



Monitoring the Dynamic Networks of Stock Returns with an Application to the Swedish Stock Market

Elena Farahbakhsh Touli^{1,2} · Hoang Nguyen³ · Olha Bodnar⁴ 

Accepted: 21 April 2024
© The Author(s) 2024

Abstract

In this paper, two approaches for measuring the distance between stock returns and the network connectedness are presented that are based on the Pearson correlation coefficient dissimilarity and the generalized variance decomposition dissimilarity. Using these two procedures, the center of the network is determined. Also, hierarchical clustering methods are used to divide the dense networks into sparse trees, which provide us with information about how the companies of a financial market are related to each other. We implement the derived theoretical results to study the dynamic connectedness between the companies in the Swedish capital market by considering 28 companies included in the determination of the market index OMX30. The network structure of the market is constructed using different methods to determine the distance between the companies. We use hierarchical clustering methods to find the relation among the companies in each window. Next, we obtain a one-dimensional time series of the distances between the clustering trees that reflect the changes in the relationship between the companies in the market over time. The method from statistical process control, namely the Shewhart control chart, is applied to those time series to detect abnormal changes in the financial market.

Keywords Dynamic network · Hierarchical clustering tree · Stock returns · Tree distance · Swedish capital market

✉ Olha Bodnar
olha.bodnar@oru.se

¹ Department of Mathematics, Stockholm University, Stockholm, Sweden

² Department of Occupational Health, Psychology and Sports Sciences, University of Gävle, Gävle, Sweden

³ Department of Management and Engineering, Linköping University, Linköping, Sweden

⁴ School of Business, Örebro University, Örebro, Sweden

1 Introduction

A financial market is often considered as a network where the nodes are companies and the links among nodes represent the connectedness (Diebold & Yılmaz, 2015). The connectedness of financial assets plays an important role for policymakers and forecasters, especially during recessions and crises (see Minoiu et al., 2015; Bouri et al., 2021; among others). However, there is a controversy on how to define and measure the connectedness as well as how to keep track of changes in it. Vandewalle et al. (2001), Onnela et al. (2004), Bonanno et al. (2004), and Chi et al. (2010) use the Pearson correlation of financial returns as a measure of connectedness in the network. This measure is symmetric and may be subject to the choice of the sample size. Recently, Diebold and Yılmaz (2014) introduce an asymmetric measure of connectedness based on the effects of a shock from one node to other nodes.

In this study, we consider two common measures of connectedness in the networks, namely the Pearson Correlation Coefficient Dissimilarity (PCCD) and the Generalized Variance Decomposition Dissimilarity (GVDD), and derive the distance between stock returns. First, we contribute to the literature by comparing the center of the network determined by these measures. Second, we use a hierarchical clustering method to divide the dense networks into sparse trees. Using the tree representation of the network, it is easier to analyze which companies are more related to each other. Third, we monitor the real-time changes in the tree distance as a signal of changes in the financial network. Lastly, we analyze the 28 biggest companies listed on Sweden's stock exchange to illustrate the pros and cons of the considered connectedness measures.

The center of the network is one of the major interests as it is a node or group of nodes that is most influenced by other nodes in the network. Sensoy and Tabak (2014) define the center of the network as the node with the largest number of adjacent edges or with the largest number of shortest paths going through. Diebold and Yılmaz (2014) analyze the volatility connectedness and define the centrality in terms of the net transmitter of shock. On the other hand, we model the financial returns as a network graph where the distance in the graph is based on the connectedness/dissimilarity measures. Hence, the center of the network can be considered as the node with the shortest distance to the furthest node. We also analyze how the center changes and evolves over time.

Another interesting feature of the financial network is to identify which neighbors or groups of companies are most related to each other. Mantegna (1999) takes advantage of the Pearson correlation to calculate the distance matrix of stock returns and cluster them into groups via a minimum spanning tree. Jung and Chang (2016) apply an agglomerative hierarchical clustering of the Pearson correlation and the partial correlation in the Korean stock market. They find that the traditional sector classifications are insufficient to determine the proximity of companies. Raffinot (2017) proposes hierarchical clustering methods for asset allocation and shows the advantage of hierarchical clustering over the classical asset allocation method. Similar to Mantegna (1999) we use a hierarchical clustering method to convert the networks of companies to rooted trees. The hierarchical clustering algorithm takes

advantage of the distance in the network to merge the most similar nodes into a cluster. The tree structure also highlights the main difference of the implied network based on the two connectedness measures.

In Diebold and Yılmaz (2014), the changes in the network are monitored by using a total sum of connectedness. But that measure ignores the entity of the total changes and somehow misinterprets the real changes in the network structure. Except Jarochnokanan et al. (2022), most of the literature focus on a static clustering method. To get over this shortcoming, we use a tree distance method (see Smith, 2020) based on the generalized Robinson–Foulds distance (see Nye et al., 2006). The generalized Robinson–Foulds distance compares two trees and pairing splits in one tree with similar splits in the other. Hence, we obtain a daily series of distances due to the changes in the tree structures. To the best of our knowledge, we are the first to measure the tree distance in finance and recommend the abnormal change as a warning signal. The results can also be generalized to different stock markets.

Using the returns on stocks traded on the Swedish capital market, we analyze the network of financial returns for 5 years from 2017 to 2022. We consider a rolling window of three months to calculate the daily measures of connectedness. It appears that *Investor*, a Swedish investment company, is the center of the network most of the time for both considered connectedness measures. However, there is a quite difference in hierarchical clustering trees between the two measures. In general, the companies in the same sector are closer together, but the links between sectors are diverse between these two methods. Using both methods, *Swedish Match* has been determined as the center in the least times, namely, 0 times by using the PCCD method and 5 times by the GVDD method. Interestingly, *Swedish Match* is the only company from the Swedish capital market index (OMX30) that belongs to the Consumer Defensive sector. Therefore, it is expected to have a minor impact only on the rest of the companies from OMX30, which has also been documented by the application of the two proposed methods. Finally, we observe that the tree distance computed using Diebold & Yılmaz (2014) is in a higher magnitude and more volatile than the one by the Pearson correlation. However, the tree distances are correlated and in line with the high volatility period of stock returns.

The rest of the paper is organized as follows. Section 2 describes the two common measures of connectedness or dissimilarity. Section 3 introduces the network construction of financial returns by applying the considered measures. Here, we outline the notations of a center in a network as well as describe how to form a hierarchical clustering tree and compute a tree distance. An empirical illustration is presented in Sect. 4 and conclusions are reached in Sect. 5.

2 Dissimilarity Measure of Stock Returns

In this section, we present two common connectedness measures of financial returns based on the Pearson correlation and variance decomposition.

Let $p_{i,t}$ be the closing price of stock i at day t and let $r_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1})$ denote the log-return on stock i on day t . The Pearson correlation coefficients (PCC)

between asset returns i and j at time t is defined using data over M periods from $t_0 = t - M + 1$ up to t as

$$\rho_{i,j}^t = \frac{\sum_{s=t_0}^t (r_{i,s} - \bar{r}_i)(r_{j,s} - \bar{r}_j)}{\sqrt{\sum_{s=t_0}^t (r_{i,s} - \bar{r}_i)^2 \sum_{s=t_0}^t (r_{j,s} - \bar{r}_j)^2}}, \quad (1)$$

where $\bar{r}_i = \frac{1}{M} \sum_{s=t_0}^t r_{i,s}$ and $\bar{r}_j = \frac{1}{M} \sum_{s=t_0}^t r_{j,s}$ are the sample means of the returns on the i and j stocks, respectively, computed over the last M periods. The correlation coefficient receives a value in the range $[-1, 1]$ and it represents the linear dependence between two financial returns. Then, the dissimilarity $h_{i,j}^{t,PCCD}$ between two stock returns i and j at time t can be written as

$$h_{i,j}^{t,PCCD} = \sqrt{2(1 - \rho_{i,j}^t)}. \quad (2)$$

This definition of the dissimilarity satisfies the three axioms that define a metric $\mathbf{H}^{t,PCCD} = [h_{i,j}^{t,PCCD}]$ (see Mantegna, 1999) where $0 \leq h_{i,j}^{t,PCCD} \leq 2$. The dissimilarity expresses the level at which the stocks are correlated (e.g., Onnela et al., 2003).

The Pearson Correlation Coefficient Dissimilarity (PCCD) in (2) measures a relationship among variables and has been employed in many studies, for example, Vandewalle et al. (2001), Onnela et al. (2004), Bonanno et al. (2004) Chi et al. (2010), among others. The PCCD only considers the pairwise linear correlation but ignores other non-linearities created by the time-varying correlations. It is also a non-directional measure that makes it difficult to distinguish the asymmetric effect of one firm from another. Alternatively, Diebold and Yilmaz (2014) propose a measure of the similarity based on the variance decomposition associated with a VAR model which helps to overcome the limitation of the PCCD. The similarity matrix is created based on the shares of forecast error of the returns which allows the measurement of how many percentages of the forecast error of one variable are caused by another variable.

Following Diebold & Yilmaz (2014), a VAR model of order p is used to model the dynamic behavior of asset returns expressed as

$$\mathbf{r}_t = \mathbf{B}_1 \mathbf{r}_{t-1} + \dots + \mathbf{B}_p \mathbf{r}_{t-p} + \boldsymbol{\Sigma}^{1/2} \boldsymbol{\epsilon}_t, \quad (3)$$

where \mathbf{r}_t is a n -dimensional vector of demeaned asset returns; \mathbf{B}_j is a $n \times n$ variate matrix of regression coefficients with $j = 1, \dots, p$; $\boldsymbol{\Sigma}$ is a $n \times n$ covariance matrix that describes the interaction between the components of the error process; $\boldsymbol{\epsilon}_t$ is a n -dimensional vector of error terms that follows a white noise process with zero mean vector and identity covariance matrix.

We rewrite the VAR model (3) in the moving average (MA) representation as

$$\begin{aligned} \mathbf{r}_t &= \boldsymbol{\Theta}(L) \boldsymbol{\Sigma}^{1/2} \boldsymbol{\epsilon}_t, \\ \boldsymbol{\Theta}(L) &= (\mathbf{I} - \mathbf{B}_1 L - \dots - \mathbf{B}_p L^p)^{-1} = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 L + \boldsymbol{\Theta}_2 L^2 + \dots, \end{aligned} \quad (4)$$

where L is the lag operator that is $L\mathbf{r}_t = \mathbf{r}_{t-1}$. To compute the MA representation, the Cholesky factor of the covariance matrix Σ is commonly used together with the generalized variance decomposition (GVD) framework of Koop et al. (1996) and Pesaran and Shin (1998). Note that the GVD helps to produce variance decompositions that are invariant to the order of the variables. Diebold and Yilmaz (2014) consider the standardized variance decomposition matrix $\hat{\mathbf{H}}^t = [\hat{h}_{ij}^t]$ as the shares of the K -step-ahead error variances in forecasting \mathbf{r}_i due to the shocks to \mathbf{r}_j ,

$$\hat{h}_{ij}^t = \frac{v_{ij}^t}{\sum_{j=1}^n v_{ij}^t} \quad \text{with} \tag{5}$$

$$v_{ij}^t = \frac{\sigma_{jj}^{-1} \sum_{k=0}^{K-1} (\mathbf{e}'_i \Theta_k \Sigma \mathbf{e}_j)^2}{\sum_{k=0}^{K-1} (\mathbf{e}'_i \Theta_k \Sigma \Theta'_k \mathbf{e}_i)},$$

where \mathbf{e}_j is a n -dimensional vector with the j th element unity and zeros elsewhere and σ_{jj} is the j -th diagonal element of Σ . Therefore, \hat{h}_{ij}^t measures the connectedness between i and j stock returns based on the variance decomposition of forecast errors. And due to the construction, $\sum_{j=1}^n \hat{h}_{ij}^t = 1, \forall i = 1, \dots, n$ and $\hat{h}_{ij}^t \in [0, 1]$. To convert this variance decomposition matrix or the similarity matrix to a dissimilarity matrix $\mathbf{H}^{t,GVDD} = [h_{ij}^{t,GVDD}]$, we use the following formula:

$$h_{ij}^{t,GVDD} = \sqrt{2(1 - \hat{h}_{ij}^t)}. \tag{6}$$

In this case, if two companies are more related to each other, then $h_{ij}^{t,GVDD}$ is smaller.

3 Graphs

In this section, we consider the network of financial returns using the graph theory. A graph or a network $G(V, E)$ contains a set of vertices (V) and the relation between the nodes which is indicated by E (see, e.g., Cormen et al. 2001; Bondy and Murty 1976). In graphs, sometimes there is more than one edge between some nodes, and also there may be a loop, i.e., an edge from a node to itself. If there is no loop nor multiple edges in the graph we have a *simple graph*. If the relation is symmetric we have an *undirected graph*, otherwise, we have a *directed graph*. If the edges have some numeric values (or weights) we have a *weighted graph*. Based on the measures proposed in Sect. 2, the smaller weights (or lengths) of the edges mean that they are more similar. For each graph, we have some subgraphs as well. A subgraph of a graph $G(V, E)$ is a graph $G' = (V', E')$ such that $V' \subset V$ and $E' \subset E$. Another characteristic of a graph is connectedness. We say that an undirected graph G is *connected*, if and only if there is a path between all the nodes in the graph. A similar definition is also presented for a directed graph which is called *strongly connected graph*. For further definitions and properties from graph theory, we refer to Cormen et al. (2001).

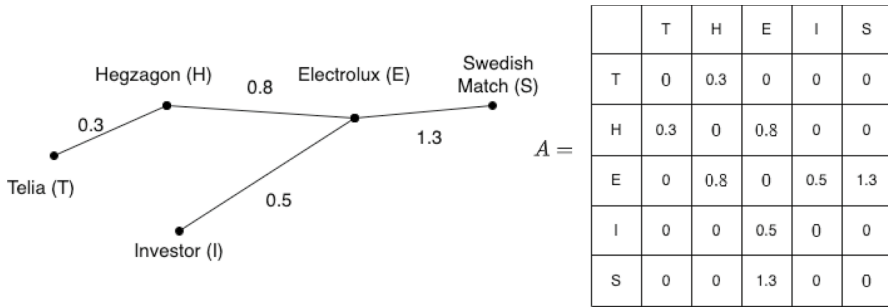


Fig. 1 Undirected graph consisting of five vertices together with the adjacency matrix \mathbf{A} . In the case of the PCCD, the elements A_{ij} of the adjacency matrix \mathbf{A} correspond to the artificial elements of $\mathbf{H}^{t,PCCD}$ for some t which are random numbers between 0 and 2

3.1 Adjacency Matrix

One way to describe a graph is by an adjacency matrix. For a given graph with n vertices V_1, V_2, \dots, V_n , the adjacency matrix A is an n by n matrix such that the numbers of rows and columns are equal to the number of vertices V . Then, the element (i, j) in the adjacency matrix is w_{ij} (or 0) if there is (or is not) an edge between vertices V_i and V_j with the weight of w_{ij} . The adjacency matrix is symmetric if the graph is undirected which means that $A_{ij} = A_{ji}$. In Fig. 1, an illustration of a graph is presented, which is constructed by using the artificial returns on five stocks traded on the Swedish capital market.

Contrariwise, for any squared matrix \mathbf{A} we can construct a graph in which the number of vertices of the graph is equal to the number of columns of the matrix, and between any two vertices V_i and V_j we add an edge with the weight of A_{ij} . This matrix \mathbf{A} is the adjacency matrix of the graph. In the following, we consider two dissimilarity measures $\mathbf{H}^{t,PCCD}$ in (2) and $\mathbf{H}^{t,GVDD}$ in (6) as the adjacency matrices to identify the connectedness between Swedish companies. The adjacency matrix that we obtain from the PCCD is a symmetric matrix and we construct an undirected graph. Similarly, we construct an asymmetric adjacency matrix and a directed graph by using the GVDD measure.

3.2 Distance Matrix

For a graph, we can construct the distance matrix by using the adjacency matrix. The distance between two vertices V_i and V_j is the summation of weights of all the edges in the shortest path from V_i to V_j . The distance matrix D for a graph G is a squared matrix such that the number of columns is the number of vertices of the graph and each element D_{ij} indicates the distance from vertex V_i to V_j . If the graph is undirected, the distance matrix is a symmetric matrix, otherwise, the matrix is asymmetric. Figure 2 depicts the undirected graph and the distance matrix obtained by using the results presented in Fig. 1.

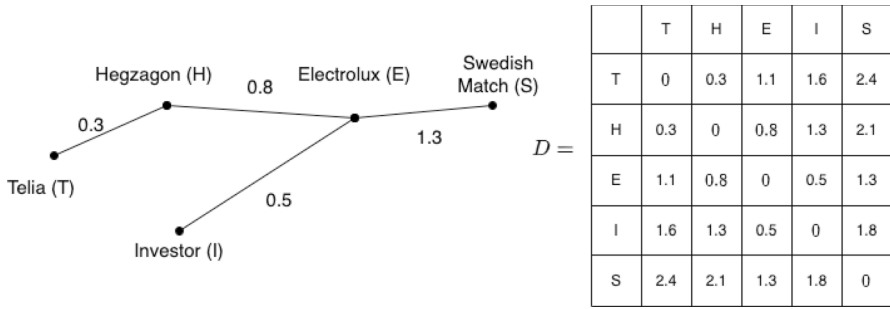


Fig. 2 An undirected graph consisting of five vertices together with the distance matrix D

In this section, we are interested in graphs and the clustering of the networks that we make by using the variance decomposition matrix. Networks and graphs are two important topics in the field of statistics and finance and have attracted lots of attention (see Mantegna 1999; Diebold and Yılmaz 2014; Cerbo and Taylor 2021; Touli and Lindberg 2022).

3.3 Center of a Graph

There are some characteristics of graphs that we use in this paper. One of the important properties of the graph is the center of the graph which we define below (see also Wasserman and Faust (1994) for more information and definition related to the center of the graph).

In a network of companies, we consider that the center of the network is a vertex (or a set of vertices) in the graph that has a minimum value of the maximum distances from it (them) to other vertices. For finding the center of a graph, we add a column (or a row) to the distance matrix of the graph, called *max*, whose elements max_i indicate the maximum distance from the i -th vertex to other vertices. The center of the graph is then the vertex (a set of vertices) that has the minimum value at the *max* column.

Figure 3 illustrates the computation of the center of the undirected graph of Fig. 2, which appears to be *Electrolux*.

Besides, based on the results discussed in the previous section, it can be proven that the PCCD is a distance. Therefore, at each window t and for any pair of i and j , $h_{ij}^{t,PCCD}$ is the shortest distance between two vertices V_i and V_j . Therefore, for the PCCD method, the distance matrix is equal to the adjacency matrix.

3.4 Hierarchical Clustering of a Data

As the graph of the financial assets is a presentation of dense network structure, analyzing and reporting their properties based on different connectedness measures can be very difficult. Instead, we work with a tree clustering that indicates the relation between companies. There are many methods to cluster the data, for example, flat

Fig. 3 Determination the center of the graph in Fig. 2. The added column to the distance matrix \mathbf{D} is *max*. The minimum value of this column, depicted in red, specifies *Electrolux* as the center of the graph

$D =$

	T	H	E	I	S	Max
T	0	0.3	1.1	1.6	2.4	2.4
H	0.3	0	0.8	1.3	2.1	2.1
E	1.1	0.8	0	0.5	1.3	1.3
I	1.6	1.3	0.5	0	1.8	1.8
S	2.4	2.1	1.3	1.8	0	2.4

clustering and hierarchical clustering (see Dasgupta 2016; Wang and Wang 2020). We focus on hierarchical clustering in this paper because of some advantages of the method. The first advantage of using hierarchical clustering is that we do not need to indicate the number of clusters before starting the clustering. Another advantage is that the structure of the cluster is a tree and therefore we can use some properties of trees, such as the distance between trees. In the hierarchical clustering tree, the leaves correspond to the firms and each internal node corresponds to a cluster such that all the data in one cluster are indicated by the leaves of the subtree rooted at the internal node.

There are two methods for hierarchical clustering: *agglomerative* and *divisive* (see Dasgupta 2016; Wang and Wang 2020). The algorithm for the divisive method is more complicated than the one for the agglomerative method. Moreover, most of them are NP-hard to compute which means that there is no known polynomial time algorithm for implementing them. As such, we make use of agglomerative methods. We first start with vertices that are more similar to each other and then merge them until we reach the groups that are less similar to each other. At last, we merge even those groups that are completely different from each other. In all kinds of hierarchical clustering, all the groups are merged eventually.

In this paper, we choose to work with a single linkage clustering algorithm that is efficient and suitable for the symmetric distance. When the matrix is asymmetric, then we consider the max between the element uv and vu in the asymmetric matrix and we convert it to a symmetric matrix. Then, the methods that exist for symmetric matrices, are employed (see Carlsson et al. 2018 for details).

3.5 Distance Between Trees

In the previous section, different types of clustering methods were introduced for the stock returns. As in each period, a hierarchical clustering tree is obtained and the changes in the hierarchical clustering tree can be summarized by using the distances between the trees. We start this section by introducing some methods for calculating the distance between trees.

The *tree edit distance* and the *tree alignment distance* are two distances that are primarily defined between trees (see Bille, 2005). Furthermore, the *interleaving distance* and the *Frechet-like distance* are defined between *merge trees*¹ (see e.g., Morozov et al. 2013; Touli 2021). Recently, the interleaving distance was generalized in Touli and Wang (2022), which proposed a fixed parameter tractable algorithm for finding the interleaving distance between two merge trees. The generalized Robinson–Foulds metrics for comparing and finding the similarity between *phylogenetic trees*² has been worked by M.R. Smith. The practical computation of the distance between trees can be performed by using the R package *TreeDist* (see Smith, 2020).

3.5.1 Robinson–Foulds Distance

Robinson–Foulds distance is a distance that is defined on unrooted labeled trees. Each edge in a tree is a *bridge*³ that divides the leaves of a labeled tree into two groups such that there is no overlapping between them. The Robinson–Foulds algorithm counts the number of splits in one tree that do not exist in another one (see Bogdanowicz and Giaro 2011; Smith 2020). In other words, it is defined by

$$d_{RF}(T_1, T_2) = \frac{1}{2} |\psi(T_1) \ominus \psi(T_2)|$$

such that $\psi(T_1)$ is the set of all splits related to edges of T_1 , similar for $\psi(T_2)$. Also, for two sets A and B , $A \ominus B = (A \setminus B) \cup (B \setminus A)$.

Since the set of rooted trees is a subset of trees, we can also use the above definition for labeled rooted trees. Moreover, the hierarchical clustering trees are labeled rooted trees. As such, we can find the dissimilarity between them or in other words the distance between them by using the Robinson–Foulds distance.

The Robinson–Foulds method does not provide an acceptable result when there is a small change in trees, for example, when the difference between the two trees T_1 and T_2 is that just one leaf in T_1 moves in T_2 like Fig. 4. In this case, the Robinson–Foulds distance returns a very large number that indicates that the two trees are not similar. Therefore, the *generalized Robinson–Foulds* method was introduced in Smith (2020).

Smith (2020) introduced three information-based distances between the phylogenetic trees. As the distance between the rooted trees indicates the relationship between the clustered markets, the *clustering information distance* is the most suitable one that we can use here. The phylogenetic trees and the hierarchical trees are very similar in structure. In both of them, the leaves of the tree save the information about the data. In the phylogenetic trees, we have the names of species, while the

¹ From Morozov et al. (2013) a merge tree is a rooted tree with a real-valued function which is defined on the tree. The function is monotonically decreasing from the root of the leaves.

² A phylogenetic tree is a rooted labeled tree such that the tree indicates the evolutionary relationships between different species.

³ In a graph, a bridge is an edge that if we cut it, the graph is divided into two separate graphs.

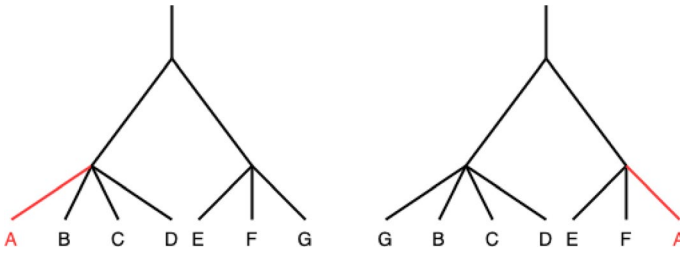


Fig. 4 Two trees with similar structures (just one leaf has been changed), but with large Robinson–Foulds distance

names of the companies are used in the hierarchical tree that we have constructed from the relationship between these companies. Also, in phylogenetic trees, the nearest common ancestor of similar species is closer to them rather than the different species. In the hierarchical clustering trees, we have a similar situation as well. Namely, if two companies are more related, then they are merged faster than the ones that are more different. Therefore, in this work, we use the distance that is defined on the phylogenetic trees for finding the distance between the hierarchical clustering.

4 Empirical Illustration

In this section, we consider 28 Swedish companies. We analyze the network structure of the asset returns through the PCCD method and the GVDD method. Then, we find the center of the networks. Also, by using hierarchical clustering and the information distance between rooted trees, we investigate the changes in the hierarchical trees.

We first take the adjusted closing prices of 28 Swedish companies from Yahoo Finance for 5 years, from March 31st, 2017 to March 30th, 2022. In the analysis, a moving window of three months is employed. Commonly, three months have 63 open days and, therefore, we consider the first 63 days as the first window. Then we shift by one day and the second window starts from day two and ends at day 64, and so on. For each window we use two methods to find the adjacency matrices for these companies: (i) PCCD which constructs a symmetric matrix and therefore undirected graph, and (ii) GVDD which constructs an asymmetric matrix and a directed graph, respectively.

4.1 Networks of Financial Returns

Using the two proposed methods of determining the adjacency matrix, we find the center of the graphs at each window and compute the frequency of each company to be the center during the past 5 years.

Figures 5 and 6 depict the centers of the network determined in each window by applying the GVDD method and the PCCD method, respectively. In both figures,

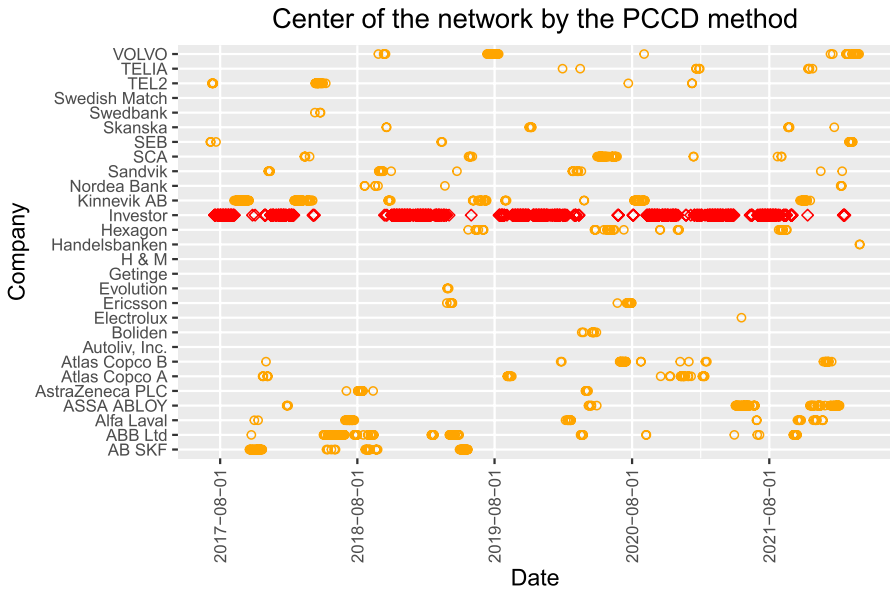


Fig. 5 The center of the network during the period from March 31st, 2017 to March 30th, 2022 by using the PCCD method

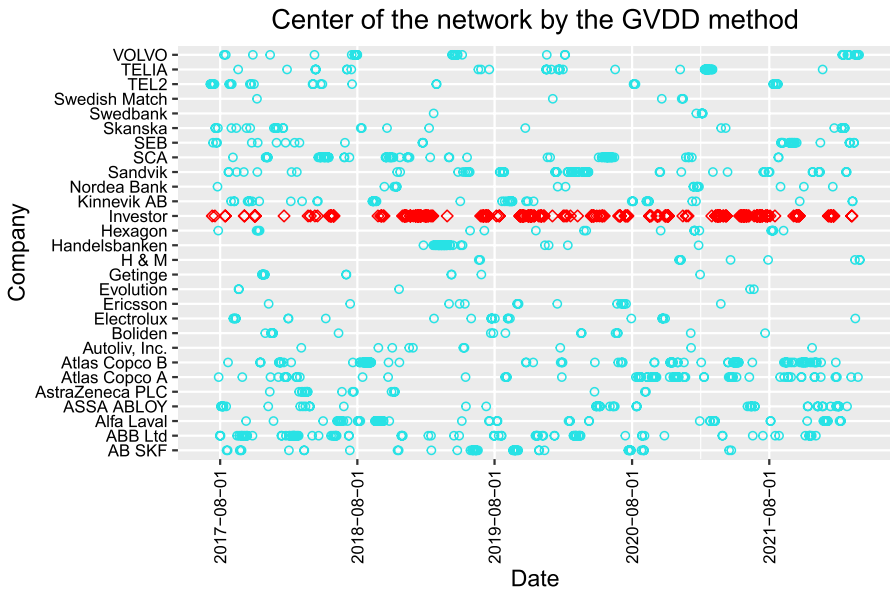


Fig. 6 The center of the network during the period from March 31st, 2017 to March 30th, 2022 by using the GVDD method

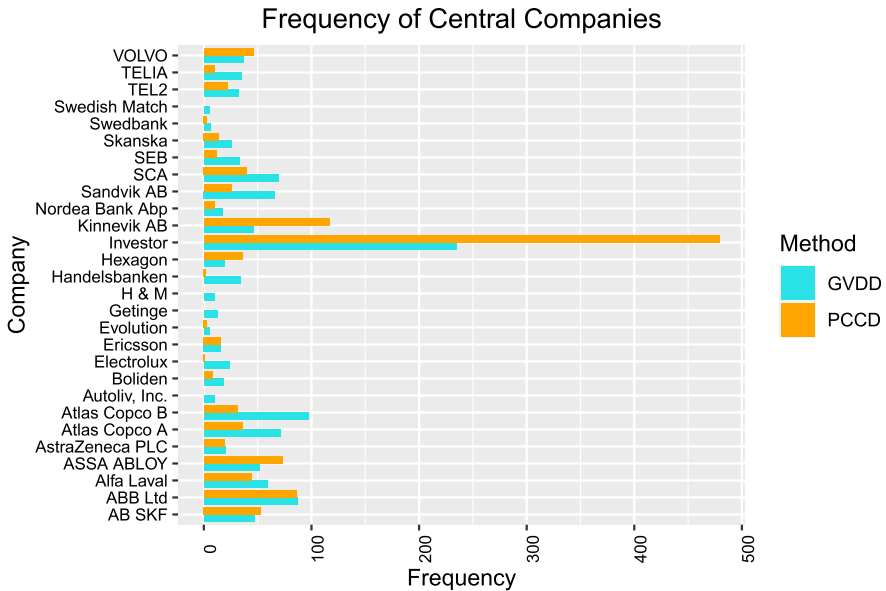


Fig. 7 The frequency of the companies to be the center of the graphs using the PCCD method (blue) and the GVDD method (yellow)

we observe that the center of the network is time-dependent with a larger number of changes present when the GVDD approach is used. Also, the number of companies determined as the center of the network is larger for the GVDD method. In the case of the PCCD approach, the center of the network shows more stable behavior. Finally, both approaches select the company *Investor* in the majority of times.

By definition, the center of the graphs is a company that has the most influence on stock returns of all the other companies in the shortest time. Figure 7 presents the absolute frequencies of each company to be the center of the network. The computations are performed by the GVDD method and the PCCD method.

In Fig. 7, we observe that the highest frequency of the center happens for the company called *Investor* by both methods. It means that during the considered period, *Investor* was most of the time the center of the companies between the 28 companies that we chose. Therefore, as *Investor* is an investment company, the financial industry has the most influence on all the other companies during the considered time in Sweden. The companies with the second and third highest frequencies are *Atlas Copco B* and *ABB Ltd* for the GVDD method and *Kinnevik AB* and *ABB Ltd* for the PCCD method, respectively. Furthermore, *Swedish Match*, *H&M*, *Getinge*, and *Autoliv, Inc.* have never been determined as the center of the network by the PCCD methods, while all companies have been chosen at least one time to be the center of the network by the GVDD method with *Swedish Match* having the smallest absolute frequency equal 5.

If we consider the data from 31 January 2020 to 31 July 2020, which is the time that COVID-19 was started, then *Svenska Cellulosa Aktiebolaget (SCA)* was the most popular center using the PCCD method and *Investor* remained

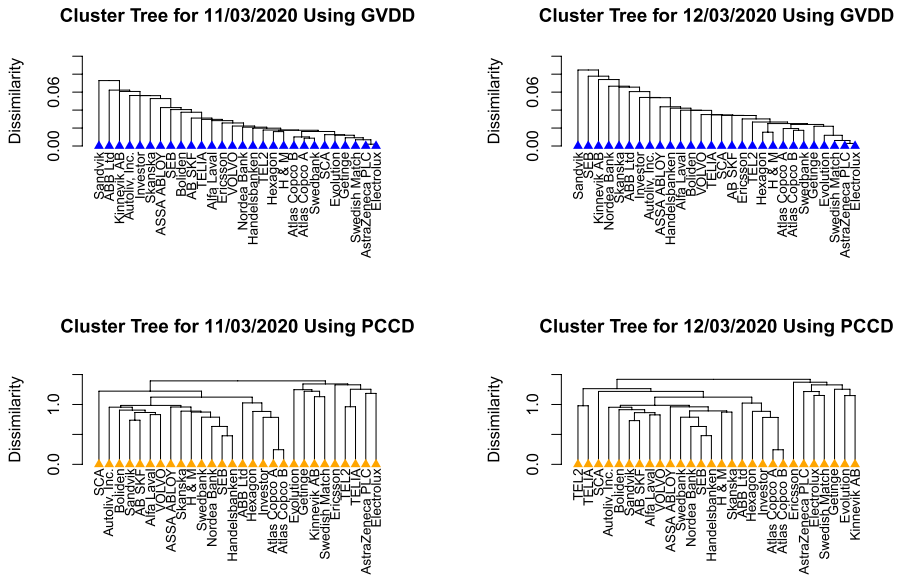


Fig. 8 Hierarchical clustering trees for window number 670 (March 11th, 2020) when the GVDD method (first row) and the PCCD method (second row) are used. Hierarchical clustering trees obtained by both methods for the window number 671 are depicted in the second column. During the last 5 years, these days had the most different hierarchical trees by using the GVDD method

the most popular center by the GVDD method. Also, from 1 October 2020 to 10 May 2021, which corresponds to the time when the Coronavirus Delta variant was dominant, *Investor* was indicated to be the most popular center by both methods. Finally, from 1 December 2021 until the last day in the data *ASSA ABLOY AB* was the most popular center by using both methods.

4.2 Hierarchical Clustering Tree

Figures 8 and 9 depict the hierarchical clustering trees computed for subsequent days by the PCCD method and the GVDD method. We chose the date that observed the largest changes in the information distance between two hierarchical trees. It took place on 10/02/2021 in the case of the PCCD method and on 11/03/2020 in the case of the GVDD method.

By looking at the figures, we see that there is a difference between the considered two methods, especially between those hierarchical clustering trees which correspond the largest differences in two consequent days by each of the methods. The height of the hierarchical distance happened on two different days. Furthermore, we see that on average the height of the hierarchical trees in the PCCD method is larger than the one in the GVDD method.

4.3 Distance Between Hierarchical Clustering Trees

In the previous section, by using the single linkage hierarchical clustering for the PCCD method and for the GVDD method we construct the hierarchical clustering of the stock data at each window. Using the distance defined in Sect. 3.5, we compute the distance between the hierarchical trees in this section sequentially. The results are depicted in Fig. 10.

The three lines in Fig. 10 demonstrate the behavior of an autoregressive process. To study this effect and also to investigate possible (lag) relationships between the three-time series we fit a vector autoregressive model (VAR) to these series. First, the order of the autoregressive model is chosen by using the Hannan and Quinn model selection criteria (see, Hannan & Quinn, 1979), which results in two. Second, we fit a VAR(2) model to $(\{PCCD_t\}, \{GVDD_t\}, \{OMX_t\})$, which leads to the following multivariate model

$$\begin{pmatrix} PCCD_t \\ GVDD_t \\ OMX_t \end{pmatrix} = \begin{pmatrix} 1.72767^{***} \\ 5.03862^{***} \\ 0.00146 \end{pmatrix} + \begin{pmatrix} 0.20913^{***} & 0.04571^* & -10.91973^* \\ 0.19580^{***} & 0.31478^{***} & -25.64787^{**} \\ 0.00009 & -0.00011 & -0.04697 \end{pmatrix} \begin{pmatrix} PCCD_{t-1} \\ GVDD_{t-1} \\ OMX_{t-1} \end{pmatrix} \\ + \begin{pmatrix} 0.22751^{***} & 0.01034 & 3.84118 \\ 0.14021^{**} & 0.00758 & 11.12222 \\ 0.00014 & -0.00011 & -0.01427 \end{pmatrix} \begin{pmatrix} PCCD_{t-2} \\ GVDD_{t-2} \\ OMX_{t-2} \end{pmatrix} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a white noise process with covariance matrix given by

$$\Sigma = \begin{pmatrix} 4.26874 & 1.21287 & -0.00024 \\ 1.21287 & 11.87265 & -0.00321 \\ -0.00024 & -0.00321 & 0.00014 \end{pmatrix}.$$

In the model equation, the coefficients denoted with ‘***’ are statistically significant at 0.1%, ‘**’—at 1%, ‘*’—at 5%, and ‘ ’—at 10%. We observe that the current values of tree distances constructed by using the PCCD and GVDD methods are positively correlated with their previous values. Moreover, the values obtained by using the PCCD method have also an impact on future values obtained for both the PCCD method and the GVDD method at lag 2. While the previous distances are positively correlated with the future ones, the values of the Swedish capital market index, OMX, have a negative significant impact at lag 1. To this end, we note the OMX index cannot be predicted by any of the distances considered in the study nor by the previous values of the index itself.

In Fig. 10, we compare the PCCD and GVDD methods and we see that, in general, the application of the GVDD method leads to larger values of the distances between the trees over time, which means that in the GVDD methods, the hierarchical clustering trees differ more than in the case when the hierarchical clustering trees are constructed by using the PCCD method. Also, the average value of the distances is larger when the GVDD method is used in comparison to the PCCD method. Both plots in Fig. 11 depict the series with the distances together with the mean line and 5× standard deviation (SD) line. The comparison of the pattern of

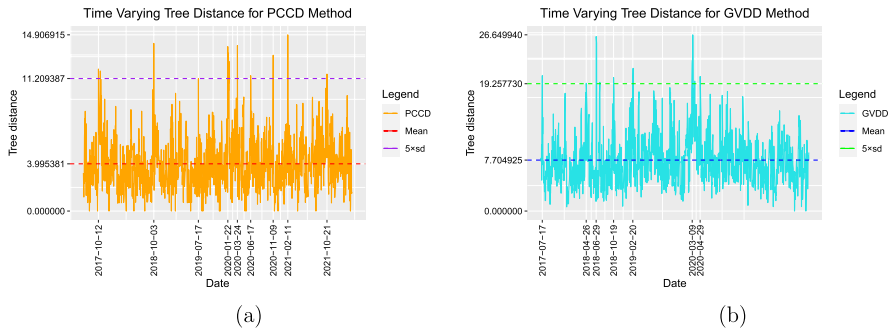


Fig. 11 **a** Distances between the hierarchical clustering trees computed by using the PCCD method. The purple dashed line indicates $5\times SD$ and the red dashed line is the mean line. **b** Distances between the hierarchical clustering trees were obtained by using the GVDD method. The green dashed line indicates $5\times SD$ and the blue dashed line is the mean line

computed distances with $5\times SD$ corresponds to the application of the Shewhart control chart for detecting changes in statistical process control (see e.g., Psarakis and Papaleonida 2007; Bodnar and Schmid 2011; Bisiotis et al. 2022). We see that most of the time the jumps of magnitude larger than $5\times SD$ happen after the beginning of the year 2020 in both plots. Also, the number of jumps that are higher than $5\times SD$ is larger when the GVDD method is used. Moreover, 7 out of 20 jumps that are above the green line happened in 2020 during the COVID-19 time.

5 Conclusion

Specifying and monitoring the structure of the financial market is an important research topic with direct applications to real-life problems. The knowledge about the center of the capital market is useful in the determination of its stability while clustering the companies on the capital market provides us information about the connectedness of the companies traded on it. The above-mentioned topics are treated in the literature by constructing a graph and determining its center as well as by finding the hierarchical clustering trees.

In the paper, two methods are compared for determining the network of companies traded on the Swedish capital market. While the first approach, the PCCD method, is based on a symmetric adjacency matrix, the second one, the GVDD method, employs an asymmetric adjacency matrix. Both methods indicate the company *Investor* as the center of the Swedish capital market in most of the considered cases. On the other side, the company *Swedish Match* shows the largest dissimilarity to all other companies traded on the Swedish stock exchange. Finally, computing the distances between the hierarchical clustering trees we found that most of the changes in the structure of the Swedish capital market happen at the beginning of 2020, i.e., during the first COVID-19 wave.

Acknowledgements The authors would like to thank Professor Hans Amman and the two anonymous Reviewers for their constructive comments that improved the quality of this paper.

Author Contributions The contribution of each author is proportional to the number of authors.

Funding Open access funding provided by Örebro University. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

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

References

- Bille, P. (2005). A survey on tree edit distance and related problems. *Theoretical Computer Science*, 337(1–3), 217–239.
- Bisiotis, K., Psarakis, S., & Yannacopoulos, A. N. (2022). Control charts in financial applications: An overview. *Quality and Reliability Engineering International*, 38(3), 1441–1462.
- Bodnar, O., & Schmid, W. (2011). CUSUM charts for monitoring the mean of a multivariate Gaussian process. *Journal of Statistical Planning and Inference*, 141(6), 2055–2070.
- Bogdanowicz, D., & Giaro, K. (2011). Matching split distance for unrooted binary phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(1), 150–160.
- Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B*, 38(2), 363–371.
- Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications*. New York: Elsevier.
- Bouri, E., Cepni, O., Gabauer, D., & Gupta, R. (2021). Return connectedness across asset classes around the COVID-19 outbreak. *International Review of Financial Analysis*, 73, 101646.
- Carlsson, G., Mémoli, F., Ribeiro, A., & Segarra, S. (2018). Hierarchical clustering of asymmetric networks. *Advances in Data Analysis and Classification*, 12(1), 65–105.
- Cerbo, L. F. D., & Taylor, S. (2021). Graph theoretical representations of equity indices and their centrality measures. *Quantitative Finance*, 21(4), 523–537.
- Chi, K. T., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, 17(4), 659–667.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). The MIT Press.
- Dasgupta, S. (2016). A cost function for similarity-based hierarchical clustering. In *Proceedings of the forty-eighth annual ACM symposium on theory of computing*, STOC '16 (pp. 118–127), New York, NY, USA. Association for Computing Machinery. ISBN 9781450341325. <https://doi.org/10.1145/2897518.2897527>.
- Diebold, F. X., & Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1), 119–134.
- Diebold, F. X., & Yilmaz, K. (2015). *Financial and macroeconomic connectedness: A network approach to measurement and monitoring*. Oxford University Press.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2), 190–195.

- Jaroonchokanan, N., Termsaithong, T., & Suwanna, S. (2022). Dynamics of hierarchical clustering in stocks market during financial crises. *Physica A: Statistical Mechanics and its Applications*, *607*, 128183.
- Jung, S. S., & Chang, W. (2016). Clustering stocks using partial correlation coefficients. *Physica A: Statistical Mechanics and its Applications*, *462*, 410–420.
- Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, *74*(1), 119–147.
- Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, *11*(1), 193–197.
- Minoiu, C., Kang, C., Subrahmanian, V., & Berea, A. (2015). Does financial connectedness predict crises? *Quantitative Finance*, *15*(4), 607–624.
- Morozov, D., Beketayev, K., & Weber, G. H. (2013). Interleaving distance between merge trees. In *Workshop on topological methods in data analysis and visualization: Theory, algorithms and applications*.
- Nye, T. M., Lio, P., & Gilks, W. R. (2006). A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics*, *22*(1), 117–119.
- Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. (2003). Asset trees and asset graphs in financial markets. *Physica Scripta*, *2003*(T106), 48.
- Onnela, J.-P., Kaski, K., & Kertész, J. (2004). Clustering and information in correlation based financial networks. *The European Physical Journal B*, *38*(2), 353–362.
- Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, *58*(1), 17–29.
- Psarakis, S., & Papaleonida, G. (2007). Spc procedures for monitoring autocorrelated processes. *Quality Technology & Quantitative Management*, *4*(4), 501–540.
- Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, *44*(2), 89–99.
- Sensoy, A., & Tabak, B. M. (2014). Dynamic spanning trees in stock market networks: The case of Asia-Pacific. *Physica A: Statistical Mechanics and its Applications*, *414*, 387–402.
- Smith, M. R. (2020). Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. *Bioinformatics*, *36*(20), 5007–5013.
- Touli, E. F. (2021). Fréchet-like distances between two rooted trees. *Journal of Algorithms and Computation*, *53*(1), 1–12.
- Touli, E. F., & Lindberg, O. (2022). Relative clustering coefficient. *Journal of Algorithms and Computation*, *54*, 99–108.
- Touli, E. F., & Wang, Y. (2022). FPT-algorithms for computing Gromov–Hausdorff and interleaving distances between trees. *Journal of Computational Geometry*, *13*, 89–124.
- Vandewalle, N., Brisbois, F., & Tordoior, X. (2001). Non-random topology of stock markets. *Quantitative Finance*, *1*(3), 372.
- Wang, D., & Wang, Y. (2020). An improved cost function for hierarchical cluster trees. *Journal of Computational Geometry*, *11*(1), 283–331.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge University Press.

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