

**UNIVERSITÀ POLITECNICA DELLE MARCHE**  
**Facoltà di Economia G. Fuà**

---

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

Ph.D. degree in Economics - XXXII cycle



Doctoral Thesis

**Network metrics for FinTech services: an application  
on robot-advisors and crypto assets**

Author:

**Gloria Polinesi**

Supervisors:

**Prof. Francesco Maria Chelli**

**Prof. Paolo Giudici**

---

# Abstract

The last decade has witnessed the rapid development of a broad range of Financial Technology (FinTech) services. This work shows innovations in two functions of finance, savings and payments, focusing on robot-advisory platforms and crypto prices.

Robot-advisors involve the provision of online automated investment services with virtually no human contact. For this reason, they may reduce costs and improve the quality of the service, making user involvement more transparent. However, this digital platforms may underestimate market risks, especially when highly correlated assets are being considered, leading to a mismatch between investors' expected and actual risk.

In this perspective, the main goal is to enhance robot-advisory portfolio allocation, taking users' preference into account. In particular, this work demonstrates how random matrix theory and network models can be combined to construct investment portfolios that provide lower risks and higher returns with respect to standard Markowitz portfolios.

According to digital currencies that allow online payments to be sent directly from one party to another without going through a financial institution, this work analyses the dynamics of crypto asset prices and, specifically, how price information is transmitted among different bitcoin market exchanges, and between bitcoin markets and traditional ones.

The methodology adopted groups bitcoin prices from different exchanges, as well as classic assets, by enriching the correlation based minimal spanning tree algorithm with a preliminary filtering method based on the random matrix approach.

To this aim, main empirical findings are: i) bitcoin exchange prices are positively related with each other and, among them, the largest exchanges, such as Bitstamp, drive the prices; ii) bitcoin exchange prices are not affected by classic asset prices while their volatilities are, with a negative and lagged

effect.

Some of the most known techniques can be combined in order to reach up completely different goals, from one side the construction of asset allocation model and to the other the detection of mechanisms of price information between traditional and new financial products.

# Acknowledgements

Vorrei approfittare di queste brevi righe per ringraziare il Professor Francesco Maria Chelli e il Professor Paolo Giudici per la professionalità con la quale mi hanno aiutato durante lo svolgimento della tesi ma soprattutto li ringrazio per avermi insegnato ad affrontare le nuove sfide che si sono presentate durante il mio percorso.

Ringrazio la Professoressa Maria Crisitna Recchioni che, insieme al Professor Chelli, mi ha supportato e sopportato con grande professionalità (e pazienza!).

Per gli aiuti e i consigli tecnici ricevuti ringrazio i professori Riccardo Lucchetti, Giulio Palomba e Claudia Pigni, insieme al coordinatore Marco Cuculelli.

Un doveroso ringraziamento va a Andrea Bucci, Alessandro Spelta e Luca Pedini per avermi aiutato concretamente nella realizzazione della tesi.

Un ringraziamento caloroso ai miei compagni di corso Diego, Francesco, Chiara, Luca e Taslim per le risate e i pianti condivisi durante questa bella quanto faticosa avventura.

Grazie a Sabrina, collega e amica, che mi è stata ed è sempre vicina.

Grazie alla mia famiglia, a Matteo e a tutti i miei amici che hanno avuto un peso determinante nel conseguimento di questo obiettivo, punto di arrivo e di partenza. GRAZIE a tutti di cuore!



# Contents

<b>Abstract</b>	<b>1</b>
<b>Acknowledgements</b>	<b>3</b>
<b>1 Literature Review</b>	<b>8</b>
1.1 Introduction . . . . .	8
1.2 Techniques to “filter” correlation matrix . . . . .	11
1.2.1 Hierarchical clustering algorithm . . . . .	12
1.2.2 Extensions of the MST: different algorithms . . . . .	15
1.3 Correlation analysis and portfolio optimization . . . . .	18
1.3.1 Alternative asset allocation strategies . . . . .	19
1.3.2 Centrality and optimal weights . . . . .	20
1.4 Conclusion . . . . .	22
<b>2 Robot-advisors and portfolio model</b>	<b>25</b>
2.1 Introduction and literature review . . . . .	25
2.2 Methodology . . . . .	29
2.2.1 The random matrix approach . . . . .	29
2.2.2 The minimal spanning tree . . . . .	31
2.2.3 The eigenvector centrality . . . . .	32
2.2.4 Portfolio model . . . . .	33
2.3 Empirical findings . . . . .	35
2.3.1 Random matrix theory and network topology . . . . .	37
2.3.2 Portfolio construction . . . . .	41
2.4 Conclusion . . . . .	49
<b>3 Crypto price drivers</b>	<b>51</b>
3.1 Introduction and literature review . . . . .	51
3.2 Proposal . . . . .	53
3.3 Data . . . . .	55
3.4 Empirical findings . . . . .	57
3.5 Conclusion . . . . .	67

<i>CONTENTS</i>	6
<b>Appendix A</b>	<b>70</b>





# Literature Review

## 1.1 Introduction

The Financial Stability Board defines the Financial Technology (FinTech) as technologically-enabled financial innovation that could result in new business models, applications, processes or products with an associated material effect on financial markets, financial institutions and the provision of financial services.

During the last decade, FinTech innovations are increasing exponentially in many facets of finance, from making payments, to saving, borrowing, managing risks, and getting financial advice from automated consultants known as robot-advisors. Total global investment in FinTech companies reportedly increased from US\$ 9 billion in 2010 to over US\$25 billion in 2016. Venture capital investment has also risen steadily, from US\$0.8 billion in 2010 to US\$13.6 billion in 2016 (Fortnum *et al.*, 2017).

The Financial Stability Board (FSB) in its two recent reports (FSB (2017a), FSB (2017b)) identifies three common drivers of FinTech: shifting consumer preferences on the demand side; evolving technology and changing financial regulation on the supply side. The first one concerns higher customer expectations for convenience, speed, cost and “user-friendliness, the second concerns advances in technology, mainly related to big data and mobile technology and the third regards the increased frequency of changes in regulatory and supervisory requirements.

Unfortunately, rapid growth and innovation in Fintech context were not well supported by regulation. First, robot-advisors that build personal portfolio on the basis of algorithms that take into account information provided by investors filling out online questionnaire have been indicted of underestimating the investors risk profile. Investors information is relative to age, risk tol-

erance and aversion, net income, family status and this obtaining is a legal requirement specified by the MIFID directive.

Regulatory Authorities impose on robot-advisors a best matching between risk profile resulting from online questionnaire and assigned portfolio. In this perspective, our work that proposes a revisited technique of constructing portfolios including network parameters contributes to improve this relevant issue.

Also cryptocurrencies, which allow online payments to be sent directly from one party to another without going through a financial institution <sup>1</sup>, fall outside traditional regulation (Gregoriou & Nian, 2015).

The key role of information contained in correlation matrix of financial asset returns is the crucial point of this work both to construct portfolios tailor made for investors and for the detection of price drivers, specifically for robot-advisor and cryptocurrency fields.

For this reason, the purpose of this first chapter is to show the state of the art of methods concerning cluster algorithms applied on correlation matrices in order to analyze interconnectedness of assets in the financial markets. The literature relative to clustering algorithms on correlation matrices of stock returns time series is very amply and it stems from the seminal paper of Mantegna (1999). In fact this author investigates the correlation coefficient matrix to detect the hierarchical organization present inside the stock market: distance matrix based on correlation matrix is used to determine the minimal spanning tree (MST) connecting the N stocks considered in the time horizon T. With only a few exceptions Mantegna shows that groups are homogeneous with respect to industry and often also subindustry sectors, meaning that stocks belonging to the same sector or subsector are driven by the same economic factors.

Marti *et al.* (2017) in their review list methodological extensions for improving algorithm, distance and other aspects raised in literature with respect to the paper of Mantegna, however this work remains the more important in order to detect structural organization in financial markets.

Tumminello *et al.* (2005) extend Mantegna by filtering correlation matrix through a network which is a topological generalisation of the MST. This network is known as planar maximally filtered graph (PMFG) which retains the same hierarchical properties of the MST but adds more complex topo-

---

<sup>1</sup>“Establishing a definition of cryptocurrencies is no easy task. Much like blockchain, cryptocurrencies has become a buzzword to refer to a wide array of technological developments that utilize a technique better known as cryptography. In simple terms, cryptography is the technique of protecting information by transforming it (i.e. encrypting it) into an unreadable format that can only be deciphered (or decrypted) by someone who possesses a secret key.” (Houben & Snyers, 2018).

logical structures, such as loops and cliques.

The main purpose of these previous researches is to look for hierarchical structure of financial assets (especially stocks) in the complex financial market and on the basis of this methodology some authors construct investment portfolios as explained in Section 2.2.4.

For example, Tola *et al.* (2008) show how the use of clustering algorithms can improve the reliability of the portfolios in terms of the ratio between predicted and realized risk. Once they filtered correlation matrix using clustering techniques (specifically single and average linkage methods) and the alternative approach based on random matrix theory (RMT) they construct portfolios by using the resulting covariance matrix in the Markowitz's solution.

León *et al.* (2017) first place stocks into given groups applying hierarchical clustering technique on correlation matrix then select best assets in each cluster relying on a Sharpe Ratio optimization.

Raffinot (2017) constructs groups of stocks using hierarchical clustering algorithms then builds portfolios by assigning capital equally to each cluster, inside each cluster allocation is equally-weighted.

Quite different clustering methods are used by other authors, always considering the same distance measure based on Pearson correlation coefficient.

Ren *et al.* (2017) show a portfolio strategy that consider the time-varying structures of MST networks in the Chinese stock markets.

Instead, Zhan *et al.* (2015) investigate four different portfolio selection strategies and one of these concerns correlation clusters constructed by using Neighbor-Net graph.

Before them, other authors like Rea & Rea (2014) presented a method to visualize the correlation matrix using neighbor-Net networks (Bryant & Moulton, 2004), seizing relationships in the financial stock market.

This work presents a common part based on literature review related to hierarchical clustering algorithms and RMT exploited to detect topological structure of financial market since the beginning paper of the 1999 and two empirical applications in the FinTech world.

These applications show how known techniques could be implied in pioneering contexts for the development of a portfolio model specific for robot-advisors able to diversify investments at individual level and for the investigation of cryptocurrencies price drivers.

The structure of the thesis is as follows: next sections of this Chapter presents literature review relative to hierarchical clustering algorithms and extensions to "filter" correlation matrix of financial assets considered in a specific time horizon. Chapter 2 proposes a portfolio optimization approach specific for robot-advisory platforms that takes RMT filter and network centrality explic-

itly into account. Finally, Chapter 3 concludes by showing an application of RMT and graph tree on the correlation matrix of cryptocurrencies and “classical” assets such as gold and oil, with the aim to discover interconnectedness among crypto prices and between crypto prices and classic prices.

## 1.2 Techniques to “filter” correlation matrix

Time series is one of the many instruments to represent data and this type of data is present in a variety of fields, from brain activity to financial area. There are many reasons for which researchers apply clustering techniques to time series data.

Zhang *et al.* (2011) remark three main objectives implied in order to catch different similarities between time series: in time, in shape and in change. Similarity in time means that time series are grouped together when they move similarly time by time; there is similarity in shape if time series share common trends or sub patterns. Instead, similarity in change means that time series shown similarity of fitted parameters referred to fitted underlying models, which may be different.

For the nature of this type of data, cluster analysis of time series requires peculiar techniques.

Mantegna (1999) and other authors following him use a raw data approach for returns time series. A single observation (day, week or month) of the time series represents a characteristic of the element and stocks are grouped together when they are correlated: the Pearson correlation coefficient  $c_{ij}$  quantifies the degree of interdependence between pair of financial assets.

Clustering algorithms allow to extract leading information about aspects of structural organization from correlation matrix of return time series whereas correlation matrices can be represented as complete graphs where the notion of hierarchy lacks (de Prado, 2016).

Clustering tools, spectral methods (theory of random matrix) and correlation based graph represent the algorithms in order to extract information from complex systems of the correlation matrices. In fact, correlation matrices are subjected to the non stationary market conditions and “measurement noise” due to the finite length of time series that make the analysis difficult without appeal these filtering tools.

Furthermore, these procedure could improve portfolios diversification, as explained in Section 2.2.4.

### 1.2.1 Hierarchical clustering algorithm

This paragraph aims to show the use of hierarchical clustering algorithm in order to filter correlation matrices of return time series reducing the number of parameters. The analysis considers both a static financial market and a complex system that evolves over time.

Mantegna's first work dates back to the 1999 and this author investigates the correlation matrix to detect the hierarchical organization of stocks in financial market. In a ultrametric space, the graph minimal spanning tree (MST) connecting stocks reveals a topological arrangement of financial market that has an important meaning from an economic point of view. MST, defined as the minimum structure in terms of sum of distances between nodes, groups stocks homogeneous with respect to the economic sector of underlying companies.

Also Tumminello *et al.* (2010) confirm that elements (or nodes) share information according to the communities they belong to and communities are organized in a nested structure. Hierarchical clustering algorithms allow to detect this complex structure.

Spelta & Araújo (2012) qualify the minimal spanning tree as the corresponding representation of a fully-connected system (network) where sparseness replaces completeness in a suitable way.

Steps in order to draw MST can be summarized as follow. One starts from the correlation matrix of the time series of  $N$  stock returns, computed as difference of the logarithm of stock prices in the time horizon  $T$ <sup>2</sup>.

$$r_i(t) = \log P_i(t) - \log P_i(t-1) \quad (1.1)$$

The elements of correlation matrix for each pair of stocks

$$c_{ij} = \frac{E(r_i r_j) - E(r_i)E(r_j)}{\sigma_i \sigma_j} \quad (1.2)$$

converted in distance elements:

$$d_{ij} = \sqrt{2 - 2c_{ij}} \quad (1.3)$$

MST is based on the distance matrix computed in this way and this tree graph allows to shrink links connecting stocks from  $\frac{N(N-1)}{2}$  (total number

---

<sup>2</sup>Prices and returns of stocks, in general of the financial assets considered, can be daily, weekly, monthly or yearly.

of parameters in the distance matrix) to  $N - 1$ . Kruskal's (Kruskal, 1956) and Prim's (Prim, 1957) algorithms are the most known algorithms to build MST.

Kruskal's algorithm operates by building a MST one vertex at a time: at each step it adds the edge, by increasing distances, with least distance that does not form a loop; if the graph is not connected, a minimal spanning forest is founded.

Differently from the above algorithm, the Prim's one works by building a MST one vertex at a time, from an arbitrary starting vertex, at each step adding the edge with least distance, that connects a new vertex to the graph. For further details on algorithms to derive MST see Moret & Shapiro (1991). In general, minimal spanning trees allow to detect the hierarchical organization in sectors and subsectors of stocks, but it is known in literature that result changes if frequency of data changes.

Bonanno *et al.* (2001) demonstrate that decreasing the time horizon (i.e. from daily to intraday frequency) correlation between pairs of stocks decreases by affecting the hierarchical organization: minimal spanning tree moves from a clustered and structured set to a simpler set. This result is known as *EPPS* effect, for more details see Epps (1979), Münnix *et al.* (2010).

MST is the spanning tree associated to Single Linkage Clustering Algorithm (SLCA). However the MST retains some information that the single linkage dendrogram throws away (Raffinot, 2017).

This latter tests some variants of SL: complete linkage (CL), average linkage (AL) and Ward's Method (WM).

All these algorithms follow an agglomerative approach based on aggregating clusters in bottom-up fashion, they differ for the way of grouping elements into clusters. At each clustering algorithm corresponds an ultrametric matrix, used as a "filter" of the original one.

- SL, at each step it combines two clusters that contain the closest (minimum distance) pair of elements. Let  $C_1$  and  $C_2$  two clusters, the linkage function is described by:

$$Dist(C_1, C_2) = \min_{c_1 \in C_1, c_2 \in C_2} dist(c_1, c_2)$$

- CL, it works in a converse way with respect to SL. In fact, at each step combines two clusters that hold the farthest (maximum distance) pair of elements:

$$Dist(C_1, C_2) = \max_{c_1 \in C_1, c_2 \in C_2} dist(c_1, c_2)$$

- AL, it considers the distance between two clusters as the average distance between pair of elements belonging to those clusters:

$$\frac{1}{|C_1||C_2|} \sum_{c_1 \in C_1} \sum_{c_2 \in C_2} dist(c_1, c_2)$$

- WM, at each stage, two clusters merge if they provide the smallest increase of the squared error:

$$Dist(C_1, C_2) = dist(\{C_1\}, \{C_2\}) = \|C_1 - C_2\|$$

Different dendrograms or hierarchical trees are associated to these algorithms. Other authors concentrate on the MST as a characteristic tree graph for the description of the correlation matrix.

For example, the work of Onnela *et al.* (2003b) emphasizes the aspects already presented by previous authors but in the same time it complains about the fact that minimal spanning tree (or simply “asset tree”) only represents a static average of an evolving complex system. For this reason these authors explore the asset tree dynamics computing the correlation matrix for each rolling window of width  $T$  and draw MST for each period considered in order to see how the structure of the minimal spanning tree changes over time.

They demonstrate that the basic structure of MST is very robust with respect to time but during market crisis it shrinks due to the strong global correlation which makes the assets behaviour very homogeneous.

Mean occupation layer is used as a summary statistics in order to monitor the time evolution of the asset tree. It represents the layer of the tree where, on average, the mass of stocks is concentrated. During the 1987 stock market crash, the structure of MST becomes flat and the value of the mean occupation layer decreases.

Also Spelta & Araújo (2012) propose a measure called *residuality* coefficient that compares the relative strengths of the connections above and below a threshold distance value of the tree in order to asses structural changes of MST over time.

Matesanz & Ortega (2015) draw MSTs in each temporal window considered with the aim of evaluating temporal changes of time series of countries debt-to-GDP ratio. They calculate agglomerative coefficient (Kaufman & Rousseeuw, 2009) of each temporal tree and cophenetic correlation (Sokal & Rohlf, 1962) between hierarchical trees of different times. Agglomerative coefficient close to 1 implies a highly nested structure of tree instead cophenetic correlation gives an idea of how similar is the grouping structure between two different hierarchical tree.

During market crises started at the year 2008 the value of the agglomerative coefficient is much less than 1 and hierarchical trees for overlapping windows are not correlated.

Although works refer to different time periods and type of data considered, results confirm what said above: structure of hierarchical trees tends to be flat and different from others during market crisis.

A critical aspect in considering dynamic MSTs is represented by the fact that the choice of the temporal windows (number and length) is arbitrary as asserted by Marti *et al.* (2017).

There is not a predominant methodology but the “naive” one consists in:

- Computing Pearson correlations on a rolling window, which width can be linked, for example, to the duration of the investment in the case of cluster analysis for portfolio construction.
- Drawing a network or a tree on the rolling empirical correlation matrix.

A trade off exists between too noisy and too smoothed data for small and large window widths, respectively (see Onnela *et al.* (2003a) for detail).

Some works try to solve problems linked to Pearson correlation coefficient considering nonlinear correlations for example Spearman correlation coefficient (Miccichè *et al.* (2003) and Hartman & Hlinka (2018)), conditional Spearman correlation coefficient (Durante *et al.*, 2014) that is conditional to the behavior in the tail of assets after fitting of a copula-based time series model. Baitinger & Papenbrock (2017) introduce mutual information measure able to detect linear and nonlinear dependences between different asset classes.

### 1.2.2 Extensions of the MST: different algorithms

According to the work of Marti *et al.* (2017), this part aims to list some of different algorithms used to replace the minimal spanning tree and its corresponding clusters, with the goal to improve the seminal work of Mantegna (1999).

These algorithms have both hierarchical, with correlated graphs, and not hierarchical nature. This latter consider a spectral method based on the study of eigenvalues of correlation matrices.

From the side of hierarchical algorithms, Tumminello *et al.* (2005) introduce a graph for filtering correlation matrix that preserves the hierarchical organization of the minimal spanning tree but comprising more information.



This graph is known as planar maximally filtered graph (PMFG) and it represents an extension of the MST. The basic difference between the two is the number of links considered: MST contains  $N - 1$  links instead PMFG  $3(N - 2)$ , where  $N$  is the number of nodes in the graph.<sup>3</sup>

Therefore PMFG holds the hierarchical skeleton of the minimal spanning tree but it is enriched with loops and cliques. As explained in Tumminello *et al.* (2010) a clique of  $k$  elements is a complete subgraph that links all  $k$  elements. Due to the Kuratowski's theorem, PMFG can only have cliques of 3 and 4 elements. The number of 3-cliques and 4-cliques that can be built is  $3N - 8$  and  $N - 3$  respectively.

Tumminello *et al.* (2007a) investigate the planar graph considering two different sampling time horizons of return time series: PMFG at 5 min time horizon (intradaily time scale) and PMGF at daily time horizon.

In general, they note that PMFG selects links which higher values of  $\rho_{i,j}$  between pairs of stocks but the arrangement of the PMFG is also affected by the EPPS effect: the inter-sector correlation increases while intra-sector correlation decreases by decreasing the frequency of data. This means that PMFG at daily time horizon detects connections inside the same sector instead PMFG at intradaily time horizon catch up correlations between different sector.

An other network-filtering method for correlation matrices is the triangulated maximally filtered graph (TMFG) proposed by Massara *et al.* (2016). Starting from a 4-clique, TMFG produces planar graphs by optimizing an objective function, also known as "score function", that represents the sum of the weights of the edges.

Tumminello *et al.* (2007b) introduce the average linkage minimum spanning tree (ALMST) that is the spanning tree associated to the average linkage clustering algorithms (ALCA).

They show that ALMST detects economic sectors and subsectors in the network slightly better than the MST. Instead, the average value of the reliability of links relative to minimal spanning tree is slightly greater with respect to the average linkage minimum spanning tree.

As a global measure of the reliability of links in a graph, authors consider the average of bootstrap values. The bootstrap value represents the times number in which the link appears in the  $MST_i$  and  $ALMST_i$  drawn for each replica  $i$  randomly constructed selecting rows from the original correlation matrix.

---

<sup>3</sup>In the case of planar filtered graphs the genus is equal to 0. According to the definition in Tumminello *et al.* (2005), the genus is a topologically invariant property of a surface defined as the largest number of nonisotopic simple closed curves that can be drawn on the surface without separating it, i.e., the number of handles in the surface.

Recently, Musmeci *et al.* (2015) have introduced the directed bubble hierarchical tree (DBHT), a novel clustering algorithm based on the topological structure of the PMFG. Differently from other hierarchical techniques, DBHT first identifies clusters and then set the hierarchy intra and inter groups. From the not hierarchical side, random matrix theory (RMT) is the main approach in order to investigate the structure of return correlation matrices of financial assets.

The basic idea of RMT is testing eigenvalues such that  $\lambda_k < \lambda_{k+1}$  of an empirical correlation matrix which elements derive from equation (1) against the null hypothesis given by eigenvalues of a same size random Wishart matrix  $\mathbf{R} = \frac{1}{T} \mathbf{A} \mathbf{A}^T$ , where  $\mathbf{A}$  is an  $N \times T$  matrix containing  $N$  time series of length  $T$  which elements are independent, identically distributed random variables with zero mean and variance  $\sigma^2 = 1$ .

The random correlation matrix of this set of variables in the limit  $T \rightarrow \infty$  is the identity matrix, when  $T$  is finite the correlation matrix is in general different from identity matrix.

The theory of random matrices allows to prove that in the limit  $N \rightarrow \infty$  and  $T \rightarrow \infty$  with a fixed ratio  $Q = \frac{T}{N} \geq 1$  and  $\sigma^2 = 1$ , the eigenvalue spectral density is given by:

$$f(\lambda) = \frac{T}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (1.4)$$

where  $\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}$ .

The theory of random matrices has a long history (Mehta (1990)), first results in the financial context are known from works of Galluccio *et al.* (1998), Laloux *et al.* (1999) and Plerou *et al.* (1999).

Coronnello *et al.* (2005) also compare eigenvalues of empirical correlation matrix and random correlation matrix by imposing  $\sigma^2 = 1 - \frac{\lambda_1}{N}$  that allows to take into account the behaviour of the first eigenvalue, the “market mode”. Authors have already tested the finite size effect on  $\lambda_+$ , Plerou *et al.* (2002) demonstrate that the eigenvalues deviating from random matrix convey meaningful information stored in the correlation matrix.

Information can be extract from eigenvalues that are higher then  $\lambda_+$  (deviating eigenvalues) and it involves correlations between stocks that belong to the same industry or geographical area, instead the “bulk” of eigenvalues agree with RMT reveals the random correlations.

Coronnello *et al.* (2005) introduce a measure that quantifies the role of a

given sector<sup>4</sup>  $s$  in explaining the composition of each eigenvector  $u^k$ :

$$X_s^k = \sum_{i=1}^n P_{si} [u_i^k]^2 \quad (1.5)$$

Deviating eigenvectors associated to deviating eigenvalues are characterized by prominent sectors, consequently they present higher values of  $X_s^k$ .

The robustness of these results has been checked also testing the statistical properties of the eigenvalues of original correlation matrix for three known universal properties displayed by gaussian orthogonal ensemble (GOE) matrices: the distribution of nearest-neighbor eigenvalue spacings, the distribution of next-nearest neighbor eigenvalue spacings and the long-range eigenvalue correlations.

Onnela *et al.* (2004) remark that random matrix theory offers an interesting comparative perspective with respect to hierarchical clustering techniques.

Tola *et al.* (2008) compare filtering procedures of correlation matrix based on RMT and hierarchical clustering methods.

Some authors use RMT to filter correlation matrix and construct MST on this filtered matrix because, in order to extract the structure hidden in large correlation matrices, trees are easier to interpret with respect to the inspection of large matrices. By using this procedure, Miceli & Susinno (2004) obtain a clusterization per strategies of the Hedge Funds returns, according to the definition of strategies provided by Lhabitant & Learned (2002). Also Conlon *et al.* (2007) confirmed this result.

### 1.3 Correlation analysis and portfolio optimization

The hierarchical organization behind the financial market represented as a complex network contributes to solve portfolio optimization problem.

Onnela *et al.* (2003b) show that the assets of classic Markowitz portfolio are always located on the outer leaves of the minimal spanning tree, although the topological structure of the tree changes with time. Once again correlation matrix plays a fundamental role in the financial sector, in fact empirical researches have explored potential diversification benefits across asset classes by studying their centrality in the network.

Chi *et al.* (2010) confirm the importance of correlation analysis in portfolio

---

<sup>4</sup>If data considered are stock returns.

optimization for both international and domestic market investments.

Výrost *et al.* (2018) suggest network-based asset allocation strategies aimed to improve risk-return profiles relative to portfolios. This work is based on the study of Peralta & Zareei (2016) who exploit the negative correlation between the centrality of assets within a complex financial network and optimal weights under the Markowitz model.

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 peripheral portfolios.

It is quite clear the importance of centrality and peripherality measures in order to distinguish central and non-central vertices in the network, explained in detail in the Section 1.3.2.

In the Section 1.3.1, we introduce all the asset allocation based on network strategies employed as alternative or improvement of Markowitz framework.

### 1.3.1 Alternative asset allocation strategies

A large amount of works has contributed to the study of portfolios by using alternative or improving methods with respect to the original one of Markowitz (1952).

These methods include neural networks, genetic algorithms, random matrix theory filtering and hierarchical clustering. Among them, the latter is one of the most efficient methods for the selection of stocks for optimal portfolios (Ren *et al.*, 2017).

The use of “improved covariance” matrix estimators as an alternative to the sample estimator is considered an important approach for enhancing portfolio optimization (Pantaleo *et al.*, 2011).

In particular, we consider works relative to portfolio strategies based on RMT filtering procedure and ultrametric matrices associated to the hierarchical clustering algorithms applied to the empirical correlation matrices.

Lopez de Prado (2016) explains the Hierarchical Risk Parity (HRP) approach for the portfolios construction starting from the hierarchical tree inside the correlation matrix of investments. The HRP method not only exploits the information contained in the covariance matrix without requiring its inversion or positive definitiveness, fundamental condition for the Markowitz’ solution, also it reaches a better diversification than minimum-variance portfolios for a very similar level of risk.

Plerou *et al.* (2002) use the filtering correlation matrix resulting from random matrix approach in the minimum variance model proposed by Markowitz.

They show that for these portfolios the realized risk is more closer to the predicted one.

Years after, Conlon *et al.* (2007) demonstrate that RMT is found to greatly reduce the difference between the predicted and realized risk of a portfolio, leading to an improved risk profile for a fund of hedge funds.

Tola *et al.* (2008) construct portfolio by solving Markowitz solution but cleaning correlation matrix of stock returns through random matrix approach, single linkage and average linkage. They show that clustering algorithms and RMT filtering procedure improve the reliability of the Markowitz portfolio in terms of the ratio between predicted and realized risk.

Also Tumminello *et al.* (2010) prove that using the empirical correlation matrix leads to a dramatic underestimation of the real risk. In fact, they demonstrate that the risk of the optimized portfolio obtained using a “filtered” correlation matrix is more stable, although the real risk is always larger than the predicted one.

Furthermore, Gera *et al.* (2016) show that using filtered covariance matrices (RMT and MST) the original Markowitz solution is outperformed in terms of standard portfolio performance measures.

Previous works consider Pearson correlation coefficient, other researches focus their attention on different correlations.

For example, Baitinger & Papenbrock (2017) apply an active portfolio management framework (see Grinold & Kahn (1999)) that consider interconnectedness between assets based on mutual information measure, Durante *et al.* (2014) and Durante *et al.* (2015) propose to group time series of returns according to assets behaviour in tail (loss events) and then construct portfolios in order to manage risk during crisis scenario.

Härdle *et al.* (2018) pick satellite (or peripheral) assets according to their Adaptive Lasso Quantile Regression (ALQR) coefficients, that provide the information concerning the dependence between core portfolio and satellites at different tail events.

Satellite-assets with negative ALQR coefficient are included in a new portfolio with core assets and the rebalancing of weights occurs appealing different techniques (basic, naive and hybrid).

### 1.3.2 Centrality and optimal weights

This section aim to describe the main centrality/peripherality measures and the negative relationship between the centrality of a node, in various ways considered, and its optimal weight in a portfolio based on clustering strategy. For a graph  $G(V, E)$ , where  $V$  is the set of vertices (nodes) and  $E$ , the set of edges between nodes, the most common measures of centrality and periph-

erality are:

- **Degree** of a node is the number of edges connected to that node. The formula is:

$$\sum_{v \in V} \deg(v) = 2|E| \quad (1.6)$$

- **Betweenness centrality** is the ratio between the number of times which a vertex lies in the shortest path of pairs of nodes and the number of shortest paths of these two nodes. The betweenness of a node  $i$  with respect to nodes  $j, l$ , with  $i, j, l \in V$ , can be computed as follow:

$$\sum_{j \neq i \neq l} \frac{n_{j,i,l}}{n_{j,l}} \quad (1.7)$$

- **Closeness centrality** of a node  $i$  is the reciprocal of the sum of the length of the shortest paths between the node and all other nodes  $j$  in the graph. It represents the reciprocal of the farness and for a node  $i$  it is computed as:

$$C(i) = \sum_j \frac{1}{d_{i,j}} \quad (1.8)$$

- **Eigenvector Centrality** is a qualitative index of centrality and it measures the influence of a node in the network. This measure assumes that the centrality of a node  $i$  is proportional to the centrality of its neighbor nodes: being connected few times to highly central nodes may be more relevant than having a lot of links with less central nodes. Let  $\mathbf{D}$  the adjacency matrix of a graph  $G(V, E)$ , the centrality of all vertices is obtained by finding the greatest eigenvalue  $\lambda \in \mathbb{R}$  such that for a vector  $\mathbf{x} \in \mathbb{R}^M$ :

$$\mathbf{D}\mathbf{x} = \lambda\mathbf{x} \quad (1.9)$$

In a recent work Vÿrost *et al.* (2018) assert that a vertex that is very strongly connected to others may be viewed as risky because any negative market movement is able to influence not just the asset, but also its neighbors.

In fact, they add constraints based on the negative correlation between weight of an asset and its centrality coming from different network graphs (MST, PMFG and threshold graph) to the classical benchmark portfolio based on return maximization and risk minimization strategies.

They use the long-run correlation coefficient, which is based on the estimator of the heteroskedasticity and autocorrelation consistent variance-covariance matrix introduced by Andrews (1991).

The performance of the benchmark portfolio is always lower than those of a portfolio that takes into account the centrality of a node in the network.

Peralta & Zareei (2016) have already shown that optimal portfolio's weights of assets are negatively correlated to their centrality and positively correlated to the individual performance of an asset, under two different portfolios strategies that minimize the variance of portfolio *m-var* and *M-var*, the latter differ because it considers a given level of expected returns.

The individual performance of assets is measured by standard deviation in the first strategy and Sharpe Ratio in the second one, instead, the systemic performance is represented by the nodes' eigenvector centrality.

The dependence between performance (individual and systemic) of an asset  $i$  and its optimal weight in the portfolio considering *m-var* and *M-var* strategies can be explained by these two OLS equations:

$$w_{i,m-var}^* = \beta_0 + \beta_1 Centrality_i + \beta_2 Std_i + \epsilon_i$$

$$w_{i,M-var}^* = \beta_0 + \beta_1 Centrality_i + \beta_2 SharpeRatio_i + \epsilon_i$$

The most result obtained by Peralta & Zareei (2016) show that higher values of centrality implies lower values of weights in an optimal portfolio under both strategies, in a static and dynamic (among different rolling time windows) temporal dimension.

Pozzi *et al.* (2013) first draw filtered graphs (MST and PMFG) from the stock complex network in order to evaluate the centrality of assets, then they compare portfolios made with central, peripheral and randomly selected stocks: peripheral portfolios perform always better than the others.

Different works show the importance of constructing portfolios taking into account the role of a node in the network regardless the type of assets considered.

Results also outlined by Baitinger & Papenbrock (2016), where interconnect-edness risk of assets can be quantified by their respective centrality scores according to the centrality measures listed above.

## 1.4 Conclusion

In this first chapter, we have presented hierarchical clustering and spectral methods in order to highlight stronger correlations between time series returns of financial assets. These methods allow to filter information in complex datasets by building sparse networks or trees but retaining relevant edges.

These tools not only provide useful information about the hierarchical organization behind asset markets but if employed in the portfolio construction

they allow to reach investments more performing in term of realized risk. The relationship between centrality of a node/asset and its optimal weight in the portfolio become a crucial aspect in the perspective of optimal investments.

This part is the preliminary one of a research that aims to filter correlation matrix of Exchange Traded Funds (ETFs), type of data never used in this research field.

In fact, in the second Chapter, we will show that applying random matrix approach to the correlation matrix of ETF returns and then drawing a minimal spanning tree as in the work of Miceli & Susinno (2004), clusters obtained represent specific class of ETFs according to the classification per classes provided by the Italian Stock Exchange.

The information coming from clusterization of correlation matrix is used to construct investment portfolios. This latter part, that is based on the works of Vÿrost *et al.* (2018) and Bai *et al.* (2009), adds network constraints and random matrix theory to the Markowitz optimization model. Starting from the idea that filtering techniques are tailored to manage risk of portfolio, farther we exploit the information provided by robot-advisors online questionnaires.

Instead, in the third Chapter, the application of RMT and MST to correlation matrix of crypto and real assets allow to detect transmission channels of price information. Specifically, how price information is transferred among different bitcoin market exchanges, and between bitcoin markets and traditional ones.





# Network models to improve robot-advisory portfolio management

## 2.1 Introduction and literature review

FinTechs potential springs from its possibility to unbundle banking activity into its main functions of settling payments, performing maturity transformation and asset allocation (Carney, 2017).

In the last few years, FinTech innovations have increased exponentially by delivering new payments and lending methods, and by penetrating the insurance sector and asset management activity.

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, they allow to act quickly and in the first person. According to Statista, in 2019, the masses managed by automatic consultancy are estimated around 980 billion dollars and forecasts state they will reach over 2,552 billion in 2023<sup>1</sup>.

The rapid growth of FinTech activities, and of robot-advisors in particular, has determined the emergence of new financial risks. In this perspective, robot-advisors that build personalized portfolios on the basis of automated algorithms have been suspected of underestimating investors' risk profile in their actual assets' allocations. Indeed, on one side, the user could not understand the mechanisms underlying portfolios construction and, on the other

---

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

side, computational models that are employed to build portfolios are often simplified, and do not take multivariate dependencies between asset returns properly into account.

We remark that our work is related to a recent stream of papers in the operations research literature, dedicated to the extension of Markowitz portfolio allocation theory, taking network models and, more generally, graph theory models into account.

The paper closest to our work is Clemente *et al.* (2019), who approach asset allocation problems within a network framework. They modify the objective function of the classical minimum variance portfolio taking into account not only the volatility of individual assets but also their network interconnectiveness. Both aspects are considered including an  $N$ -square interconnectiveness matrix computed by means of a local clustering coefficient, calculated on three different correlation networks, and of a diagonal matrix whose elements represent the contribution of the standard deviation of each asset with respect to the total standard deviation.

The three alternative correlation networks considered in the paper are based on both linear and non linear dependence structures, obtained applying Pearson, Kendall and tail correlations. Based on their empirical results, the authors conclude that network-based portfolios lead to portfolios that are more diversified and that show best performances and higher values of risk-adjusted measures 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 the Markowitz objective function. Our additional contribution, from a theoretical viewpoint, is a more interpretable measure of interconnectiveness, based on the notion of network centrality, and on the community structure present in a correlation network. In addition, we extend the applicability of the proposal to Exchange Traded Funds, extensively used in robot-advisory and in many forms of “passive investment platforms.

Another work related to our research is Boginski *et al.* (2014), who exploit the concept of *clique* relaxations in weighted graphs to find profitable “well-diversified” portfolios. In this paper, 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. The work is based on the paper of Boginski *et al.* (2006) in which models of *clique* relaxations are proposed, exploiting the previous work of Pattillo *et al.* (2013). A third paper related to our research is He & Zhou (2011), who modify the optimal allocation 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 of Tversky & Kahneman (1992).

Other extensions of the Markowitz portfolio model that involve operation research and graphical models, as in our proposal, consider multiobjective evolutionary algorithms (MOEAs), as in Metaxiotis & Liagkouras (2012).

For example, 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 portfolio) and allocation shares (which determines the fraction of the capital invested in each asset).

Further researches within the same stream concern local search, simulated annealing, tabu search and genetic algorithm, as in Schaerf (2002), Crama & Schyns (2003), Shoaf & Foster (1996) and Branke *et al.* (2009). In this framework, Ehrgott *et al.* (2004) present 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.

And, finally, Doerner *et al.* (2004) introduce Pareto Ant Colony Optimization as an effective meta-heuristic solution, proposing a two-stage procedure that first identifies the solution space of all efficient portfolios and then locates the best solution within that space.

We contribute to the above literature on graphical network models by proposing a novel network-based asset allocation strategy, which takes into account the network centrality and the community structure of assets in the objective function of Markowitz portfolio allocation.

More precisely, we propose to exploit the information embedded into similarity networks. Indeed, understanding the structure of the similarity networks among assets (see Mantegna & Stanley (1999)) is instrumental for recognizing the origin and the distribution of their returns and to build portfolios more robust against adverse shocks hitting the economy. Similarity patterns between asset returns can be extracted from distance matrices derived from pairwise correlations among returns' time series and these patterns can reveal how asset performances are related to the topology of the network. To account for such topological information, we rely on centrality measures introduced in network theory (see Newman (2018)). More specifically we adopt a global centrality measure that provides information on the position of each node relative to all other nodes in the network; namely the eigen-

vector 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.

The aim of the paper is to show how the inclusion of such centrality measure into portfolio generative models improves their performances in terms of risk/returns. Moreover, we propose a novel 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 an hypothetical investor. Our purpose is to demonstrate that topological methods based on correlation matrices, not yet employed by robot-advisory platforms, can generate new portfolio allocation strategies able to systematically outperform the benchmark index and other naive investment strategies in terms of the implied 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 stocks on a considered time horizon. Tumminello *et al.* (2005) extend Mantegna (1999) with a topological generalization 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, clustering techniques on correlation matrices combined with filtering approaches based on the random matrix theory (RMT) improve the reliability of portfolios in terms of expected and realized risk. Other contributions that follow Tola *et al.* (2008) are León *et al.* (2017), Raffinot (2017), Ren *et al.* (2017) and Zhan *et al.* (2015). To summarize, this literature looks for a hierarchical structure in stock returns and, based on such structure, it constructs Markowitz portfolios.

In this paper, we follow the previous stream of literature by adding 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. A single ETF embeds the characteristics of both a fund and a stock, allowing investors to exploit the strengths of both instruments: diversification and risk reduction peculiar to funds; flexibility and information transparency of real-time trading of stocks. From a methodological viewpoint, we propose a portfolio optimization approach different from what proposed by Tola *et al.*

(2008), taking network centrality explicitly into account in the Markowitz model. In this work, we do not rely only on a typical indicator of diversification, such as the pairwise covariance between assets returns, but we apply a topological measure which embeds also higher order information on assets' behavior to build suitable portfolios.

The empirical findings obtained from the application of our proposed method confirm the validity of the approach revealing that such methodology can constitute a new instrument in robot-advisor toolboxes.

The structure of the chapter 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 similarity network among assets. Furthermore, Section 2 presents a new portfolio optimization strategy that embeds topological information extracted from the network through 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.

## 2.2 Methodology

### 2.2.1 The random matrix approach

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). In a nutshell, RMT aims at separating the “systematic” 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:  $\lambda_k < \lambda_{k+1}; k = 1, \dots, n$ , 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 random variables, with zero mean and unit variance.

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

$$f(\lambda) = \frac{T}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad (2.1)$$

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 “reconstructs” the correlation matrix applying a singular value decomposition based only on the eigenvectors that correspond to eigenvalues that are greater than  $\lambda_+$ .

In other words, RMT eigendecomposes a correlation matrix of time series by returning a filtered correlation matrix (see Eom *et al.* (2009)). The empirical correlation matrix can be used to extract the observed eigenvalues and, then, through Equation (2.1) to check whether some of them reject the null hypothesis. Finally, the filtered correlation matrix is built retaining only those eigenvectors which reject the null hypothesis.

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 systematic information related to the market structure. This corresponds, in this framework, to the identification of empirically constructed variables that drive the EFTs’ system being the number of surviving eigenvalues the effective characteristic dimension of this economic space<sup>2</sup>. 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:

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

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

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

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 RMT the filtered correlation matrix which is given by:

$$\mathbf{C}' = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T, \quad (2.2)$$

where

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

---

<sup>2</sup>With respect to the principal component analysis (PCA), which is a shrinkage method with the same goal but where the threshold for the eigenvalues must be defined in a deterministic way, in RMT the number of “deviating” eigenvalues used to reconstruct the “filtered” correlation matrix depends from the value of the ratio  $Q$ .

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

### 2.2.2 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 from such matrix. To accomplish this purpose we apply the minimal spanning tree (MST) representation of EFT returns similarities (see Mantegna & Stanley (1999); Bonanno *et al.* (2003); Spelta & Araújo (2012)).

In particular, Mantegna (1999) has demonstrated that this particular type of graph, derived from the correlations between stock returns, reveals a topological arrangement of the assets 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:

$$d_{ij} = \sqrt{2 - 2c'_{ij}}. \quad (2.3)$$

and the pairwise distances can be organised in a distance matrix  $\mathbf{D} = \{d_{ij}\}$ , which can be used to draw the MST<sup>3</sup>.

The MST is derived using the 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 the MST allows to reduce the number of links between assets from  $\frac{N(N-1)}{2}$  to  $N - 1$  and the sum of those links provide the minimum weight of the graph.

More precisely, to build the MST we initially consider  $N$  clusters corresponding to the  $N$  ETFs. Then, at each subsequent step, two clusters  $l_i$  and  $l_j$  are merged into a single cluster if:

$$d(l_i, l_j) = \min_{i,j} \{d(l_i, l_j)\}$$

with the distance between clusters being defined as:

$$\hat{d}(l_i, l_j) = \min_{p,q} \{d_{pq}\}$$

---

<sup>3</sup>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 structures, but not in terms of performance.



with  $p \in l_i$  and  $q \in l_j$ . The steps are repeated until a single cluster emerges. In the following, we use the symbols  $\hat{d}$  and  $\hat{\mathbf{D}}$  to denote the distances representing the MST derived from the fully connected network  $\mathbf{D}$ .

In order to detect how financial relationships evolve over time we follow Spelta & Araújo (2012) by employing the *residuality* coefficient measure ( $R$ ) that compares the relative strengths of the connections above and below a threshold value, in formula:

$$R = \frac{\sum_{d_{i,j} > L} d_{i,j}^{-1}}{\sum_{d_{i,j} \leq L} d_{i,j}^{-1}} \quad (2.4)$$

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

We expect that during crisis phases the structure of the MST reinforces in the topological sense, thus impacting on the number of redundant elements that characterize the distinct time periods. Such structural changes, due to the emergence of high correlated positions (synchronization) in the network, affect the behavior of the *residuality* coefficient. Overall,  $R$  decreases when a network becomes less sparse (the number of links increases), and vice-versa.

### 2.2.3 The eigenvector centrality

Understanding the structure of the similarity network, and in particular determining which nodes act as hubs in the network, is key for assessing how EFT returns behave in a multidimensional space and thus to construct portfolios that suitably take into account risk/return trade-off.

The research in network theory has dedicated a huge effort to developing measures of interconnectedness, aimed to detect the most important player 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 group’s 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 & Fortunato (2008)).

Spectral centrality measures include the eigenvector centrality (Bonacich (2007)), Katzs centrality (Katz (1953)), PageRank (Brin & Page (1998)), hub and authority centralities (Kleinberg (1999)). These measures are feedback, also known as global, centrality measures and provide information on the position of each node relative to all other nodes in the network.

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 equation:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^N \hat{d}_{i,j} x_j \quad (2.5)$$

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

The previous equation can be rewritten for all nodes more compactly using matrix notations:

$$\hat{\mathbf{D}}\mathbf{x} = \lambda\mathbf{x} \quad (2.6)$$

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

Note that, in general, there will be many different eigenvalues  $\lambda$  for which a solution to the previous equation exists. However, the additional requirement that all the elements of the eigenvector must 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 similarity networks based on distances between objects, the higher the centrality score 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).

## 2.2.4 Portfolio model

In this Section, we show how similarity networks and topological measures can be combined into a portfolio optimization framework and how they can contribute to improve portfolio performances.

Correlations between stocks play a central role in investment theory and risk management being key elements for optimization problems as in Markowitz (1952) portfolio theory. Thus, correlation based graphs could be very useful for analyzing the interactions between financial markets and building optimal

investment strategy. Onnela *et al.* (2003b) indeed have shown that the assets picked by Markowitz portfolios (Markowitz (1952)) are always located on the outer nodes of a MST, i.e. the portfolios are mainly composed by assets that lay in the periphery of the network and not in its core. Moreover, Pozzi *et al.* (2013) have shown how an efficient portfolio strategy benefits from the knowledge of the hierarchical structure of the market: portfolios including central assets are characterized by a greater risk and a worse performance with respect to portfolios including peripheral assets. Vÿrost *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 & Zareei (2016) which have found a negative relationships between assets' centrality within a complex financial network and the optimal weights under the Markowitz model.

Other authors have gone beyond the above remarks by proposing novel portfolio optimization 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. They have shown 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, have provided 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 the one associated with 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 optimization function itself, rather than applying Markowitz to the filtered and simplified correlation matrix. In our case we minimize the following constrained objective function:

$$\min_{\mathbf{w}} \mathbf{w}^T \mathbf{COV}' \mathbf{w} + \gamma \sum_{i=1}^n x_i w_i \quad (2.7)$$

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 returns of the portfolio, the parameter  $\gamma$  represents a risk aversion coefficient,  $x_i$  is the eigenvector centrality associated with ETF  $i$  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 assets such as to minimize risk and maximize total return by offsetting the synchronous trends of the individual securities. In a nutshell, assets that constitute the portfolio must be uncorrelated or, rather, not perfectly correlated. Within the minimization problem we are facing, the component derived from the MST structure is related to additional risks that eventually, an investor want to minimize. In other words, by increasing the value of  $\gamma$  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  $\gamma$  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  $\gamma$ ) translates in a portfolio composed by systemically riskier assets that lay in the central body of the network thus avoiding peripheral ETFs.

## 2.3 Empirical findings

The data set we consider is composed by 92 ETFs returns' time series traded over the period January 2006 - February 2018 (3173 daily observations) <sup>4</sup>. Table 2.1 shows the classification of the 92 ETFs in 11 asset classes, according to the classification per class of ETFs provided by the Italian Stock Exchange. From Table 2.1 note that the Emerging Market asset classes are the most frequent, followed by Corporate ETFs. Table 2.2 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 2.2 note that the mean value of the returns is around 0 for each asset class, consistently with the efficient market hypothesis suggested by Malkiel & Fama (1970). Differently, the value assumed by 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.

---

<sup>4</sup>The Author thank ModeFinance, a European ECAI, for the data

	ETF class	Number of ETFs
1	Aggregate Bond	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 2.1: **ETFs by asset classes.** Number of Exchange Traded Funds for each class.

	ETF class	Mean	St. Dev.	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 2.2: **ETF classes summary statistics.** Summary statistics for Exchange Traded Funds classes compositions: mean, standard deviation and kurtosis computed for the whole dataset.

To compare the behavior of the returns during financial crises and expansive market phases, the data set has been divided in two chronologically successive batches, from 2006 to 2012 (crisis) and from 2013 to 2018 (post-crisis). Figure 2.1 provides temporal boxplots for ETFs' returns, grouped by their asset class, as reported in Table 2.1. Figure 2.1 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 feature explains why their overall standard deviation, reported in 2.2, is the highest.

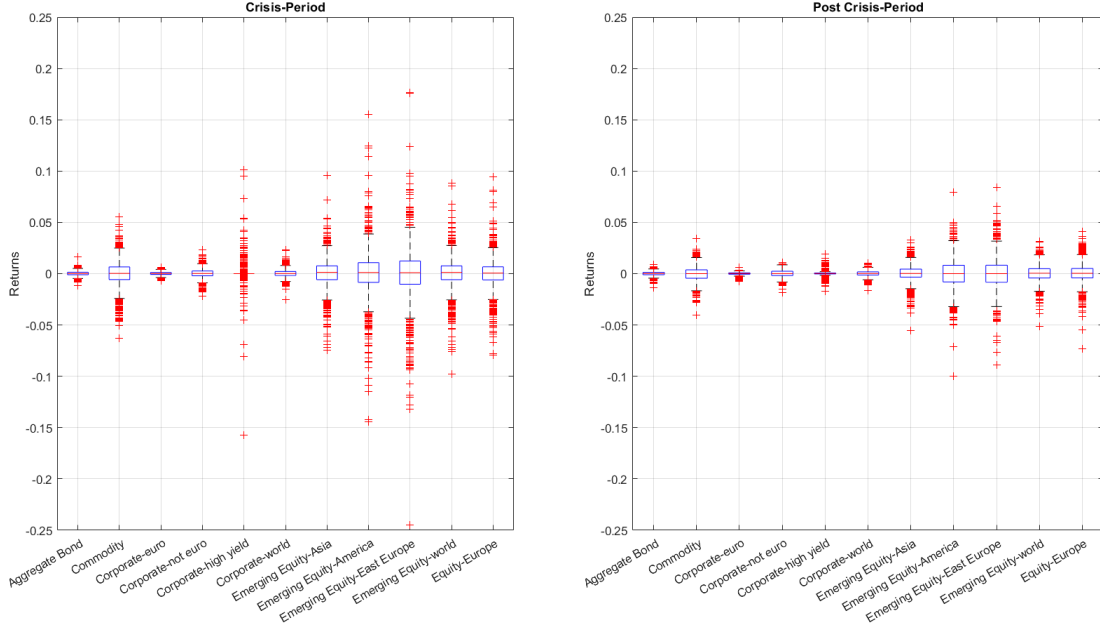


Figure 2.1: **Summary plots for ETF class returns.** Two different periods are compared: crisis (2006-2012) and post crisis (2012-2013).

### 2.3.1 Random matrix theory and network topology

To apply RTM 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 overlapping windows. For each time window, we 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 “filter” correlation matrix applying the RTM approach and the consequent portfolio, which is validated in an out-of-sample fashion using the last month of the window.

Figure 2.2 shows the ordered eigenvalue distribution, for both the empirical and the random correlation matrices, for the last time window of the data set (March 2017- January 2018) as representative of the procedure.

Figure 2.2 shows that most of the data of the eigenvalues' distribution lies between  $\lambda_-$  and  $\lambda_+$ , 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  $\lambda_+$  represent the effective characteristic dimension of the economic space described by the correlation

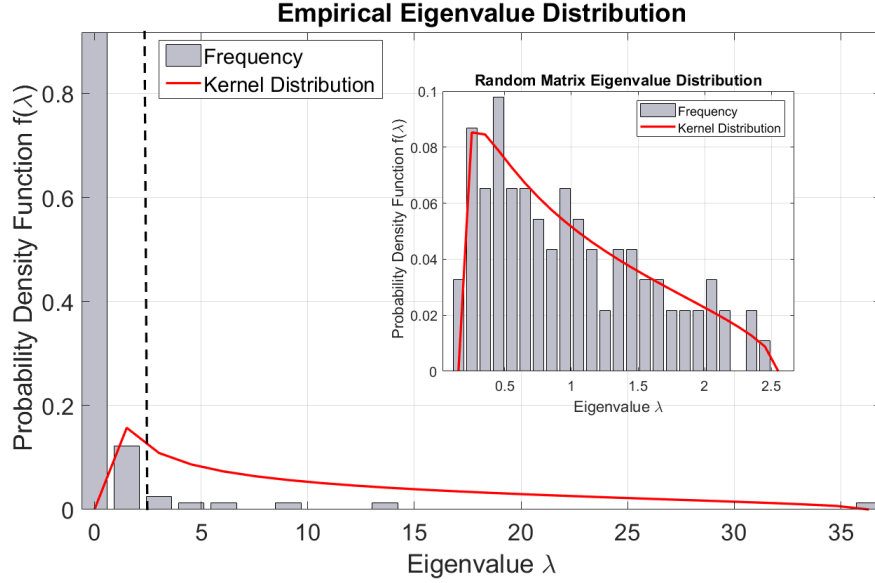


Figure 2.2: **Eigenvalue distribution** In the figure, the red line shows the kernel density of the eigenvalues associated to the empirical correlation matrix  $\mathbf{C}$ , in the main graph, and of the random correlation matrix  $\mathbf{R}$ , in the smaller graph. The dashed vertical line indicates the threshold value  $\lambda_+$  which separates the “signal” eigenvalues from the “noise” ones.

matrix. Similar considerations occur for the other considered time windows. According to Equation 1.5 that measures the role of each class of ETFs in the composition of the eigenvectors, Figure 2.3 shows the value  $X_s^k$  for each eigenvector associated to the six deviating eigenvalues of the last time window (March 2017- January 2018).

In Figure 2.3, peaks indicating the prominent role of one or few classes in determining the dynamics of these eigenvectors.

As described in the methodological Section, if, for each time window, we reconstruct the correlation matrix using only the eigenvectors that correspond to the larger eigenvalues, we obtain a sequence of filtered correlation matrices which can be used to improve the minimal spanning tree representation of the ETFs’ similarity network. Figure 2.4 reports for both the filtered and the unfiltered correlation matrices and 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.

From Figure 2.4 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

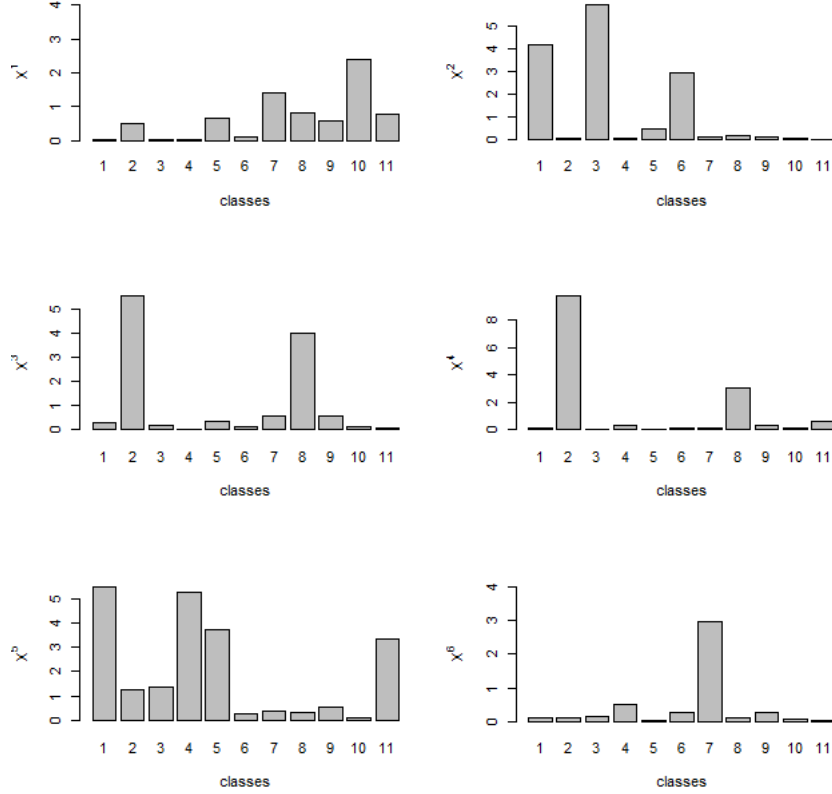


Figure 2.3: **Eigenvector composition.** In the Figure, x-axis represents the classes of ETFs according to the classification per class provided by the Italian Stock Exchange, instead the y-axis shows the value  $X_s^k$  computed for the six deviating eigenvector.

the other hand, the MSTs based on the empirical correlation matrices seem to almost always repeat the same structure: the ETF labeled EIMI-IM, belonging to the Asia Emerging Market class, is for most of the time the central node.

We finally evaluate how the MSTs dynamically change over time. To this aim, we employ, as a summary measure of each MST, the Max link, i.e. 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 building procedure. Figure 2.5 shows the evolution of these two quantities over the considered period.

From Figure 2.5 note that, during the 2008 financial crisis, the Max link sharply decreases, due to the decrease of most distances between ETF re-



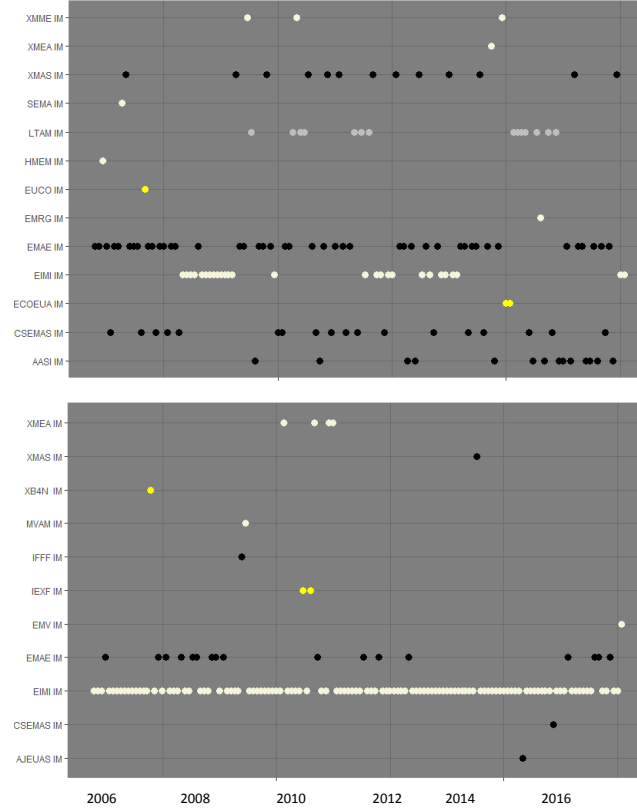


Figure 2.4: **Central nodes of MSTs, along time windows.** The figure reports the ETF node that has the highest degree in the MST representation, in each of the 140 time windows, considering the filtered correlation matrix (top) and the empirical correlation matrix (bottom). Node colors represent the belonging class of ETFs: Corporate (yellow), Emerging Market Asia (black), America (grey), World (beige).

turns. This can be explained by the increased correlations between all returns, which synchronize 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

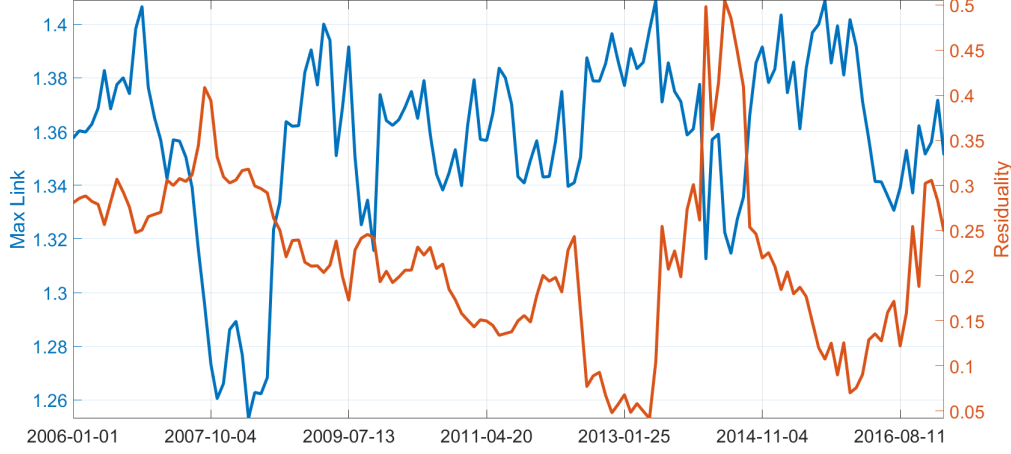


Figure 2.5: **MST thresholds and *residuality* coefficients.** The blue line shows the Maximum link distance (Max), while the red line shows the *residuality* coefficient, whose values are reported respectively on the left and right-y axis.

a limited number of links.

To better understand the previous findings, Figure 2.6 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 2.6 reflects how correlations increase during the crisis phase, leading to the growth of the number of links in the network. In addition, 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.

### 2.3.2 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 2.7), where the value of  $\gamma$  is set a priori accordingly

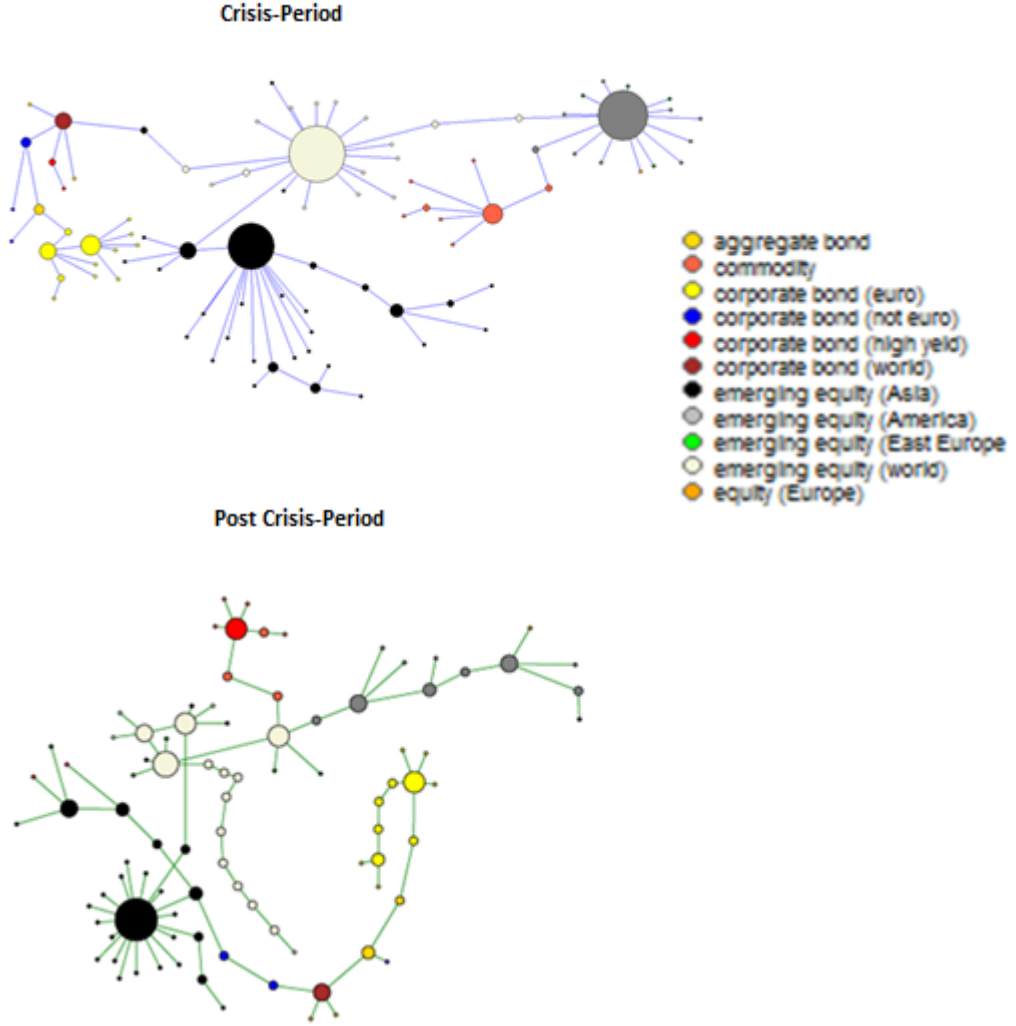


Figure 2.6: **Minimal spanning tree drawn from the RMT filtered correlation matrix for crisis and post-crisis periods.** The nodes in the figure indicate ETFs, the size of the node represents their degree centrality. Colors indicate different asset classes, as reported in the legend.

to the level of the risk aversion of a hypothetical investor. A high value of  $\gamma$  indicates that, in the desired allocation of financial assets, more central ETFs (such as the Emerging Markets ones) have higher weights.

Portfolio returns, and the associated Profit & Loss, are computed using the last month of each time window, in an out-of-sample manner. 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 cumulate each monthly portfolio return, from December 2006 to February 2018 taking re-balancing costs of 10 bps into account .

Figure 2.7 represents the cumulative returns obtained by performing investment strategies based on different values of  $\gamma$ , using the model reported in Equation 2.7. The figure also reports for comparison, the portfolio Profit & Loss of a “naive” (equally weighted) strategy as well as a benchmark portfolio performance, the MSCI Index. Moreover, we also compute the performances obtained by employing the non filtered covariance matrix and the one filtered with the Glasso<sup>5</sup> method of Friedman *et al.* (2008). We report in Table 2.3 the annual Profit & Loss of each competing strategy.

---

<sup>5</sup>The sparsity parameter  $\rho$  has been set to 0.01 as in the reference paper.

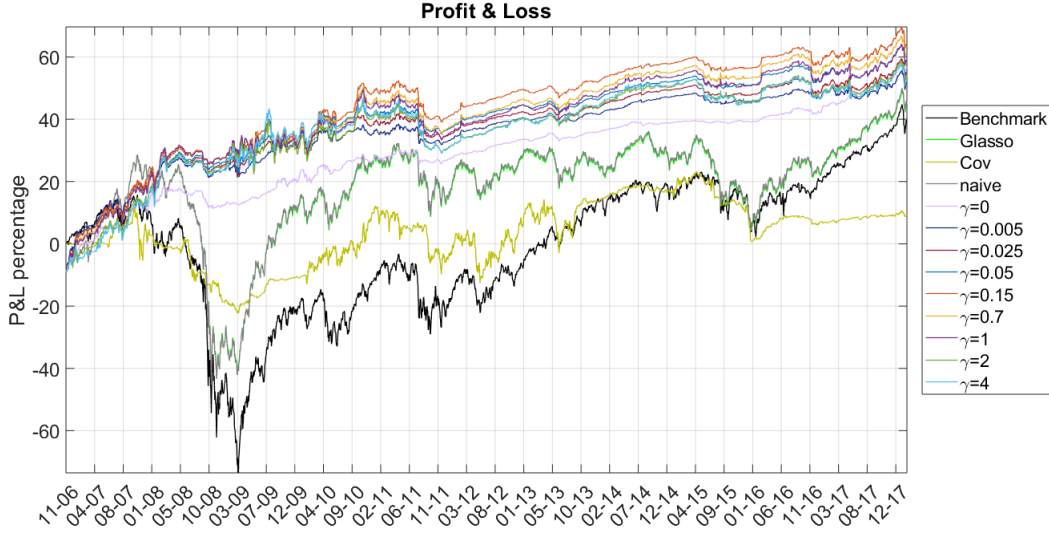


Figure 2.7: **Cumulative returns for different portfolio strategies.** The plot reports the cumulative Profits & Losses obtained by using our model based on different values of  $\gamma$ , the “naive” strategy portfolio (orange line), the MSCI benchmark index (blue line), the Glasso covariance matrix (Glasso) and the standard covariance matrix (not filtered) indicated with Cov.

Figure 2.7 highlights that our proposed model performs better than the benchmark index, the “naive” portfolio strategy, the standard Markowitz portfolio and the portfolio obtained with the sparse Glasso covariance matrix. All of our strategies win in terms of end of sample cumulative returns, regardless the coefficient of the individual risk propensity. Notice that the portfolio based on the non filtered covariance matrix produces the worst performance thus the RMT filter appears to be a fundamental condition for having adequate asset diversification in investment portfolios.

Looking in more detail, during the crisis period (2007 – 2009) our strategy produces higher returns with respect to the competitor portfolios, since it suffers less from financial draw-down. However, it is not able to intercept 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 fact is clearly shown in Table 2.3, the annual loss suffered during 2008 is significantly lower for our strategy with respect to the competitor portfolios.

To provide further insights on portfolio compositions, we report in Figure 2.8 the dynamic of the portfolio weights for ETF classes, considering  $\gamma = 0.7^6$ .

<sup>6</sup>Results for the other  $\gamma$  coefficients are qualitatively the same.

year	Bench.	glasso	naive	cov	$\gamma = 0$	$\gamma = .005$	$\gamma = .025$	$\gamma = .05$	$\gamma = .15$	$\gamma = .7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
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

Table 2.3: **Annual cumulative Profits & Losses** . The table shows the cumulative Profits & Losses of portfolios under different strategies. All the values are expressed in %.

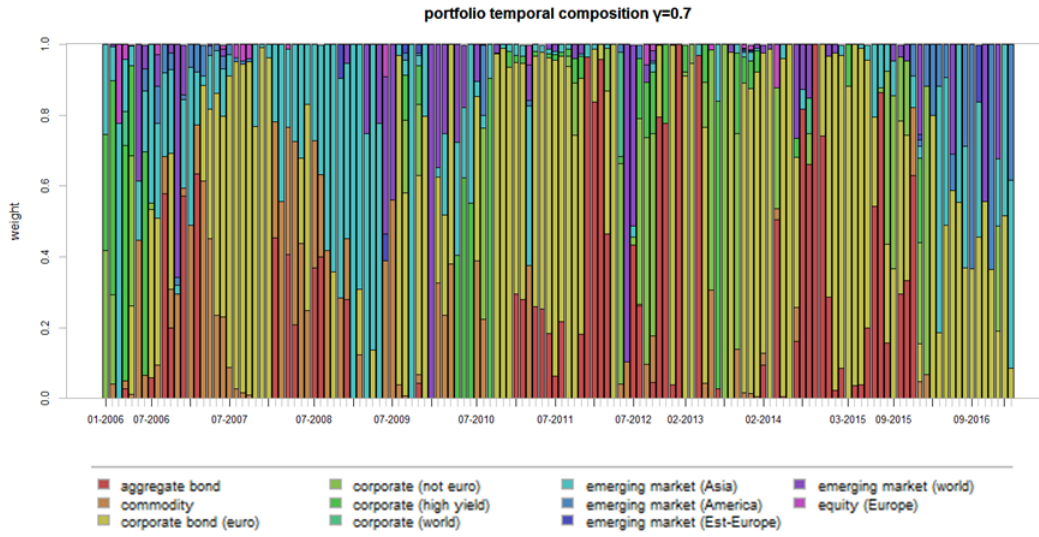


Figure 2.8: **Portfolio weights along time**. The figure reports the portfolio weights associated with a risk aversion coefficient equal to  $\gamma = 0.7$

From Figure 2.8 it is clear that during crisis times the weight of the ETFs belonging to Emerging Equity classes is the highest. 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), the  $\alpha$  of the Capital Asset Pricing Model (CAPM), the Value at Risk

(VaR) and Conditional VaR (CVaR).

Table 2.4 specifically refers to the yearly Sharpe Ratio, defined as the ratio between the mean value of the excess returns and its standard deviation.

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

Table 2.4: **Annual Sharpe Ratio.** The table shows the Sharpe Ratio of portfolios under different strategies. All measures are computed relatively to the benchmark strategy.

Table 2.4 highlights how, during market crises (as in 2008), the Sharpe Ratio of our portfolio strategy is higher with respect to the “naive” one and to the Sharpe Ratio obtained with the Glasso shrinkage method, due to the higher returns produced in this specific phase. The subsequent rebound of 2009 is not captured by our strategy and the low Sharpe Ratio reflects this feature. However, the worst values of the Sharpe Ratio are associated to the portfolio derived for the non-filtered covariance matrix.

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 2.5 describes the  $\alpha$  coefficient, which reflects portfolio extra/under performances with respect to the benchmark. Table 2.5 shows that our portfolios outperform the benchmark strategy reporting values greater than 0. Moreover they are also generally better than the “naive” and “Glasso” portfolios. Only during the growing rebound phase of 2009 do our strategies under perform the “naive” portfolio.

Differently from portfolio performance measures based on returns, those focused on risk compare only our strategies with respect to benchmark and “naive” portfolio. Table 2.6 specifically refers to Value at Risk. Table 2.6 highlights that our portfolio strategies, although becoming more risky during the crisis period (proportionally to risk aversion), present a lower risk than the benchmark portfolio and the “naive” one.

Table 2.7 reports the values of the CVaR of the different portfolio strategies. This measure introduced by Rockafellar *et al.* (2000) quantifies the potential

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

Table 2.5: **Annual portfolio  $\alpha$** . The table shows  $\alpha$  of the CAPM model of portfolios under different strategies. All measures are computed in relation to the benchmark strategy and all the reported values are multiplied by a scale factor of 100.

year	Bench.	glasso	naive	cov	$\gamma = 0$	$\gamma = .005$	$\gamma = .025$	$\gamma = .05$	$\gamma = .15$	$\gamma = .7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	0.35	0.68	0.68	0.11	0.35	0.58	0.61	0.64	0.76	1.12	1.31	1.27	1.15
2007	1.46	1.24	1.25	1.36	0.75	0.76	0.78	0.79	0.79	0.85	0.85	0.89	0.93
2008	3.73	3.04	3.12	1.10	0.51	0.54	0.57	0.59	0.58	0.58	0.56	0.55	0.56
2009	2.35	1.38	1.42	0.41	0.30	0.47	0.76	0.96	1.29	1.50	1.59	1.62	1.64
2010	1.70	1.37	1.38	1.46	0.53	0.65	0.70	0.73	0.76	0.82	0.98	1.00	1.02
2011	2.29	1.63	1.64	1.35	0.42	0.52	0.56	0.57	0.67	0.70	0.68	0.64	0.74
2012	1.35	0.94	0.95	1.55	0.11	0.12	0.12	0.13	0.14	0.15	0.15	0.15	0.18
2013	0.88	0.79	0.79	1.34	0.14	0.24	0.26	0.25	0.26	0.27	0.28	0.28	0.31
2014	1.07	0.78	0.79	0.38	0.09	0.09	0.10	0.14	0.16	0.24	0.27	0.30	0.32
2015	1.27	1.09	1.09	0.80	0.10	0.23	0.23	0.26	0.31	0.42	0.41	0.46	0.52
2016	1.23	1.12	1.13	0.38	0.22	0.26	0.22	0.29	0.29	0.27	0.27	0.29	0.29
2017	0.48	0.47	0.47	0.22	0.20	0.46	0.47	0.50	0.50	0.50	0.50	0.53	0.56
2018	2.04	1.11	1.11	0.37	0.69	0.87	1.05	1.14	1.27	1.27	1.27	1.27	1.27

Table 2.6: **VaR**. The table shows annual Value at Risk of portfolios under different strategies for a confidence interval of 95 %. All the values are expressed in absolute terms multiplied by a scale factor of 100.

extreme losses in the tail of a distribution of possible returns. Results are in line with those presented in Table 2.6: our strategies over perform in terms of expected losses the “naive” and benchmark portfolios, except for the last year considered in the analysis.



year	Bench.	glasso	naive	cov	$\gamma = 0$	$\gamma = .005$	$\gamma = .025$	$\gamma = .05$	$\gamma = .15$	$\gamma = .7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	0.74	0.95	0.95	0.12	0.39	1.48	1.65	1.68	1.75	1.89	1.90	1.80	1.40
2007	1.79	2.14	2.08	2.42	1.02	1.07	1.07	1.10	1.20	1.29	1.28	1.28	1.44
2008	5.56	4.36	4.43	1.79	0.92	1.01	1.08	1.15	1.17	1.16	1.16	1.15	1.17
2009	3.33	2.16	2.23	0.58	0.54	0.62	1.14	1.57	2.03	2.09	2.19	2.38	2.30
2010	2.48	1.83	1.84	2.00	0.73	0.95	1.10	1.17	1.21	1.44	1.50	1.52	1.55
2011	3.35	2.15	2.16	2.32	0.68	0.79	0.86	0.90	1.11	1.16	1.19	1.22	1.24
2012	1.67	1.22	1.23	2.26	0.15	0.18	0.20	0.23	0.24	0.25	0.25	0.27	0.29
2013	1.45	1.25	1.25	1.92	0.26	0.36	0.38	0.38	0.38	0.41	0.41	0.44	0.52
2014	1.35	0.97	0.96	0.45	0.17	0.18	0.20	0.22	0.26	0.36	0.39	0.45	0.48
2015	2.00	1.52	1.53	1.14	0.14	0.36	0.39	0.40	0.46	0.59	0.63	0.75	0.75
2016	1.97	1.67	1.68	0.91	0.32	0.40	0.55	0.61	0.64	0.66	0.70	0.71	0.71
2017	0.73	0.76	0.77	0.33	0.29	0.87	0.95	0.96	0.97	0.97	1.00	1.05	1.05
2018	2.94	1.47	1.47	0.45	0.89	1.26	1.48	1.68	1.95	1.95	1.95	1.95	1.95

Table 2.7: **CVaR**. The table shows annual Conditional VaR of portfolios under different strategies for a confidence interval of 95 %. All the values are expressed in absolute terms multiplied by a scale factor of 100.

year	Bench.	glasso	naive	cov	$\gamma = 0$	$\gamma = .005$	$\gamma = .025$	$\gamma = .05$	$\gamma = .15$	$\gamma = .7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	0.85	1.03	1.04	0.12	0.39	1.76	1.93	1.95	2.00	2.09	2.07	1.89	1.44
2007	2.00	2.39	2.39	3.16	1.18	1.13	1.17	1.24	1.42	1.48	1.40	1.39	1.69
2008	6.85	5.24	5.34	2.00	1.24	1.23	1.60	1.65	1.75	1.77	1.77	1.84	1.66
2009	4.07	2.87	2.95	0.66	0.68	0.68	1.49	1.96	2.28	2.21	2.26	2.80	2.58
2010	2.89	2.03	2.04	2.29	0.91	1.08	1.34	1.50	1.52	1.82	1.94	1.96	1.95
2011	4.34	2.02	2.02	2.89	0.82	1.00	1.02	1.18	1.45	1.49	1.54	1.56	1.58
2012	1.88	1.49	1.50	2.69	0.16	0.22	0.23	0.30	0.29	0.30	0.30	0.32	0.31
2013	1.48	1.32	1.32	2.22	0.31	0.41	0.44	0.44	0.42	0.52	0.53	0.60	0.75
2014	1.54	1.10	1.11	0.50	0.23	0.17	0.19	0.22	0.29	0.36	0.42	0.52	0.54
2015	2.42	1.66	1.68	1.37	0.14	0.46	0.44	0.47	0.51	0.69	0.69	0.87	0.84
2016	2.12	2.01	2.02	1.08	0.41	0.47	0.52	0.59	0.69	0.73	0.75	0.88	0.86
2017	0.82	0.85	0.85	0.40	0.35	0.80	0.83	0.84	0.84	0.86	0.87	0.87	0.86
2018	3.01	1.50	1.50	0.47	0.89	1.35	1.57	1.76	2.01	2.01	2.01	2.01	2.01

Table 2.8: **VaR**. The table shows annual Value at Risk of portfolios under different strategies for a confidence interval of 99 %. All the values are expressed in absolute terms multiplied by a scale factor of 100.

year	Bench.	glasso	naive	cov	$\gamma = 0$	$\gamma = .005$	$\gamma = .025$	$\gamma = .05$	$\gamma = .15$	$\gamma = .7$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
2006	1.05	1.19	1.19	0.13	0.40	2.26	2.43	2.44	2.45	2.46	2.37	2.06	1.51
2007	2.32	2.91	2.92	3.72	1.30	1.46	1.47	1.54	1.82	2.09	2.08	2.06	2.02
2008	7.26	5.73	5.83	2.84	1.56	1.83	2.03	2.16	2.25	2.18	2.15	2.02	1.93
2009	4.56	3.21	3.33	0.87	0.84	0.87	1.68	2.60	3.27	3.28	3.29	3.53	3.55
2010	3.04	2.34	2.34	3.43	1.12	1.42	1.68	1.83	1.84	1.98	2.12	2.12	2.25
2011	4.71	3.32	3.33	3.20	1.09	1.32	1.50	1.53	1.97	2.00	2.02	2.06	1.86
2012	2.04	1.64	1.65	3.12	0.22	0.25	0.32	0.46	0.48	0.47	0.47	0.48	0.47
2013	2.27	2.08	2.08	2.85	0.46	0.60	0.67	0.66	0.64	0.66	0.66	0.69	0.87
2014	1.75	1.21	1.21	0.55	0.33	0.33	0.42	0.43	0.48	0.58	0.66	0.79	0.79
2015	3.10	2.39	2.40	1.76	0.23	0.58	0.72	0.73	0.76	0.90	0.90	0.97	0.98
2016	3.16	2.32	2.34	1.91	0.53	0.71	1.20	1.37	1.45	1.49	1.50	1.52	1.50
2017	1.09	1.29	1.30	0.43	0.43	1.74	1.97	1.98	2.03	2.15	2.23	2.32	2.31
2018	3.18	1.58	1.58	0.52	0.92	1.57	1.80	1.95	2.17	2.17	2.17	2.17	2.17

Table 2.9: **CVaR**. The table shows annual Conditional VaR of portfolios under different strategies for a confidence interval of 99 %. All the values are expressed in absolute terms multiplied by a scale factor of 100.

Tables 2.8 and 2.9 show annual values of VaR and CVaR for a given confidence interval of 99%, results are slightly different with respect to Tables 2.6 and 2.7. However, RMT approach applied to the covariance matrix allow the construction of portfolios that are not only more remunerative but also less risky.

## 2.4 Conclusion

In this paper we have shown how similarity networks can be used to improve robot-advisory services, generating portfolios with better risk/return performances.

In particular, we have demonstrated how random matrix theory together with a topological approach based on the minimal spanning tree can be used to 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 assets traded by 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, 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 assets but it also incorporates aversion towards systemic risk by managing the coefficient  $\gamma$ .

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

Future extensions of this work will be purposed at investigating the relationship between the risk propensity deriving from the robot-advisor online questionnaires and the parameter  $\gamma$ , together with the inclusion of other measures of network centrality in the Markowitz objective function.



# Crypto price discovery through correlation networks

## 3.1 Introduction and literature review

The research literature on crypto currencies is relatively new, but is constantly growing. After the seminal technical paper of Nakamoto (2008), the paper by Dwyer (2015) examines the economics and financial properties of cryptocurrencies, and the paper by Corbet *et al.* (2019) provides a systematic review of the literature that has been developed after 2008 on cryptocurrencies as financial assets.

Within such literature, a relevant stream of research concerns the study of the dynamics of cryptocurrency market prices, either from an internal viewpoint (between cryptocurrency markets) or in relationship with other “classic” market prices.

While some papers investigate this issue from a univariate statistical approach, focusing on bitcoin prices, very few consider a multivariate statistics viewpoint, which deals with the interconnectedness among crypto prices and between crypto prices and classic prices.

A noticeable exception is the paper by Corbet *et al.* (2018), who analyses the relationships among alternative cryptocurrencies: Bitcoin, Litecoin and Ripple, and show that they are strongly interconnected, demonstrating similar patterns of returns and volatility. A related paper is Ciaian & Rajcaniova (2018) who analyses the relationship between the bitcoin and sixteen alternative coin prices, and found that they are indeed interdependent, but independent from exogenous factors.

Corbet *et al.* (2018) also analyse interdependence between crypto prices and a variety of other financial assets such as gold, bonds and stocks. They

found that the volatility of cryptoassets is substantially higher than that of traditional assets, and that cryptocurrencies are rather isolated from other assets, thus showing a diversification benefit. Dyhrberg (2016) and Bouri *et al.* (2017) reach similar conclusions, thus confirming that cryptocurrencies are rather isolated from classical assets. Note however that the same authors conclude that such isolation emerges in the short run, but not in the long run and, thus, the evidence on the diversification benefit is not conclusive.

Understanding price interconnectedness is important not only to describe relationships between different asset prices, but also to understand whether prices in different markets quickly react to each other or, in other words, whether markets are efficient. The paper by Brandvold *et al.* (2015) is the first one that addresses this question, studying the price discovery process in bitcoin markets, by means of the econometric methodologies of Hasbrouck (1995) and Gonzalo & Granger (1995). Using data from seven exchanges, in the period from April 2013 to February 2014, they find that Mtgox (bankrupting shortly after the sampled period) and BTC-e are the price setters. Pagnottoni *et al.* (2018) extends their analysis to the period January 2014 to March 2017, and found an increased role of Chinese exchanges.

A related work is the paper of Urquhart (2016) who specifically analyzes whether bitcoin markets are efficient, using price return data from August 2010 through July 2016: they cannot confirm the efficient market hypothesis. However, another study (Nadarajah & Chu, 2017) reveals that a power transformation of bitcoin returns can be concluded as “weakly efficient” and, thus, the evidence on bitcoin market efficiency is not conclusive.

Our contribution is to develop a novel multivariate statistical model to study cryptocurrency price dynamics, aimed at acquiring further empirical evidence on whether bitcoin prices from different exchanges are strongly interrelated, as in an integrated and efficient market, following the paper by Brandvold *et al.* (2015); but also whether such interactions are affected by “exogenous” prices of classical assets, as in the paper of Corbet *et al.* (2019). In other words, we aim to know if the bitcoin, whose capitalization is now substantial, is still an investment diversifier and whether the bitcoin markets are efficiently integrated.

Besides shedding more light on the diversification and efficiency property of bitcoin prices, we extend Corbet *et al.* (2018), Brandvold *et al.* (2015), and the related papers, by modelling price interconnectedness with correlation network models, as in the recent paper of Giudici & Abu-Hashish (2019). However, differently from the previous authors, instead of inserting correlation networks into a Vector Autoregressive model, which requires strong distributional assumptions, we follow a non parametric clustering model,

based on the minimal spanning tree (MST) approach proposed by Mantegna (1999). The MST approach will be extended with a preliminary random matrix filtering, that improves its interpretability.

The paper is organized as follows: Section 2 contains our proposed model; Section 3 presents the available data; Section 4 the empirical application of the proposed model to the data; Section 5 contains some concluding remarks.

## 3.2 Proposal

In this section we present our methodological contribution: a clustering method for market prices, based on the minimal spanning tree approach proposed by Mantegna, empowered by the random matrix theory approach. Mantegna (1999) proposed the minimal spanning tree (MST) to detect the hierarchical organization of stock prices in financial markets, using their correlation matrix. Spelta & Araújo (2012) further qualified the MST as a network structure between a group of nodes, representing different time series, whose edges minimize the pairwise distances between each pair of nodes. In other words, an MST can be seen as a parsimonious representation of a network model, in which sparseness replaces completeness in a suitable way. More formally, consider  $N$  financial assets, for which we observe the corresponding price time series:  $(P_i, i = 1, \dots, N)$ , each of which is a vector of prices observed in  $T$  different time periods:  $P_i = (P_i(t), t = 1, \dots, T)$ . From the price time series we can obtain  $N$  return time series,  $(r_i, i = 1, \dots, N)$ , as follows:

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

From the return time series we can calculate the correlation matrix  $\mathbf{C}$ , whose elements  $c_{ij}$  are defined by:

$$c_{ij} = \frac{E(r_i r_j) - E(r_i)E(r_j)}{\sigma(r_i)\sigma(r_j)},$$

where  $E(\circ)$  indicates the mean value and  $\sigma(\circ)$  the standard deviation of each return time series. From the correlation matrix  $\mathbf{C}$  we can then calculate the distance between any two asset returns,  $d_{ij}$ , as follows:

$$d_{ij} = \sqrt{2 - 2\rho_{ij}},$$

a function which ranges between  $(0, 2)$ , with  $d_{ij} = 0$  when  $\rho_{ij} = 1$  and  $d_{ij} = 2$  when  $\rho_{ij} = -1$ . It assumes that, for any pair of asset return time series, the higher the correlation, the lower the distance.

Let then  $\mathbf{D} = (d_{ij}, i = 1, \dots, N; j = 1, \dots, N)$  be a matrix which contains

all pairwise distances. We can associate to the distance matrix a network  $G = (V, W)$ , with vertices  $V$  that correspond to the  $N$  asset return time series and with connection weights  $W$  between them which correspond to the  $\frac{N(N-1)}{2}$  pairwise distances  $d_{ij}$ .

The minimal spanning tree (MST) proposed by Mantegna (1999) is based on the distance matrix  $\mathbf{D}$ . It reduces the number of weights that can connect the  $N$  nodes, from  $\frac{N(N-1)}{2}$  to  $N - 1$ . It does so through a hierarchical clustering algorithm which associates to each node only another one, that is minimally distant from it, under the constraint of avoiding loops between groups of nodes.

We remark that the network structure simplification induced by a minimal spanning tree may be too drastic, especially if based on random noise rather than on actual distances between nodes. To overcome this problem, in this paper we suggest of preprocessing the correlation matrix and, therefore, the distance matrix, before applying the minimal spanning tree method.

The necessity to improve the MST representation was pointed out by Tumminello *et al.* (2005), who introduced the planar maximally filtered graph (PMFG), which preserves the hierarchical structure of the MST, but with a more complex structure. Indeed, given a set of  $N$  time series, a MST contains  $N - 1$  links whereas a PMFG contains  $3(N - 2)$  links.

Here we aim to improve the MST without enriching its structure but, rather, working on its input: the distance matrix. To achieve this goal, we employ the random matrix theory approach (RMT), proposed by Onnela *et al.* (2004) and Tola *et al.* (2008), preprocessing the correlation matrix by removing the noise contained in it.

The rationale behind the random matrix theory approach is to employ each empirical eigenvalue  $(\lambda_k, k = 1, \dots, N)$  obtained from the correlation matrix  $\mathbf{C}$ , as a test statistic for the null hypothesis that the correlation matrix is a random Wishart matrix  $\mathbf{C}' = \frac{1}{T}\mathbf{A}\mathbf{A}^T$ , where  $\mathbf{A}$  is a  $N \times T$  matrix containing  $N$  time series of length  $T$ , whose elements are independent and identically distributed “white noise” random variables, with zero mean and unit variance.

To actually implement the test, we need a statistical distribution. Marchenko & Pastur (1967) showed that, under the null hypotheses,  $\lambda_1 = \dots = \lambda_N = \lambda$ , and that the asymptotic density of  $\lambda$ , for a fixed  $Q = \frac{T}{N} \geq 1$ , as  $N \rightarrow \infty$  and  $T \rightarrow \infty$ , is given by:

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

where  $\lambda \in (\lambda_-, \lambda_+)$ , with  $\lambda_+ = \sigma^2 + \frac{1}{Q} + \sqrt{\frac{1}{Q}}$  and  $\lambda_- = 1 + \frac{1}{Q} - \sqrt{\frac{1}{Q}}$ . From the above density, it follows that, when  $\lambda_k \geq \lambda_+$ , the null hypotheses is rejected, as the  $k$ -th empirical eigenvalue cannot be an eigenvalue from a random Wishart matrix.

From an operational viewpoint, if the eigenvalues are ordered from the largest to the smallest, we can retain only those that exceed  $\lambda_+$  and reconstruct the correlation matrix, through singular value decomposition, using only the eigenvectors corresponding to them. Doing so, as suggested by Plerou *et al.* (2002) we “filter” the correlation matrix.

From an empirical viewpoint, Miceli & Susinno (2004) show that, when the random matrix approach is applied as outlined before, the minimal spanning tree leads to a grouping of assets that better correspond to “typical” investment strategies. Our aim is different: we would like to verify whether the application of the RMT on the correlation matrix between bitcoin exchange and classical market prices produces a minimal spanning tree that can shed light on what drives bitcoin prices: endogenous or exogenous factors.

### 3.3 Data

In this Section, we describe the analyzed data. We consider, without loss of generality, the most important cryptocurrency: the bitcoin, whose relative price will be taken with respect to the US dollar. With no further loss of generality, and to reduce volatility issues, we consider daily prices, obtained at the end of the day.

Our first research question is to assess whether bitcoin prices in different exchange markets are correlated with each other, thus exhibiting “endogenous” price variations. To understand this question, we have chosen a set of representative exchange markets, for which price data is available, in a sufficiently long period of time. Specifically, we have selected eight exchange markets, representative of different geographic locations, which represent about 60% of the total daily volume trades. They are reported in Table 3.1, along with the corresponding market shares. For each exchange market, we have collected daily data for a time period that goes from May, 18th, 2016, to April 30th, 2018.

Our second research question is to understand whether bitcoin price variations can also be explained by exogenous classical market prices. To evaluate this issue, we have obtained daily data on some of the most important asset prices: Gold, Oil and SP500; as well as on the exchange rates USD/Yuan and USD/Eur. Similarly to what done for bitcoin prices, we have considered, as daily price, the market closing price. When jointly considering bitcoin and



Exchange	Market share
Bitfinex	42%
Coinbase	6%
Bitstamp	5%
Hitbtc	3%
Gemini	2%
itBit	1%
Kraken	0.5%
Bittrex	0.5%

Table 3.1: **Exchange markets by daily trading volumes.** Source: <https://coin.market/markets/info>.

“standard” markets, one issue to be solved is that, while Bitcoins are traded 24 hours per day and 7 days per week, standard markets have closing times and days. We have overcome this issue keeping standard market prices constant at the last closing time, during market closure.

Figure 3.1 presents the time evolution of the Bitcoin prices, in the considered time period.

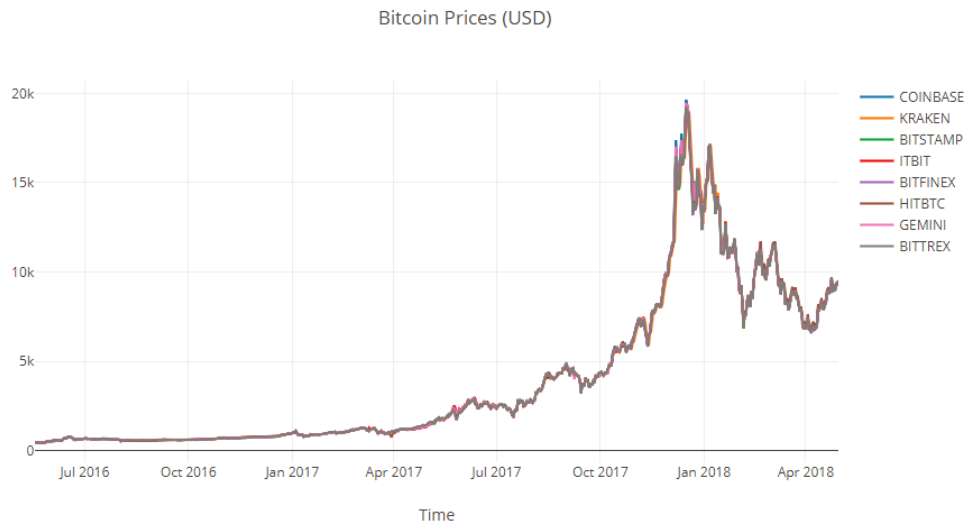


Figure 3.1: **The time series plot of Bitcoin prices.**

From Figure 3.1 note the well known 2017 rise in Bitcoin prices, from a minimum of about 430 dollars per bitcoin to a maximum of almost 20,000 dollars, followed by a high volatility in 2018. Note the slight differences between prices, which shows that bitcoin prices in different market exchanges

are not perfectly aligned. To better understand the latter finding, some summary statistics on the considered data are presented in Table 3.2.

Price	Mean	St. Dev.	Min	Max	Kurtosis
Bitfinex Bitcoin	3899.56	4274.46	435.61	19187.12	3.22
Coinbase Bitcoin	3919.05	4318.98	438.38	19650.01	3.22
Bitstamp Bitcoin	3899.04	4286.02	439.62	19187.78	2.50
HitBtc Bitcoin	3916.19	4297.17	436.36	19095.30	3.77
Gemini Bitcoin	3910.38	4306.36	437.57	19475.90	2.90
ItBit Bitcoin	3907.13	4300.32	438.61	19357.97	2.67
Kraken Bitcoin	3890.18	4272.55	433.50	19356.91	2.05
Bittrex Bitcoin	3893.83	4269.89	421.11	19261.10	2.53
Gold	1275.57	52.34	1128.42	1366.38	7.02
Oil	48.67	3.16	39.51	54.45	18.98
SP500	2414.78	212.308	2000.54	2872.87	11.86
USDYuan	6.67	0.19	6.26	6.96	4.85
USDEur	0.88	0.04	0.80	0.96	4.53

Table 3.2: **Summary statistics for bitcoin and classic asset prices.** Means, standard deviations, minimum and maximum values are all expressed in dollars.

Table 3.2 confirms the slight differences in bitcoin prices along the considered market exchanges: the means and the standard deviations are slightly different, and more so are the extreme statistics.

With respect to classical assets, such as Gold and Oil, the volatility of bitcoin prices is much higher: respectively, about 80 and 1400 times higher.

Even with respect to SP500, the volatility of Bitcoin prices is about 20 times higher. Instead, exchange rates are, as well known, much less volatile than bitcoin prices. These results are in line with the available literature (see e.g. Corbet *et al.* (2019)).

Finally, looking at the last column in Table 3.2 note that bitcoin prices have values of kurtosis quite similar among each other, and lower than those of the classical assets.

### 3.4 Empirical findings

The aim of this section is to apply our proposed model to verify whether bitcoin prices from different exchanges are strongly interrelated with each other and whether such “endogenous” interactions are affected by “exoge-

nous” prices of classical assets.

Figure 3.2 presents, by means of a heatmap, all pairwise correlations between the considered asset prices, in the considered time period. Positive correlations are marked in blue, and negative correlations in red, with stronger colors indicating higher correlations (in absolute values).

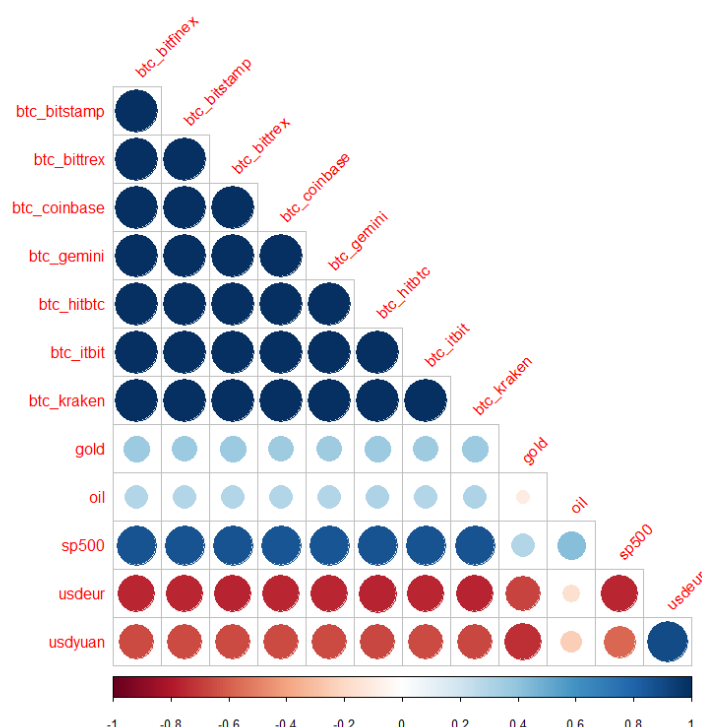


Figure 3.2: Correlation matrix between prices.

From Figure 3.2 note that the correlations between different exchange prices are quite high, revealing that markets are highly correlated and synchronized, resulting in a strong endogenous driver of price variation. On the other hand, correlations with “real” asset prices, such as gold and oil, are low, a result in line with the literature, that considers bitcoins as potential diversification assets (see e.g. Corbet *et al.* (2018)). However, the correlation with the SP500 index is positive and those with the exchange rates are negative, a

result that seems to conflict with the reference literature.

To better understand the implications of Figure 3.2, Giudici & Abu-Hashish (2019) analyzed similar bitcoin price data using partial correlation networks. This because pairwise correlation may be inflated by correlations that may arise from a common relationship with third variables. Their empirical findings show that bitcoin prices on one hand, and “classic” asset prices on the other hand, form two rather distinct clusters of connections, which are highly interconnected inside. They also show the high centrality of two of the largest bitcoin exchanges: Bitfinex and Bitstamp, which thus emerge as “price setters”. They also find that the link between the two clusters is given by the Hitbtc exchange, which is affected both by standard asset prices and by other exchange market prices.

Here we take a different approach to improve the empirical findings that can be obtained from the correlation matrix in Figure 3.2. We derive the minimal spanning tree of the correlation matrix, introduced in Section 2. The obtained MST is shown in Figure 3.3.

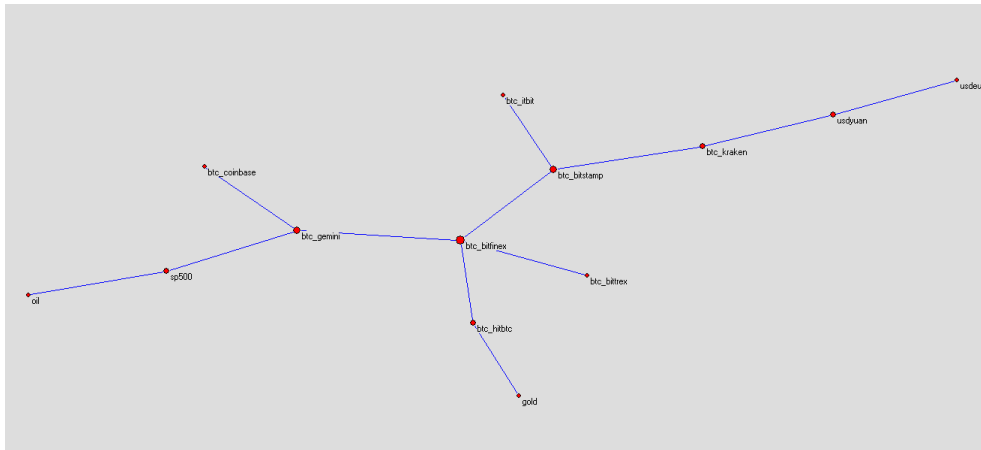


Figure 3.3: **Minimal spanning tree between prices.**

From Figure 3.3 note that Bitfinex and Bitstamp have a pivotal role, as found in Giudici & Abu-Hashish (2019). However, the MST reveals more insights. For example, it shows that, while Bitfinex is “closer” to real financial assets, such as Gold, SP500 and Oil, Bitstamp is more related with exchange rates. In addition, Hitbtc separates Bitfinex from Gold. Note also that smaller exchanges are more peripheral.

Indeed, the advantage of MST models, with respect to correlation network models, is that they provide a “hierarchical” split of the prices (nodes), showing them in order of distances (weights), calculated from their correlations.

Table 3.3 reports the weights corresponding to the application of the MST algorithm to the considered data.

			weight
1	btc_coinbase	btc_gemini	0.207968897069462
2	btc_bitfinex	btc_gemini	0.252298438331244
3	btc_bitstamp	btc_itbit	0.279684526027965
4	btc_bitstamp	btc_bitfinex	0.314896774631822
5	btc_bitfinex	btc_hitbtc	0.380165409259575
6	btc_bitfinex	btc_bittrex	0.458640635735809
7	btc_kraken	btc_bitstamp	0.596065889298915
8	usdeur	usdyuan	0.989859235886413
9	sp500	oil	1.23032152766612
10	gold	btc_hitbtc	1.35470479294313
11	btc_gemini	sp500	1.3728902044797
12	usdyuan	btc_kraken	1.3951809819423

Table 3.3: **Adjacency matrix from MST.** Indirect links between nodes from the distance matrix.

From Table 3.3, note that the “closest” nodes are those between bitcoin price exchanges, as expected: their pairwise connections correspond to the first seven edges of the MST. The following edge is placed between the two exchange rates, then between SP500 and oil.

Last, the procedure finds three edges that break the “separation” between crypto and classical asset prices: the first one relates Hitbtc with Gold; the second one Gemini with SP500; the last one UsdYuan with Kraken. These latter results are quite meaningful, as they characterise the “local” behaviour of specific exchanges, a phenomena already found in Giudici & Abu-Hashish (2019).

We now verify whether the application of the random matrix theory approach, before implementing the minimal spanning tree, can extract further empirical findings from the correlation matrix. The results are shown in Figure 3.4.

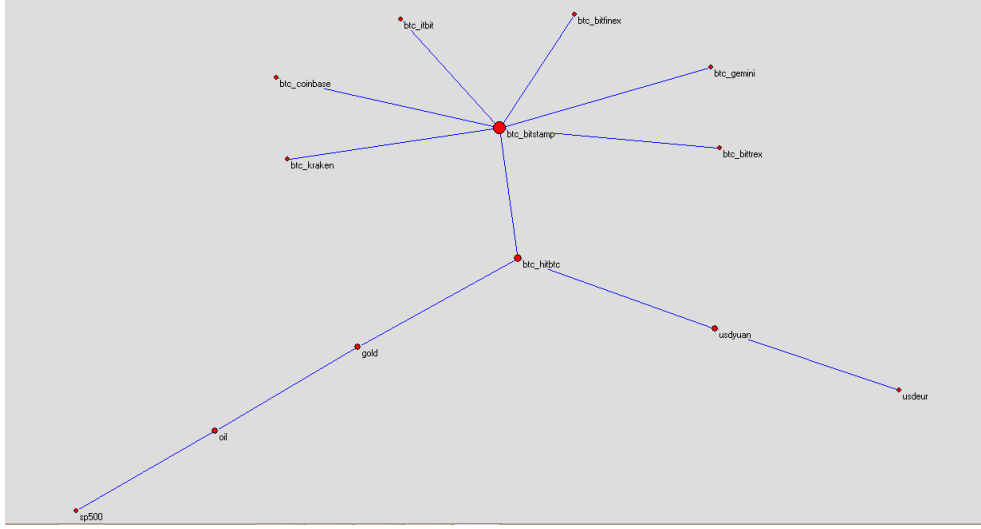


Figure 3.4: **Minimal spanning tree between prices, after random matrix theory filtering.**

From figure 3.4 note that the filtered MST provides a graphical structure that is simpler than that obtained in 3.3, without the application of RMT filtering. On one hand, bitcoin exchange prices form a “star” configuration, with Bitstamp at the center, confirming its role of price setter; while Bitfinex, probably because of its relatively high volatility, is not found to be central. On the other hand, all classic asset prices are separated from bitcoin prices, pointing towards a “diversification benefit” of bitcoins with respect to them, a result fully in line with the existing literature. Note also that the MST well separates the role of “real” assets, such as SP500, Oil and Gold, from “financial” assets such as the exchange rates.

To summarize, filtering the correlation matrix with the random matrix approach leads to a minimal spanning tree that, with respect to the unfiltered one, is simpler and which leads to empirical findings that: i) do not indicate a significant correlation between crypto prices and exogenous price drivers, from classical markets, consistently with the literature; ii) indicate that exchange prices have a strong endogenous source, which specifically come from the largest and least volatile exchanges, such as Bitstamp.

We can draw more interpretation examining the distance weights corresponding to the joint application of the RMT and MST, in Table 3.4.

			weight
1	btc_gemini	btc_bitstamp	0.288433056210555
2	btc_bitstamp	btc_bitfinex	0.292594189179681
3	btc_coinbase	btc_bitstamp	0.292645290798001
4	btc_bitstamp	btc_itbit	0.359144767683509
5	btc_hitbtc	btc_bitstamp	0.39147736801806
6	btc_bitstamp	btc_bittrex	0.440520251858754
7	btc_kraken	btc_bitstamp	0.600090764365521
8	usdeur	usdyuan	0.759135861596123
9	sp500	oil	1.14667657820831
10	oil	gold	1.2914392737851
11	gold	btc_hitbtc	1.35598797995048
12	usdyuan	btc_hitbtc	1.39185686202477

Table 3.4: **Adjacency matrix from RMT+MST.** Indirect links between nodes from the distance matrix based on the filtered correlation matrix.

Comparing Table 3.4 with Table 3.3, the previously discussed findings are confirmed. Again the seven closest pairs of nodes concern bitcoin exchange prices, indicating a strong presence of endogenous price variation; in addition, in Table 3.4, all pairs contain the Bitstamp node, indicating its centrality. A further difference is that the connections between crypto prices and classic prices reduce to two, and they both involve Hitbtc. This result is more in line with what obtained in Giudici & Abu-Hashish (2019) about the role of Hitbtc as a “separator” between classic and crypto assets.

To assess the robustness of our empirical findings, we now verify whether the found tree structure is stable over time. For this purpose, Figure 3.5 shows the MST obtained in each of nine one-year rolling periods, after the application of RMT. The first one starts from 18/05/2016, the following are shifted ahead by one month, until the eighth one which starts on 18/02/2017.

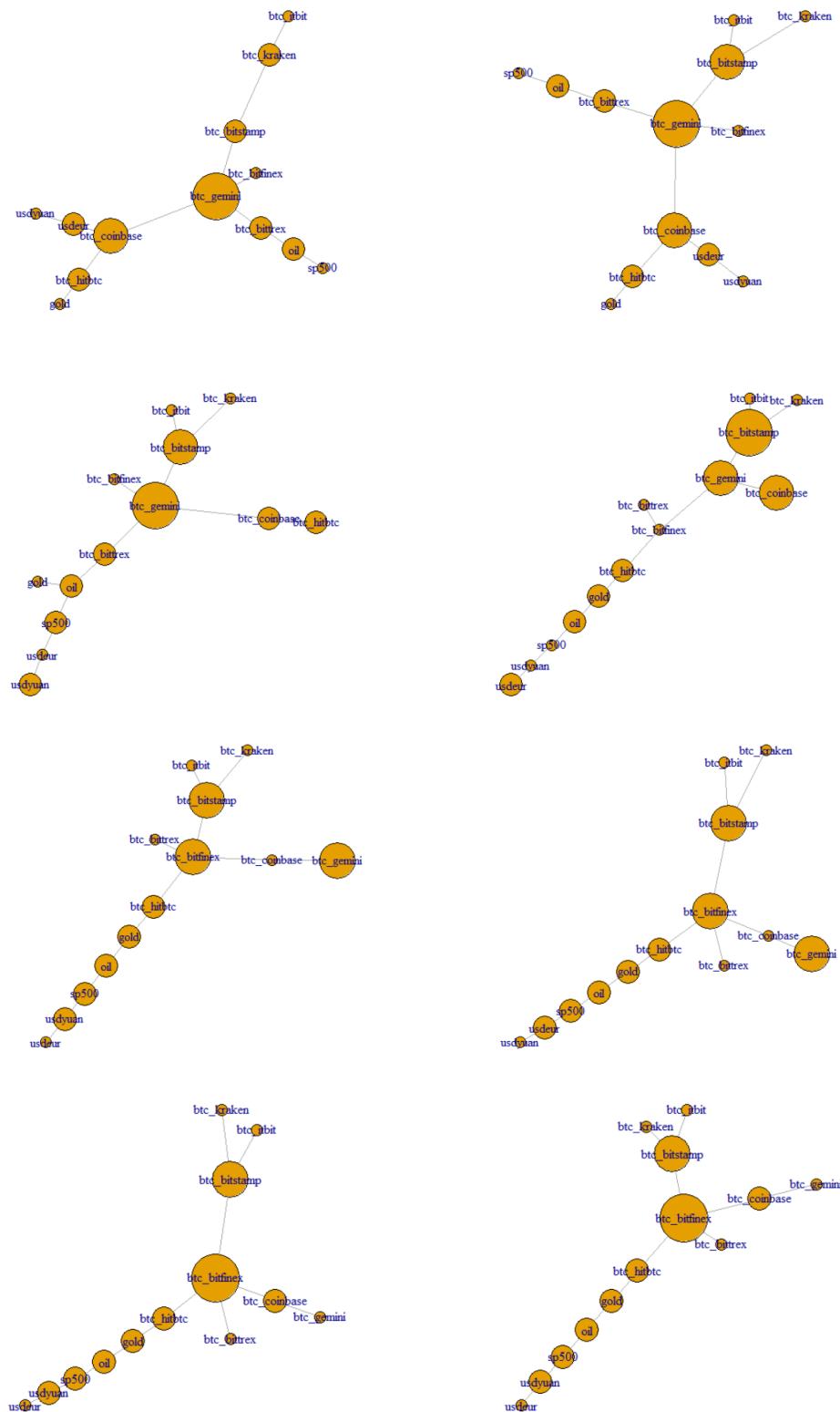


Figure 3.5: Time evolution of the the minimal spanning tree.



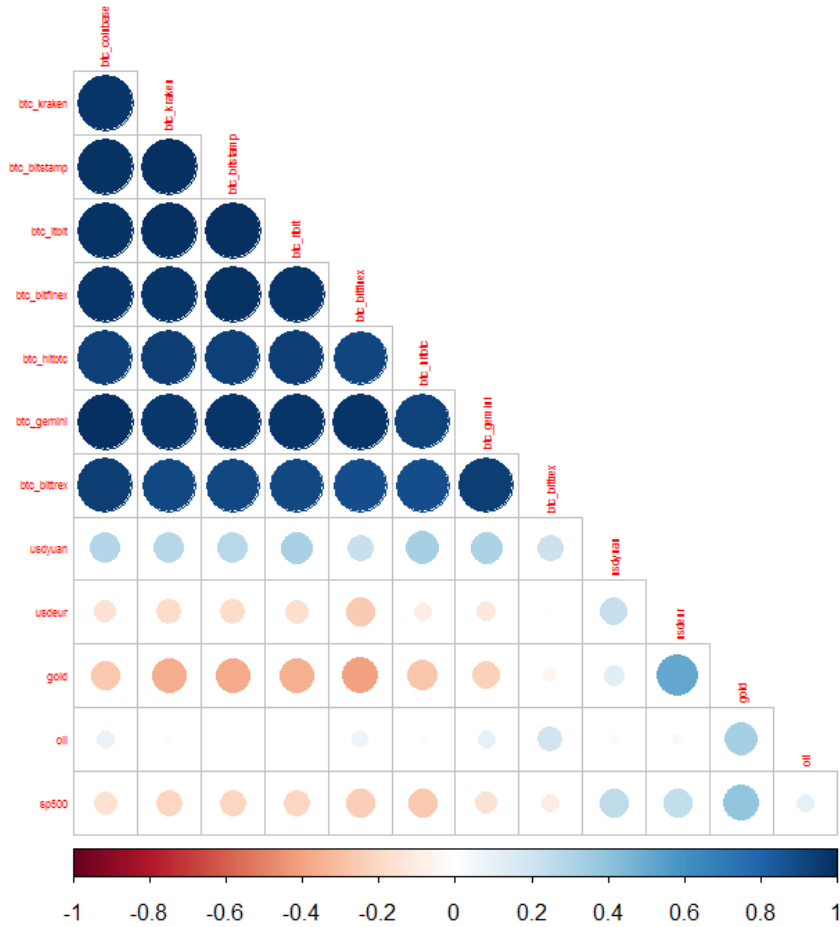
Figure 3.5 shows that the MST configuration is quite stable over time, particularly from the third period onwards, as all graphs show a configuration similar to the “static” one in Figure 3.4. From a theoretical viewpoint, we remark that, when random matrix theory is applied before the application of the minimal spanning tree the results are “stabilized, as RMT filters out noise. In fact, comparing the different time periods in Figure 3.5, the spanning trees do not change sensibly, even during bubble periods. We have indeed applied the test for crypto bubbles suggested in Cheah & Fry (2015) and Hafner (2018), obtaining that the December 2017 period shows a significant bubble. However, Figure 3.5 shows that the two correlation networks at the bottom of Figure 3.5, which fully contain the bubble period, do not show an evident structural change.

We have conducted a further robustness test on the time dynamics of our results. From Table 3.2 the kurtosis observed for the bitcoin prices is smaller than that of classical assets, and this may justify the use of an unconditional variance.

We have however assumed that the unconditional variance is different from the realized one and we have calculated pairwise correlations not among returns, as before, but among volatilities, to see what could drive the volatility dynamics, rather than the price dynamics.

In particular, we have postulated the existence of a negative correlation between the realized macroeconomic volatility and the realized volatility of bitcoin prices, as suggested by Christian *et al.* (2018).

These authors report that the two months lagged SP500 realized volatility may be a useful predictor for the bitcoin volatility. Following this suggestion, we have calculated the pairwise correlations between all bitcoin exchange volatilities, and all classical assets volatilities (lagged by two months), and reported them in Figure 3.6.

Figure 3.6: **Volatility correlation plot.**

From Figure 3.6, note that the correlation between classic assets and bitcoins is constant across different exchanges. In particular, the bitcoin volatility is negatively correlated with that of the SP500 index. This indicates a further, important, empirical finding: the volatility of classic asset prices negatively affects the volatility of bitcoin prices, with a delay.

This result, that confirms Christian *et al.* (2018) can be better seen in Figure 3.7, which reports in the same graph the realized volatilities of the Bitcoin Bitstamp price and for the SP500 index.

In the Figure 3.7, both volatilities have been normalized, being the volatility of the bitcoin price much higher (10 times more on average).

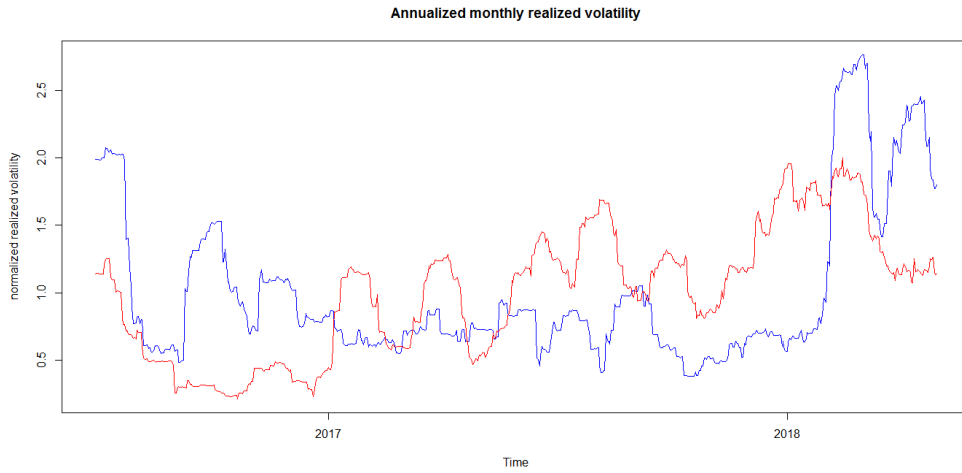


Figure 3.7: **Realized volatilities for SP500 (blue) and Bitcoin Kraken (red), normalized.**

From Figure 3.7 the two months lagged effect of the volatility of the SP500 index on the bitcoin price volatility of the Bitstamp exchange, is evident. Similar results hold for all other exchange prices, consistently with the found price setter nature of the Bitstamp exchange.

As a last robustness exercise on our empirical findings, we compare, on the same data, our method with the planar maximally filtered graph and with the Granger causality network, suggested in Billio *et al.* (2012).

Figure 3.8 and 3.9 give the results from the application of these two methodologies.

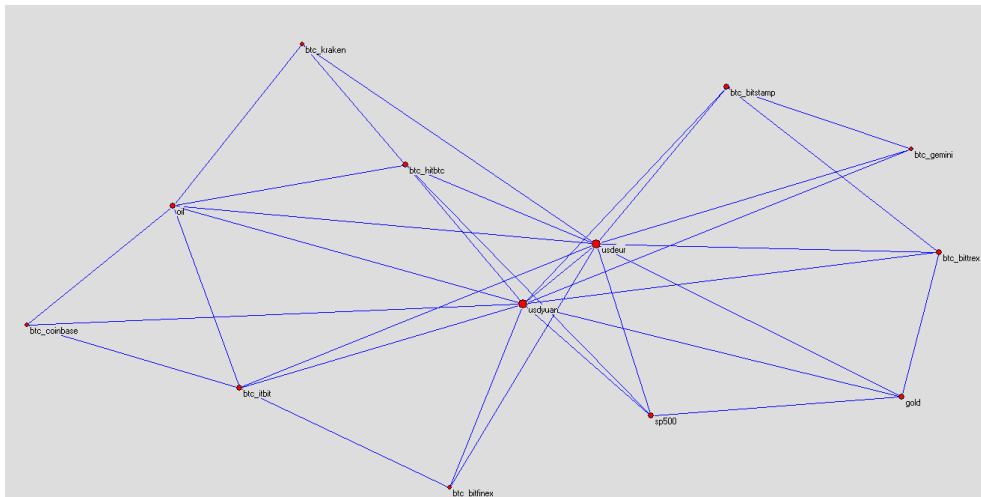


Figure 3.8: Planar maximally filtered graph.

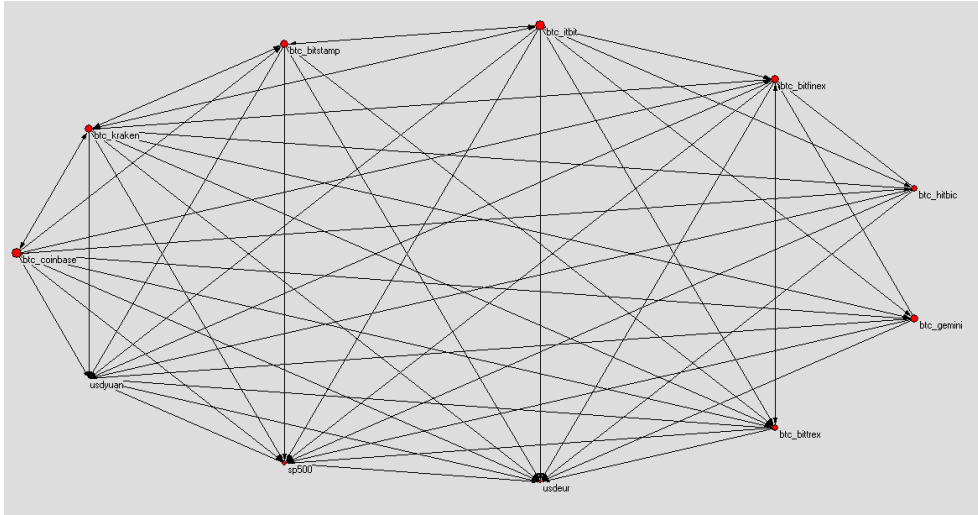


Figure 3.9: **Interconnectedness graph obtained from the application of Granger causality networks.**

Figure 3.9 and 3.8 show that, as expected, the granger causality network graph and the planar maximally filtered graph are more connected than our minimal spanning tree graphs.

They also show that the connections found with the MST are also significant present in the planar maximally filtered graph and in the Granger causality network graph. Upon comparison with Giudici & Abu-Hashish (2019) the same connections are also present in their partial correlation graph.

All these findings lead to the conclusion that the relationships found by our MST graphs are consistently found using other methods and, therefore, the interpretation drawn upon their findings are quite robust.

### 3.5 Conclusion

We have proposed a new statistical model for the explanation of what drives the bitcoin prices. The model is based on the correlation matrix between the observed returns, which is first filtered from noise, applying the random matrix theory and, then, employed to derive a clustering structure among prices, applying the minimal spanning tree.

The main methodological contribution consists of combining the random matrix theory approach with the minimal spanning tree one, with the aim to assess the bitcoin price drivers.

Empirical findings show that bitcoin prices from different exchanges have a strong endogenous driver of variation: they are highly interrelated, as in an

efficiently integrated market, consistently with the literature. In addition, we found that the largest and least volatile exchanges (such as Bitstamp) are the most important price setters.

Our results also confirm the literature in showing that bitcoin prices are unrelated with classical market prices, thus bringing further support to the “diversification benefit” property of crypto assets. In addition, we found that the volatility of classic assets affects negatively, and with a time lag, the volatility of bitcoin prices.

Finally, our empirical findings are robust, with respect to the consideration of different time periods, that also include bubbles, and are consistent with those obtained from different methodologies aimed at measuring interconnectedness between market prices.

We believe that the main beneficiaries of our results may be regulators and supervisors, aimed at preserving financial stability, as well as investors of crypto assets, who should be protected against the negative sides of FinTech innovations (higher risks) while keeping their positive sides (lower costs and better user experience). For a general discussion of this point see also Giudici (2018).

Future work requires acquiring more data, on other bitcoin exchanges, and on other crypto assets, to further assess the validity of the obtained conclusions, and possibly obtain further findings. From a methodological viewpoint, it may be worth considering modeling assets returns with generalized extreme value distributions (as in Calabrese & Giudici (2015)), which can take high volatility into account; or with Bayesian models (as in Figini & Giudici (2011)), which can incorporate expert information into the model. It would also be important to consider the implications of our results in terms of asset allocation.



# Appendix A

	ETF ticker	ETF class
1	IEAG IM	Aggregate Bond
2	XBAG IM	Aggregate Bond
3	XBAE IM	Aggregate Bond
4	GAGG IM	Aggregate Bond
5	DJCOMEX IM	Commodity
6	CRB IM	Commodity
7	XDBC IM	Commodity
8	CRWE IM	Commodity
9	CCUSAS IM	Commodity
10	WCOA IM	Commodity
11	CRBA IM	Commodity
12	CMOD IM	Commodity
13	R1JKEX IM	Corporate-euro
14	ICOV IM	Corporate-euro
15	IEAC IM	Corporate-euro
16	SE15 IM	Corporate-euro
17	XBLC IM	Corporate-euro
18	XB4N IM	Corporate-euro
19	EUCO IM	Corporate-euro
20	ECOEUA IM	Corporate-euro
21	IEXF IM	Corporate-euro
22	ECRP IM	Corporate-euro
23	PSFE IM	Corporate-euro
24	LQDE IM	Corporate-not euro
25	USCO IM	Corporate-not euro
26	LUSC IM	Corporate-not euro

	ETF ticker	ETF class
27	JNKE IM	Corporate-high yield
28	WING IM	Corporate-high yield
29	CORP IM	Corporate-world
30	SAFRI IM	Emerging Equity-Asia
31	AEJ IM	Emerging Equity-Asia
32	INDI IM	Emerging Equity-Asia
33	KOR IM	Emerging Equity-Asia
34	MAL IM	Emerging Equity-Asia
35	TWN IM	Emerging Equity-Asia
36	APEX IM	Emerging Equity-Asia
37	FXC IM	Emerging Equity-Asia
38	IKOR IM	Emerging Equity-Asia
39	ITWN IM	Emerging Equity-Asia
40	IFFF IM	Emerging Equity-Asia
41	XMKO IM	Emerging Equity-Asia
42	XMAS IM	Emerging Equity-Asia
43	XMTW IM	Emerging Equity-Asia
44	XNIF IM	Emerging Equity-Asia
45	XAXJ IM	Emerging Equity-Asia
46	CI2 IM	Emerging Equity-Asia
47	CSKR IM	Emerging Equity-Asia
48	CSEMAS IM	Emerging Equity-Asia
49	XMIN IM	Emerging Equity-Asia
50	XCS3 IM	Emerging Equity-Asia
51	XCS4 IM	Emerging Equity-Asia
52	AASI IM	Emerging Equity-Asia
53	TAI IM	Emerging Equity-Asia
54	EMAE IM	Emerging Equity-Asia
55	XCHA IM	Emerging Equity-Asia
56	AJEUAS IM	Emerging Equity-Asia
57	RQFI IM	Emerging Equity-Asia
58	CASH IM	Emerging Equity-Asia
59	CHNA IM	Emerging Equity-Asia
60	CNAA IM	Emerging Equity-Asia
61	BRA IM	Emerging Equity-America
62	LATAM IM	Emerging Equity-America
63	IBZL IM	Emerging Equity-America
64	LTAM IM	Emerging Equity-America
65	XMLA IM	Emerging Equity-America



	ETF ticker	ETF class
66	XMBR IM	Emerging Equity-America
67	XMEX IM	Emerging Equity-America
68	CSMXCP IM	Emerging Equity-America
69	ALAT IM	Emerging Equity-America
70	BRZ IM	Emerging Equity-America
71	TUR IM	Emerging Equity-East Europe
72	ITKY IM	Emerging Equity-East Europe
73	XMRC IM	Emerging Equity-East Europe
74	RDXS IM	Emerging Equity-East Europe
75	EMKT IM	Emerging Equity-world
76	IEEM IM	Emerging Equity-world
77	XMEM IM	Emerging Equity-world
78	XMEA IM	Emerging Equity-world
79	XSFR IM	Emerging Equity-world
80	SEMA IM	Emerging Equity-world
81	AEEM IM	Emerging Equity-world
82	EMMV IM	Emerging Equity-world
83	EMRG IM	Emerging Equity-world
84	EMGEAS IM	Emerging Equity-world
85	EMMEUA IM	Emerging Equity-world
86	EIMI IM	Emerging Equity-world
87	MXFS IM	Emerging Equity-world
88	EMV IM	Emerging Equity-world
89	MVAM IM	Emerging Equity-world
90	HMEM IM	Emerging Equity-world
91	XMME IM	Emerging Equity-world
92	SXXPIEX IM	Equity-Europe

Table A.1: Summary statistics for ETFs daily returns.



# References

- Anderberg, Michael R. 1973. Cluster analysis for applications. *Probability and Mathematical Statistics*, New York: Academic Press, 1973.
- Andrews, D. 1991. Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica*, **59**(3), 817–858.
- Bai, Zhidong, Liu, Huixia, & Wong, Wing-Keung. 2009. Enhancement of the applicability of Markowitz’s portfolio optimization by utilizing random matrix theory. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, **19**(4), 639–667.
- Baitinger, Eduard, & Papenbrock, Jochen. 2016. Interconnectedness risk and active portfolio management.
- Baitinger, Eduard, & Papenbrock, Jochen. 2017. Interconnectedness Risk and Active Portfolio Management: The Information-Theoretic Perspective.
- Beenakker, Carlo WJ. 1997. Random-matrix theory of quantum transport. *Reviews of modern physics*, **69**(3), 731.
- Billio, Monica, Getmansky, Mila, Lo, Andrew W, & Pelizzon, Lorian. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, **104**(3), 535–559.
- Boginski, Vladimir, Butenko, Sergiy, & Pardalos, Panos M. 2006. Mining market data: a network approach. *Computers & Operations Research*, **33**(11), 3171–3184.
- Boginski, Vladimir, Butenko, Sergiy, Shirokikh, Oleg, Trukhanov, Svyatoslav, & Lafuente, Jaime Gil. 2014. A network-based data mining ap-

- proach to portfolio selection via weighted clique relaxations. *Annals of Operations Research*, **216**(1), 23–34.
- Bonacich, Phillip. 2007. Some unique properties of eigenvector centrality. *Social networks*, **29**(4), 555–564.
- Bonanno, Giovanni, Lillo, Fabrizio, & Mantegna, Rosario N. 2001. High-frequency cross-correlation in a set of stocks.
- Bonanno, Giovanni, Caldarelli, Guido, Lillo, Fabrizio, & Mantegna, Rosario N. 2003. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, **68**(4), 046130.
- Bouri, E., Azzi, G., & Dyhrberg, A.H. 2017. On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics: The Open-Access, Open-Assessment E-Journal*, **11**(2017-2).
- Brandvold, M., Molnr, P., Vagstad, K., Valstad, A., & Ole, C. 2015. Price discovery on Bitcoin exchanges. *Journal of International Financial Markets, Institutions and Money*, **36**(C), 18–35.
- Branke, Jürgen, Scheckenbach, B, Stein, Michael, Deb, Kalyanmoy, & Schmeck, Hartmut. 2009. Portfolio optimization with an envelope-based multi-objective evolutionary algorithm. *European Journal of Operational Research*, **199**(3), 684–693.
- Brin, Sergey, & Page, Lawrence. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, **30**(1-7), 107–117.
- Bryant, David, & Moulton, Vincent. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution*, **21**(2), 255–265.
- Calabrese, R., & Giudici, P. 2015. Estimating bank default with generalised extreme value regression models. *Journal of the Operational Research Society*, 1–10.
- Carney, Mark. 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, Francesco, Scozzari, Andrea, & Tardella, Fabio. 2013. A new method for mean-variance portfolio optimization with cardinality constraints. *Annals of Operations Research*, **205**(1), 213–234.

- Cheah, E.T., & Fry, J. 2015. Speculative bubbles in bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, **130**, 32–36.
- Chi, K Tse, Liu, Jing, & Lau, Francis CM. 2010. A network perspective of the stock market. *Journal of Empirical Finance*, **17**(4), 659–667.
- Christian, Conrad, Custovic, Anessa, & Ghysels, Eric. 2018. Long- and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, **11**(2), 23.
- Ciaian, P., & Rajcaniova, M. 2018. Virtual relationships: Short-and long-run evidence from bitcoin and altcoin markets. *Journal of International Financial Markets, Institutions and Money*, **52**, 173–195.
- Clemente, Gian Paolo, Grassi, Rosanna, & Hitaj, Asmerilda. 2019. Asset allocation: new evidence through network approaches. *Annals of Operations Research*, 1–20.
- Conlon, Thomas, Ruskin, Heather J, & Crane, Martin. 2007. Random matrix theory and fund of funds portfolio optimisation. *Physica A: Statistical Mechanics and its applications*, **382**(2), 565–576.
- Corbet, Shaen, Meegan, Andrew, Larkin, Charles, Lucey, Brian, & Yarovaya, Larisa. 2018. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, **165**, 28–34.
- Corbet, Shaen, Lucey, Brian, Urquhart, Andrew, & Yarovaya, Larisa. 2019. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, **62**, 182–199.
- Coronnello, C, Tumminello, M, Lillo, F, Micciche, S, & Mantegna, ROSARIO NUNZIO. 2005. Sector identification in a set of stock return time series traded at the London Stock Exchange. *arXiv preprint cond-mat/0508122*.
- Crama, Yves, & Schyns, Michaël. 2003. Simulated annealing for complex portfolio selection problems. *European Journal of operational research*, **150**(3), 546–571.
- de Prado, Marcos López. 2016. Building Diversified Portfolios that Outperform Out of Sample. *Journal of Portfolio Management*, **42**(4), 59.

- Doerner, Karl, Gutjahr, Walter J, Hartl, Richard F, Strauss, Christine, & Stummer, Christian. 2004. Pareto ant colony optimization: A metaheuristic approach to multiobjective portfolio selection. *Annals of operations research*, **131**(1-4), 79–99.
- Durante, Fabrizio, Pappadà, Roberta, & Torelli, Nicola. 2014. Clustering of financial time series in risky scenarios. *Advances in Data Analysis and Classification*, **8**(4), 359–376.
- Durante, Fabrizio, Pappadà, Roberta, & Torelli, Nicola. 2015. Clustering of time series via non-parametric tail dependence estimation. *Statistical Papers*, **56**(3), 701–721.
- Dwyer, G.P. 2015. The economics of Bitcoin and similar private digital currencies. *Journal of Financial Stability*, **17**(apr), 81–91.
- Dyhrberg, A.H. 2016. Bitcoin, gold and the dollar: A GARCH volatility analysis. *Finance Research Letters*, **16**, 85 – 92.
- Ehrgott, Matthias, Klamroth, Kathrin, & Schwehm, Christian. 2004. An MCDM approach to portfolio optimization. *European Journal of Operational Research*, **155**(3), 752–770.
- Eom, Cheoljun, Oh, Gabjin, Jung, Woo-Sung, Jeong, Hawoong, & Kim, Seunghwan. 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.
- Epps, Thomas W. 1979. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, **74**(366a), 291–298.
- Figini, S., & Giudici, P. 2011. Statistical merging of rating models. *Journal of the Operational Research Society*, **62**, 1067–1074.
- Fortnum, Dennis, Pollari, I, Mead, W, Hughes, B, & Speier, A. 2017. The Pulse of Fintech Q4 2016: Global analysis of investment in fintech. *Amstelveen: KPMG*.
- Friedman, Jerome, Hastie, Trevor, & Tibshirani, Robert. 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)*.
- Fund, IM. *Fintech and Financial Services: Initial Considerations*.
- Galluccio, Stefano, Bouchaud, Jean-Philippe, & Potters, Marc. 1998. Rational decisions, random matrices and spin glasses. *Physica A: Statistical Mechanics and its Applications*, **259**(3-4), 449–456.
- Gera, Imre, Bánhelyi, Balázs, & London, András. 2016. Testing the markowitz portfolio optimization method with filtered correlation matrices. *Pages 44–47 of: Proceedings of the Middle-European Conference on Applied Theoretical Computer Science*.
- Giudici, P. 2018. Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance. *Frontiers in Artificial Intelligence*, **1**.
- Giudici, Paolo. 2003. *Applied data mining*.
- Giudici, Paolo, & Abu-Hashish, Iman. 2019. What determines bitcoin exchange prices? a network VAR approach. *Finance Research Letters*, **28**, 309–318.
- Gonzalo, J., & Granger, C. 1995. Estimation of Common Long-Memory Components in Cointegrated Systems. *Journal of Business & Economic Statistics*, **13**(1), 27–35.
- Gregoriou, Greg N, & Nian, Lam Pak. 2015. Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data. *The Journal of Wealth Management*, **18**(2), 96.
- Grinold, R., & Kahn, R. 1999. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk*. McGraw-Hill Library of Investment and Finance. McGraw-hill.
- Guhr, Thomas, Müller-Groeling, Axel, & Weidenmüller, Hans A. 1998. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, **299**(4-6), 189–425.
- Hafner, C.M. 2018. Testing for bubbles in cryptocurrencies with time-varying volatility. *Journal of Financial Econometrics*, **to appear**.

- Härdle, Wolfgang Karl, Lee, David Kuo Chuen, Nasekin, Sergey, & Petukhina, Alla. 2018. Tail Event Driven ASset allocation: evidence from equity and mutual funds markets. *Journal of Asset Management*, **19**(1), 49–63.
- Hartman, David, & Hlinka, Jaroslav. 2018. Nonlinearity in stock networks. *arXiv preprint arXiv:1804.10264*.
- Hasbrouck, J. 1995. One Security, Many Markets: Determining the Contributions to Price Discovery. *The Journal of Finance*, **50**(4), 1175–1199.
- He, Xue Dong, & Zhou, Xun Yu. 2011. Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science*, **57**(2), 315–331.
- Houben, Robby, & Snyers, Alexander. 2018. Cryptocurrencies and blockchain: Legal context and implications for financial crime, money laundering and tax evasion.
- Katz, Leo. 1953. A new status index derived from sociometric analysis. *Psychometrika*, **18**(1), 39–43.
- Kaufman, Leonard, & Rousseeuw, Peter J. 2009. *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons.
- Kim, H, Maurer, Raimond, & Mitchell, Olivia S. 2016. When and How to Delegate? A Life Cycle Analysis of Financial Advice.
- Kleinberg, Jon M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, **46**(5), 604–632.
- Kruskal, Joseph B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, **7**(1), 48–50.
- Laloux, Laurent, Cizeau, Pierre, Bouchaud, Jean-Philippe, & Potters, Marc. 1999. Noise dressing of financial correlation matrices. *Physical review letters*, **83**(7), 1467.
- León, Diego, Aragón, Arbey, Sandoval, Javier, Hernández, Germán, Arévalo, Andrés, & Nino, Jaime. 2017. Clustering algorithms for Risk-Adjusted Portfolio Construction. *Procedia Computer Science*, **108**, 1334–1343.



- Lhabitant, François-Serge, & Learned, Michelle. 2002. Hedge fund diversification: How much is enough? *The Journal of Alternative Investments*, **5**(3), 23–49.
- Lopez de Prado, Marcos. 2016. Building diversified portfolios that outperform out-of-sample.
- Malkiel, Burton G, & Fama, Eugene F. 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, **25**(2), 383–417.
- Mantegna, Rosario N. 1999. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, **11**(1), 193–197.
- Mantegna, Rosario N, & Stanley, H Eugene. 1999. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press.
- Marchenko, Vladimir Alexandrovich, & Pastur, Leonid Andreevich. 1967. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, **114**(4), 507–536.
- Markowitz, Harry. 1952. Portfolio selection. *The journal of finance*, **7**(1), 77–91.
- Marti, Gautier, Nielsen, Frank, Bińkowski, Mikołaj, & Donnat, Philippe. 2017. A review of two decades of correlations, hierarchies, networks and clustering in financial markets. *arXiv preprint arXiv:1703.00485*.
- Massara, Guido Previde, Di Matteo, Tiziana, & Aste, Tomaso. 2016. Network filtering for big data: Triangulated maximally filtered graph. *Journal of complex Networks*, **5**(2), 161–178.
- Matesanz, David, & Ortega, Guillermo J. 2015. Sovereign public debt crisis in Europe. A network analysis. *Physica A: Statistical Mechanics and its Applications*, **436**, 756–766.
- Mehta, Madan Lal. 1990. Random matrices.
- Mehta, Madan Lal. 2004. *Random matrices*. Vol. 142. Elsevier.
- Metaxiotis, Konstantinos, & Liagkouras, Konstantinos. 2012. Multiobjective evolutionary algorithms for portfolio management: A comprehensive literature review. *Expert Systems with Applications*, **39**(14), 11685–11698.

- Miccichè, Salvatore, Bonanno, Giovanni, Lillo, Fabrizio, & Mantegna, Rosario N. 2003. Degree stability of a minimum spanning tree of price return and volatility. *Physica A: Statistical Mechanics and its Applications*, **324**(1-2), 66–73.
- Miceli, Maria-Augusta, & Susinno, Gabriele. 2004. Ultrametricity in fund of funds diversification. *Physica A: Statistical Mechanics and its Applications*, **344**(1-2), 95–99.
- Moret, Bernard ME, & Shapiro, Henry D. 1991. An empirical analysis of algorithms for constructing a minimum spanning tree. *Pages 400–411 of: Workshop on Algorithms and Data Structures*. Springer.
- Münnix, Michael C, Schäfer, Rudi, & Guhr, Thomas. 2010. Impact of the tick-size on financial returns and correlations. *Physica A: Statistical Mechanics and its Applications*, **389**(21), 4828–4843.
- Musmeci, Nicolo, Aste, Tomaso, & Di Matteo, Tiziana. 2015. Relation between financial market structure and the real economy: comparison between clustering methods. *PloS one*, **10**(3), e0116201.
- Nadarajah, S., & Chu, J. 2017. On the inefficiency of Bitcoin. *Economics Letters*, **150**(6-9).
- Nakamoto, Satoshi. 2008. Bitcoin: A peer-to-peer electronic cash system. [www.bitcoin.org](http://www.bitcoin.org).
- Newman, M. 2018. *Networks*. OUP Oxford.
- Onnela, J-P, Chakraborti, Anirban, Kaski, Kimmo, & Kertesz, Janos. 2003a. Dynamic asset trees and Black Monday. *Physica A: Statistical Mechanics and its Applications*, **324**(1-2), 247–252.
- Onnela, J-P, Chakraborti, Anirban, Kaski, Kimmo, Kertesz, Janos, & Kanto, Antti. 2003b. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, **68**(5), 056110.
- Onnela, J-P, Kaski, Kimmo, & Kertész, Janos. 2004. Clustering and information in correlation based financial networks. *The European Physical Journal B*, **38**(2), 353–362.
- Pagnottoni, P., Dirk, B., & Dimpfl, T. 2018. Price Discovery on Bitcoin Markets. *Working Paper*.

- Pantaleo, Ester, Tumminello, Michele, Lillo, Fabrizio, & Mantegna, Rosario N. 2011. When do improved covariance matrix estimators enhance portfolio optimization? An empirical comparative study of nine estimators. *Quantitative Finance*, **11**(7), 1067–1080.
- Pattillo, Jeffrey, Youssef, Nataly, & Butenko, Sergiy. 2013. On clique relaxation models in network analysis. *European Journal of Operational Research*, **226**(1), 9–18.
- Peralta, Gustavo, & Zareei, Abalfazl. 2016. A network approach to portfolio selection. *Journal of Empirical Finance*, **38**, 157–180.
- Perra, Nicola, & Fortunato, Santo. 2008. Spectral centrality measures in complex networks. *Physical Review E*, **78**(3), 036107.
- Plerou, Vasiliki, Gopikrishnan, Parameswaran, Rosenow, Bernd, Amaral, Luís A Nunes, & Stanley, H Eugene. 1999. Universal and nonuniversal properties of cross correlations in financial time series. *Physical review letters*, **83**(7), 1471.
- Plerou, Vasiliki, Gopikrishnan, Parameswaran, Rosenow, Bernd, Amaral, Luis A Nunes, Guhr, Thomas, & Stanley, H Eugene. 2002. Random matrix approach to cross correlations in financial data. *Physical Review E*, **65**(6), 066126.
- Potters, Marc, Bouchaud, Jean-Philippe, & Laloux, Laurent. 2005. Financial applications of random matrix theory: Old laces and new pieces. *arXiv preprint physics/0507111*.
- Pozzi, Francesco, Di Matteo, Tiziana, & Aste, Tomaso. 2013. Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, **3**, 1665.
- Prim, Robert Clay. 1957. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, **36**(6), 1389–1401.
- Raffinot, Thomas. 2017. Hierarchical Clustering-Based Asset Allocation. *The Journal of Portfolio Management*, **44**(2), 89–99.
- Rea, Alethea, & Rea, William. 2014. Visualization of a stock market correlation matrix. *Physica A: Statistical Mechanics and its Applications*, **400**, 109–123.

- Ren, Fei, Lu, Ya-Nan, Li, Sai-Ping, Jiang, Xiong-Fei, Zhong, Li-Xin, & Qiu, Tian. 2017. Dynamic portfolio strategy using clustering approach. *PloS one*, **12**(1), e0169299.
- Rockafellar, R Tyrrell, Uryasev, Stanislav, *et al.* 2000. Optimization of conditional value-at-risk. *Journal of risk*, **2**, 21–42.
- Schaerf, Andrea. 2002. Local search techniques for constrained portfolio selection problems. *Computational Economics*, **20**(3), 177–190.
- Sharpe, William F. 1994. The sharpe ratio. *Journal of portfolio management*, **21**(1), 49–58.
- Shoaf, Jaqueline S, & Foster, James A. 1996. A Genetic Algorithm Solution to the Efficient Set Problem: A Technique for Portfolio Selection Based on the Markowitz Model.”. *Pages 571–573 of: Proceedings of the Decision Sciences Institute Annual Meeting.*
- Sokal, Robert R, & Rohlf, F James. 1962. The comparison of dendrograms by objective methods. *Taxon*, **11**(2), 33–40.
- Spelta, Alessandro, & Araújo, Tanya. 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, Vincenzo, Lillo, Fabrizio, Gallegati, Mauro, & Mantegna, Rosario N. 2008. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, **32**(1), 235–258.
- Tulino, Antonia M, Verdú, Sergio, *et al.* 2004. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, **1**(1), 1–182.
- Tumminello, Michele, Aste, Tomaso, Di Matteo, Tiziana, & Mantegna, Rosario 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, Michele, Di Matteo, Tiziana, Aste, Tomaso, & Mantegna, Rosario N. 2007a. Correlation based networks of equity returns sampled at different time horizons. *The European Physical Journal B*, **55**(2), 209–217.

- Tumminello, Michele, Coronello, Claudia, Lillo, Fabrizio, Micciche, Salvatore, & Mantegna, Rosario N. 2007b. Spanning trees and bootstrap reliability estimation in correlation-based networks. *International Journal of Bifurcation and Chaos*, **17**(07), 2319–2329.
- Tumminello, Michele, Lillo, Fabrizio, & Mantegna, Rosario N. 2010. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, **75**(1), 40–58.
- Tversky, Amos, & Kahneman, Daniel. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, **5**(4), 297–323.
- Urquhart, A. 2016. The inefficiency of Bitcoin. *Economic Letters*, **148**(80–82).
- Výrost, Tomas, Lyócsa, Štefan, & Baumöhl, Eduard. 2018. Network-based asset allocation strategies. *The North American Journal of Economics and Finance*.
- Woodside-Oriakhi, Maria, Lucas, C, & Beasley, John E. 2011. Heuristic algorithms for the cardinality constrained efficient frontier. *European Journal of Operational Research*, **213**(3), 538–550.
- Wu, Jianshe, Li, Xiaoxiao, Jiao, Licheng, Wang, Xiaohua, & Sun, Bo. 2013. Minimum spanning trees for community detection. *Physica A: Statistical Mechanics and its Applications*, **392**(9), 2265–2277.
- Zhan, Hannah Cheng Juan, Rea, William, & Rea, Alethea. 2015. An Application of Correlation Clustering to Portfolio Diversification. *arXiv preprint arXiv:1511.07945*.
- Zhang, Xiaohang, Liu, Jiaqi, Du, Yu, & Lv, Tingjie. 2011. A novel clustering method on time series data. *Expert Systems with Applications*, **38**(9), 11891–11900.