Cluster fusion-fission dynamics in the Singapore stock exchange

Teh, Boon Kin; Cheong, Siew Ann

2015

Teh, B. K., & Cheong, S. A. (2015). Cluster fusion-fission dynamics in the Singapore stock exchange. *The European Physical Journal B*, 88(10), 263-.

<https://hdl.handle.net/10356/82803>

<https://doi.org/10.1140/epjb/e2015-60456-y>

© 2015 EDP Sciences, Societa Italiana di Fisica. This is the author created version of a work that has been peer reviewed and accepted for publication in *The European Physical Journal B*, published by Springer-Verlag on behalf of EDP Sciences, Societa Italiana di Fisica. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at:
[<http://dx.doi.org/10.1140/epjb/e2015-60456-y>].

Downloaded on 26 Apr 2023 22:12:23 SGT

Cluster Fusion-Fission Dynamics in the Singapore Stock Exchange

Boon Kin Teh · Siew Ann Cheong

Received: date / Accepted: date

Abstract In this paper, we investigate how the cross-correlations between stocks in the Singapore Stock Exchange (SGX) evolve over 2008 and 2009 within overlapping one-month time windows. In particular, we examine how these cross-correlations change before, during, and after the Sep-Oct 2008 Lehman Brother Crisis. To do this, we extend the complete-linkage hierarchical clustering algorithm, to obtain robust clusters of stocks with stronger intracluster correlations, and weaker intercluster correlations. After we identify the robust clusters in all time windows, we visualize how these change in the form of a fusion-fission diagram. Such a diagram depicts graphically how the cluster sizes evolve, the exchange of stocks between clusters, as well as how strongly the clusters mix. From the fusion-fission diagram, we see a giant cluster growing and disintegrating in the SGX, up till the Lehman Brothers Crisis in September 2008 and the market crashes of October 2008. After the Lehman Brothers Crisis, clusters in the SGX remain small for few months before giant clusters emerge once again. In the aftermath of the crisis, we also find strong mixing of component stocks between clusters. As a result, the correlation between initially strongly-correlated pairs of stocks decay exponentially with average life time of about a month. These observations impact strongly how portfolios and trading strategies should be formulated.

Keywords Lehman Brothers Crisis · Singapore Stock Exchange · hierarchical clustering · interaction hierarchical clustering · mixing · persistence

1 Introduction

A magnitude-9 submarine earthquake hit Japan on 11 March 2011, triggering a massive tsunami that swept over eastern Japan and killing 15,000 people [1]. Combining massive data sets and computational models, earth scientists pieced together an exquisite picture of what happened at the fastest time scales [2–4]. This is considered by many a leap forward in our understanding of earthquake dynamics and processes. With the help of high-resolution data, earth scientists are generally confident of long-term (decades to centuries) and mid-term (years) earthquake assessments [5]. However, short-term (days to months) forecasting remain controversial [6]. Many have liken market crashes to earthquakes [7–18]. We cannot help but wonder whether data and models can point out early warning signatures of crises in the financial market?

Financial markets have been around for many centuries, and we have collected significantly more data on market crashes than we have for large earthquakes. Yet we are less confident with short-term prediction of market crashes than with short-term prediction of earthquakes. Compared to the data on earthquakes, the data we have available on financial markets is more complete, larger in volume, and higher in time resolution, thanks to the fact that all major exchanges in the world are computerized. But still there are

Boon Kin Teh

1. Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Republic of Singapore.

2. Complexity Institute, Nanyang Technological University, Block 2 Innovation Centre, Level 2 Unit 245, 18 Nanyang Drive, Singapore 637723, Republic of Singapore.

E-mail: S130005@e.ntu.edu.sg

Siew Ann Cheong

1. Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Republic of Singapore.

2. Complexity Institute, Nanyang Technological University, Block 2 Innovation Centre, Level 2 Unit 245, 18 Nanyang Drive, Singapore 637723, Republic of Singapore.

E-mail: cheongsa@ntu.edu.sg

major financial downturns, such as the Asian financial crisis in 1997, the Subprime Crisis in 2007, and the Euro Zone Crisis in 2009 [19–21]. Why is this so? In contrast to the quantitative Gutenberg-Richter Law, Omori’s Law, and Bath’s Law [22–24], the stylized facts that we have discovered about financial markets, for example the fat tails in short-time financial returns [25–27], are mostly qualitative in nature. In addition, financial markets are highly heterogeneous, highly nonlinear, and strongly interacting [28, 29], and also notoriously non-stationary at human time scales [30, 31]. Before we can tackle the problem of short-term prediction of market crashes, we need to first find clever ways to analyze and understand the constant rapid changes within financial markets.

In the econophysics literature a lot of work has been done to understand the dynamics of financial market. In 1999, Laloux et al. found that the noise components in financial returns can be described by the theory of random Wishart matrices [32–34], and for the very first time allows us to discriminate between signal and noise. Following this, Mantegna showed that the signals in the financial returns actually point to the hierarchical organization and structures [35], which can be discovered by constructing the MST and hierarchical clustering [36–39]. Indeed, other groups confirmed the existence of these hierarchical structures, which they discovered across different financial markets using more sophisticated methods [40–42]. More importantly, researchers found that there are structural changes in the financial markets, such as their distributions of returns and their MST structures [43–47] after the financial crisis. In addition, researchers also discovered precursor patterns prior to financial crises, such as an increase in the volatility of financial returns and the cross-correlations between stocks [48–50], as well as an increase in the information dissipation level [51]. These observations not only can be found in stock markets, but also for interest rate swap and currency exchange [52].

Although there are now many previous studies on how the financial hierarchy changes before and after the financial crisis [53–55] our understanding of such changes remain fragmentary, as only a few previous works focus on the dynamics of the hierarchical structures [56–58]. In this paper our goal is hence to study the dynamics of hierarchy changes systematically. To do this, we construct a series of dynamical cross-correlation matrices in sliding windows using high-frequency data (Section 2.1). Candidates for financial clusters are first determined using complete-linkage hierarchical clustering (CLHC) in Section 2.2, and thereafter the robust clusters are identified using the *interaction hierarchical clustering* (IHC) procedure that we developed (Section 2.3). In this extension of the CLHC, we cluster CLHC clusters by looking at the off-diagonal (interaction) elements of the correlation matrix between two CLHC clusters. Finally, after CLHC clustering and subsequent IHC recluster of the CLHC clusters, we explain how thresholds can be identified to obtain the most robust clusters (Section 2.4). We then introduce the *fusion-fission diagram* as a visualization tool to help us understand how these robust clusters evolve from time window to time window (Section 2.5). In Section 3, we describe our findings on the Singapore market across 2008 and 2009. We observed the emergence of giant cluster that grow and dissipate quasiperiodically, and finally culminating in the October crashes that were seen also in other world markets. We also observed in the fusion-fission diagram strong mixing between clusters and hence proceeded to quantify this degree of mixing (Section 4.1). Finally, we look at how persistent the clusters are, and how long small clusters of stocks remain strongly-correlated across repeated episodes of fusions and fissions (Section 4.2). We found that the mixing degree in the Singapore Stock Exchange is high, and on average, a pair of stocks remain strongly-correlated only for about two months. These findings have important implications on how trading strategies should be designed on the Singapore Stock Exchange. We conclude our paper in Section 5.

2 Data and Methods

2.1 Data

In this study, we work with data from the Singapore Stock Exchange (SGX). There are three reasons for this choice. First, even though the model market econophysicsts study is the New York Stock Exchange (NYSE), we are much more familiar with the stocks listed on the SGX, and the profiles and outlooks of their companies. Second, the SGX is a small market with fewer than 1000 stocks, making it possible to do a full cross-section analysis. Third, Singapore has a very open economy, and thus the SGX reacts readily to global market movements. Based on the Straits Times Index (STI) that tracks the 30 top performing stocks listed on the SGX, we know that the bourse reacted to the Lehman Brothers Crisis, losing 45% of its value from early August to the end of October 2008. Therefore, we chose the period of study to be from January 2008 to December 2009, roughly centered around the Lehman Brother Crisis.

Tick-by-tick data for more than 800 SGX stocks were downloaded from the Thomson-Reuters Tick History Database (Url: <http://thomsonreuters.com/tick-history/>). We first filtered out stocks with very low trading frequencies, before extracting from the tick-by-tick data price time series at five-minute intervals. After cleaning the data, 561 stocks remains, with 40,240 time intervals corresponding to 503 trading days.

To understand how the SGX dynamics evolve across the market crashes, we break the 40,240-point time series into one-month time windows that overlap with their predecessors and successors by two weeks. Each time window contains more than 2,000 data points, which is sufficient for computing reliable cross-correlation coefficients, as $Q = T/N = 2000/561 \approx 4 > 1$ [59]. The overlapping time windows also help us track hierarchy changes in the market in a smooth fashion.

2.2 Hierarchical Clustering

In this study we compute the Pearson correlation coefficient between stock prices. Given two stocks i and j , with price time series $\mathbf{P}_i = (P_i(1), P_i(2), \dots, P_i(t), \dots, P_i(T))$ and $\mathbf{P}_j = (P_j(1), P_j(2), \dots, P_j(t), \dots, P_j(T))$, their *Pearson* or *linear* cross-correlation is given by

$$C_{ij} = C_{ji} = \frac{\text{cov}(\mathbf{P}_i, \mathbf{P}_j)}{\sigma_i \sigma_j}, \quad (1)$$

where σ_i is the standard deviation of $\mathbf{P}_i$, σ_j is the standard deviation of $\mathbf{P}_j$ and $\text{cov}(\mathbf{P}_i, \mathbf{P}_j)$ is the covariance between $\mathbf{P}_i$ and $\mathbf{P}_j$. These cross-correlations can be organized into a cross-correlation matrix $\mathbf{C}$ and visualized as a color map. In such color maps, red represents strong correlations, blue indicates strong anti-correlations, and green means nearly uncorrelated.

Figure 2(a) shows the color map of an actual cross-correlation matrix, which contains hidden structures that we want to discover. In the literature we find many correlation filtering methods to discover these hidden structures: threshold filtering [60,61], the minimum spanning tree (MST) [35], and the planar maximally filtered graph (PMFG) [62]. In addition, there are also cluster analysis methods that can discover hidden hierarchy and organization, for example centroid-based clustering [63], connectivity based clustering (hierarchical clustering) [64–66] and distribution-based clustering [67]. Each method has its pros and cons, and therefore we must pick a method based on the purpose it is best suited for. In this paper we chose to perform hierarchical clustering, because this fits most naturally into the ‘chemical reaction’ picture described in Section 1 and our recent paper [68].

In agglomerative hierarchical clustering [69,35] of N elements, we start with N clusters containing one element each. We then merge the two most similar clusters to form a larger cluster. This process is repeated until we end up with a single cluster of N elements. To perform hierarchical clustering, we need a measure of *dissimilarity* or *distance* between elements, as well as a measure of dissimilarity or distance between clusters. We define the distance matrix between elements as

$$D_{ij} = 1 - C_{ij} \quad (2)$$

where $\mathbf{C}$ is the cross-correlation matrix. This definition of distance is more common in the computer sciences community but not in the econophysics community as it does not satisfy the ‘ultrametric triangle inequality’. Nevertheless, it is straightforward to show that the same hierarchies are obtained whether we use $D_{ij} = 1 - C_{ij}$ or $D_{ij} = \sqrt{2(1 - C_{ij})}$. The reason we chose Equation (2) as the distance measure is because it is more convenient to deal with a linear scale for the robustness analysis that we will introduce in Section 2.3. There are three common measures of distances, called linkages, between clusters defined based on the distances between elements. The *single linkage* or *minimum linkage* (SLHC), defined as the minimum distance between an element in the first cluster and an element in the second cluster, is most commonly used for phylogenetic analysis, where we suspect that all elements descended from the same ancestral element through series of small changes. Also commonly used for grouping variables influenced by common factors is the *complete linkage* or *maximum linkage* (CLHC), defined as the maximum distance between an element in the first cluster and an element in the second cluster. Finally, there is the *average linkage* (ALHC), defined as the average between distances between elements in the first cluster and elements in the second cluster. As its name suggests, the average linkage produces clusters that are intermediate between those produced by single and complete linkage. We illustrate in Figure 1 hierarchical clustering using CLHC, SLHC, and ALHC.

The cross-correlation matrix in Figure 2(a) contains much more information than its hierarchical clustering dendrogram suggests. In order not to lose this information, we reorder the rows and columns of the cross-correlation matrix using the linear ordering suggested by the dendrogram. The reordering process is called *seriation* [70–72], and we see that the seriated cross-correlation matrix in Figure 2(b) is more structured and we can identify a few clusters just through visual inspection. This is evidence of self-organization within the financial market.

However, although CLHC do separate stocks into groups based on their cross-correlations, the organization is not very clean. Ideally we should have strongly-correlated (red) diagonal blocks and weakly-correlated (blue) off-diagonal blocks. However, as Figure 2(b) shows, there remains strong cross-correlations in the

Fig. 1: Five elements at positions 1, 3, 6, 10, 14.5 along a one-dimensional chain, and how they are hierarchically clustered using different algorithms: (a)-(d) complete-linkage (CLHC), (e)-(h) single-linkage (SLHC), and (i)-(l) average linkage (ALHC). The Euclidean distance between elements 1 and 2 is 2, which is the shortest amongst all pairs of elements. Thus in all three algorithms, they are merged to form the first cluster in (a), (e), and (i). In the CLHC, the linkage between element 3 and the cluster $\{1, 2\}$ is $\max\{6 - 1 = 5, 6 - 3 = 3\} = 5$ whereas the linkage between element 3 and element 4 is $10 - 6 = 4$. We therefore merge elements 3 and 4 to (b) form the next cluster. In contrast, for the SLHC, the linkage between element 3 and the cluster $\{1, 2\}$ is $\min\{6 - 1 = 5, 6 - 3 = 3\} = 3$, which is smaller than the distance of 4 between elements 3 and 4. Hence we merge element 3 with cluster $\{1, 2\}$ to (f) form the larger cluster $\{1, 2, 3\}$. For the ALHC the linkage between cluster $\{1, 2\}$ and element 3 is the average distance between element 3 with elements 1 and 2, hence 4 units, and in this case the element can either merge with cluster $\{1, 2\}$ or with element 4. In (j), we show the ALHC merging cluster $\{1, 2\}$ with element 3. For all three algorithms, the agglomeration process is iterated until we end up with a single cluster of all five elements.

off diagonal blocks of the CLHC seriated matrix. These are strong correlations overlooked by the CLHC, SLHC, and ALHC algorithms which considers a single distance between clusters to decide whether they are to be merged. These strong off-diagonal correlations contain information on how clusters interact with each other, and are thus important. In our previous studies on the Hong Kong Stock Exchange and the SGX, we found clusters of local stocks and a distinct cluster of Chinese stocks. These interact to form a supercluster comprising both local and Chinese stocks [68]. In order to retain such information, we perform clustering on the CLHC clusters, focusing on the off-diagonal correlations, and call this higher-level clustering procedure *interaction hierarchical clustering*.

2.3 Interaction Hierarchical Clustering

As we illustrate in Figure 3, the objective of interaction hierarchical clustering (IHC) is to reorder the sequence of clusters, to obtain stronger diagonal blocks and weaker off-diagonal blocks. In order to do this, we first need to find the robust clusters from the seriated correlation matrix and then regroup those robust clusters according to their inter-cluster correlations.

The first step in IHC is to obtain the preliminary structure and also extract the robust clusters. A cluster is considered robust if the components in the cluster remain unchanged over a wide range of depth in the dendrogram. We quantify this notion of robustness $R(\mathbf{cl}_{i \cup j})$ of a cluster, $\mathbf{cl}_{i \cup j}$ as the linkage difference between the hierarchy level where it is first formed and the hierarchy level where it is merged with another

Fig. 2: The cross-correlation matrix (a) for 561 stocks in the SGX from Jan 2008 to Dec 2009, where red indicates strong correlation, blue indicates strong anti-correlation and green means nearly uncorrelated. As the stocks are arranged in alphabetical order no structure can be discerned. After we perform CLHC, and rearrange the financial stocks in the order of the dendrogram, the structure in the CLHC-seriated matrix (b) becomes obvious. The diagonal blocks are red, indicating strong intra-cluster correlations. However, we also see red off-diagonal blocks. This suggests strong interactions between correlations between clusters, which ideally should be grouped together or at least ordered close to each other. In the IHC-seriated matrix (c) where we reordered the CLHC clusters to account for their interactions, we see less red color outside of the diagonal blocks, which are now more consistently red (i.e. intracluster correlations are more homogeneous).

Fig. 3: (a) An artificial cross-correlation matrix with elements drawn from three different Gaussian distributions. The objects the matrix represents are organized into four distinct clusters (red diagonal blocks). In addition, these clusters are also ‘interacting’ with one and other (orange off-diagonal blocks). In this particular example, the first cluster is ‘interacting’ with the third cluster, whereas the second cluster is ‘interacting’ with the fourth cluster. Our goal is to reorder the objects so that the stronger cross-correlations are as close to the diagonal as possible, and the weaker cross-correlations are as far away from the diagonal as possible. In this example, this can be achieved by swapping the orders of the second cluster (larger) and the third cluster (smaller). However, neither CLHC, SLHC nor ALHC manage to do so, as their seriated matrix remained the same as the original cross-correlation matrix. Using the interaction hierarchical clustering (IHC) procedure described in Section 2.3, we find the IHC-seriated matrix shown in (b) achieving the desired outcome.

cluster. In other words, if $\min(H(\mathbf{cl}_{i \cup j}))$ is the linkage distance where $\mathbf{cl}_{i \cup j}$ is formed from the merger of $\mathbf{cl}_i$ and $\mathbf{cl}_j$, and $\max(H(\mathbf{cl}_{i \cup j}))$ is the linkage distance where $\mathbf{cl}_{i \cup j}$ is merged with another cluster, then the robustness of $\mathbf{cl}_{i \cup j}$ is

$$R(\mathbf{cl}_{i \cup j}) = \max(H([\mathbf{cl}_{i \cup j}])) - \min(H([\mathbf{cl}_{i \cup j}])). \quad (3)$$

Furthermore, we only consider $\mathbf{cl}_k$ as a robust cluster if the following criteria are met.

1. $R(\mathbf{cl}_k)$ need to be greater than a robustness length threshold.
2. $\mathbf{cl}_k$ must form at $H(\mathbf{cl}_k)$ lesser or equal to the cluster cutoff point
3. No $\mathbf{cl}_k$ can form at $H(\mathbf{cl}_k)$ lesser or equal to the cluster unbreak point.

In this research, we use 0.05 as the robustness length threshold and set the cluster cutoff point and cluster unbreak point at 1 and 0.3 respectively. Clusters with distance above 1 are anti-correlated, thus they should

not grouped into one cluster. Meanwhile, clusters with distance below 0.3 are strongly correlated and should not be broken up into smaller clusters. In Figure 4(a) we illustrate the *robustness length*, *cluster cutoff point*, and *cluster unbreak point*. In Figure 4(a), we also see how the 30 elements are organized into nine robust clusters based on the criteria we defined. These will be clustered a second time using the IHC to give the dendrogram in Figure 4(b).

Fig. 4: A schematic description (a) for *robustness length*, *cluster cutoff point*, and *cluster unbreak point*. A cluster is considered robust if its robustness length is greater than a threshold, as it retains its identity over a large range of cutoffs. However, we do not consider clusters that are formed above the cluster cutoff point ($D = 1$) as robust, as their components are already anti-correlated. Also, we do not admit cutoffs below the cluster unbreak point, because clusters that are so strongly-correlated should not be broken into smaller clusters. In (b), we show the IHC dendrogram of the 9 robust CLHC clusters, which form the leaves of the dendrogram.

Once we have the robust clusters, the next step is to perform hierarchical clustering treating these robust clusters as the objects. To do this, we need to construct the distance matrix $\mathbf{ID}$ for robust clusters. We define the interaction distance (ID_{ij}) between two robust clusters $\mathbf{cl}_i$ and $\mathbf{cl}_j$ as the average inter-cluster pair distance

$$ID(\mathbf{cl}_i, \mathbf{cl}_j) = \frac{1}{N_i N_j} \sum_{k_i=1}^{N_i} \sum_{k_j=1}^{N_j} D(\mathbf{cl}_i(k_i), \mathbf{cl}_j(k_j)), \quad (4)$$

where $D(\mathbf{cl}_i(k_i), \mathbf{cl}_j(k_j))$ is the distance between the k_i th element in robust cluster $\mathbf{cl}_i$ and the k_j th element in robust cluster $\mathbf{cl}_j$. Like the hierarchical clustering of stocks described in Section 2.2 the pair of clusters with shortest distance according to Equation (4) will be merged first to form a bigger cluster. This process is iterated until only one cluster remains, and we end up with the IHC dendrogram (see Section 2.4). As we can see comparing Figure 2(b) and Figure 2(c), diagonal blocks are better defined in the IHC seriated matrix than in the CLHC seriated matrix. As those clusters with strong interaction (short ID) but far apart under CLHC (due to its history dependence and definition of distance) are now brought closer together by IHC.

2.4 Cluster Identification

Because the CLHC clusters are reordered and regrouped, we must think of them as candidate robust clusters. To find the more well-defined robust clusters, we need to redo the clustering analysis after performing IHC. So far we have been using a bottom-up approach to group elements and clusters together, now let us use a top-down approach to determine the robust clusters. The standard way to do this is to cut the dendrogram at a cutoff threshold and clusters that remain are considered the robust clusters. However, a natural question we must ask is whether there is a way to unambiguously specify this threshold.

Before answering the question, let us consider the artificial correlation matrix shown in Figure 5(a). In this artificial correlation matrix, we find four robust clusters (RC_1, RC_4, RC_3, RC_2) with interaction blocks (IB_1, IB_2, IB_3, IB_4). Its IHC dendrogram is shown in Figure 5(b). In general, the lower the cutoff the more robust clusters we obtain. In this process, the most important compromise we have to make is: if we break a cluster up into two clusters, then some intra-cluster correlations will become inter-cluster

Fig. 5: In (a) we show an artificial cross-correlation matrix with four robust clusters: RC_1, RC_4, RC_3 , and RC_2 . Interactions between robust clusters are weak, except for that between RC_1 and RC_4 . In (b), we show the IHC dendrogram for this artificial correlation matrix, where we distinguish between three regions: Region 1 with two clusters, Region 2 and Region 3 with three and four clusters respectively. As the cutoff is lowered, more clusters will be formed, and previously intracluster correlations become intercluster correlations. In (c), we show the normalized distributions of existing inter-cluster correlations (green), newly intercluster correlations (red), and remaining intra-cluster correlations (blue). The top panel of (c) corresponds to the situation encountered when we lower the cutoff from Region 1 to Region 2. The bottom panel of (c) corresponds to the cutoff being lowered from Region 2 to Region 3.

correlations. For example, when we lower the cutoff in Figure 5(b) from Region 1 to Region 2, we go from two clusters, the first represented by the correlations in $\{RC_1, RC_4, RC_3, IB_1, IB_2\}$, and the second represented by the correlations in $\{RC_2\}$, to three clusters $\{RC_1, RC_4, IB_1\}$, $\{RC_3\}$, and $\{RC_2\}$. Here, note that the IB_2 correlations which were considered intra-cluster correlations in Region 1 has become inter-cluster correlations in Region 2. Similarly, when we lower the cutoff from Region 2 to Region 3, the IB_1 correlations will also go from being intra-cluster correlations to inter-cluster correlations.

Fig. 6: In this figure, we show the 35 CLHC clusters and their member stocks in a given time window. After selecting the cutoff using the procedure described in Section 2.4, we obtain seven robust clusters from the IHC dendrogram of these 35 CLHC clusters. Each robust cluster is then represented by a circle whose radius is proportional to the number of member stocks. In addition, we also color each circle based on the percentages of member stocks that have been randomly assigned the colors red, green, and blue.

More importantly, as Figure 5(b) illustrates, when the cutoff is progressively lowered, we find initially the intra-cluster correlations that becomes inter-cluster correlations are those that are statistically more similar to the inter-cluster correlations than they are to the intra-cluster correlations. Eventually, the cutoff becomes so low that the intra-cluster correlations that become inter-cluster correlations are statistically more similar to the intra-cluster correlations than they are to the inter-cluster correlations. With this understanding, we select the optimum cutoff to be the one at which the new inter-cluster correlations switch from being more similar to existing inter-cluster correlations to being more similar to the remaining intra-cluster correlations. There are many ways to measure this statistical similarity. Here we use the overlaps between the normalized distribution of new inter-cluster correlations with the normalized distribution of existing inter-cluster correlations and the normalized distribution of remaining intra-cluster correlations as the measure of statistical similarity. With this we are now able to reliably identify the robust clusters within the sliding time windows over the period Jan 2008 to Dec 2009.

2.5 Fusion-Fission Diagram

As we are interested in the dynamics of the hierarchical organization of the financial market, we need to examine these over many sliding time windows. The large number of sliding windows makes keeping track of the clusters and their time evolutions tedious and difficult. Thus, to visualize the dynamics of these clusters, we introduce the *fusion-fission diagram*. Similar diagrams have been introduced by Mantegna et al.[73] in the econophysics and sociophysics literature. In our fusion-fission diagram, we show some of the largest robust clusters, represented by circles whose radii are directly proportional to the sizes of the robust clusters. We also show how many elements from each robust cluster in time window t go to robust clusters in time window $t + 1$. In addition, we assign each stock randomly to one of the colors red, green, and blue, and thereafter color the circles based on the percentages of red, green, and blue elements. To illustrate how we construct the fusion-fission diagram, we show in Figure 6 the IHC dendrogram for 35 CLHC clusters. By cutting the dendrogram according to the procedure described in Section 2.4, we ended up with seven robust clusters. Finally, the flow of the elements from one robust cluster into another robust cluster is indicated by a line, whose thickness is proportional to the number of elements that flow between them.

With a fusion-fission diagram, we are able to condense a tremendous amount of information on the market dynamics into a simple diagram that is visually meaningful. Specifically, the number of circles and their sizes tell us how stocks are organized at any time, while the thicknesses of the lines tell us how strong the mixing dynamics is in the financial market. This mixing information is also encoded in the colors of the clusters. Such a diagram thus helps us visually detect key events in the stock market before and after the financial crisis.

3 Result and Discussion: Dynamics of SGX Financial Clusters

In this study, we are interested in the description of financial market dynamics in terms of clusters. In their recent paper, Munnix et al. found that the US stock market switches abruptly between different ‘states’, which can be distinguished from each other by their correlation structures [74]. More recently, Musmeci et al. found using the directed bubble hierarchical tree (DHBT) method strong mixing in the compositions of clusters in the US stock market, and proceeded to measure the rates at which the market reorganizes itself, and also how persistent strong correlations between stocks are [47]. However, they stop short of presenting this information in the form of a fusion-fission diagram, which we feel gives the most compact visual representation of the complexity and non-stationarity in financial markets. This dynamics is reminiscent of John Holland’s ideas that complex adaptive systems achieve perpetual novelty by taking different combinations of a small set of building blocks [75]. To tell a compelling story of how the Lehman Brother Crisis unfolded in the SGX, we split our two-year data set into one-month time windows that overlap with their predecessors and successors by two weeks. In total, we have 46 such time windows from Jan 2008 to Dec 2009. We then perform IHC within each time window to determine the robust clusters, and construct the fusion-fission diagram of the SGX within this time period.

Within this period, two crises impacted the Singapore stock market: the Subprime Mortgage Crisis that origin from U.S and then spread to the rest of the world. Because of this shock, Asian countries experienced a short recession culminating in an Asian Correction where regional stock prices plunged further. From Figure 7(a), we that after Jan 2008 the Straits Times Index reached its peak on 6 May 2008 with 3248.75. Thereafter, the index slid, and reached its lowest for 2008 on 27 Oct 2008 at 1600.29, as a result of the Lehman Brothers Crisis. After that, there are weak signs of recovery, but due to the Asian Correction, the index again drops from the after-crisis peak 1924.87 (5 Jan 2009) to the lowest of 1456.95 (9 Mar 2009). Fortunately, the Singapore market eventually recovered with a bull run after the Asian Correction. Over

the period 9 Mar to 3 Aug 2009, the index saw an 85% rise to 2681.64. In Figures 7(b) and 7(c), we show the IHC-seriated correlation matrix before, at the lowest, and after the two financial crises.

From the fusion-fission diagram shown in Figure 8, we find repeated episodes of smaller clusters agglomerating into a giant cluster, which thereafter fragmenting into smaller clusters. In particular, after the fission episode in May 2008, the giant cluster persisted till Oct 2008 (the Lehman Brothers Crisis), where we find a sustained series of fissions that lasted till Mar 2009 (the Asian Correction). After Mar 2009, a giant cluster re-emerges, interrupted by fission episodes that were closer and closer together. This suggests that the Lehman Brothers Crisis and Asian Correction are not separate events, but merely the beginning and end of an extended crisis phase. Consistent with this critical transition interpretation is the slowing down of the dynamical time scale before the Lehman Brothers Crisis, and the quickening of the dynamical time scale after the Asian Correction. During the recovery phase, we also observed the formation and persistence of a giant cluster, but its size is smaller than the giant cluster formed in Oct 2008.

Fig. 7: (a) Straits Times Index downloaded from Yahoo Finance, <https://sg.finance.yahoo.com>. The vertical dashed lines are correspond to the lows index due to the Lehman Brothers Crisis on 27 Oct 2008 (red) and the Asian Correction on 9 Mar 2009 (blue), and also the peak on 5 Jan 2009 after the Lehman Brothers Crisis (green). (b) IHC-seriated cross-correlation matrices before (Jun/2008 and Jul/2008), during (Sept/2008 and Oct/2008), and after (Nov/2008, Dec/2008) the Lehman Brothers Crisis where the Straits Times Index reached a low on 27 Oct 2008. (c) IHC seriated cross-correlation matrixes dynamics of the Asian Correction, before (Jan/2009 and Feb/2009), during (Mar/2009 and May/2009) and after (Jun/2009 and Jul/2009). For the Lehman Brothers Crisis, we see a giant cluster growing as we approached the market crash. After the financial crisis the giant cluster fragmented into smaller clusters. However, the picture is less clear with the Asian Correction.

Fig. 8: The fusion-fission diagram of the SGX from Jan 2008 to Dec 2009. This diagram is composed of three parts: (I) the innermost polar plot of the STI, where the index value is given by the radial distance, and additionally color coded such that blue is the highest (3249), green is intermediate (2353), and red is the lowest (1457). In this plot we also show as dashed lines the date (27 Oct 2008, cyan) when the STI is at its lowest due to the Lehman Brothers Crisis, and on 9 Mar 2009 (magenta) due to the Asian Correction; (II) the start and end dates of the time window, color coded by the average value of the STI within the time window; and (III) the robust clusters within each time window, and the exchanges of members with robust clusters of the previous and subsequent time windows.

4 Further Analysis: Mixing Degree and Persistence

As we can see from Figure 8, the fusion-fission diagram allows us to take what Gell-Mann called a *crude look at the whole* [76]. But it also helps us formulate the right detailed questions to ask. First, based on the visual representation of member stock exchanges between clusters, we realized that there are episodes of extensive mixing. Second, related to mixing and based on the colors of the clusters, we suspected that stocks do not remain in the same clusters for very long. Therefore, in Section 4.1, we will quantify the degree of mixing in the SGX, based on the robust clusters identified, and then in Section 4.2 quantify the persistence of strong correlations between stocks.

4.1 Mixing Degree

4.1.1 Adjusted Rand Index

To measure the mixing degree, we can use the well known Adjusted Rand Index (ARI), proposed by Hubert and Arabie [77]. This is the extension of Rand Index (RI), developed by Rand [78] to measure the similarity between two lists of clusters. Given the elements set $\mathbf{S} = \{E_1, E_2, \dots, E_N\}$ and two cluster lists as $\mathbf{CL}_1 = \{\mathbf{Cl}_1^1, \mathbf{Cl}_2^1, \dots, \mathbf{Cl}_u^1\}$ and $\mathbf{CL}_2 = \{\mathbf{Cl}_1^2, \mathbf{Cl}_2^2, \dots, \mathbf{Cl}_v^2\}$, where $\cup_{i=1}^u \mathbf{Cl}_i^1 = \cup_{j=1}^v \mathbf{Cl}_j^2 = \mathbf{S}$. The RI measures the percentages of pairs of elements that remain in same cluster in both lists ($\{E_m, E_n\} \in \mathbf{Cl}_a^1$ and $\{E_m, E_n\} \in \mathbf{Cl}_b^2$ for arbitrary a and b) and remain in different clusters in both lists (Pair $\{E_m, E_n\}$ which $E_m \in \mathbf{Cl}_a^1, E_n \in \mathbf{Cl}_b^1$ and $E_m \in \mathbf{Cl}_c^2, E_n \in \mathbf{Cl}_d^2$ for arbitrary a, b, c , and d that $a \neq b$ and $c \neq d$). Because of this, RI is highly dependent on the number of clusters and their distribution, as some combinations may by chance have higher index values. The ARI corrected this by removing the expected index and standardize it for the sake of comparison. The ARI is defined as

$$ARI(\mathbf{CL}_1, \mathbf{CL}_2) = \frac{Index - Expected\ Index}{Max\ Index - Expected\ Index} = \frac{\sum_{ij} \binom{n_{ij}}{2} - \frac{\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}}{\binom{N}{2}}}{\frac{1}{2} \left(\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2} \right) - \frac{\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}}{\binom{N}{2}}}, \quad (5)$$

where N is the number of elements, n_{ij} is the number of shared elements between cluster $\mathbf{Cl}_i^1$ and $\mathbf{Cl}_j^2$, $n_{i\cdot}$ is number of elements in cluster $\mathbf{Cl}_i^1$, and $n_{\cdot j}$ is number of elements in cluster $\mathbf{Cl}_j^2$. If the second list of clusters is obtained by randomly re-clustering the first list of clusters, the ARI is 0. In contrast, when the second list of clusters is identical to the first list of clusters, the ARI is 1. Since $ARI = 0$ when the elements are randomly reclustered, and $ARI = 1$ when they remain in the same clusters, the mixing degree can be defined as $1 - ARI$.

When the mixing degree changes from time window to time window, we want to know how significant this change is. Therefore, we estimate the uncertainty in the mixing degree through Monte Carlo simulations. First, we estimate the empirical transition probabilities between clusters in adjacent windows. Suppose at time t the cluster list is $\mathbf{CL}_t = \{\mathbf{Cl}_1^t, \dots, \mathbf{Cl}_u^t\}$, cluster list $\mathbf{CL}_{t+1} = \{\mathbf{Cl}_1^{t+1}, \dots, \mathbf{Cl}_v^{t+1}\}$ at time $t + 1$, and $\Gamma(\mathbf{Cl}_i^t, \mathbf{Cl}_j^{t+1})$ is the transition probability from $\mathbf{Cl}_i^t$ to $\mathbf{Cl}_j^{t+1}$, defined as

$$\Gamma(\mathbf{Cl}_i^t, \mathbf{Cl}_j^{t+1}) = \frac{n(\mathbf{Cl}_i^t \cap \mathbf{Cl}_j^{t+1})}{n(\mathbf{Cl}_i^t)}, \quad (6)$$

where $n(\mathbf{Cl})$ is the number of elements within $\mathbf{Cl}$. In some sense, this is the maximum-likelihood Markov chain model of the observed reclustering from $\mathbf{CL}_t$ to $\mathbf{CL}_{t+1}$. In each Monte Carlo simulation, we then go through each cluster $\mathbf{Cl}_i^t$, and choose which clusters $\mathbf{Cl}_j^{t+1}$ its elements go to. After all elements have been reclustered, we compute the $1 - ARI$ mixing degree. After many Monte Carlo simulations, we have a distribution of $1 - ARI$, whose mean is our observed mixing degree, and whose standard deviation tells us how much spread we expect from the Markov chain model of reclustering between time windows t and $t + 1$.

Figure 9 shows the mixing degree given by $1 - ARI$ (blue), and the results agreed with our expectations of high mixing degree after the financial crisis and low mixing degree during the bullish and bearish markets. In early 2008, the mixing degree was high and oscillating, and then started to decrease from Jul 2008 as SGX experienced the financial downturn. More importantly, right after the 2 Oct 2008-31 Oct 2008 window (lowest STI due to Lehman Brothers Crisis), we observe high mixing degree over a few months, followed by low mixing degree from 2 Mar 2009-31 Mar 2009 window (lowest STI due to Asian Correction then bullish market right after the correction) and back to oscillating high mixing degree in late 2009. In addition, we

also learned that when the mixing degree is high in SGX, market participants were forced to update their portfolios more frequently, as we will see in Section 4.2 for the lifetimes of SGX clusters. Moreover, as we can see most of the clusters are actually very small, and we are more interested in the mixing between the large clusters. Thus, we introduce another measurement for mixing degree named *Degree Of Mixing*, that is able to measure the mixing degree for a single cluster.

Fig. 9: Mixing degree of robust SGX clusters, determined using different methods: (green) DOM using all robust clusters, (red) DOM using three largest robust clusters, and (blue) $1 - ARI$ based on all robust clusters. The vertical dashed lines indicate the windows where the STI is at the lowest due to the Lehman Brothers Crisis (2 Oct 2008 to 31 Oct 2008, cyan) and the Asian Correction (2 Mar 2009 to 31 Mar 2009, magenta).

4.1.2 Degree of Mixing, DOM

Suppose the cluster lists at time t and $t+1$ are $\mathbf{CL}_t = \{\mathbf{CI}_1^t, \dots, \mathbf{CI}_u^t\}$ and $\mathbf{CL}_{t+1} = \{\mathbf{CI}_1^{t+1}, \dots, \mathbf{CI}_v^{t+1}\}$. Further suppose that we are only interested in the mixing between the first three largest clusters in each list, $\mathbf{CL}_t' = \{\mathbf{CI}_{Largest}^t, \mathbf{CI}_{2^{nd} largest}^t, \mathbf{CI}_{3^{rd} largest}^t\}$ and $\mathbf{CL}_{t+1}' = \{\mathbf{CI}_{Largest}^{t+1}, \mathbf{CI}_{2^{nd} largest}^{t+1}, \mathbf{CI}_{3^{rd} largest}^{t+1}\}$, which are significant clusters in term of size. However, ARI is a global measurement for mixing, and does not tell us what the mixing is locally for each cluster in $\mathbf{CL}_{t+1}'$. In order to deal with this situation, we thus introduce the Degree Of Mixing (DOM) measure that can be used for individual clusters $\mathbf{CI}_j^{t+1}$ starting from the cluster list $\mathbf{CL}_t$.

Intuitively, the mixing degree of $\mathbf{CI}_j^{t+1}$ is highest when its elements are equally contributed (size normalized) by all clusters from $\mathbf{CL}_t$, and lowest when it is identical to or a subset of one of the clusters from $\mathbf{CL}_t$ (see Figure 10). Because of this, we define DOM as the average of the Mixing Degree (MD) for all clusters within $\mathbf{CL}_{t+1}$, where the MD for $\mathbf{CI}_j^{t+1}$ relative to $\mathbf{CL}_t$ is defined as

$$MD(\mathbf{CI}_j^{t+1}, \mathbf{CL}_t) = \sqrt[N(\mathbf{CL}_t)]{\prod_{i=1}^{N(\mathbf{CL}_t)} f\left(\frac{n(\mathbf{CI}_i^{In_t} \cap \mathbf{CI}_j^{In_{t+1}})/n(\mathbf{CI}_i^{In_t})}{\sum_{i=1}^{N(\mathbf{CL}_t)} n(\mathbf{CI}_i^{In_t} \cap \mathbf{CI}_j^{In_{t+1}})/n(\mathbf{CI}_i^{In_t})}\right)} = \sqrt[N(\mathbf{CL}_t)]{\prod_{i=1}^{N(\mathbf{CL}_t)} f(x)} \quad (7)$$

In this equation, $N(\mathbf{CL}_t)$ is the number of clusters in $\mathbf{CL}_t$, $n(\mathbf{CI}_i^{In_t})$ is the number of elements within $\mathbf{CI}_i^{In_t}$. We also distinguish between two subclusters $\mathbf{CI}_i^{In_t}$ and $\mathbf{CI}_i^{Ext_t}$ of $\mathbf{CI}_i^t$, where elements of $\mathbf{CI}_i^{In_t}$ appear in both $\mathbf{CL}_t$ and $\mathbf{CL}_{t+1}$, while elements of $\mathbf{CI}_i^{Ext_t}$ do not appear in $\mathbf{CL}_{t+1}$. Similarly, $\mathbf{CI}_j^{t+1} = \{\mathbf{CI}_j^{In_{t+1}}, \mathbf{CI}_j^{Ext_{t+1}}\}$,

where all the element within $\mathbf{C}\mathbf{I}_j^{In_{t+1}}$ appear in both $\mathbf{C}\mathbf{L}_t$ and $\mathbf{C}\mathbf{L}_{t+1}$. Finally, $f(x)$ is given as

$$f(x) = \frac{x(x-1)\exp(-(A-x)^2)}{\alpha(\alpha-1)\exp(-(A-\alpha)^2)}; \alpha = \frac{1}{N(\mathbf{C}\mathbf{L}_t)}; A = \frac{\alpha^3 - \alpha^2 - \alpha + \frac{1}{2}}{\alpha^2 - \alpha} \quad (8)$$

so that $f(0) = 0$, $f(1) = 0$, and $f(\alpha) = 1$ and also its derivative with respect to x is $f_x(x < \alpha) > 0$, $f_x(x = \alpha) = 0$, and $f_x(x > \alpha) < 0$.

There are three reasons why we define MD for $\mathbf{C}\mathbf{I}_{t+1}^j$ relative to $\mathbf{C}\mathbf{L}_t$ in this way: (1) it normalized against the cluster sizes of all clusters within cluster list $\mathbf{C}\mathbf{L}_t$, so that for maximum mixing degree, the larger cluster will contribute a larger portion in cluster $\mathbf{C}\mathbf{I}_{t+1}^j$; (2) it also normalized against the sizes of clusters in $\mathbf{C}\mathbf{L}_t$, such that is bounded between 0 and 1; (3) the $f(x)$ peaks at $1/N(\mathbf{C}\mathbf{L}_t)$, which corresponds to equal contributions from all the clusters in $\mathbf{C}\mathbf{L}_t$, and decreases monotonically when x deviates from $1/N(\mathbf{C}\mathbf{L}_t)$. Due to this, the geometric mean of $f(x)$ will peak at 1 if it is well mixed, and decreases from 1 monotonically if there are deviations. Other than $1 - ARI$, Figure 9 also shows the mixing degree measured by DOM using all clusters (green) and DOM using three largest clusters (red). As with for ARI , we run 5000 Monte Carlo simulations for each time window in the manner described in Section 4.1.1 to estimate the standard error of the DOM using all clusters and DOM using three largest clusters. Comparing these two measurements, we noticed that right after the Lehman Brothers Crisis, the mixing degree measured by DOM with all clusters is lower than that measured by DOM that only consider three largest clusters, which means that the strong mixing is mainly due to the large clusters. Meanwhile, during the bullish market the both DOM mixing degrees are close to each other, suggesting that the mixing degree due to large clusters is comparable to that of small clusters. All in all, the general trends are similar to $1 - ARI$, as the DOM measurements also show low mixing degrees during the bearish and bullish market, and high mixing degree right after the Lehman Brothers Crisis. We suspect these observations are due to behavioral synchronization of market participants during bullish and bearish markets, but after the financial crisis, there is a larger diversity in the market participant responses towards market recovery.

Fig. 10: A schematic illustration for mixing degree based on DOM for two adjacent windows: (a) before normalized and (b) normalized against the cluster sizes. Each cluster $\mathbf{C}\mathbf{I}$ can be broken down to two sub-cluster as $\mathbf{C}\mathbf{I}^{In}$ and $\mathbf{C}\mathbf{I}^{Ext}$, where $\mathbf{C}\mathbf{I}^{In}$ consists of elements that appear in both windows. In contrast, elements within $\mathbf{C}\mathbf{I}^{Ext}$ only appear in one of the window and will be excluded when we compute the DOM . At maximum mixing degree, it is equally likely for an element in a $\mathbf{C}\mathbf{L}_t$ cluster to move to any cluster in $\mathbf{C}\mathbf{L}_{t+1}$, thus the number of elements from cluster $\mathbf{C}_i^{In_t}$ transferred to cluster $\mathbf{C}_j^{In_{t+1}}$ is proportional to size of cluster $\mathbf{C}_i^{In_t}$ (a). Else, after normalization against the cluster sizes, $\mathbf{N}\mathbf{C}_j^{In_{t+1}}$ receives equal contributions from clusters $\mathbf{N}\mathbf{C}_i^{In_t}$, $i = \{1, 2, 3\}$.

4.2 Persistence

In Section 4.1, we see high mixing degrees among SGX clusters from Jan 2008 to Dec 2009, as most of the time the mixing degree is above 0.7. This leads naturally to a question about the persistence of the clusters, such as how long a pair of stocks will remain together in the same cluster. This question is particularly important for portfolio selection, due to the important role played by the correlation in risk diversification.

Whenever the correlation coefficient between two financial stocks is less than 1, there is a potential for diversifying the risks. Yet, correlations are frequently unstable, thus portfolios need to be updated from time to time, using the lifetimes of these correlations as a gauge on how frequently this is to be done.

For simplicity, we determine the persistence of the clusters by measuring the lifetime of a pair of financial stocks that continuously remain within the same cluster. This is done by counting the number of pairs that remain in the same clusters across several time windows. For example, stocks FS_A and FS_B are within the same cluster at time t_n , t_{n+1} , and t_{n+2} ; but broken up into two different financial clusters at t_{n+3} and back to same financial cluster again at t_{n+4} . Based on this information, FS_A and FS_B contribute one pair of stocks from t_n to t_{n+2} , zero pair of stocks in t_{n+3} , and one pair of stocks in t_{n+4} , hence contribute one pair of stocks that last for 3 time windows and another one last for 1 time window. The longer the stocks stay together the more persistent their correlation, which implies the market participants prefer to trade these stocks simultaneously.

Fig. 11: Distribution of the number of stock pairs that continuously remain within a same cluster across different duration. In this semilog plot, we show the double exponential fit $f(t) = A \exp(-t/\tau_1) + B \exp(-t/\tau_2)$ (red dashed curve) to the data (black solid curve), and compare it against a single exponential fit $f(t) = C \exp(-t/\tau_A)$ (blue dashed curve, offset). The double exponential decay function fits the data better and suggests that there are two scales governing the decay of correlations between stocks.

Since there are 46 time windows and there are 561 financial stocks, thus we can track in total $561 \times 561 \times 46$ different pairs of stocks from Jan 2008 to Dec 2009. Figure 11 shows the distribution of the number of pairs remaining within a same cluster continuously across different duration. On a semilog scale, we see that the distribution can be explained by a double exponential decay with two different time scales. Since each time window is one month long with two-week overlaps with the time windows before and after it, the data points are two weeks apart. The double exponential decay function suggests that there are two groups of correlations, the first with a short lifetime (2.94 ± 0.20 weeks) and long lifetime (5.85 ± 0.84 weeks). This implies that if a portfolio consists of stocks with short-lived correlations, the portfolio needs to be updated at least once every three weeks. Similarly, for a portfolio consisting stocks with long-lived correlations, the portfolio needs to be updated once every six weeks. In general, the high mixing degree among SGX clusters and short-lived correlations between stocks suggests that trading strategies need to be updated frequently.

5 Conclusion

In this study, we extended the hierarchical clustering method to an Interaction Hierarchical Clustering method (IHC), which is a two-stage hierarchical clustering process that gives robust clusters with strong intracuster correlations and weak intercluster correlations. In addition, we introduce the use of a fusion-fission diagram to summarize information on the number of robust clusters, sizes of robust clusters, and the exchange of elements between robust clusters. This fusion-fission diagram helps us visualize the correlation dynamics in the Singapore Stock Exchange (SGX).

We studied 561 financial stocks in SGX from Jan 2008 to Dec 2009. In these two years, SGX had experienced the effects of the Lehman Brothers Crisis (Oct 2008) and the Asian Correction (Mar 2009) followed by a bullish recovering market. Based on the fusion-fission diagram, we observed the emergence

of a giant cluster before the financial crisis, which remained persistently large until the Lehman Brothers Crisis. Right after the Lehman Brothers Crisis we find rapid mixing between a large number of small to mid-sized clusters, which pointed to a wide spectrum of expectations and responses from the market participants. This strong mixing ended with the Asian Correction of Mar 2009. Finally, we quantified the mixing between robust clusters and the persistence of correlations between pairs of stocks. We found that there are two groups of correlations, one with a life time of three weeks, and the other with a life time of six weeks. Together, these observations suggest that SGX market participants must update their portfolios frequently.

Author contributions

S.A.C. conceived the study, B.K.T. conducted the study, S.A.C. and B.K.T analysed the results, S.A.C. and B.K.T. wrote and reviewed the manuscript.

References

1. Jess Baker. Japan 2011 earthquake and tsunami disaster. *The Weather Channel, News*, 2014.
2. Mourad Zeghal and Ahmed-W Elgamal. Analysis of site liquefaction using earthquake records. *Journal of Geotechnical Engineering*, 120(6):996–1017, 1994.
3. Satoshi Ide, Annemarie Baltay, and Gregory C Beroza. Shallow dynamic overshoot and energetic deep rupture in the 2011 mw 9.0 Tohoku-Oki earthquake. *Science*, 332(6036):1426–1429, 2011.
4. Susumu Kurahashi and Kojiro Irikura. Source model for generating strong ground motions during the 2011 off the Pacific coast of Tohoku earthquake. *Earth, Planets and Space*, 63(7):571–576, 2011.
5. Thomas H Jordan and Lucile M Jones. Operational earthquake forecasting: some thoughts on why and how. *Seismological Research Letters*, 81(4):571–574, 2010.
6. Ian Main. Is the reliable prediction of individual earthquakes a realistic scientific goal. *Nature*, 397(1), 1999.
7. James A Feigenbaum and Peter GO Freund. Discrete scale invariance in stock markets before crashes. *International Journal of Modern Physics B*, 10(27):3737–3745, 1996.
8. Nicolas Vandewalle, Marcel Ausloos, Ph Boveroux, and A Minguet. How the financial crash of October 1997 could have been predicted. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2):139–141, 1998.
9. Didier Sornette. Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2522–2529, 2002.
10. Fabrizio Lillo and Rosario N Mantegna. Power-law relaxation in a complex system: Omori law after a financial market crash. *Physical Review E*, 68(1):016119, 2003.
11. Faruk Selçuk. Financial earthquakes, aftershocks and scaling in emerging stock markets. *Physica A: Statistical Mechanics and its Applications*, 333:306–316, 2004.
12. Fabrizio Lillo and Rosario N Mantegna. Dynamics of a financial market index after a crash. *Physica A: Statistical Mechanics and its Applications*, 338(1):125–134, 2004.
13. Faruk Selçuk and Ramazan Gençay. Intraday dynamics of stock market returns and volatility. *Physica A: Statistical Mechanics and its Applications*, 367:375–387, 2006.
14. Philipp Weber, Fengzhong Wang, Irena Vodenska-Chitkushev, Shlomo Havlin, and H Eugene Stanley. Relation between volatility correlations in financial markets and Omori processes occurring on all scales. *Physical Review E*, 76(1):016109, 2007.
15. Guo-Hua Mu and Wei-Xing Zhou. Relaxation dynamics of aftershocks after large volatility shocks in the SSEC index. *Physica A: Statistical Mechanics and its Applications*, 387(21):5211–5218, 2008.
16. Bikas K Chakrabarti, Arnab Chatterjee, and Pratip Bhattacharyya. Two-fractal overlap time series: Earthquakes and market crashes. *Pramana*, 71(2):203–210, 2008.
17. Alexander M Petersen, Fengzhong Wang, Shlomo Havlin, and H Eugene Stanley. Quantitative law describing market dynamics before and after interest-rate change. *Physical Review E*, 81(6):066121, 2010.
18. Fulvio Baldovin, Francesco Camana, Michele Caraglio, Attilio L Stella, and Marco Zamparo. Aftershock prediction for high-frequency financial markets dynamics. In *Econophysics of Systemic Risk and Network Dynamics*, pages 49–58. Springer, 2013.
19. Robert Wade and Frank Veneroso. The Asian Crisis: the high debt model versus the Wall Street-Treasury-IMF complex. *New Left Review*, 228(1): 3–24, 1998.
20. Graciela L Kaminsky and Carmen M Reinhart. The twin crises: the causes of banking and balance-of-payments problems. *American Economic Review*, 89(3): 473–500, 1999.
21. Anthony Sanders. The Subprime Crisis and its role in the financial crisis. *Journal of Housing Economics*, 17(4):254–261, 2008.
22. Fusakichi Omori. On the after-shocks of earthquakes. *The Journal of the College of Science, Imperial University, Japan*, 7(2):111–200, 1895.
23. Beno Gutenberg and Charles F Richter. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4):185–188, 1944.
24. Markus Båth. Lateral inhomogeneities of the upper mantle. *Tectonophysics*, 2(6):483–514, 1965.
25. Asli Demirgüç-Kunt and Ross Levine. Stock market development and financial intermediaries: stylized facts. *The World Bank Economic Review*, 10(2):291–321, 1996.
26. Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.
27. Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. Stylized facts of financial markets and market crashes in minority games. *Physica A: Statistical Mechanics and its Applications*, 294(3):514–524, 2001.

28. Arnoud WA Boot and Anjan V Thakor. Financial system architecture. *Review of Financial Studies*, 10(3):693–733, 1997.
29. Robert J Shiller. Human behavior and the efficiency of the financial system. *Handbook of Macroeconomics*, 1(Part C):1305–1340, 1999.
30. Zhuanxin Ding, Clive WJ Granger, and Robert F Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.
31. Mico Loretan and Peter CB Phillips. Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance*, 1(2):211–248, 1994.
32. Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467, 1999.
33. John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1-2): 32–52, 1928.
34. Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483, 1967.
35. Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197, 1999.
36. RN Mantegna. Information and hierarchical structure in financial markets. *Computer Physics Communications*, 121:153–156, 1999.
37. Jukka-Pekka Onnela, Anirban Chakraborti, Kimmo Kaski, Janos Kertesz, and Antti Kanto. Asset trees and asset graphs in financial markets. *Physica Scripta*, 2003(T106):48, 2003.
38. Michele Tumminello, Fabrizio Lillo, and Rosario N Mantegna. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1):40–58, 2010.
39. Khaldoun Khashanah and Linyan Miao. Dynamic structure of the US financial systems. *Studies in Economics and Finance*, 28(4):321–339, 2011.
40. Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, and Rosario N Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130, 2003.
41. Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, Salvatore Miccichè, Nicolas Vandewalle, and Rosario Nunzio Mantegna. Networks of equities in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):363–371, 2004.
42. Laurent Laloux, Pierre Cizeau, Marc Potters, and Jean-Philippe Bouchaud. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3(03):391–397, 2000.
43. J-P Onnela, A Chakraborti, K Kaski, and J Kertész. Dynamic asset trees and portfolio analysis. *The European Physical Journal B-Condensed Matter and Complex Systems*, 30(3):285–288, 2002.
44. J-P Onnela, Anirban Chakraborti, Kimmo Kaski, and Janos Kertesz. Dynamic asset trees and Black Monday. *Physica A: Statistical Mechanics and its Applications*, 324(1):247–252, 2003.
45. Yiting Zhang, Gladys Hui Ting Lee, Jian Cheng Wong, Jun Liang Kok, Manamohan Prusty, and Siew Ann Cheong. Will the US economy recover in 2010? A minimal spanning tree study. *Physica A: Statistical Mechanics and its Applications*, 390(11):2020–2050, 2011.
46. Siew Ann Cheong, Robert Paulo Forna, Gladys Hui Ting Lee, Jun Liang Kok, Woei Shyr Yim, Danny Yuan Xu, and Yiting Zhang. The Japanese economy in crises: A time series segmentation study. *Economics: The Open-Access, Open-Assessment E-Journal*, 6(2012-5):1–81, 2012.
47. Nicoló Musmeci, Tomaso Aste, and Tiziana Di Matteo. Risk diversification: A study of persistence with a filtered correlation-network approach. *Network Theory in Finance*, 1(1): 77–98, 2015.
48. Ip-Wing Yu, Kang-Por Fung, and Chi-Sang Tam. Assessing financial market integration in Asia-equity markets. *Journal of Banking & Finance*, 34(12):2874–2885, 2010.
49. Feng Guo, Carl R Chen, and Ying Sophie Huang. Markets contagion during financial crisis: A regime-switching approach. *International Review of Economics & Finance*, 20(1):95–109, 2011.
50. Leonidas Sandoval and Italo De Paula Franca. Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 391(1):187–208, 2012.
51. R Quax, D Kandhai, and PM Sloot. Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series. *Scientific Reports*, 3:1898–1898, 2012.
52. Dimitris Kenourgios, Aristeidis Samitas, and Nikos Paltalidis. Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. *Journal of International Financial Markets, Institutions and Money*, 21(1):92–106, 2011.
53. Michael J Naylor, Lawrence C Rose, and Brendan J Moyle. Topology of foreign exchange markets using hierarchical structure methods. *Physica A: Statistical Mechanics and its Applications*, 382(1):199–208, 2007.
54. Mustafa Keskin, Bayram Deviren, and Yusuf Kocakaplan. Topology of the correlation networks among major currencies using hierarchical structure methods. *Physica A: Statistical Mechanics and its Applications*, 390(4):719–730, 2011.
55. Ersin Kantar, Mustafa Keskin, and Bayram Deviren. Analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods. *Physica A: Statistical Mechanics and its Applications*, 391(7):2342–2352, 2012.
56. Daniel J. Fenn, Mason A. Porter, Stacy Williams, Mark McDonald, Neil F. Johnson, and Nick S. Jones. Temporal evolution of financial-market correlations. *Physical review E*, 84(2): 026109, 2011.
57. Daniel J. Fenn, Mason A. Porter, Peter J. Mucha, Mark McDonald, Stacy Williams, Neil F. Johnson, and Nick S. Jones. Dynamical clustering of exchange rates. *Quantitative Finance*, 12(10):1493–1520, 2012.
58. Nicol Musmeci, Tomaso Aste, and T. Di Matteo Relation between Financial Market Structure and the Real Economy: Comparison between Clustering Methods. *PloS ONE*, 10(3): e0116201, 2015. Doi:10.1371/journal.pone.0116201.
59. Yakov Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, 2002.
60. Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
61. Dong-Hee Kim and Hawoong Jeong. Systematic analysis of group identification in stock markets. *Physical Review E*, 72(4):046133, 2005.
62. Michele Tumminello, Tomaso Aste, Tiziana Di Matteo, and Rosario N Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10421–10426, 2005.
63. Zhexue Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3):283–304, 1998.

64. Robin Sibson. SLINK: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16(1):30–34, 1973.
65. Daniel Defays. An efficient algorithm for a complete link method. *The Computer Journal*, 20(4):364–366, 1977.
66. Hans-Peter Kriegel, Peer Kröger, Jörg Sander, and Arthur Zimek. Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):231–240, 2011.
67. Stuart Lloyd. Least squares quantization in PCM. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.
68. Boon Kin Teh, Yik Wen Goo, Tong Wei Lian, Wei Guang Ong, Wen Ting Choi, Mridula Damodaran, and Siew Ann Cheong. The Chinese Correction of Feb 2007: How financial hierarchies change in a market crash. *Physica A: Statistical Mechanics and its Applications*, 424:225–241, 2015.
69. Lawrence Hubert. Approximate evaluation techniques for the single-link and complete-link hierarchical clustering procedures. *Journal of the American Statistical Association*, 69(347):698–704, 1974.
70. Robert C Dunnell. Seriation method and its evaluation. *American Antiquity*, pages 305–319, 1970.
71. Innar Liiv. Seriation and matrix reordering methods: An historical overview. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 3(2):70–91, 2010.
72. Innar Liiv, Rain Opik, Jaan Ubi, and John Stasko. Visual matrix explorer for collaborative seriation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1):85–97, 2012.
73. Luca Marotta, Salvatore Micciché, Yoshi Fujiwara, Hiroshi Iyetomi, Hideaki Aoyama, Mauro Gallegati, and Rosario N Mantegna. Bank-firm credit network in Japan. An analysis of a bipartite network. *PloS ONE*, 10(5): e0123079, 2015. Doi:10.1371/journal.pone.0123079.
74. Michael C Münnix, Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H Seligman, Thomas Guhr, and H Eugene Stanley. Identifying states of a financial market. *Scientific Reports*, 2:644, 2012.
75. John H Holland. *Signals and boundaries: Building blocks for complex adaptive systems*. MIT Press, 2012.
76. Murray Gell-Mann. Lets call it plectics. *Complexity*, 1(5):96, 1995.
77. Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.
78. William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.