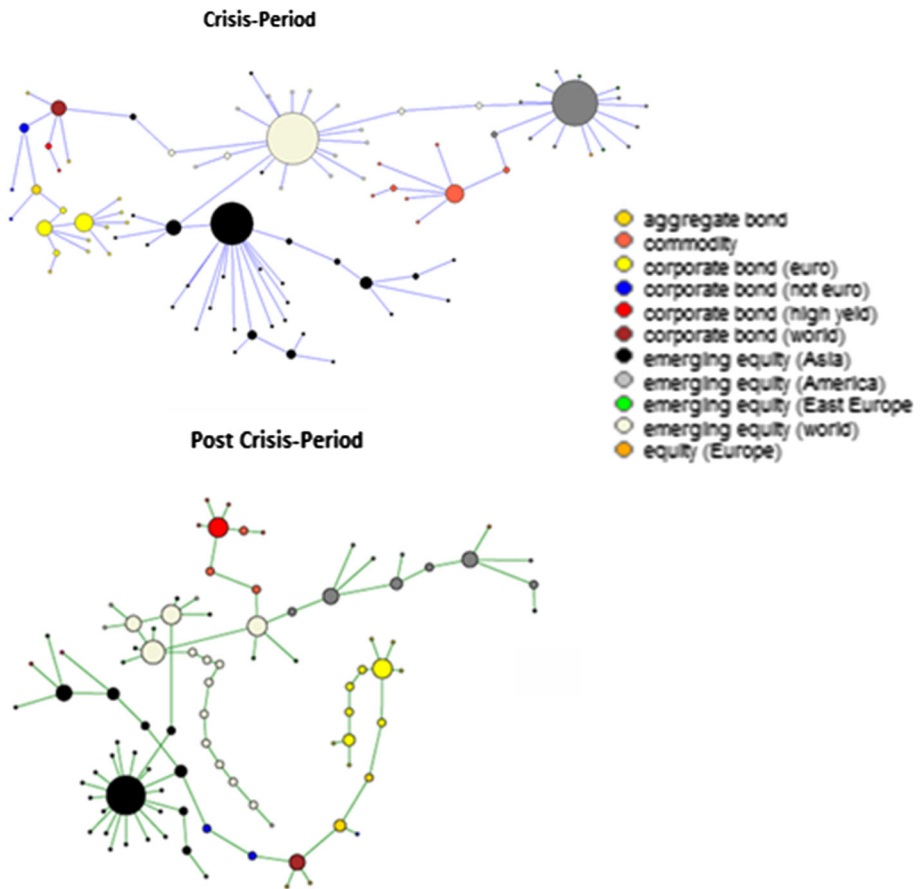
**Fig. 3** Central ETFs in the MST network representation along time. The figure reports, for each of the 140 time windows, the ETF that has the highest degree in the MST representation, using the filtered correlation matrix (top) and the unfiltered empirical correlation matrix (bottom). Node colors represent the asset classes: Corporate (yellow), Emerging Market Asia (black), America (grey), World (beige)

correlation matrix is inserted as an additional measure of risk in the objective function of Markowitz optimisation problem.

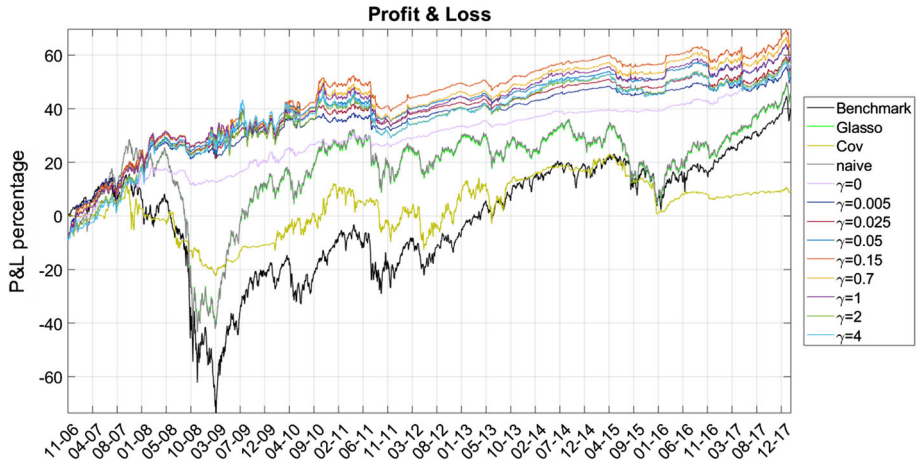
The optimal portfolio weights are obtained minimizing the constrained objective function in Eq. 9, for which the value of  $\gamma$  is set a priori, accordingly to the level of risk aversion of a



**Fig. 4** MST thresholds and residuality coefficient. 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 axes



**Fig. 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



**Fig. 6** Cumulative returns for different portfolio strategies. The plot reports the cumulative Profit and Loss obtained using our proposal based on different values of  $\gamma$ , the “naive” strategy portfolio (orange line), the MSCI benchmark index (blue line) and, for robustness, our proposal based on an unfiltered correlation matrix, rather than RMT (Cov) and our proposal based on a Glasso regularisation, rather than MST (Glasso)

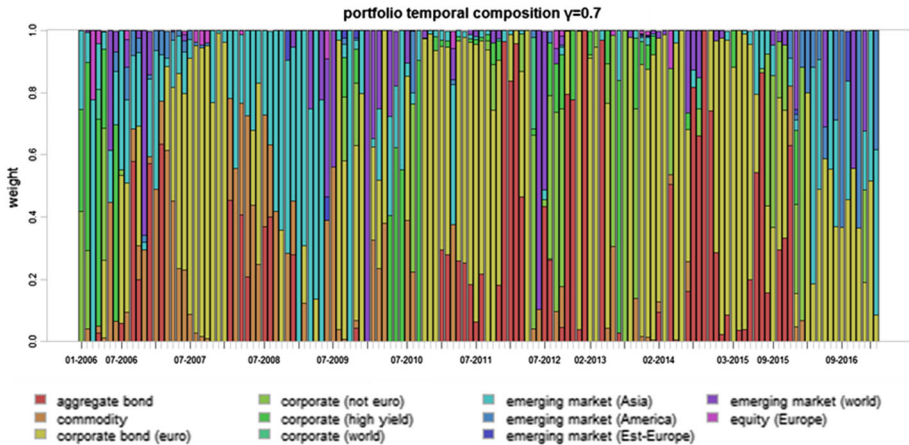
hypothetical investor. A high value of  $\gamma$  indicates that, in the desired allocation, more central ETFs (such as the Emerging Markets ones) will have higher weights.

Once the optimal portfolio weights are derived, we can calculate portfolio returns, and the associated Profit and Loss, for the time windows described in the previous subsections. More precisely, we use rolling windows in each of which the last month acts as an out-of-sample month to predict. The remaining eleven months of observations are used as a build-up period, to apply RMT, MST, compute asset return centralities and obtain the consequent optimal portfolio weights. We then calculate the return which corresponds to the optimal portfolio over the next month, weighting each ETF with the obtained weights. Finally, we cumulate each monthly portfolio return, from December 2006 to February 2018, taking a re-balancing cost of 10 basis points into account.

According to the described computational procedure, Fig. 6 presents the cumulative returns obtained with different values of  $\gamma$ , using the model in Eq. 9. The figure also reports the portfolio Profit and Loss of a “naive” (equally weighted) strategy as well as the performance of a benchmark, the MSCI Index. We also compute the performances obtained by employing a non filtered correlation matrix and those obtained with a Glasso regularisation method of Friedman et al. (2008), rather than with the MST approach<sup>3</sup> The results in Fig. 6 are summarised in Table 3, which presents the annual Profit and Loss of each competing strategy.

Figure 6 highlights that our proposed model performs better than the benchmark index, the “naive” portfolio strategy, the standard Markowitz portfolio (which correspond to  $\gamma = 0$  and the portfolios obtained either without RMT or with the sparse Glasso regularisation. All of our strategies win in terms of end of sample cumulative returns, regardless the coefficient of the individual risk propension, ranging from  $\gamma = 0.05$  to  $\gamma = 4$ . Note that the portfolio based on the non filtered covariance matrix produces the worst performance and, therefore, the RMT filter appears to be a fundamental condition for having adequate asset diversification in investment portfolios.

<sup>3</sup> The sparsity parameter  $\rho$  has been set to 0.01 as in the reference paper.



**Fig. 7** Portfolio weights along time. The figure reports the portfolio weights (for each asset class) associated with a risk aversion coefficient equal to  $\gamma = 0.7$

Looking in more detail, during the crisis period (2007–2009) our strategy produces higher returns with respect to the competitor portfolios, as it correctly takes high order risks into account. However, it is not able to capture the growing rebound at the end of 2009. More generally, during non-crisis times, our strategy, despite producing positive returns, can not reach the performance of the other portfolios. This because, during normal times, it probably excessively overweights high order risks.

These results are indeed consistent with our proposed modification to the Markowitz' algorithm. We expect that, during crisis times, when high order interconnections among financial assets strongly increase, our strategy produces higher returns with respect to competitor portfolios, and that it will do especially so for higher values of  $\gamma$ . This intuition is confirmed: in the year 2008 profits increase, particularly when  $\gamma$  is high, and so does the annual Sharpe Ratio, as shown in Table 4.

To provide further insights on portfolio compositions, we report in Fig. 7 the dynamic of the portfolio weights for ETF classes, considering  $\gamma = 0.7$ .<sup>4</sup>

From Fig. 7 it is clear that during crisis times the weight of the ETFs belonging to Emerging Equity classes is the highest. During non-crisis times, Emerging ETFs become less important, and their role is taken by other ETFs; in particular with Corporate ones.

To gain further 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), the  $\alpha$  of the Capital Asset Pricing Model (CAPM), the Value at Risk (VaR) and the Conditional VaR (CVaR).

Table 4 specifically refers to the yearly Sharpe Ratio, defined as the ratio between the mean value of the excess returns and its standard deviation. Table 4 highlights how, during market crises (as in 2008), the Sharpe Ratio of our portfolio strategy is higher with respect to the “naive” one with respect to the Sharpe Ratio obtained with the Glasso regularisation method. The subsequent rebound of 2009 is not captured by our proposal and the lower Sharpe Ratio obtained under different values of  $\gamma$  reflect this feature. Notice that the worst values of the Sharpe Ratio are associated with the portfolio derived using a non-filtered covariance matrix:

<sup>4</sup> Results for the other  $\gamma$  coefficients are qualitatively the same.

**Table 3** Annual cumulative profits and losses

Year	Bench.	Glasso	Naive	Cov	$\gamma = 0$	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	2.82	6.07	6.08	-0.18	6.46	10.53	8.74	8.67	8.54	7.57	7.62	5.67	7.13
2007	7.14	26.11	26.17	3.81	16.21	18.24	19.99	19.95	20.39	20.05	19.86	19.41	17.44
2008	-52.97	-53.84	-54.26	-22.87	-1.27	2.06	2.68	3.51	4.29	5.64	5.90	6.41	6.13
2009	48.71	51.67	52.97	9.50	8.04	7.24	6.27	5.56	5.87	4.92	4.51	5.64	9.42
2010	13.31	12.56	12.54	23.01	4.65	2.38	4.43	6.87	10.52	9.45	8.97	7.82	5.46
2011	-6.55	-10.21	-10.20	-14.13	-0.50	-2.42	-3.09	-3.95	-5.81	-6.39	-6.93	-7.69	-8.43
2012	15.04	11.57	11.54	11.29	4.88	5.41	5.38	5.50	6.02	6.59	6.77	7.14	7.05
2013	21.37	-1.53	-1.54	7.60	1.72	0.72	0.86	1.11	1.53	1.54	1.73	2.09	2.75
2014	2.80	0.99	0.98	2.59	2.22	3.99	3.99	3.94	3.91	4.14	4.22	4.66	4.49
2015	-1.75	-10.49	-10.50	-4.89	0.21	-1.19	-1.23	-0.87	-1.03	-1.22	-1.83	-3.11	-3.77
2016	5.09	8.61	8.60	-3.36	2.91	1.66	1.45	1.76	1.49	1.60	1.47	1.34	1.49
2017	14.25	14.91	14.88	2.55	7.72	3.05	4.53	5.04	5.05	5.43	5.62	5.90	5.75
2018	0.86	1.26	1.27	-0.87	0.86	-0.53	-0.96	-1.29	-1.55	-1.65	-1.72	-1.93	-2.25

The table shows the portfolios' cumulative profits and losses. All the values are expressed in %



**Table 4** Annual Sharpe ratio

Year	Glasso	Naive	Cov	$\gamma = 0$	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	0.37	0.37	-0.08	0.54	0.42	0.33	0.33	0.31	0.28	0.28	0.22	0.30
2007	0.14	0.14	0.02	0.16	0.17	0.18	0.18	0.17	0.16	0.16	0.15	0.12
2008	-0.10	-0.10	-0.13	-0.02	0.03	0.03	0.04	0.04	0.06	0.06	0.06	0.06
2009	0.15	0.14	0.11	0.15	0.12	0.07	0.05	0.04	0.03	0.03	0.03	0.05
2010	0.07	0.07	0.09	0.08	0.03	0.05	0.08	0.11	0.09	0.08	0.07	0.05
2011	-0.05	-0.05	-0.06	-0.00	-0.04	-0.05	-0.06	-0.08	-0.08	-0.08	-0.09	-0.09
2012	0.09	0.09	0.04	0.30	0.29	0.25	0.19	0.20	0.20	0.20	0.20	0.19
2013	-0.01	-0.01	0.04	0.09	0.03	0.03	0.04	0.05	0.04	0.05	0.06	0.06
2014	0.01	0.01	0.05	0.17	0.25	0.24	0.22	0.20	0.14	0.14	0.13	0.12
2015	-0.07	-0.07	-0.05	0.02	-0.05	-0.05	-0.03	-0.04	-0.03	-0.04	-0.06	-0.07
2016	0.05	0.05	-0.04	0.12	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03
2017	0.21	0.21	0.08	0.30	0.05	0.07	0.08	0.08	0.08	0.08	0.07	0.07
2018	0.08	0.08	-0.11	0.10	-0.05	-0.08	-0.10	-0.11	-0.12	-0.12	-0.13	-0.15

The table shows the Sharpe Ratio of portfolios obtained from different strategies. All measures are computed relatively to the benchmark strategy

**Table 5** Annual portfolio  $\alpha$

year	Glasso	Naive	Cov	$\gamma = 0$	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	0.13	0.13	0.00	0.19	0.29	0.23	0.22	0.22	0.19	0.19	0.13	0.16
2007	0.08	0.08	0.00	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.06
2008	-0.11	-0.11	-0.08	-0.00	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03
2009	0.14	0.14	0.03	0.03	0.03	0.02	0.01	0.01	0.01	0.00	0.01	0.02
2010	0.03	0.03	0.08	0.01	0.00	0.01	0.02	0.04	0.03	0.03	0.02	0.01
2011	-0.03	-0.03	-0.05	0.00	-0.01	-0.01	-0.01	-0.02	-0.02	-0.02	-0.03	-0.03
2012	0.02	0.02	-0.00	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03
2013	-0.04	-0.04	-0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2014	0.00	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
2015	-0.04	-0.04	-0.02	0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.01	-0.01	-0.01
2016	0.02	0.02	-0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
2017	0.03	0.03	0.01	0.02	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
2018	0.02	0.02	-0.03	0.02	-0.02	-0.03	-0.04	-0.05	-0.05	-0.05	-0.06	-0.06

The table shows the  $\alpha$  coefficient of the CAPM model for the portfolios generated from the alternative strategies. All the measures are computed relative to the benchmark strategy and all the reported values are multiplied by a scale factor of 100

**Table 6** VaR measures

Year	Bench.	Naive	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	1.12	1.10	0.65	0.70	0.72	0.82	0.85	0.86	0.87	0.90
2007	1.61	1.49	0.67	0.73	0.75	0.76	0.78	0.78	0.78	0.78
2008	3.45	2.60	0.35	0.49	0.59	0.69	0.79	0.84	0.92	0.91
2009	1.58	1.26	0.49	0.64	0.83	1.00	1.19	1.27	1.39	1.40
2010	1.45	1.14	0.51	0.57	0.60	0.64	0.72	0.75	0.78	0.84
2011	1.95	1.43	0.50	0.53	0.55	0.68	0.72	0.75	0.77	0.77
2012	1.65	1.30	0.15	0.16	0.17	0.18	0.20	0.22	0.23	0.23
2013	1.03	0.79	0.17	0.17	0.17	0.18	0.19	0.19	0.21	0.25
2014	0.96	0.92	0.20	0.21	0.22	0.26	0.29	0.29	0.30	0.31
2015	0.87	0.71	0.11	0.13	0.15	0.18	0.23	0.25	0.28	0.30
2016	1.32	1.10	0.18	0.20	0.20	0.26	0.31	0.33	0.38	0.42
2017	1.30	1.14	0.16	0.16	0.16	0.18	0.21	0.23	0.28	0.30
2018	0.61	0.75	0.51	0.60	0.62	0.64	0.67	0.70	0.75	0.76

The table shows the annual Value at Risk for the portfolios related to the different strategies. All the values are expressed in absolute terms multiplied by a scale factor of 100

**Table 7** CV<sub>αR</sub>

Year	Bench.	Naive	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	1.36	1.36	0.86	0.92	0.95	1.09	1.13	1.13	1.12	1.12
2007	1.95	1.79	0.82	0.91	0.94	0.96	0.97	0.97	0.97	0.98
2008	4.21	3.25	0.43	0.59	0.70	0.81	0.92	1.00	1.11	1.08
2009	1.99	1.52	0.62	0.80	1.02	1.22	1.46	1.57	1.73	1.73
2010	1.88	1.35	0.62	0.70	0.75	0.80	0.91	0.94	0.97	1.04
2011	2.42	1.72	0.58	0.62	0.65	0.84	0.89	0.93	0.98	0.93
2012	2.02	1.56	0.19	0.24	0.37	0.45	0.49	0.51	0.52	0.53
2013	1.34	1.03	0.20	0.20	0.20	0.21	0.22	0.23	0.26	0.40
2014	1.27	1.13	0.25	0.27	0.28	0.34	0.38	0.38	0.40	0.49
2015	1.09	0.83	0.17	0.18	0.19	0.23	0.28	0.31	0.36	0.38
2016	1.67	1.31	0.22	0.25	0.25	0.32	0.38	0.40	0.47	0.51
2017	1.62	1.34	0.20	0.20	0.22	0.24	0.31	0.34	0.40	0.44
2018	0.82	0.92	0.65	0.78	0.81	0.85	0.90	0.94	1.00	1.02

The table shows annual Conditional VaR derived from the competing strategies. All the values are expressed in absolute terms multiplied by a scale factor of 100

**Table 8** Annual information ratio

Year	Glasso	Naive	Cov	$\gamma = 0$	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	2.67	2.65	-3.82	4.43	6.31	4.47	4.33	4.03	3.07	3.09	0.97	2.69
2007	2.47	2.48	-0.28	0.80	1.21	1.41	1.40	1.42	1.36	1.34	1.25	1.05
2008	0.10	0.08	0.81	1.44	1.56	1.59	1.62	1.65	1.69	1.69	1.71	1.70
2009	1.73	1.83	-0.91	-0.84	-0.85	-0.96	-1.01	-1.00	-1.08	-1.10	-1.03	-0.81
2010	0.66	0.66	0.71	-0.05	-0.25	-0.06	0.17	0.54	0.40	0.34	0.22	0.01
2011	-0.11	-0.11	-0.21	0.45	0.31	0.26	0.20	0.06	0.05	0.02	-0.01	-0.06
2012	0.04	0.04	-0.29	-0.42	-0.33	-0.32	-0.29	-0.21	-0.17	-0.17	-0.15	-0.16
2013	-2.83	-2.83	-1.29	-1.96	-2.07	-2.06	-2.02	-1.95	-1.94	-1.92	-1.89	-1.79
2014	-0.50	-0.50	-0.18	-0.20	0.09	0.10	0.11	0.13	0.15	0.16	0.21	0.19
2015	-0.80	-0.81	-0.02	0.33	0.16	0.16	0.19	0.17	0.14	0.08	-0.05	-0.11
2016	0.28	0.28	-0.72	-0.09	-0.22	-0.24	-0.20	-0.22	-0.20	-0.21	-0.22	-0.20
2017	0.39	0.40	-2.77	-1.18	-2.03	-1.59	-1.46	-1.42	-1.35	-1.32	-1.30	-1.34
2018	1.37	1.38	-0.53	0.44	-0.59	-0.87	-1.05	-1.14	-1.17	-1.19	-1.25	-1.36

The table shows the Information Ratio of portfolios under different strategies. All measures are computed relatively to the benchmark strategy

this explains, once more, the importance of the preliminary processing done by the RMT approach.

The value of the CAPM  $\alpha$  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. Table 5 describes the  $\alpha$  coefficient, which reflects portfolio extra/under performances with respect to the benchmark. Table 5 shows that our portfolios outperform the benchmark strategy, as they all report values greater than 0. They are also generally better than the “naive” and “Glasso” portfolios. Only during the growing rebound phase of 2009 our strategies seem to under perform.

Our portfolio allocation method can also be compared with other methods in terms of risk. Table 6 specifically refers to the Value at Risk. Table 6 highlights that our portfolio strategies, although becoming more risky during the crisis period (proportionally to risk aversion), has an overall risk that is lower than that of the benchmark portfolio and that of the “naive” one.

Table 7 reports the values of the CVaR of the different portfolio strategies. This measure, introduced by Rockafellar et al. (2000), quantifies the potential extreme losses in the tail of the return distribution. The results are in line with those presented in Table 6: our strategies over perform in terms of expected losses the “naive” and benchmark portfolios, except for the last year considered in the analysis.

Finally, Table 8 reports the values of the information ratio which, differently from the Sharpe ratio, takes into account the tracking error. The results are in line with those obtained with the calculation of the Sharpe ratio.

## 4 Conclusions

In the paper we have shown how to improve robot-advisory portfolio allocation, typically based on several asset returns, highly correlated with each other, and especially during crisis times.

In particular, we have demonstrated how to build high performing portfolios by means of a mix of data processing strategies, that include: (i) filtering the correlation matrix among asset returns with the Random Matrix Theory approach; (ii) calculating the correlation network centrality of each asset return, after the application of the Minimal Spanning tree approach; (iii) selecting portfolio weights including in Markowitz’ optimisation function the network centralities, thereby taking into account high order interconnection risks.

The application of our proposal to the observed returns of a set of Exchange Traded Funds (ETFs), which are representative of the assets traded by robot-advisors, shows that our proposal leads to higher returns, or to lower risks, when compared to standard portfolios, especially during crisis times.

We therefore believe that our proposal could be relevant, for robot-advisors aimed at improving their services, while maintaining the accessibility to a large audience of potential investors; but also for regulators and supervisors, aimed at measuring and preventing the under estimation of market risks coming from the adoption of robot-advisory financial consulting.

The research work could be extended in several directions, and we believe it would also be extremely important to apply what proposed to other datasets and robot-advisory settings and, in particular, to those concerning crypto assets (see e.g. Giudici and Pagnottoni 2020) and foreign exchanges (see e.g. Giudici et al. 2021).

**Acknowledgements** This research has received funding from the European Union’s Horizon 2020 research and innovation program “FIN-TECH: A Financial supervision and Technology compliance training pro-

gramme” under the Grant Agreement No. 825215 (Topic: ICT-35-2018, Type of action: CSA). The authors thank the editor and the two anonymous referees for useful comments and suggestions. They also thank the company Moneyfarm, for having provided the data.

**Funding** Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Avdjiev, S., Giudici, P., & Spelta, A. (2019). Measuring contagion risk in international banking. *Journal of Financial Stability*, 42, 36–51.
- Bai, Z., Liu, H., & Wong, W. (2009). Multiobjective evolutionary algorithms for portfolio management: A comprehensive literature review. *Mathematical Finance*, 19, 639–667.
- Beenakker, C. W. (1997). Random-matrix theory of quantum transport. *Reviews of Modern Physics*, 69(3), 731.
- Boginski, V., Butenko, S., & Pardalos, P. M. (2006). Mining market data: A network approach. *Computers & Operations Research*, 33(11), 3171–3184.
- Boginski, V., Butenko, S., Shirokikh, O., Trukhanov, S., & Lafuente, J. G. (2014). A network-based data mining approach to portfolio selection via weighted clique relaxations. *Annals of Operations Research*, 216(1), 23–34.
- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555–564.
- Bonanno, G., Caldarelli, G., Lillo, F., & Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4), 046130.
- Branke, J., Scheckenbach, B., Stein, M., Deb, K., & Schmeck, H. (2009). Portfolio optimization with an envelope-based multi-objective evolutionary algorithm. *European Journal of Operational Research*, 199(3), 684–693.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.
- Bun, J., Bouchaud, J., & Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, 666, 1–109.
- Carney, M. (2017). 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*.
- Cesarone, F., Scozzari, A., & Tardella, F. (2013). A new method for mean-variance portfolio optimization with cardinality constraints. *Annals of Operations Research*, 205(1), 213–234.
- Clemente, G. P., Grassi, R., & Hitaj, A. (2019). Asset allocation: New evidence through network approaches. *Annals of Operations Research*, 299, 1–20.
- Conlon, T., Ruskin, H. J., & Crane, M. (2007). Random matrix theory and fund of funds portfolio optimisation. *Physica A: Statistical Mechanics and its applications*, 382(2), 565–576.
- Couillet, R. (2015). Robust spiked random matrices and a robust G-MUSIC estimator. *Journal of Multivariate Analysis*, 140, 139–161.
- Crama, Y., & Schyns, M. (2003). Simulated annealing for complex portfolio selection problems. *European Journal of Operational Research*, 150(3), 546–571.
- Doerner, K., Gutjahr, W. J., Hartl, R. F., Strauss, C., & Stummer, C. (2004). Pareto ant colony optimization: A metaheuristic approach to multiobjective portfolio selection. *Annals of Operations Research*, 131(1–4), 79–99.
- Ehrgott, M., Klamroth, K., & Schwehm, C. (2004). An MCDM approach to portfolio optimization. *European Journal of Operational Research*, 155(3), 752–770.
- El Karoui, N. (2010). High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Annals of Statistics*, 38, 3487–3566.

- El Karoui, N. (2013). On the realized risk of high-dimensional Markowitz portfolios. *SIAM Journal of Financial Mathematics*, 4, 737–783.
- Eom, C., Oh, G., Jung, W.-S., Jeong, H., & Kim, S. (2009). 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.
- Fraha, C., & Jaekel, U. (2005). Random matrix theory and robust covariance matrix estimation for financial data. Preprint, [arXiv:physics/0503007](https://arxiv.org/abs/physics/0503007).
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441.
- FSB. (2017a). Financial stability implications from fintech: Supervisory and regulatory issues that merit authorities' attention. June, Basel.
- FSB. (2017b). Fintech credit. Financial Stability Board Report (27 June, 2017).
- Giudici, P., & Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network VAR approach. *Finance Research Letters*, 28, 309–318.
- Giudici, P., & Pagnottoni, P. (2020). Vector error correction models to measure connectedness of bitcoin exchange markets. *Applied Stochastic Models in Business and Industry*, 36, 95–109.
- Giudici, P., & Polinesi, G. (2021). Crypto price discovery through correlation networks. *Annals of Operations Research*, 229(1–2), 443–457.
- Giudici, P., Leach, P., & Pagnottoni, P. (2021). Libra or librae? basket based stablecoins to mitigate foreign exchange volatility spillovers. *Finance Research Letters*, 2021, 102054.
- Guhr, T., Müller-Groeling, A., & Weidenmüller, H. A. (1998). Random-matrix theories in quantum physics: Common concepts. *Physics Reports*, 299(4–6), 189–425.
- He, X. D., & Zhou, X. Y. (2011). Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science*, 57(2), 315–331.
- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1), 39–43.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632.
- León, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., & Nino, J. (2017). Clustering algorithms for risk-adjusted portfolio construction. *Procedia Computer Science*, 108, 1334–1343.
- Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1), 193–197.
- Mantegna, R. N., & Stanley, H. E. (1999). *Introduction to econophysics: Correlations and complexity in finance*. Cambridge University Press.
- Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4), 507–536.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91.
- Metaxiotis, K., & Liagkouras, K. (2012). Multiobjective evolutionary algorithms for portfolio management: A comprehensive literature review. *Expert Systems with Applications*, 39(14), 11685–11698.
- Newman, M. (2018). *Networks*. Oxford University Press.
- Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5), 056110.
- Pattillo, J., Youssef, N., & Butenko, S. (2013). On clique relaxation models in network analysis. *European Journal of Operational Research*, 226(1), 9–18.
- Peralta, G., & Zareei, A. (2016). A network approach to portfolio selection. *Journal of Empirical Finance*, 38, 157–180.
- Perra, N., & Fortunato, S. (2008). Spectral centrality measures in complex networks. *Physical Review E*, 78(3), 036107.
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., & Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6), 066126.
- Potters, M., Bouchaud, J.-P., & Laloux, L. (2005). Financial applications of random matrix theory: Old laces and new pieces. [arXiv preprint arxiv:physics/0507111](https://arxiv.org/abs/physics/0507111).
- Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: Better to invest in the peripheries. *Scientific Reports*, 3, 1665.
- Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, 44(2), 89–99.
- Ren, F., Lu, Y.-N., Li, S.-P., Jiang, X.-F., Zhong, L.-X., & Qiu, T. (2017). Dynamic portfolio strategy using clustering approach. *PLoS ONE*, 12(1), e0169299.



- Rockafellar, R. T., Uryasev, S., et al. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2, 21–42.
- Schaerf, A. (2002). Local search techniques for constrained portfolio selection problems. *Computational Economics*, 20(3), 177–190.
- Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.
- Shoaf, J. S., & Foster, J. A. (1996). A genetic algorithm solution to the efficient set problem: A technique for portfolio selection based on the Markowitz model. In *Proceedings of the decision sciences institute annual meeting* (pp. 571–573).
- Spelta, A., & Araújo, T. (2012). The topology of cross-border exposures: Beyond the minimal spanning tree approach. *Physica A: Statistical Mechanics and its Applications*, 391(22), 5572–5583.
- Tola, V., Lillo, F., Gallegati, M., & Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32(1), 235–258.
- Tulino, A. M., Verdú, S., et al. (2004). Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1), 1–182.
- Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). 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.
- Tumminello, M., Lillo, F., & Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1), 40–58.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
- Výrost, T., Lyócsa, Š, & Baumöhl, E. (2018). Network-based asset allocation strategies. *The North American Journal of Economics and Finance*, 47, 516–536.
- Woodside-Oriakhi, M., Lucas, C., & Beasley, J. E. (2011). Heuristic algorithms for the cardinality constrained efficient frontier. *European Journal of Operational Research*, 213(3), 538–550.
- Zhan, H. C. J., Rea, W., & Rea, A. (2015). An application of correlation clustering to portfolio diversification. arXiv preprint [arXiv:1511.07945](https://arxiv.org/abs/1511.07945).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.