



# Studying the correlation structure based on market geometry

Chun-Xiao Nie<sup>1</sup>

Received: 20 January 2019 / Accepted: 10 October 2020 / Published online: 27 October 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Network methods can extract the structure of financial correlation matrices, and market geometry reconstructs the correlation relationship by constructing a vector set in the Euclidean space. This study uses a geometric perspective to analyse financial networks and examine the relationship between correlation structures and geometric conditions. Based on the concept of Euclidean space, we can naturally define geometric concepts such as stock vector and inner product between stocks. The analysis reveals that the structure of the financial correlation network is significantly affected by geometric conditions. We use stock market data to construct networks with different structures, such as a network with a hub node. We find that some stocks with small vector norms have an important effect on changes in network structure. In addition, we define a dimension to describe the correlation information included in the subspace of the market space and find that the dynamics of the dimension are related to the market state. This paper establishes a way to study network structure through market geometry, thereby providing a new method of correlation analysis.

**Keywords** Correlation matrix · Complex network · Stock norm · Financial crisis

## 1 Introduction

To study large-dimensional correlation matrices, how to extract the hidden structure of the matrix is an important issue. Random matrix theory (RMT) and complex network theory provide some effective methods (Laloux et al. 1999; Plerou et al. 1999; Mantegna 1999; Tumminello et al. 2005; Yang and Yang 2008; Tse et al. 2010). RMT-based analysis reveals that the financial correlation matrix includes patterns different from theoretical predictions (Laloux et al. 1999; Plerou et al. 1999). Theoretical and empirical analysis supports the widespread use of RMT in financial data analysis, such

---

✉ Chun-Xiao Nie  
niechunxiao2009@163.com; niechunxiao@mail.zjgsu.edu.cn

<sup>1</sup> School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China

as constructing portfolios (Plerou et al. 2002; Joël et al. 2017). Furthermore, empirical analysis has shown that the dynamics of the correlation matrix are helpful for understanding the overall pattern of the market, such as market status (Thomas et al. 2009; Ahmet et al. 2013; Junior and Franca 2012; Zheng et al. 2012). In particular, analyzing the relationship between the dynamics of correlation structure and crises, such as constructing indicators of the financial crisis based on correlations (Ahmet et al. 2013; Junior and Franca 2012; Zheng et al. 2012), is an important issue.

In addition to directly studying the spectrum or properties of the correlation matrix, an important method used in recent years is converting the correlation matrix into a network and examining its structure (Mantegna 1999; Tumminello et al. 2005; Yang and Yang 2008; Tse et al. 2010). Consistent with the concepts used in previous studies, a network of this type is called a correlation-based network (Tumminello et al. 2007). A correlation network can be constructed in various ways, including through several common methods. One of them is to construct a financial correlation-based network using a minimum spanning tree algorithm (MST) (Mantegna 1999). MST has no closed loop and can be used to extract the hierarchical structure in the market (Mantegna 1999). In general, researchers have proposed a method for constructing a planar maximally filtered graph (PMFG) that includes some small subgraph structures, such as 3- and 4-cliques (Tumminello et al. 2005). In particular, research suggests that PMFGs in the market may include some communities, a type of structure that has been extensively studied in complex network research (Fortunato 2010; Malliaros and Vazirgiannis 2013).

Existing studies have deeply analysed various structures of MST and PMFG, such as the degree distribution (Antonios and Argyrakis 2007; Wiliński et al. 2013; Nie et al. 2016) and the community structure (Zhao et al. 2016; Vodenska et al. 2016). Various markets have been extensively studied by converting the time series into financial networks. For example, industry indices (Giuseppe et al. 2013), international indices (Kumar and Deo 2012), exchange rates (Matesanz and Ortega 2014), etc., have been systematically analysed. By examining the structure extracted by the network, we can analyse the correlation matrix in depth, which enables us to discuss the topics of risk management and investment portfolios. For example, Pozzi et al. (2013) analysed the risk and investment performance of core and peripheral nodes in the correlation network and found that portfolios based on peripheral nodes performed better. Empirical analysis has also revealed that network topology can be used to predict the returns of asset portfolios (Eng-Uthaiwat 2018). Peralta and Zareei (2016) constructed an analysis framework combining the Markowitz theory with a network representation and found that the network structure helps enhance the performance of asset portfolios. Sandhu et al. (2016) introduced the curvature of the network and used it to analyse systemic risks. Moreover, empirical analysis has shown that a financial crisis significantly affects the fractal structure of the correlation network (Nie and Song 2019).

In addition to methods based on graph theory, the threshold method can be used to convert the correlation matrix into a network (Yang and Yang 2008; Tse et al. 2010). Researchers have combined PMFG and threshold methods to construct networks and study changes in the correlation structure of financial markets (Nie and Song 2018). The current study mainly examines the structure of MST and PMFG from a geometric perspective.

In the study of financial MST, the correlation coefficient can be converted into the distance between two time series (Mantegna 1999), which allows the correlation matrix to be analysed from a geometric perspective (Mendes et al. 2003; Araújo and Louçã 2007; Araújo et al. 2013; Araújo and Spelta 2014; Eleutério et al. 2014; Echaust and Just 2013). By defining the distance in the stock set, it can be embedded in a Euclidean space, where each stock corresponds to a Euclidean vector (Mendes et al. 2003). It should be noted that the dimension of these vectors is not the length of the return series, which is usually of a lower dimension.

Mendes et al. (2003) used the stochastic geometry technique to embed  $N$  stocks in an  $N - 1$ -dimensional Euclidean space. Furthermore, some studies have extended the application of this method. Based on the data of the constituent stocks of the S&P500 index, Araújo and Louçã investigated changes in the market space and introduced an index to analyse nine financial crises (Araújo and Louçã 2007). Araújo et al. (2013) applied multivariate kurtosis to characterize the distortion of market geometry and found that it can be used to portray the emergence of crises. In addition, the researchers analysed the international interbank market and found that the data can be represented well in the low-dimensional space (Araújo and Spelta 2014). From an application perspective, market geometry can be used for portfolio analysis (Eleutério et al. 2014). In addition to the aforementioned technique, some researchers use multidimensional scaling (MDS) analysis to examine market geometry (Echaust and Just 2013). MDS is a commonly used technique for analyzing financial data, which can be implemented through different definitions of similarity or distance (Groenen and Franses 2000; Machado et al. 2011; Yin and Shang 2014; Esmalifalak et al. 2015; Fernández-Avilés et al. 2020). In this study, we apply the MDS technique to construct the market geometry, in which the stock set with the distance is displayed in a Euclidean space.

In summary, the correlation coefficient matrix can be filtered into a network or directly embedded in a Euclidean space. However, to the best of our knowledge, no research has focused on the possible relationship between the two approaches. This study focuses on combining the two methods to analyse the structures in the correlation matrix. To characterize the network structures, we use the influence-strength (IS) and Rényi index to identify the influence of the nodes and the macrostructure of the network, respectively (Jung et al. 2006; Eliazar 2011). The IS extracts the neighbouring information of a node to portray its influence on other nodes; this approach has been widely used in financial network analysis (Jung et al. 2006; Gałazka 2011; Wang and Xie 2015; Wang et al. 2018; Zhu et al. 2016; Nie 2020). The Rényi index, also known as the normalized Rényi entropy, was originally used to examine the heterogeneity of the distribution (Eliazar 2011). The Rényi index of the network is defined in the degree distribution and has been used to study correlation-based networks (Nie et al. 2016; Nie and Song 2019).

In addition to microscopic (IS) and macroscopic structural indicators (Rényi index), we use an algorithm to detect an important mesoscale structure (community) of the network. Many researchers have proposed community detection algorithms based on different methods (Fortunato 2010; Malliaros and Vazirgiannis 2013). Here, we choose a classic community detection algorithm proposed by Newman, which is calculated by optimizing the modularity (Newman 2004). Modularity is an indicator used to

characterize the saliency of a community structure, and a value greater than 0.3 means that the structure is significant (Clauset et al. 2004).

The rest of this paper is organized as follows. We first review the MDS algorithm and the indicators used to examine the correlation network. Second, we define some basic concepts of market geometry based on Euclidean space. Third, we analyse the relationship between the geometric concept and the network structure. Fourth, we construct networks with different geometric conditions and analyse the differences in topological structures. In addition, we examine the relationship between norms and market factor. Finally, we analyse the dynamics of the market subspace and dimensions.

## 2 Data and method

### 2.1 Data

For the empirical analysis, we selected the constituents of the indices in the US and UK markets, from which the daily closing price series for each selected stock was extracted. Some stocks were removed due to missing data in the considered period. For the S&P500 index, we selected price data for 448 stocks from 3/1/2007 to 31/12/2010 (Table 6). To show the network clearly, we selected the constituent stocks of the S&P100 index (Table 7), which are also included in the S&P500 index, for analysis.

To investigate the long-term dynamics of the correlation matrix, we selected 78 stocks from the S&P500 index from 2/1/1985 to 31/12/2012 (Table 8). Moreover, we selected the daily closing price series of the 80 constituent stocks of the FTSE 100 Index, each of which ranges from 3/1/2005 to 29/12/2017 (Table 9).

All data sets analysed in this study are extracted from Yahoo! Finance (<https://finance.yahoo.com/>). The constituent stocks included in each dataset are listed in the “Appendix”. In addition, we use software Pajek to visualize the correlation-based networks of this paper (<http://mrvar.fdv.uni-lj.si/pajek/>).

### 2.2 Reconstructing the market in a Euclidean space

We use the MDS algorithm to reconstruct the market in a Euclidean space (Borg and Groenen 2005). We discuss the geometry of  $n$  stocks ( $S = \{i | i = 1, \dots, n\}$ ), each of which corresponds to a price series ( $\{P_i(t)\}$ ). The return series  $\{R_i(t)\}$  is generated by equation  $R_i(t) = \log(P_i(t+1)) - \log(P_i(t))$ . We then calculate the correlation coefficient between each pair of stocks (Eq. (1)) as well as the distance (Eq. (2)) (Mantegna 1999). In Eq. (1), the symbol  $\langle \rangle$  indicates a calculation of the average. The MDS algorithm uses the distance matrix  $D = [D(i, j)]$  of the  $n$  stocks (Borg and Groenen 2005; Echaust and Just 2013).

$$\rho(i, j) = \frac{\langle R_i(t)R_j(t) \rangle - \langle R_i(t) \rangle \langle R_j(t) \rangle}{\sqrt{\langle R_i(t)^2 - \langle R_i(t) \rangle^2 \rangle \langle R_j(t)^2 - \langle R_j(t) \rangle^2 \rangle}} \quad (1)$$

$$D(i, j) = \sqrt{2(1 - \rho(i, j))} \quad (2)$$

Here, we use the following classic algorithm to reconstruct the market (Borg and Groenen 2005; Echaust and Just 2013).

1. We calculate the matrix  $J = I - n^{-1}I_1I_1'$ , where  $I_1$  is the column vector with elements equal to 1.
2. Based on  $J$ , the matrix  $B$  is calculated as shown in Eq. (3). Here,  $D_2 = [D_2(i, j)]$  is constructed by  $D$ , where  $D_2(i, j) = D(i, j)^2$ .
3. We need to calculate the eigenvalue decomposition  $B = Q\Lambda Q'$ . We calculate  $n$  eigenvalues  $\{\lambda_i, i = 1, \dots, n\}$  of  $B$ , and each eigenvalue  $\lambda_i$  corresponds to an eigenvector  $v_i^B$ . In addition, we assume that the eigenvalues are sorted in the descending order of the subscripts, that is,  $\lambda_i > \lambda_{i+1}$ . Here, the  $i$ -th column of the matrix  $Q$  corresponds to the  $i$ -th eigenvector  $v_i^B$  of  $B$ , and  $\Lambda$  is a diagonal matrix ( $\Lambda(i, i) = \lambda_i$ ).
4. We extract the top  $m$  eigenvalues ( $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ ) greater than 0, and construct matrix  $\Lambda_+$  ( $\Lambda_+(i, i) \in \lambda$ ). Correspondingly, the first  $m$  columns of  $Q$  are recorded as  $Q_+$ ; then, the coordinate matrix is given by  $X = Q_+\Lambda_+^{1/2}$ , where  $\Lambda_+^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2})$ .

$$B = -\frac{1}{2}JD_2J' \tag{3}$$

Since the distance between two stocks is equivalent to the Euclidean distance between the normalized return series,  $m$  can be chosen to be  $n - 1$  (Mendes et al. 2003; Araújo and Louçã 2007; Echaust and Just 2013). That is, each stock is assigned an  $n - 1$  dimensional vector so that the stock set is embedded in the Euclidean space  $R^{n-1}$ .

### 2.2.1 An example based on MDS

Here, we use a simple example to show how the MDS algorithm is applied to embed a stock set into a Euclidean space. We select the daily closing price series of five companies (APPL, ABT, ACN, AIG, ALL) from 4/1/2010 to 31/12/2010 (252 days).

First, we preprocess all price series into return series and generate a correlation matrix (Eq. (4)).

$$D = \begin{bmatrix} 0 & 1.0487 & 1.0687 & 1.1693 & 0.9982 \\ 1.0487 & 0 & 1.0490 & 1.1503 & 1.0195 \\ 1.0687 & 1.0490 & 0 & 1.1718 & 0.9812 \\ 1.1693 & 1.1503 & 1.1718 & 0 & 1.0570 \\ 0.9982 & 1.0195 & 0.9812 & 1.0570 & 0 \end{bmatrix} \tag{4}$$

Second, we calculate  $J$  (Eq. (5)) and  $B$  (Eq. (6)). The five eigenvalues of  $B$  are  $\lambda = \{0.7381, 0.5725, 0.5509, 0.4429, 0\}$ , and the eigenvectors corresponding to the four nonzero eigenvalues are as shown in Eq. (7), where each column corresponds to one eigenvalue.

$$J = \begin{bmatrix} 0.8000 & -0.2000 & -0.2000 & -0.2000 & -0.2000 \\ -0.2000 & 0.8000 & -0.2000 & -0.2000 & -0.2000 \\ -0.2000 & -0.2000 & 0.8000 & -0.2000 & -0.2000 \\ -0.2000 & -0.2000 & -0.2000 & 0.8000 & -0.2000 \\ -0.2000 & -0.2000 & -0.2000 & -0.2000 & 0.8000 \end{bmatrix} \tag{5}$$

$$B = \begin{bmatrix} 0.4602 & -0.0939 & -0.1136 & -0.1658 & -0.0869 \\ -0.0939 & 0.4517 & -0.0970 & -0.1481 & -0.1127 \\ -0.1136 & -0.0970 & 0.4548 & -0.1715 & -0.0728 \\ -0.1658 & -0.1481 & -0.1715 & 0.5753 & -0.0899 \\ -0.0869 & -0.1127 & -0.0728 & -0.0899 & 0.3623 \end{bmatrix} \tag{6}$$

$$Q = \begin{bmatrix} 0.3044 & -0.6781 & 0.4192 & 0.2680 & 0.4472 \\ 0.2384 & -0.1343 & -0.8512 & -0.0255 & 0.4472 \\ 0.3193 & 0.7141 & 0.2002 & 0.3847 & 0.4472 \\ -0.8652 & -0.0116 & -0.0122 & 0.2262 & 0.4472 \\ 0.0031 & 0.1099 & 0.2440 & -0.8534 & 0.4472 \end{bmatrix} \tag{7}$$

Finally, we calculate  $X = Q_+ \Lambda_+^{1/2}$  ( $m = 5 - 1 = 4$ ), and each row in Eq. (8) corresponds to a vector of length 4, which is the vector in the Euclidean space  $R^4$ . For example, APPL corresponds to vector (0.2615, -0.5131, 0.3112, 0.1783).

$$X = \begin{bmatrix} 0.2615 & -0.5131 & 0.3112 & 0.1783 \\ 0.2048 & -0.1016 & -0.6318 & -0.0169 \\ 0.2743 & 0.5403 & 0.1486 & 0.2560 \\ -0.7433 & -0.0088 & -0.0091 & 0.1505 \\ 0.0027 & 0.0832 & 0.1811 & -0.5679 \end{bmatrix} \tag{8}$$

### 2.3 The geometry of the stock space

Based on the MDS algorithm, the stock set is embedded in a high-dimensional Euclidean space. Each stock  $i$  corresponds to a vector  $v_i = (v_i(1), \dots, v_i(n - 1))$  in the Euclidean space  $R^{n-1}$ . This approach allows us to apply geometric concepts to analyse the structures in the correlation matrix (Halmos 1974).

First, the concepts in the Euclidean space can be naturally defined in the market geometry. Since each stock corresponds to a vector in the Euclidean space, we can naturally define the inner product between any two stock vectors as

$$(v_i, v_j) = \sum_l v_i(l)v_j(l). \tag{9}$$

We define the norm of the vector  $v_i$  as the stock norm of  $i$  (Eq. (10)).

$$N_i = \text{norm}(v_i) = \sqrt{(v_i, v_i)} \tag{10}$$

We define  $\theta(i, j)$  as the angle between stocks  $i$  and  $j$ , which can be expressed as Eq. (11).

$$\cos \theta(i, j) = \frac{(v_i, v_j)}{N_i N_j} \tag{11}$$

The distance between two stocks can be expressed by

$$D_{\text{MDS}}(i, j) = \text{norm}(v_i - v_j). \tag{12}$$

In Eq. (2),  $D(i, j)$  is essentially the Euclidean distance between the normalized return series. For example, if we assume that both time series are of length  $K$ , then the distance of Eq. (2) is equivalent to the Euclidean distance between the two  $K$ -dimensional vectors (Mantegna and Stanley 1999). Here, the distance between the reconstructed vectors shown in Eq. (12) is equal to  $D(i, j)$  of Eq. (2). However,  $v_i$  and  $v_j$  in Eq. (12) are vectors in the Euclidean space  $R^{n-1}$ , not  $K$ -dimensional vectors. Since the calculated values of Eqs. (2) and (12) are identical, we will not distinguish between  $D_{\text{MDS}}(i, j)$  and  $D(i, j)$ .

Below, we analyse the relationship between the correlation coefficient and the norm. First, based on Eq. (2), the correlation coefficient can be expressed as

$$\rho(i, j) = 1 - \frac{D(i, j)^2}{2}. \tag{13}$$

Second, from Eqs. (10) and (13), the correlation coefficient can be expressed as

$$\rho(i, j) = 1 - \left( \frac{N_i^2 + N_j^2}{2} - N_i N_j \cos \theta(i, j) \right). \tag{14}$$

We assume that  $\theta(i, j)$  is close to  $\pi/2$  and can be expressed as  $\theta(i, j) = \pi/2 - \theta_\epsilon(i, j)$ . The correlation coefficient can be re-expressed as in Eq. (15). For the term  $\cos(\frac{\pi}{2} - \theta_\epsilon(i, j))$ , we can express it as Taylor expansion, which is Eq. (16). When  $\theta_\epsilon(i, j)$  is close to 0, the correlation coefficient can be approximated as  $\rho(i, j) \approx 1 - \frac{N_i^2 + N_j^2}{2} + N_i N_j \theta_\epsilon(i, j)$ .

$$\rho(i, j) = 1 - \frac{N_i^2 + N_j^2}{2} + N_i N_j \left( \cos \left( \frac{\pi}{2} - \theta_\epsilon(i, j) \right) \right) \tag{15}$$

$$\begin{aligned} \rho(i, j) \approx & 1 - \frac{N_i^2 + N_j^2}{2} + N_i N_j \left( \theta_\epsilon(i, j) - \frac{\theta_\epsilon(i, j)^3}{6} \right. \\ & \left. + \frac{\theta_\epsilon(i, j)^5}{120} - \frac{\theta_\epsilon(i, j)^7}{5040} + O(\theta_\epsilon(i, j)^8) \right) \end{aligned} \tag{16}$$

## 2.4 Low-dimensional subspace of the stock space

Previous studies have introduced an indicator to capture the effective dimension of the market, where the effective dimension is much smaller than  $n$  (Araújo and Louçã 2007). Here, we propose a new dimension to characterize the information contained in the subspace of the financial space.

First, we calculate the Frobenius norm of a matrix  $H = [H(i, j)]$ , as shown in Eq. (17) (Leon 2010).

$$\|H\|_F = \sqrt{\sum_i \sum_j |H(i, j)|^2} \quad (17)$$

Second, by specifying a dimension  $\text{Dim}$ , we calculate the vector  $v_i^{\text{sub}} = (v_i(1), \dots, v_i(\text{Dim}))$  of each stock in the subspace and the distance  $D_{\text{Dim}}(i, j) = \text{norm}(v_i^{\text{sub}} - v_j^{\text{sub}})$  between the different stocks in the subspace. We then define  $p'$  as shown in Eq. (18), which depicts the difference between  $D_{\text{Dim}}$  and the original distance matrix  $D$ . That is, a larger  $p'$  means that the difference between the matrix  $D_{\text{Dim}}$  and  $D$  is smaller and that  $D_{\text{Dim}}$  is related to the dimension  $\text{Dim}$ . Thus,  $p'$  essentially depicts the level of correlation information contained in a subspace, and a larger  $p'$  value indicates that more correlation information is included in the subspace.

$$p' = 1 - \frac{\|D_{\text{Dim}} - D\|_F}{\|D\|_F} \quad (18)$$

Further, if we specify a  $p'$  value, we can define a dimension corresponding to  $p'$ . Here, we specify a  $p$  and define the  $p$ -dimension as

$$\text{Dim}_p = \min \left\{ \text{Dim} \mid 1 - \frac{\|D_{\text{Dim}} - D\|_F}{\|D\|_F} \geq p \right\}. \quad (19)$$

The smaller dimension means that the matrix  $D$  can be approximated well using the distance matrix  $D_{\text{Dim}_p}$ . Thus, we can compare the dimensions of different periods when  $p$  is fixed.

## 2.5 Influence-strength and Rényi index

To simplify the discussion, we use the label of the stock directly as the node label so that a network can be expressed as  $W(S, T)$ , where  $S = \{1, 2, \dots, n\}$  is the stock set, and  $T = [T(i, j)]$  is the adjacency matrix. There are a total of  $n$  nodes (stocks), the degree corresponding to node (stock)  $i$  is  $d_i = \sum_k T(i, k)$ , and the average degree of the network is  $d' = \frac{1}{n} \sum_{i=1}^n d_i$ .

Kim et al. introduced the  $IS$  to describe the influence of stock  $i$  in a network as in

$$IS_i = \sum_{j \in \Gamma_i} \rho(i, j), \quad (20)$$



where  $\Gamma_i$  is a set of nodes directly connected to node  $i$  (Kim et al. 2002).

The Rényi index is a normalized entropy that can be used to characterize randomness, taking values in the interval  $[0, 1]$  (Eliazar 2011). The Rényi index can effectively describe the macroscale structure of financial graphs (Nie et al. 2016; Nie and Song 2019). The form of the index on the network is as shown in Eq. (21), where  $q$  is a parameter. The larger the Rényi index is, the further the network deviates from a homogeneous network (Nie et al. 2016). For example, the Rényi index value of a star-like network with one hub node is close to 1. Conversely, if the degree of each node is equal to a constant, the Rényi index is zero (Nie et al. 2016). Here, we specify  $q = 2$  for analysis.

$$R(q) = 1 - \left[ \sum_{i=1}^n \left( \frac{d_i}{d'} \right)^q \cdot \left( \frac{1}{n} \right) \right]^{\frac{1}{1-q}}, q \neq 1$$

$$R(1) = 1 - \exp \left\{ - \sum_{i=1}^n \left[ \frac{d_i}{d'} \cdot \ln \left( \frac{d_i}{d'} \right) \right] \cdot \frac{1}{1-n} \right\}, q = 1$$
(21)

In addition to IS and the Rényi index, we also use the average shortest path length to describe the network structure globally. For a network, it is the average value of the off-diagonal elements of the shortest distance matrix (Boccaletti et al. 2006).

### 2.6 Relationship between the stock norm and the average correlation coefficient

We examine the relationship between the average correlation coefficient (Eq. (22)) and the stock norm. Eq. (23) can be obtained from Eqs. (14) and (22).

$$\bar{\rho}_i = \frac{\sum_{i \neq j} \rho(i, j)}{n - 1}$$
(22)

$$\bar{\rho}_i = 1 - \frac{N_i^2}{2} - \frac{1}{n - 1} \sum_{j \neq i} \frac{N_j^2}{2} + \frac{1}{n - 1} \sum_{j \neq i} N_i N_j \cos \theta(i, j)$$
(23)

Equation (23) shows that  $\bar{\rho}_i$  can be expressed as a quadratic function of  $N_i$ . Next, we analyse the non-constant items in Eq. (23). A stock set corresponds to a constant  $\sum_j \frac{N_j^2}{2}$ . Since  $\sum_{j \neq i} \frac{N_j^2}{2} \approx \sum_j \frac{N_j^2}{2}$  when  $n$  a large positive integer, the term  $\frac{1}{n-1} \sum_{j \neq i} \frac{N_j^2}{2}$  can be approximated as a constant. In addition, if most of the angles  $(\{\theta_{ij}\})$  are close to  $\pi/2$ , the term  $\frac{1}{n-1} \sum_{j \neq i} N_i N_j \cos \theta_{ij}$  is a small number.

In summary, Eq. (23) shows that the relationship between  $\bar{\rho}_i$  and  $N_i$  is a quadratic polynomial, where the coefficient of the quadratic term is -0.5 and the coefficient of the term  $N_i$  is a number with a small absolute value. The constant is approximately equal to  $\frac{\sum_j N_j^2}{2(n-1)}$ .

## 2.7 Relationship between stock norms and influence-strength

We analyse the relationship between the IS of a correlation-based network and stock norms. Applying Eqs. (14) and (20), IS can be represented as

$$IS_i = \sum_{j \in \Gamma_i} \left( 1 - \left( \frac{N_i^2 + N_j^2}{2} - N_i N_j \cos \theta(i, j) \right) \right). \quad (24)$$

Equation (24) can be further expressed as

$$\begin{aligned} IS_i = \text{Card}(\Gamma_i) - \frac{\text{Card}(\Gamma_i)N_i^2}{2} - \sum_{j \in \Gamma_j} \frac{N_j^2}{2} \\ + N_i \sum_{j \in \Gamma_j} N_j \cos \theta(i, j), \end{aligned} \quad (25)$$

where ‘Card( $\Gamma_i$ )’ denotes the cardinality of the set  $\Gamma_i$ . If most of the angles ( $\{\theta(i, j)\}$ ) are close to  $\pi/2$ , then  $\cos \theta(i, j) \approx 0$ . Equation (25) means that IS is directly related to the stock norm, as shown in Eqs. (26). Equation (26) implies that there is a negative correlation between the stock norm ( $N_i$ ) and the IS ( $IS_i$ ).

$$IS_i \approx \text{Card}(\Gamma_i) - \frac{\text{Card}(\Gamma_i)N_i^2}{2} - \sum_{j \in \Gamma_j} \frac{N_j^2}{2}. \quad (26)$$

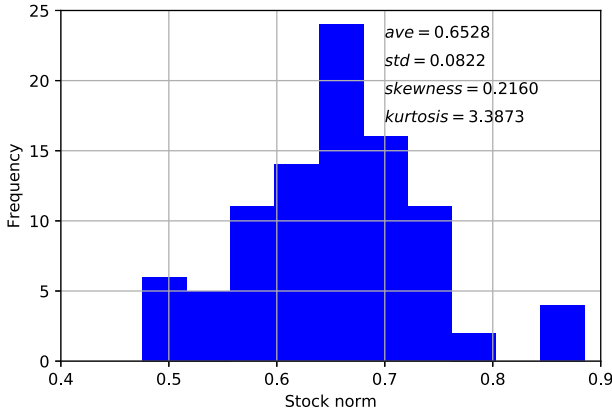
## 2.8 The $\beta$ value of the stock

For each stock, in addition to calculating network indices, we also calculated an index based on factor model. Here, since the correlation matrix is constructed directly from the logarithmic return series, we use the simple average return of the considered stocks as the market return ( $R_M(t) = \frac{1}{n} \sum_i R_i(t)$ ). Then, we estimate the  $\beta$  value for each stock  $i$  and compare the norm with the  $\beta$  value (Campbell et al. 1997). In Eq. (27),  $R_M(t)$  represents the return corresponding to the market factor,  $R_i(t)$  is the return of stock  $i$ , and  $\varepsilon_i(t)$  is a random term with a mean of 0 (Campbell et al. 1997).

$$R_i(t) = \alpha_i + \beta_i R_M(t) + \varepsilon_i(t) \quad (27)$$

## 2.9 Community

We only focus on undirected financial networks without edge weights. In this type of network, the community structure is a cluster-like structure in which there is a close relationship between nodes within the same community. However, nodes in different communities are sparsely linked or not linked. Many algorithms can be used to detect community structures, including the classical modularity-based algorithm proposed



**Fig. 1** Frequency histogram of the stock norm. Several basic statistics are listed in the figure. The minimum norm and the maximum norm are 0.4753 and 0.8846, respectively, indicating a large difference between the norms

by Newman (2004). The modularity is defined as shown in Eq. (28) (Newman 2004; Fortunato 2010), where  $C(i, j)$  takes a value of 1 if  $i$  and  $j$  belong to the same community and takes a value of 0 otherwise. The symbol  $m$  represents the number of links in the network. In general, a  $Q$  value greater than 0.3 implies that there is a significant community structure in the network (Clauset et al. 2004).

$$Q = \frac{1}{2m} \sum_{ij} \left( T(i, j) - \frac{d_i d_j}{2m} \right) C(i, j) \tag{28}$$

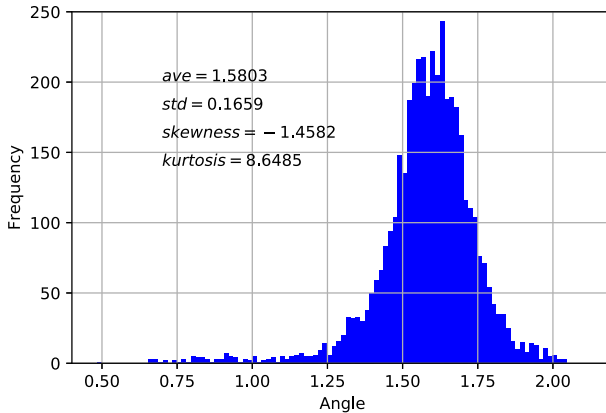
### 3 Results

#### 3.1 Empirical results of stock norm and angle

In this section, we embed the constituents of the S&P100 index into  $R^{92}$  ( $92 = n - 1 = 93 - 1$ ) and then analyse the geometric properties. Here, we only choose the data for 2008. Based on the stock vectors, we calculate the stock norm of each stock and the angle between each pair of stocks. Figures 1 and 2 present the results. Figure 1 shows that there are significant differences between the norm of some stocks. Evidently, the distribution is not homogeneous, and most of the stock norms are close to the average (0.6528), such as 66 stock norm values in the interval  $[0.5705, 0.7350]$  ( $[\text{ave} - \text{std}, \text{ave} + \text{std}]$ ).

In addition, the angle between stock vectors is found to be mainly distributed around  $\pi/2$ . In fact, most of the angles are distributed in the interval  $[1.4144, 1.7461]$  ( $[\text{ave}], \text{ave} + \text{std}]$ ).

We provide an example to analyse the effect of the angle on the correlation coefficient. It is assumed that the angle  $\theta = \text{ave} - \text{std} = 1.4144$ , where  $\text{ave}$  and  $\text{std}$  are the values in Fig. 2. Further, it is assumed that  $N_i = N_j = 0.6528$ , where 0.6528 is the



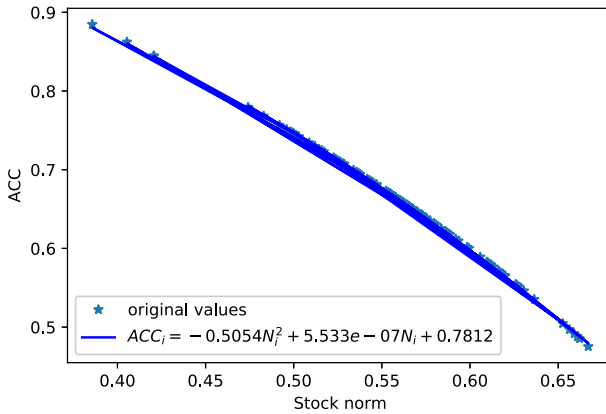
**Fig. 2** Frequency histogram of the angle between stocks. The Jarque–Bera test reports a  $p$ -value less than  $2.2e-16$ , thereby rejecting the null hypothesis of normality

average value in Fig. 1. Thus,  $\theta_\epsilon(i, j) = \pi/2 - \theta = 0.1564$  in Eq. (15). The correlation coefficient calculated directly from Eq. (14) is 0.6402. It can be calculated that  $\theta_\epsilon(i, j)^3 = 0.0038$ , which contributes little to the correlation coefficient value. Based on Eq. (16), the first-order expansion ( $\rho(i, j) \approx 1 - \frac{N_i^2 + N_j^2}{2} + N_i N_j \theta_\epsilon(i, j)$ ) is a good approximation. Since  $N_i = N_j = 0.6528$ , the term  $N_i N_j \theta_\epsilon(i, j) = 0.0666$ . If we do not consider items that include  $\theta_\epsilon$  in Eq. (15), then  $\rho(i, j) \approx 1 - (N_i^2 + N_j^2)/2 = 0.5739$ . Further, considering the first-order expansion of  $\cos(\frac{\pi}{2} - \theta_\epsilon(i, j))$  in Eq. (15),  $\rho(i, j) \approx 1 - (N_i^2 + N_j^2)/2 + N_i^2 N_j^2 \theta_\epsilon(i, j) = 0.6405$ . In summary, we find that the correlation coefficient value is mainly determined by the norm. In addition, a more accurate approximation is obtained when considering the first-order expansion of the angle term. A more accurate estimation of the correlation coefficient needs the term containing  $\theta_\epsilon$  to be considered, but the high-order term of  $\theta_\epsilon$  usually has only a small contribution.

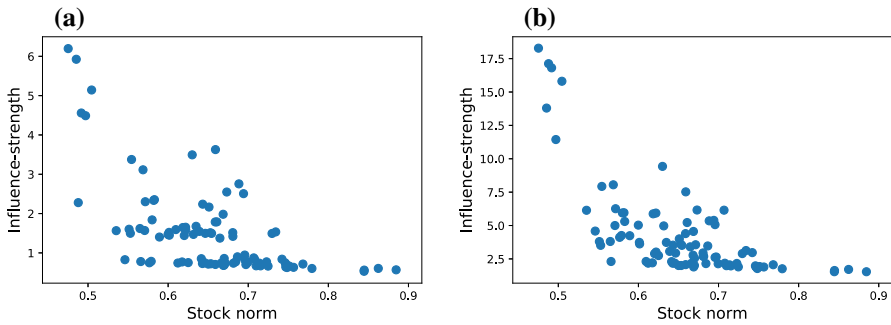
### 3.2 Empirical relationship between norms and other indicators

In this section, we discuss the relationship between norms and other indicators, including the average correlation coefficient and IS. We use the data used in Fig. 1 to calculate  $\bar{\rho}_i$ , and its relationship with  $N_i$  is shown in Fig. 3. By fitting the data, the relationship between the two variables is shown in the figure, which is a quadratic polynomial. Consistent with the analysis in the previous part of this study, the coefficient of the quadratic term is -0.5054. The constant term is 0.7812, and the calculation reveals that  $\frac{\sum_i N_i^2}{2(n-1)} = 0.7836$ , indicating only a slight difference between the two. In addition, the coefficient of  $N_i$  is close to zero ( $5.533e-07$ ), which is consistent with the previous analysis (Eq. (23)).

In summary, our analysis reveals that the relationship between the average correlation coefficient and the norm can be expressed as a quadratic function, and the results based on real data are consistent with Eq. (23).



**Fig. 3** The figure shows the relationship between the norm and the average correlation coefficient (ACC). The relationship can be well described by a quadratic function

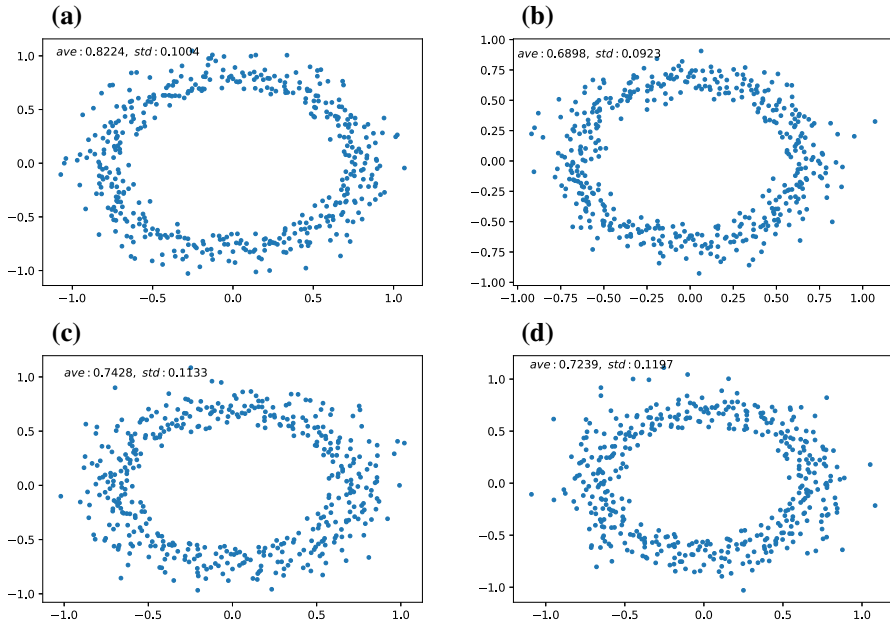


**Fig. 4** The figures show the relationship between the norm and IS, and the smaller norm corresponds to the larger IS. Subgraphs **a** (MST) and **b** (PMFG) show similar patterns in which the smaller norm corresponds to a larger IS value

Based on the data used in Fig. 1, we calculate the MST and PMFG. Figure 4 shows the relationship between IS and the norm. We find that the large IS value corresponds to the small stock norm. Moreover, since the value  $\text{Card}(I_i)$  in the relation Eq. (25) partially contributes to the IS value, we have not found a precise relationship expressed by the function similar to that in Fig. 3. However, we found a negative correlation between the norm and IS, as shown in Eq. (26).

### 3.3 Evolution of stock norms

In this section, we analyse the distribution of the stock norms in different periods. Here, we select the constituent stocks of the S&P500 index and calculate the stock norm distributions for 2007, 2008, 2009, and 2010. The stock collection includes 448 stocks, which are embedded in  $R^{447}$  ( $n - 1 = 448 - 1$ ). To clearly show the distribution of the stock norm, we use  $(x, y) = (N_i \cos \theta_i, N_i \sin \theta_i)$  to represent the point of stock



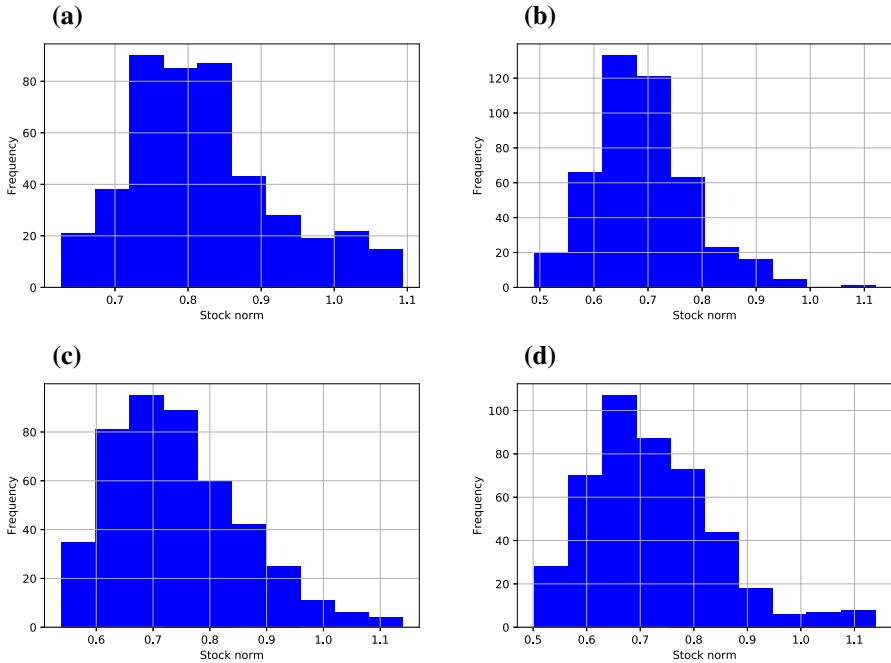
**Fig. 5** The distribution of the stock norms corresponding to different years, where the stock corresponds to a two-dimensional vector. Vectors in high-dimensional space are converted to two-dimensional vectors. The conversion does not change the stock norm

$i$  in a two-dimensional plane, where  $\theta_i$  is a number chosen randomly from the interval  $[0, 2\pi]$ .

Figure 5 shows the distribution of norms in a two-dimensional space. In the four subgraphs, the same stock  $i$  corresponds to the same  $\theta_i$ . Figure 5 shows that the distribution of the stock norms varies drastically over time. For example, the stock norms appear more evenly distributed during 2007 and 2010. During the 2008 financial crisis, the distribution is more discrete and corresponds to a larger standard deviation. The basic statistics of the norms are shown in the figure. The minimum average value of the norms occurs during the financial crisis. In addition, we show the frequency histogram of the norm for each year in Fig. 6. There are clearly some large norm values always distributed in the tail. There are also some stocks with small norms. We discuss the relationship between norms and market factors later.

In addition to the statistic of the norms, the basic statistics of the distribution of the angles are listed in Table 1, where the average of all years is close to  $\pi/2$  (1.5708) and has a negative skewness and a large kurtosis. However, there was a larger standard deviation in 2008. These calculations show that the angle between the stocks also changes over time and that most are distributed around  $\pi/2$ .

The previous calculations visually show that the norm distribution changes over time. In a stock set, stocks with a smaller norm correspond to a larger IS (Fig. 4) and have a larger ACC value (Fig. 3), suggesting that these may significantly influence the structure of the network. In the following section, we construct some networks



**Fig. 6** The subgraphs show the frequency histograms of stock norms in different years. Each stock set includes some stocks with large norms

**Table 1** The basic statistic of the angular distribution corresponding to different years

Year	Ave	SD	Skewness	Kurtosis
2007	1.5722	0.1070	-1.3836	9.7308
2008	1.5720	0.1529	-1.0295	6.8990
2009	1.5713	0.1327	-1.1870	6.7220
2010	1.5718	0.1014	-1.1676	7.7274

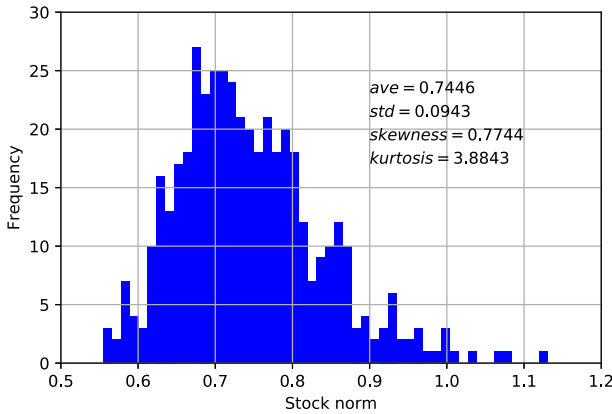
Ave: Average value  
SD: Standard deviation

based on norms and show the important influence of stocks with small norms on the structure of financial graphs.

### 3.4 Constructing networks of different structures based on the stock norm

We select the data of the constituent stocks of the S&P500 index from 3/1/2007 to 31/12/2010, which is a total of 1008 trading days. We calculate the stock norm as shown in Fig. 7. The average of the norms is 0.7446, that is, most of the stocks are distributed near the hypersphere with a radius of 0.7446.

We construct some networks based on the quantile of the empirical distribution of the stock norm. Here, we use  $Q_q$  to represent the  $q$ -quantile and select the stock corresponding to the norm in the interval  $[Q_q, N_{max}]$  to calculate the MST and PMFG.



**Fig. 7** The frequency histogram of the stock norm for 2007/1/3–2010/12/31. There are a few stocks with small norms, and some stock norms deviate far from the average

**Table 2** The  $Q_q$  values used in the calculations

$Q_{0.40}$	$Q_{0.45}$	$Q_{0.50}$	$Q_{0.55}$	$Q_{0.60}$	$Q_{0.75}$	$Q_{0.80}$	$Q_{0.85}$	$Q_{0.90}$
0.7103	0.7185	0.7320	0.7436	0.7563	0.7972	0.8125	0.8431	0.8660

Here, the maximum norm ( $N_{\max}$ ) is 1.1308. Table 2 reports the  $Q_q$  values used in the calculations.

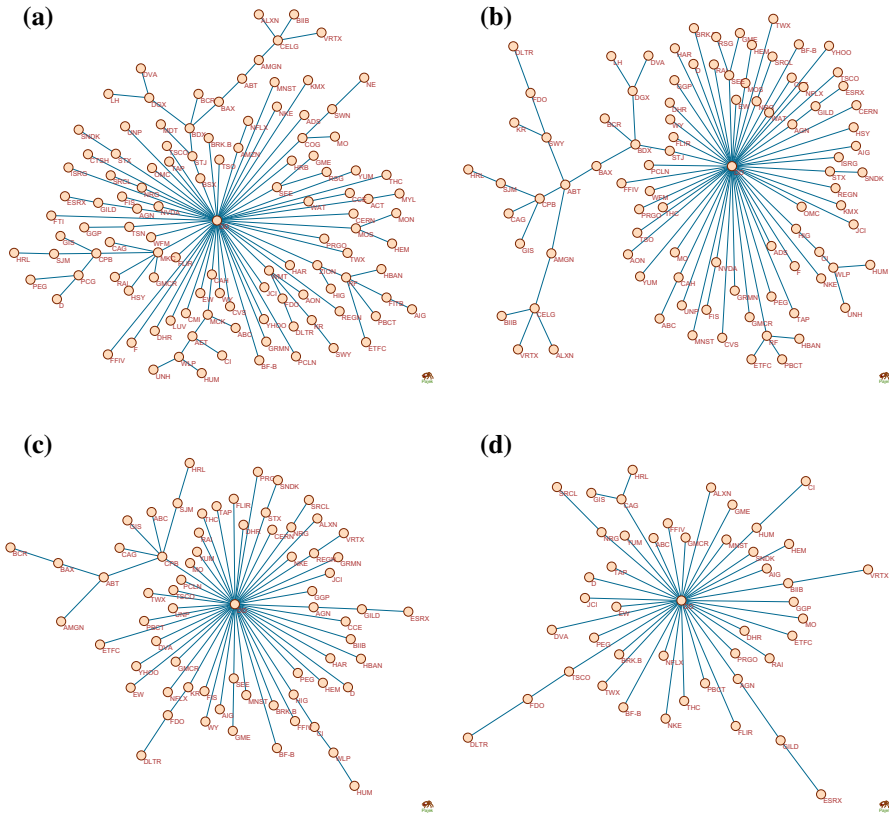
Based on the stocks selected by the interval  $[Q_q, N_{\max}]$ , we add a stock with the smallest norm (DD,  $N_{DD} = 0.5545$ ) to show its effect on the network structure. We choose  $Q_{0.75}$ ,  $Q_{0.80}$ ,  $Q_{0.85}$ , and  $Q_{0.90}$  for calculation, where the four sets of stocks include 113, 91, 68, and 46 stocks, respectively. Figures 8 and 9 show the calculated MSTs and PMFGs, respectively.

Next, based on the stock set used in Figs. 8b and 9b, we add a stock with a small norm which is only greater than the norm of DD and then calculate the MST (Fig. 10a) and PMFG (Fig. 10b). We find that the original network with only a single hub node was converted to a network with two hub nodes.

We continue to examine the network generated by stocks near the quantile  $Q_{0.50}$  of the norm distribution. We select the stocks in intervals  $[Q_{0.45}, Q_{0.55}]$  and  $[Q_{0.40}, Q_{0.60}]$  and calculate the MSTs and PMFGs, respectively (Fig. 11). The two stock sets include 44 and 90 stocks, respectively. The results reveal that both MSTs and PMFGs exhibit a similar topological structure, that is, a chain-like structure.

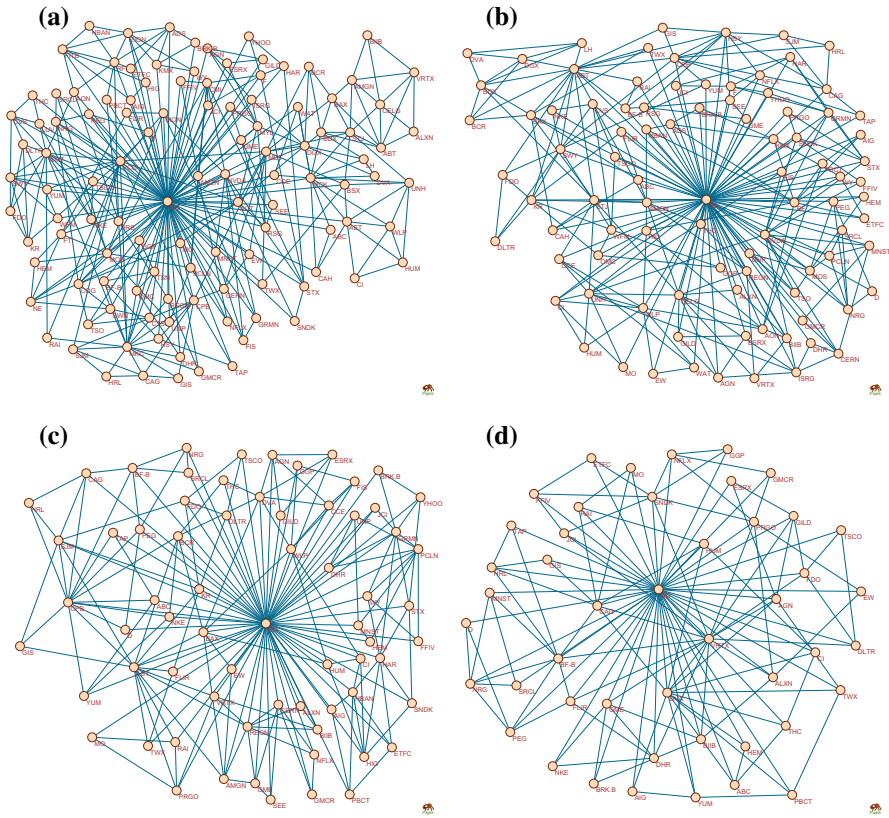
We found that changing norm-based conditions can result in networks with completely different structures. For example, Fig. 8b includes a stock with a minimum norm, while the stock set in Fig. 10a includes those with a minimum and a second small norm, such that the former includes one hub node and the latter includes two hub nodes.



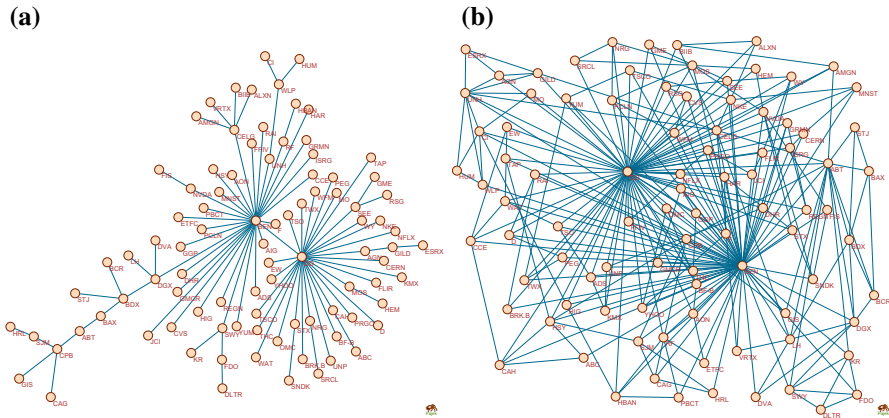


**Fig. 8** The four subgraphs show the MSTs for different intervals. **a**  $[Q_{0.75}, N_{max}]$ , **b**  $[Q_{0.80}, N_{max}]$ , **c**  $[Q_{0.85}, N_{max}]$ , **d**  $[Q_{0.9}, N_{max}]$

Table 3 reports some basic statistics such as the maximum and average. We can roughly analyse the heterogeneity of different networks from these statistics. The maximum degree of the network in Fig. 8a is 52, which is much larger than that in Fig. 11a. However, the difference in the mean value between the two is small, suggesting that the former has a higher level of heterogeneity. To show the structural changes in detail, Table 3 also presents the values of the average path length ( $L_{ave}$ ) and the Rényi index. Changes in the structure of the network can be quantified by these two indicators. The network constructed by stocks with norms near the quantile  $Q_{0.5}$  is found to have a chain-like structure, which is characterized by a significantly smaller Rényi index than that of a network with hub nodes. For example, the Rényi index values of the network in Fig. 8 are greater than 0.8, and the values of the four networks in Fig. 9 are all greater than 0.55. However, the values of the network of the four networks in Fig. 11 are all less than 0.33. In addition, the  $L_{ave}$  values of the MSTs in Fig. 11 are also found to be larger than the  $L_{ave}$  values of the MSTs in Figs. 8 and 10. Similarly, the  $L_{ave}$  values for PMFGs in Fig. 11 are all greater than 3.5, but the  $L_{ave}$  values for PMFGs in Figs. 9 and 10 are less than 2.5.



**Fig. 9** The four subgraphs show the PMFGs for different intervals. **a**  $[Q_{0.75}, N_{max}]$ , **b**  $[Q_{0.80}, N_{max}]$ , **c**  $[Q_{0.85}, N_{max}]$ , **d**  $[Q_{0.9}, N_{max}]$



**Fig. 10** The two subfigures show the financial graph generated by adding two small norm stocks. Both MST **a** and PMFG **b** contain two hub nodes, which correspond to the two small norm stocks

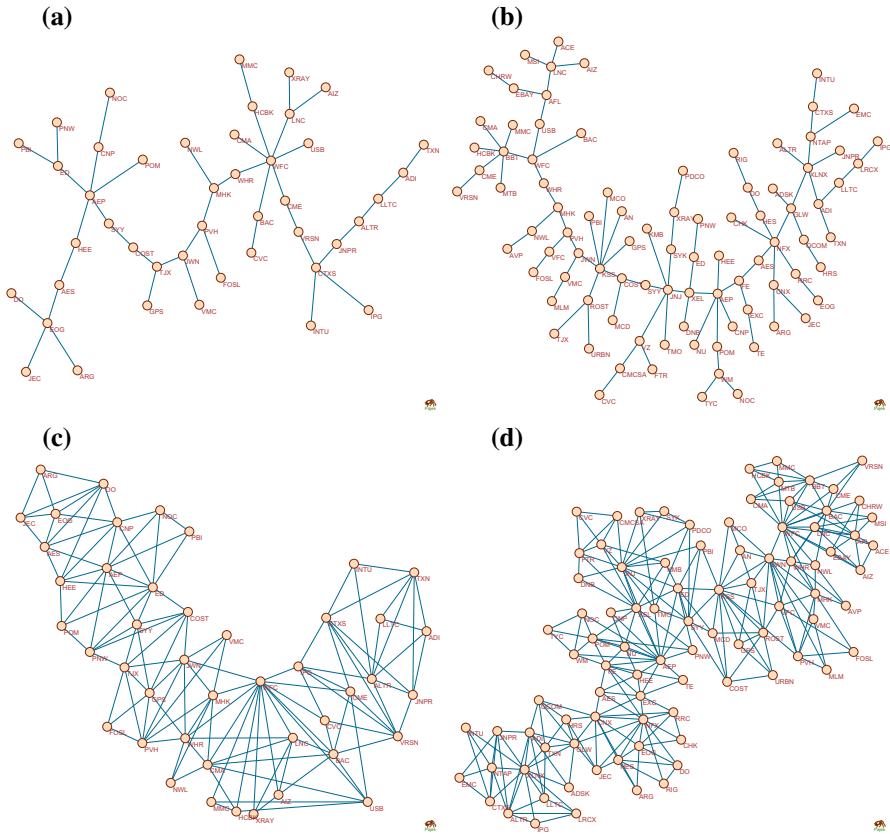


Fig. 11 The figure shows four financial graphs generated based on intervals  $[Q_{0.45}, Q_{0.55}]$  and  $[Q_{0.40}, Q_{0.60}]$ , where each graph has a chain-like topological structure

### 3.5 Analysis of the hub nodes in PMFG

The financial graphs described in the previous section reveal the high impact of stocks with small norms. Since the community structure can be directly analysed in PMFG, we examine the relationship between the community structure and the norm in more detail in this section.

Here, we select the constituents of the S&P100 index and calculated PMFG for 2008. Then, we use a classic community detection algorithm proposed by Newman (2004). This algorithm is used to divide PMFG into six communities, as shown in Fig. 12. The calculation results indicate that the module degree is 0.6021, and a value greater than 0.3 means that the community structure is significant (Clauset et al. 2004). We use the colour of the node to mark the community, and the labels corresponding to each community are shown in Table 4, in which we also list the number of stocks ( $N_C$ ) included in different communities.

In Fig. 12, each node is labelled with a company name and a stock norm. Evidently, the hub nodes in different communities are found to have smaller norm values. In

**Table 3** Comparison of structural indicators between different networks

Network	$d_M$	$d_A$	$L_{ave}$	Rényi index
Figure 8a	52.0000	1.9823	4.1963	0.8555
Figure 8b	52.0000	1.9780	4.0557	0.8791
Figure 8c	47.0000	1.9706	2.7871	0.8880
Figure 8d	36.0000	1.9565	2.4077	0.8715
Figure 9a	86.0000	5.8938	2.4425	0.6524
Figure 9b	78.0000	5.8681	2.1597	0.6528
Figure 9c	65.0000	5.8235	1.9543	0.6291
Figure 9d	45.0000	5.7391	1.8725	0.5536
Figure 10a	30.0000	1.9783	3.6730	0.8219
Figure 10b	72.0000	5.8696	2.1125	0.7297
Figure 11a	7.0000	1.9545	7.4884	0.2878
Figure 11b	7.0000	1.9778	9.0270	0.3282
Figure 11c	14.0000	5.7273	3.6744	0.1306
Figure 11d	16.0000	5.8667	4.7518	0.2154

$d_M$ : Maximum degree  
 $d_A$ : Average degree

**Table 4** The labels of different communities and the average norm and maximum norm of stocks in the community

$L_C$	Colour	$N_C$	Norm <sub>ave</sub>	Norm <sub>max</sub>
1	Green	27	0.625	0.475 (25)
2	Yellow	16	0.664	0.488 (24)
3	Red	13	0.688	0.659(10)
4	White	9	0.650	0.618(8), 0.5820(8)
5	Blue	13	0.651	0.505(22)
6	Pink	15	0.663	0.630(11)

$N_C$ : The number of nodes included in the community  $L_C$   
 $L_C$ : The community label  
 Norm<sub>ave</sub>: The average norm of the nodes in the community  
 Norm<sub>max</sub>: The norm of the node corresponding to the maximum degree. The number enclosed in parentheses is the corresponding degree

the fourth and fifth columns of Table 4, we report the average norm values of the stocks in the community and the norm values corresponding to the hub nodes. For example, there is only one hub node in community 2, with a norm of 0.488, which is significantly smaller than those of other nodes in the community. Similarly, there is only one hub node with a norm of 0.505 in community 5. In addition, there may be differences between the norms of the hub nodes of different communities. For example, the norms of several hub nodes in communities 3, 4, and 6 are all greater than 0.51, but they are still smaller than the Norm<sub>ave</sub> values. Our analysis indicates that the hub nodes in a community also correspond to smaller norms, and that the values are related to the community.

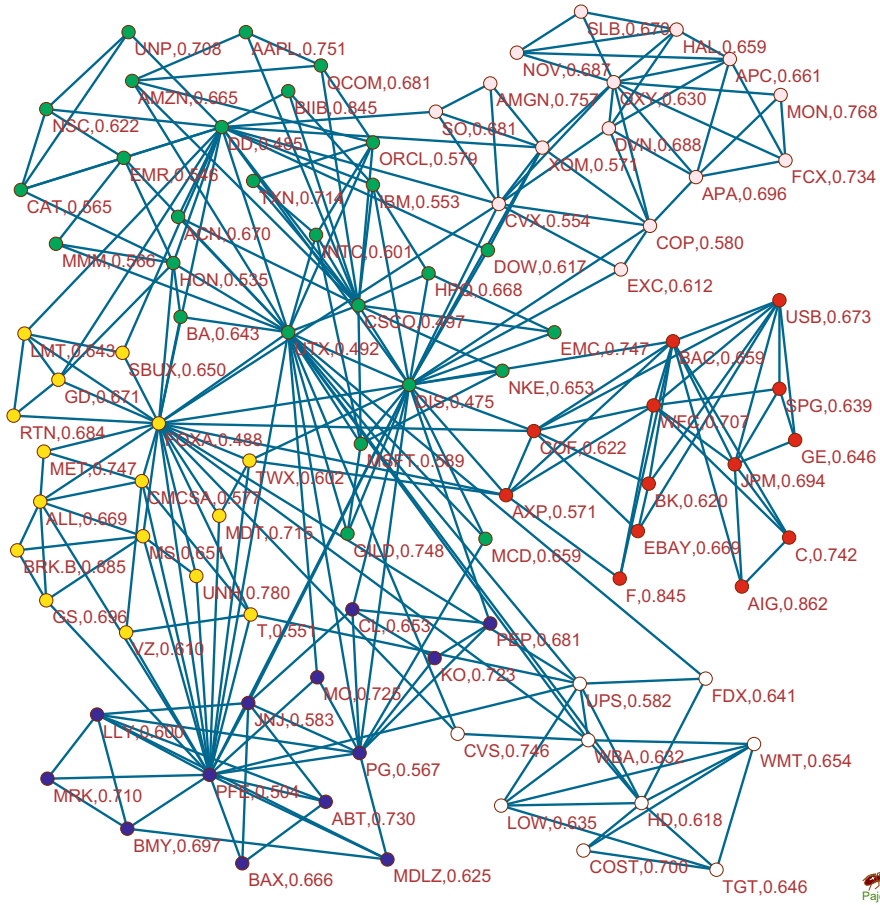
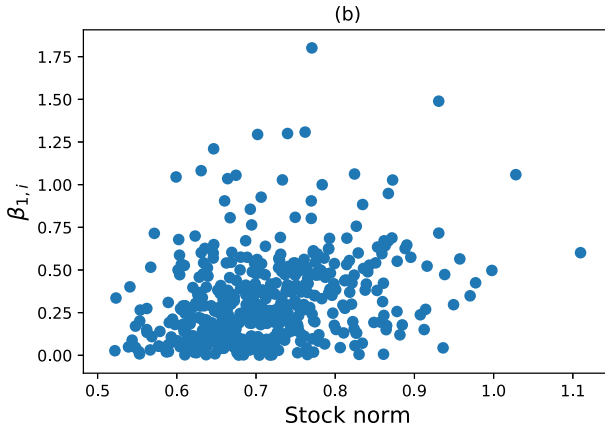


Fig. 12 The figure shows a PMFG in which different colours correspond to different communities, and each stock is labelled with a norm value

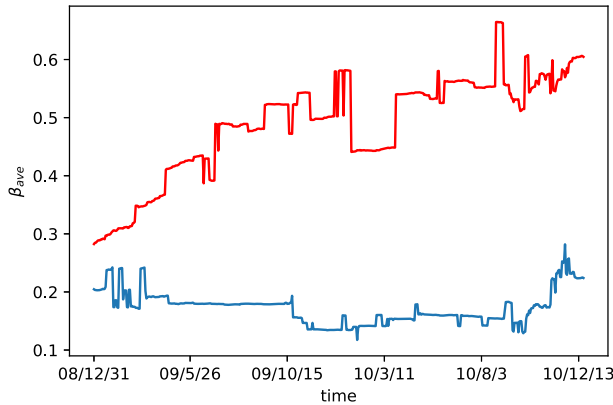
In summary, we find that the nodes of the small norm also occupy an important position in the community, which indicates that the stocks of small norms have significance in terms of the macro structure (degree distribution) and meso-structure (community structure) of the network.

### 3.6 Relationship between norm and $\beta$ value

In this section, we analyse the relationship between the norms and the  $\beta$  values. We select the data of the S&P500 index stocks from 3/1/2007 to 31/12/2010 to calculate the norm and  $\beta$  value. We analyse the difference between the  $\beta$  value and 1, that is, we calculate  $\beta_{1,i} = |\beta_i - 1|$ . The scatter plot is shown in Fig. 13. The small norm usually corresponds to a small  $\beta_{1,i}$  value, while the large norm often corresponds to a larger  $\beta_{1,i}$  value, indicating that the  $\beta$  value of the small norm is closer to 1.



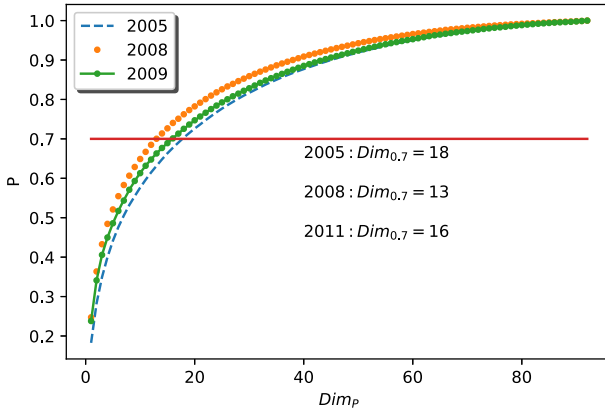
**Fig. 13** Most stocks with small norms have small  $\beta_{1,i}$ -values. Compared with small norms, large stock norms globally correspond to larger  $\beta_{1,i}$ -values



**Fig. 14** The figure shows the sequence of  $\beta_{ave}$ -values of two groups of stocks, in which significant differences can be observed. The series corresponding to small norm stocks is more stable, and most  $\beta_{ave}$  values are close to 0.15

Furthermore, we sort all the stock norms and extract the top 10 small norm stocks and the top 10 large norm stocks. Thus, for a group of stocks, we extract two groups of stocks corresponding to the small and large stock norms, respectively. We then separately calculate the average value ( $\beta_{ave}$ ) of  $\beta_{1,i}$  values of each group of stocks. We can calculate the  $\beta_{ave}$  value series through the moving window and analyse the difference between the  $\beta_{ave}$  value series. The length of the window is set to 504 trading days. Figure 14 presents the time series of  $\beta_{ave}$  values, and the  $\beta_{ave}$  value of stocks with large stock norms clearly deviate further from 0.

The analysis reveals that the  $\beta$  value corresponding to the small norm is closer to 1, which implies that there is a higher correlation between small norm stocks and the market factor.



**Fig. 15** The figure shows the relationship between dimensions and  $p$ . If  $p = 0.7$ , the  $p$ -dimension of the market in 2008 is the minimum

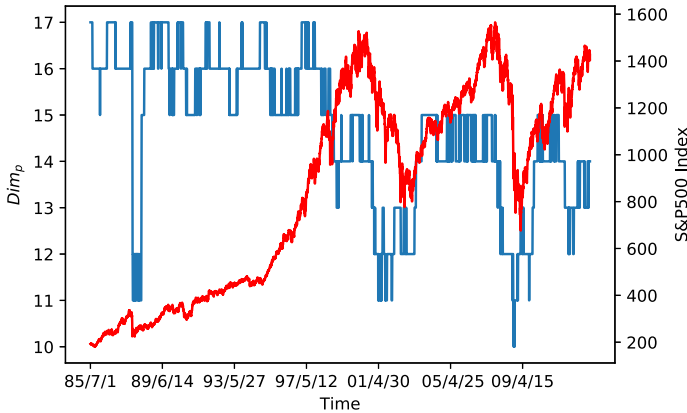
### 3.7 Relationship between subspace dimensions and financial crisis

For the market space, we apply the  $p$ -dimension to analyse the dynamics of the subspace. The results reveal that most of the correlation information can be included in a subspace, and its dimension is much smaller than the number of stocks. Here, we select the annual data of the constituent stocks of the S&P100 index.

We use the relationship shown in Eq. (18) to analyse the information included in the subspace. We specify the dimension ( $\text{Dim}$ ) and calculate  $p'$ , as shown in Fig. 15, in which the three lines correspond to 2005, 2008, and 2012, respectively. Based on Fig. 15, the main results are as follows.

First, a larger  $p'$  (Eq. (18)) means that  $\text{Dim}_p$  contains more correlation information, resulting in a smaller difference between the matrix  $D_{\text{Dim}_p}$  and  $D$ . The results indicate that although the stock set is embedded in a high-dimensional space, a smaller subspace can include the main correlation structure. Second, we find that there are differences between the corresponding curves for different years. For example, the corresponding curve in 2008 is at the top, which shows that when the dimension is small, the  $p'$  value changes more drastically. That is, when we specify the same dimension to calculate  $p$ , the corresponding matrix  $D_{\text{Dim}_p}$  for 2008 includes more correlation information. For example, according to the definition of Eq. (19), the value of  $\text{Dim}_p$  corresponding to 2008 is 13 when we specify  $p = 0.7$ . Similarly, the values of  $\text{Dim}_p$  corresponding to 2005 and 2011 are 18 and 16, respectively, which suggests that the subspace of the smaller dimension in 2008 includes more correlation information.

In summary, we find that the relationship between  $p'$  and dimensions changes over time, which implies that changes in the correlation structure of the financial market have led to changes in the subspace. To investigate the relationship between dimensions and  $p$  in detail, we specify  $p = 0.7$  below and calculate the corresponding  $\text{Dim}_p$  (Eq. (19)). We select the data in the US and UK markets and slide the time window  $[t_1, t_2]$  to calculate the  $\text{Dim}_p$  values for different time periods. Here, the calculation time window is 125 days ( $t_2 - t_1 + 1 = 125$ ), and the sliding time window is 1 day.



**Fig. 16** The figure shows the dynamics of the subspace in the US market and simultaneously marks the S&P500 index. Intuitively, when the market is in a crisis period, such as the 1987 crisis and the 2008 crisis, the  $Dim_p$  value is smaller

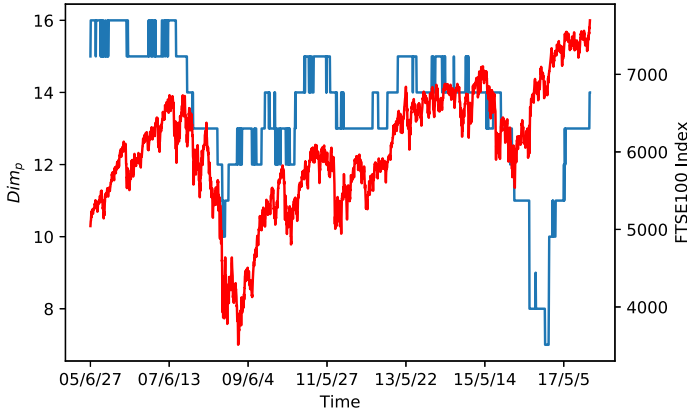
Thus, the dimension series is calculated and can be compared with the market index. Figures 16 and 17 show the results using US and UK market data, respectively. Here, Fig. 16 analyses the long-term historical data of 78 stocks. Below, we analyse the dynamics of the dimensions of the two markets.

First, we compare the  $Dim_p$  series with the S&P500 index. It can be seen that during the financial crisis, the dimension decreased significantly, that is, the period in which the index decreased significantly corresponds to a smaller dimension. For example, the corresponding dimension during the 1987 crash was significantly smaller, and the calculation indicates that the dimension of the period 22/4/1987–19/10/1987 is 15. In addition, compared with the average (14.6523),  $Dim_p$  values were significantly smaller during the bubble burst in tech stocks in 2000 and during the subprime mortgage crisis in 2008.

Second, Fig. 17 also shows empirical results similar to those shown in Fig. 16. In Fig. 17, we compare the  $Dim_p$  series to the FTSE 100 index and find that the  $Dim$  value also decreases significantly when the index drops sharply.

Table 5 shows the dimensions and market indices corresponding to some periods in detail, of which the US market index is the S&P500 index and the U.K. market index is the FTSE 100 index. Here, each index value corresponds to time  $t_2$  of the period  $[t_1, t_2]$ . Both Black Monday and the 2008 financial crisis are found to have significantly affected the change in dimensions. In addition, there is a special case in the U.K. market. The dimension has been significantly reduced since 2015, but the index rebounded significantly after reaching the local minimum during this period, and the FTSE 100 index was 5537 on 11/2/2016. This example demonstrates that during the financial crisis,  $Dim_p$  decreased significantly; however, a significant reduction in  $Dim_p$  does not necessarily correspond to a financial crisis.





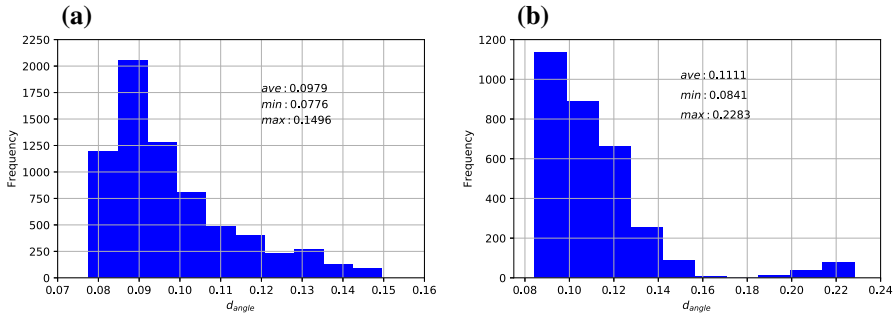
**Fig. 17** The figure shows the dynamics of the subspace in the UK market and simultaneously marks the FTSE 100 index. Overall, the  $Dim_p$  value decreases as the market falls sharply

**Table 5** The comparison between dimension and index in different periods

Market	Period	$Dim_p$	Index
U.S.	1987.4.21–1987.10.16	17	282.7
U.S.	1987.4.22–1987.10.19	15	224.84
U.S.	1987.4.23–1987.10.20	13	236.83
U.S.	1999.9.27–2000.3.24	15	1527.46
U.S.	2000.10.27–2001.4.30	11	1249.46
U.S.	2007.4.12–2007.10.9	14	1565.15
U.S.	2008.4.17–2008.10.14	10	998.01
U.K.	2007.4.12–2007.10.9	15	6615.4
U.K.	2008.4.17–2008.10.14	10	4394.2
U.K.	2016.5.18–2016.11.11	7	6730.4

### 3.8 Geometric features of the market as a vector set

In Sect. 3.1, an example shows that the angle between most stock vectors is close to  $\pi/2$ . In this section, we analyse the market reconstructed in a Euclidean space from the perspective of the angle between stock vectors. The calculations in the previous section suggest that the market’s main correlation information can be included in a low-dimensional subspace, but this space evolves over time. We use the data used in Figs. 16 and 17 to analyse the evolution of the stock angle over time. To calculate the level of the set  $\{\theta(i, j), j > i\}$  deviation from  $\pi/2$ , we calculate the global indicator as in Eq. (29), where ‘ $Card(\{\theta(i, j), j > i\})$ ’ means the cardinality of the set  $\{\theta(i, j), j > i\}$ . This indicator characterizes the difference between the angle and  $\pi/2$  in the sense of the average value. Overall, the smaller the value, the closer the stock angle is to  $\pi/2$ . Figure 18 shows that the  $d_{angle}$  value is close to  $\pi/2$  in both the U.S. and U.K. markets, which indicates that the angle between the vectors in the market vector set in a Euclidean space is generally close to  $\pi/2$ .



**Fig. 18** The frequency histograms of  $d_{angle}$ . Small average values and maximum values indicate that most of the angles are close to  $\pi/2$

$$d_{angle} = \frac{1}{\text{Card}(\{\theta(i, j), j > i\})} \sum_{j>i} \left| \theta(i, j) - \frac{\pi}{2} \right| \tag{29}$$

## 4 Discussion and conclusions

### 4.1 Discussion

The method used herein needs to define a metric, such as the distance based on the Pearson correlation coefficient. This condition limits the method to the Euclidean space. More generally, the geometry of the market defined for other non-metric relationships, such as Granger causality, needs to be further explored.

Based on the efficient market hypothesis, price movements arise from responses to information, while factor models show that stock returns can be expressed as a linear combination of economic factors. In this study, the  $p$ -dimension is essentially an indicator that expresses the size of the subspace needed to characterize specified correlation information. Thus, we suspect that the change in the dimension of the subspace is driven by market information.

The analysis reveals that the distribution of stock norms has a significant impact on the network topology, making it necessary to focus on the effects of norms in a network-based analysis. In particular, if investment indicators or risk management indicators are constructed based on the topological structure, it is necessary to carefully examine the distortion of the indicators caused by some small-norm stocks. In addition, cluster analysis based on correlation networks also needs to focus on these effects and enhance the robustness of the method.

### 4.2 Conclusions

Reconstructing the market in a Euclidean space allows us to study a group of stocks from a geometric perspective. This approach makes it possible to introduce some

geometric concepts for analysing correlation structures. The main conclusions of this study are as follows.

The angle between the vectors corresponding to the stock is close to  $\pi/2$ , forming a direct relationship between stock norms and the average correlation coefficient. Furthermore, the norm distribution is found to change over time; for example, the average norm during the financial crisis is smaller, but the standard deviation of the norm is greater. In addition, we also find that the stock with a small norm has a strong influence on the network structure. For example, we constructed networks with super hub nodes by adding small-norm stocks. In particular, we constructed several networks with a chain structure using the distribution of stock norms. In addition to conducting an analysis from the network perspective, we examine the changes in the correlation information contained in the subspace and find that  $\text{Dim}_p$  changes over time. During the financial crisis, the values of  $\text{Dim}_p$  are smaller, indicating that the subspace contains more information.

In summary, we establish the relationship between market geometry and the correlation-based network. In our study, each node in the network is given some geometric indicators, such as norm, so that the geometric perspective of the correlation-based network is established. This detailed study reveals a link between geometric conditions and the structure of the correlation-based network. The calculations herein show that the correlation structure can be effectively studied through market geometry.

## Appendix

The following tables list the abbreviations of the stocks included in the data sets.

**Table 6** List of stock abbreviations (S&P500 index)

ABT	BAC	KO	EFX	HPQ	LOW	NRG	REGN	TJX
ACE	BK	CCE	EQR	HD	MTB	NUE	RF	TMK
ACN	BCR	CTSH	ESS	HON	MAC	NVDA	RSG	TSS
ACT	BAX	CL	EL	HRL	M	ORLY	RAI	TSCO
ADBE	BBT	CMCSA	EXC	HSP	MMM	OXY	RHI	RIG
AES	BDX	CMA	EXPD	HST	MRO	OMC	ROK	FOXA
AET	BBBY	CSC	ESRX	HCBK	MAR	OKE	COL	TSN
AFL	BMS	CAG	XOM	HUM	MMC	ORCL	ROP	TYC
AMG	BRK.B	COP	FFIV	HBAN	MAS	OI	ROST	USB
A	BBY	CNX	FDO	ITW	MAT	PCG	RDC	UNP
GAS	BIIB	ED	FAST	IR	MKC	PCAR	R	UNH
APD	BLK	STZ	FDX	TEG	MCD	PLL	SWY	UPS

Table 6 continued

ARG	HRB	GLW	FIS	INTC	MHFI	PH	CRM	MLM
AKAM	BA	COST	FITB	IBM	MCK	PDCO	SNDK	UTX
AA	BXP	CCI	FE	IP	MWV	PAYX	SCG	UNM
ALXN	BSX	CSX	FISV	IPG	MDT	BTU	SLB	URBN
ATI	BMY	CMI	FLIR	IFF	MRK	PNR	STX	VFC
AGN	BRCM	CVS	FLS	INTU	MET	PBCT	SEE	VLO
ADS	BF-B	DHI	FLR	ISRG	MCHP	POM	SRE	VAR
ALL	CHRW	DHR	FMC	IVZ	MU	PEP	SHW	VTR
ALTR	CA	DRI	FTI	IRM	MSFT	PKI	SIAL	VRSN
MO	CVC	DVA	F	JBL	MHK	PRGO	SPG	VZ
AMZN	COG	DE	FOSL	JEC	TAP	PETM	SJM	VRTX
AEE	CAM	DNR	BEN	JNJ	MDLZ	PFE	SNA	VNO
AEP	CPB	XRAY	FCX	JCI	MON	PNW	SO	VMC
AXP	COF	DVN	FTR	JOY	MNST	PXD	LUV	WMT
AIG	CAH	DO	GME	JPM	MCO	PBI	SWN	WAG
AMT	KMX	DTV	GCI	JNPR	MS	PCL	STJ	DIS
ABC	CCL	DLTR	GPS	KSU	MOS	PNC	SWK	GHC
AMGN	CAT	D	GRMN	K	MSI	RL	SPLS	WM
APH	CBG	DOV	GD	KEY	MUR	PPG	SBUX	WAT
APC	CELG	DOW	GE	GMCR	MYL	PPL	HOT	WLP
ADI	CNP	DTE	GGP	KMB	NBR	PX	STT	WFC
AON	CTL	DD	GIS	KIM	NDAQ	PCP	SRCL	WDC
APA	CERN	DUK	GPC	KLAC	NOV	PCLN	SYK	WY
AIV	SCHW	DNB	GNW	KSS	NTAP	PFG	STI	WHR
AAPL	CHK	ETFC	GILD	KR	NFLX	PG	SYMC	WFM
AMAT	CVX	EMN	GS	LB	NWL	PGR	SYI	WMB
ADM	CB	ETN	GT	LLL	NFX	PLD	TROW	WEC
AIZ	CI	EBAY	GWW	LH	HEM	PRU	TGT	WYNN
T	XEC	ECL	HAL	LRCX	HEE	PEG	TE	XEL
ADSK	CINF	EIX	HOG	LM	NKE	PSA	THC	XRX
ADP	CTAS	EW	HAR	LEG	NI	PHM	TSO	XLNX
AN	CSCO	EA	HRS	LEN	NE	PVH	TXN	XL
AZO	C EMC	HIG	LUK	NBL	PWR	TXT	YHOO	
AVB	CTXS	EMR	HAS	LLY	JWN	QCOM	HSY	YUM
AVY	CLX	ESV	HCP	LNC	NSC	DGX	TRV	ZMH
AVP	CME	ETR	HCN	LLTC	NTRS	RRC	TMO	ZION
BHI	CMS	EOG	HP	LMT	NOC	RTN	TIF	-
BLL	COH	EQT	HES	L	NU	RHT	TWX	-

**Table 7** List of stock abbreviations (S&P100 index)

AAPL	BAC	COP	EMR	HD	MCD	NOV	SO	VZ
ABT	BAX	COST	EXC	HON	MDLZ	NSC	SPG	WBA
ACN	BIIB	CSCO	F	HPQ	MDT	ORCL	T	WFC
AIG	BK	CVS	FCX	IBM	MET	OXY	TGT	WMT
ALL	BMY	CVX	FDX	INTC	MMM	PEP	TWX	XOM
AMGN	BRK.B	DD	FOXA	JNJ	MO	PFE	TXN	-
AMZN	C	DIS	GD	JPM	MON	PG	UNH	-
APA	CAT	DOW	GE	KO	MRK	QCOM	UNP	-
APC	CL	DVN	GILD	LLY	MS	RTN	UPS	-
AXP	CMCSA	EBAY	GS	LMT	MSFT	SBUX	USB	-
BA	COF	EMC	HAL	LOW	NKE	SLB	UTX	-

**Table 8** The abbreviations of 78 stocks in the S&P500 index

AA	AMGN	BLL	COP	EXC	IBM	MCD	PFE	VZ
AAPL	AON	BMY	CSX	F	JNJ	MMM	RTN	WFC
ABT	APA	C	CVX	GE	JPM	MO	SLB	WHR
ADP	APD	CAG	DD	GLW	K	MRK	T	WMT
AEP	AVY	CAT	DIS	GT	KO	MSI	TAP	XOM
AET	AXP	CI	DOW	HD	KR	NUE	TE	XRX
AIG	BA	CLX	DUK	HON	LLY	OXY	TGT	-
AMAT	BAX	CMI	ED	HPQ	LMT	PBI	UNP	-
AMD	BF-B	CMS	EIX	HUM	LUV	PEP	UTX	-

**Table 9** List of stock abbreviations (FTSE 100 Index)

AAL.L	BDEV.L	CRDA.L	HSBA.L	LGEN.L	PSN.L	SBRY.L	STAN.L
ABF.L	BKG.L	CRH.L	IAG.L	LLOY.L	PERSON.L	SDR.L	STJ.L
ADM.L	BLND.L	DCC.L	IHG.L	LSE.L	RB.L	SGEL.L	SVT.L
AHT.L	BLT.L	DGE.L	IMB.L	MKS.L	RBS.L	SGRO.L	TSCO.L
ANTO.L	BNZL.L	EZJ.L	INF.L	MRW.L	REL.L	SHPL.L	TWL.L
AV.L	BPL	GFS.L	ITRK.L	NG.L	RIO.L	SKY.L	ULVR.L
AZN.L	BRBY.L	GKN.L	ITV.L	NXT.L	RR.L	SMDS.L	UU.L
BA.L	CCL.L	GSK.L	JMAT.L	OML.L	RRS.L	SMIN.L	VOD.L
BARC.L	CNA.L	HLMA.L	KGFL	PPB.L	RSA.L	SN.L	WPP.L
BATS.L	CPG.L	HMSO.L	LAND.L	PRU.L	RTO.L	SSE.L	WTB.L

## References

- Ahmet S, Yuksel S, Erturk M (2013) Analysis of cross-correlations between financial markets after the 2008 crisis. *Physica A* 392(20):5027–5045
- Antonios G, Argyrakis P (2007) Correlation study of the athens stock exchange. *Physica A* 380:399–410
- Araújo T, Louçã F (2007) The geometry of crashes: a measure of the dynamics of stock market crises. *Quant Finance* 7(1):63–74
- Araújo T, Spelta A (2014) Structural changes in cross-border liabilities: a multidimensional approach. *Physica A* 394:277–287
- Araújo T, ao Dias J, Eleutério S, Louçã F (2013) A measure of multivariate kurtosis for the identification of the dynamics of a n-dimensional market. *Physica A* 392(17):3708–3714
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: structure and dynamics. *Phys Rep* 424:175–308
- Borg I, Groenen PJF (2005) *Modern multidimensional scaling: theory and applications*. Springer, Berlin, pp 261–263
- Campbell JY, Lo AW, MacKinlay AG (1997) *The econometrics of financial markets*. Princeton University Press, Princeton, p 155
- Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E* 70:066111
- Echaust K, Just M (2013) Geometry of crises on the market of stocks listed on the Warsaw stock exchange. SSRN 2013: <https://doi.org/10.2139/ssrn.2260835>
- Eleutério S, Araújo T, Mendes RV (2014) Portfolios and the market geometry. *Physica A* 410:226–235
- Eliazar I (2011) Randomness, evenness, and Rényi's index. *Physica A* 390(11):1982–1990
- Eng-Uthaiwat H (2018) Stock market return predictability: does network topology matter? *Rev Quant Financ Acc* 51(2):433–460
- Esmalifalak H, Ajirlou AI, Behrouz SP (2015) (dis)integration levels across global stock markets: a multi dimensional scaling and cluster analysis. *Expert Syst Appl* 42:8393–8402
- Fernández-Avilés G, Montero JM, Sanchis-Marco L (2020) Extreme downside risk co-movement in commodity markets during distress periods: a multidimensional scaling approach. *Eur J Finance* 26:1207–1237
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):75–174
- Gałżka M (2011) Characteristics of the polish stock market correlations. *Int Rev Financ Anal* 20(1):1–5
- Giuseppe B, Marmi S, Mantegna RN (2013) Evolution of correlation structure of industrial indices of US equity markets. *Phys Rev E* 88(1):012806
- Groenen PJF, Franses PH (2000) Visualizing time-varying correlations across stock markets. *J Empir Finance* 7:155–172
- Halmos PR (1974) *Finite-dimensional vector spaces: second edition, 1st edn*. Springer, New York, pp 118–128
- Joël B, Bouchaud JP, Potters M (2017) Cleaning large correlation matrices: tools from random matrix theory. *Phys Rep* 666:1–109
- Jung WS, Chae S, Yang JS, Moon HT (2006) Characteristics of the Korean stock market correlations. *Physica A* 361(1):263–271
- Junior LS, Franca IDP (2012) Correlation of financial markets in times of crisis. *Physica A* 391:187–208
- Kim HJ, Lee Y, mook Kim I, Kahng B, (2002) Weighted scale-free network in financial correlations. *J Phys Soc Jpn* 71(9):2133–2136
- Kumar S, Deo N (2012) Correlation and network analysis of global financial indices. *Phys Rev E* 86(2):026101
- Laloux L, Cizeau P, Bouchaud JP, Potters M (1999) Noise dressing of financial correlation matrices. *Phys Rev Lett* 83(7):1467–1470
- Leon SJ (2010) *Linear algebra with applications*. Pearson Education Inc, London, p 235
- Machado JT, Duarte FB, Duarte GM (2011) Analysis of stock market indices through multidimensional scaling. *Commun Nonlinear Sci Numer Simul* 16:4610–4618
- Malliaros FD, Vazirgiannis M (2013) Clustering and community detection in directed networks: a survey. *Phys Rep* 533(4):95–142
- Mantegna RN (1999) Hierarchical structure in financial markets. *Eur Phys J B* 11:193–197
- Mantegna RN, Stanley HE (1999) *Introduction to econophysics: correlations and complexity in finance*. Cambridge University Press, Cambridge, pp 105–106

- Matesanz D, Ortega GJ (2014) Network analysis of exchange data: interdependence drives crisis contagion. *Qua Quan* 48(4):1835–1851
- Mendes RV, Araújo T, a FL, (2003) Reconstructing an economic space from a market metric. *Physica A* 323:635–650
- Newman MEJ (2004) Fast algorithm for detecting community structure in networks. *Phys Rev E* 69:066133
- Nie CX (2020) A network-based method for detecting critical events of correlation dynamics in financial markets. *EPL* 131(5):50001
- Nie CX, Song FT (2018) Constructing financial network based on PMFG and threshold method. *Physica A* 495:104–113
- Nie CX, Song FT (2019) Relationship between entropy and dimension of financial correlation-based network. *Entropy* 20(3):177
- Nie CX, Song FT, Li SP (2016) Rényi indices of financial minimum spanning trees. *Physica A* 444:883–889
- Peralta G, Zareei A (2016) A network approach to portfolio selection. *J Empir Finance* 38:157–180
- Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Stanley HE (1999) Universal and nonuniversal properties of cross correlations in financial time series. *Phys Rev Lett* 83(7):1471–1474
- Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE (2002) Random matrix approach to cross correlations in financial data. *Phys Rev E* 65(6):066126
- Pozzi F, Matteo TD, Aste T (2013) Spread of risk across financial markets: better to invest in the peripheries. *Sci Rep* 3(1):1665–1665
- Sandhu RS, Georgiou TT, Tannenbaum AR (2016) Ricci curvature: an economic indicator for market fragility and systemic risk. *Sci Adv* 2(5):e1501495
- Thomas C, Ruskin HJ, Crane M (2009) Cross-correlation dynamics in financial time series. *Physica A* 388(5):705–714
- Tse CK, Liu J, Lau FCM (2010) A network perspective of the stock market. *J Empir Finance* 17:659–667
- Tumminello M, Aste T, Matteo TD, Mantegna RN (2005) A tool for filtering information in complex systems. *PNAS* 102(30):10421–10426
- Tumminello M, Matteo TD, Aste T, Mantegna RN (2007) Correlation based networks of equity returns sampled at different time horizons. *Eur Phys J B* 55(2):209–217
- Vodenska I, Becker AP, Zhou D, Kenett DY, Stanley HE, Havlin S (2016) Community analysis of global financial markets. *Risks* 4:13
- Wang GJ, Xie C (2015) Correlation structure and dynamics of international real estate securities markets: a network perspective. *Physica A* 424:176–193
- Wang GJ, Xie C, Stanley HE (2018) Correlation structure and evolution of world stock markets: evidence from pearson and partial correlation-based networks. *Comput Econ* 51(3):607–635
- Wiliński M, Sienkiewicz A, Gubiec T, Kutner R, Struzik ZR (2013) Structural and topological phase transitions on the German stock exchange. *Physica A* 392(23):5963–5973
- Yang Y, Yang H (2008) Complex network-based time series analysis. *Physica A* 387:1381–1386
- Yin Y, Shang P (2014) Modified multidimensional scaling approach to analyze financial markets. *Chaos Interdiscip J Nonlinear Sci* 24:022102
- Zhao L, Li W, Cai X (2016) Structure and dynamics of stock market in times of crisis. *Phys Lett A* 380:654–666
- Zheng Z, Podobnik B, Feng L, Li B (2012) Changes in cross-correlations as an indicator for systemic risk. *Sci Rep* 2:srep00888
- Zhu Y, Yang F, Ye W (2016) Financial contagion behavior analysis based on complex network approach. *Ann Oper Res* 2016:1–19