Journal of King Saud University – Computer and Information Sciences 31 (2019) 218–228

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Network mining based elucidation of the dynamics of cross-market clustering and connectedness in Asian region: An MST and hierarchical clustering approach

Biplab Bhattacharjee [a], Muhammad Shafi [a], Animesh Acharjee [a,b,c,d,*]

[a] School of Management Studies, National Institute of Technology, Calicut, Kerala, India
[b] Department of Biochemistry, Sanger Building, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK
[c] Institute of Cancer and Genomic Sciences, Centre for Computational Biology, University of Birmingham, B15 2TT, UK
[d] Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, B15 2TT, UK

ARTICLE INFO

Article history:
Received 12 July 2017
Revised 28 October 2017
Accepted 13 November 2017
Available online 20 November 2017

Keywords:
Network filtering
Data mining
Applied Graph Theory
Financial network analysis
MST
Hierarchical Clustering

ABSTRACT

We investigate the dynamics of cross-market clustering and connectedness of the Asian capital markets in this study. We perform the cross-correlation structure analysis of the daily return data of 14 global indices belonging to the major Asian capital markets by using the sub-dominant ultrametric distance based MST and Hierarchical Clustering techniques. The study dataset is for fourteen years duration (2002–2016). A rolling window approach is used to generate 151 temporally synchronous observations. We generate MSTs and Hierarchical Clustering plots (based on average linkage distance) for these temporally synchronous observations, and visually comprehend them to decipher the cross-market cluster formation, hub node formation, and connectivity structure with hub nodes. To identify those set of Asian markets having close connectivity with India, we employed a weighted hop count method and based on its scorings the Asian indices are subsequently ranked. We also investigate the influence of the 2008 financial crisis on the connectivity and clustering patterns in the Asian indices network. We also compute the key network topological parameters to decipher the dynamically varying topological properties, and with a particular reference during financial crisis periods.

1. Introduction

In modern day theories of portfolio management, an essential element is the feature of risk diversification, be it in domestic or regional or global scales (Gao et al., 2015). Understanding this risk diversification in these three scales encompasses the interpretation of the behaviour of cluster formation & the ingredients of risk contagion in investable asset classes. The customarily utilized cross-correlation analysis is a key quantitative measure of interactive relationships among pairs of securities (Gao et al., 2015) and also for the cross-market linkages, and an investigation into this will facilitate the enhanced understanding of the dynamic mechanisms with which an integrated complex economic system operate.

As an outcome of the steady rise in the globalization and the expansion of cross-country economic activities, there has been a noticeable incline in the correlation of global financial markets. In spite of the fact that this cross-market interaction has promoted allocation of financial and economic resources in an optimal manner, it also has subsequently caused the rapid spread of financial crisis to markets possessing no significant attachment to the real cause of the crisis but rather than because of that market's linkage (in direct or indirect manner) to the source of the contagion.

For instance in the year 2007, the US subprime mortgage crisis originated because of the defaults in US mortgage lending and investment banking space (a purely domestic reason) & had swiftly transmitted its contagion to savings institutions, insurance companies, and commercial banks located all across the world in different geographies, which has subsequently resulted in a crisis of global nature. Having the capacity to depict and comprehend correlation

* Corresponding author at: Institute of Cancer and Genomic Sciences, Centre for Computational Biology, University of Birmingham, B15 2TT, UK

E-mail addresses: biplab_p140069ms@nitc.ac.in (B. Bhattacharjee), shafi@nitc.ac.in (M. Shafi), a.acharjee@bham.ac.uk (A. Acharjee).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2017.11.002
1319-1578/

structures of complex financial market systems in a visualization-friendly network diagrammatic manner with associated quantifiable network measures can accordingly assist the regulators and the market participants to gain data-driven insights for taking better economic policy decisions.

The impact of high level cross-market integration in the capital market has resulted in a change in the approach of valuation of risk for a given set of expected returns in any particular domestic market. This is on account of the covariance between risk factors associated with a single domestic market and common world factors. In this scenario, the investors cannot depend solely on the domestic economic factors and the fundamental information of the firm concerned, but also should take into account the insights about the dynamic linkage structures among cross-country financial markets. This insight is a key decider because of the fact that in the case of segmented markets, there may be nil or very limited explanatory power of common world factors on deciding the asset pricing models. In the case of these segmented markets, the disequilibrium in the risk pricing mechanism results in the formation of regionally or internationally diversified clusters, wherein the diversification benefits may arise (Setiawan, 2011).

The complex networks models based on statistical physics provides us with a quantifiable framework for investigation of systems having high number of interactive elements, which in the financial context may have elements comprising of various asset classes (such as stocks (Bhattacharjee et al., 2016; Pan and Sinha, 2007; Tabak et al., 2010), currencies, etc.), various intra-country (or inter-country (Roy and Sarkar, 2011; Wang et al., 2016)) or inter-regional (or intra-regional) markets (such as equity markets, foreign exchange markets (Jang et al., 2011), derivative markets (Lautier and Raynaud, 2012)), and financial institutions (such as banks (Georg, 2013; Reyes and Minoiu, 2011), long-term lending institutions, non-banking financiers etc.). By making use of these complex network models, researchers have been successful in both describing the topological properties and in characterizing the structure of linkages existent within these interactive financial systems.

Among the set of investable emerging market classes, the Asian capitals are relatively more attractive than other markets. Two important features that make Asian equity markets as attractive destination for portfolio diversification are: higher returns compared to developed economies (Harvey, 1994; Bekaert et al., 1998; Peter and Kannan, 2007) and relatively higher market segmentation (Dekker et al., 2001; Ng, 2002; Worthington, et al., 2003; Gérard, et al., 2003; Unteroberdoerster and Pongsaparn, 2011; Claus and Lucey, 2012). Capital markets in Asia possess characteristics that distinguish them from capital markets elsewhere in the world. Asian markets are a heterogeneous class and the inter-country differences in Asia arises because of disparities in economic determinants such as size and structure of economies, the quality of institutions, population, level of urbanization, composition of intraregional and external trade, the level of development and maturity of the domestic debt and equity markets fiscal and monetary policies, interest rates movements, budget deficits, GDP rates, and industrial structures diversity across multiple countries & regions within those countries. This diversity across the Asian countries makes them less predisposed to broad macroeconomic shocks. By embracing a complex system framework for modelling the interdependencies of an integrated Asian equity markets, we can understand the systemic level behaviour wherein the microstructural units (in this case individual scripts listed in Asian equity markets or individual Asian stock market) adapt their behaviour with respect to a dynamic changes in the internal environment (domestic conditions of the Asian countries and market) and external environment (global conditions or regional conditions), causing the generation of redesigned systemic level behaviour.

Our study begins with some investigative questions: Are there existence of close clusters among Asian markets? Are these close cluster formation stable over long periods of time or are these very weak clusters? Which are the close clusters in Asian markets? Which are the hub nodes in the Asian market network through which the spread of systemic risk may happen majorly to other markets? How is the connectivity structure of key hub nodes to the Indian equity market? Which Asian markets cluster firmly with the Indian equity market? Which Asian indices have the highest and the lowest connectivity with Indian equity market? How was the change in the clustering and connectivity structure during the 2008 financial crisis? Does occurrence of global or regional financial crisis cause any consequent changes in the topological properties of the Asian indices network?

We examine the cross-market network structure of Asian indices using 14 years daily closing prices in local currency terms of the 14 Asian countries. The 14 countries which are part of our study are Singapore, Taiwan, Hong Kong, Malaysia, Pakistan, Indonesia, Israel, India, China, Japan, Jordan, Philippines, South Korea, and Sri Lanka. We utilize sub-dominant ultrametric distance for the construction of Minimum Spanning Tree and Average Linkage Hierarchical Clustering Tree models, and also compute some key topological measures of the Asian indices network. Using these three approaches, we assess the dynamically varying trends in cross-market connectivity structure and clustering formation among these equity markets.

The paper is structured into five sections along with two Supplementary files. In section two, we discuss a brief summary of the past studies on network based cross-correlation dynamics. In Section 3, we discuss the study dataset and the analytical methodological framework we made use of for this study. Section 4 presents the results of computational data analytics on the cross-market correlation networks. In Section 5, we present our concluding remarks. In the Supplementary Table 1, we give the details of the symbols used for referring to Asian market indices. In the Supplementary Fig. 1, we exhibits the plots obtained for the elbow method, the silhouette method, and the gap statistic method.

2. Related works

Network filtering approach such as MST and data mining approach such as Hierarchical clustering have been used in many studies to decipher clustering and connectivity structures in equities. Some of the domestic markets whose equity networks have been modelled using these techniques are the Japanese markets (Jung et al., 2008), the US markets (Mantegna, 1999), the Korean markets (Jung et al., 2006) and, the markets in UK (Coelho et al., 2007). There have also been a lot of studies that investigated cross-market indices of different countries using MST and Hierarchical clustering. The study by Bonanno et al. (2000) investigated the cross-market connectivity among indices of different countries. The study by Gilmore et al. (2008) investigated the cross-market linkages among the traded securities of capital markets belonging to USA, UK and the key market indices of the European Union. The major finding of this study was that the center of market indices network was the market indices of Germany and France, and there was very weak connectivity of market indices of Eastern European countries to the core of the network. A study by Eryiğit and Eryiğit, 2009 investigated the network of 143 market indices belonging to 59 countries. This study used MST with subdominant ultrametric distance to explore the connectedness of the cross-market indices.

However, there are less number of studies on regional cross-market linkages using MST and Hierarchical clustering trees. A study by Martin Cupal et al. (2012) has examined the European capital market's correlation networks using MST at two levels, one at a national level, and the other at the multinational level. The study attempts to find the indices network topological properties. As per the author's knowledge, the only network based study to attempt an investigation of the cross-market interdependencies of indices in Asian markets was by Sensoy and Tabak (2014). This study employed Dynamic Spanning Tree Approach by using an ARMA-FIEGARCH-cDCC process (Sensoy and Tabak, 2014). However, this study (Sensoy and Tabak, 2014) was more of an initial exploratory study on Asian market interdependencies and was only able to generate a limited number of insights about inter-market dependencies. In addition to this, the past study was about trends of static nature for some chosen time-frames and the dynamic evolution of markets was not considered for investigation. Apart from this aspects, this study (Sensoy and Tabak, 2014) also has not attempted to identify observable changes in network structure during the European debt crisis period, neither it has made any attempts to characterise the set of markets which are highly connected to the hub nodes in the network, nor it has pin-pointed the set of Asian markets least prone to impacts of any global crisis. Our current study is more extensive in nature and uses both the widely cited & time-tested methods of sub-dominant ultrametric distance (as proposed by Mantegna (1999)) based MST and Hierarchical clustering and insightful data-driven network metrics. Our study differs from the earlier study (Sensoy and Tabak, 2014) in terms of the time-frames used, the methodology applied, the network metrics computed, and the India-specific connectivity and rankings identified. Our study has also adopted a unique weighted hop count analysis to assess the measure of closeness of Indian equity market to other markets in Asia. Such weighted hop count analysis (as per the author's knowledge) has not been made use in any prior complex network studies on equity markets.

3. Methodology

3.1. Data

The study data comprises of daily prices in local currencies of fourteen Asian Indices. The closing prices belong to the stock exchanges of the following Asian countries: India, China, Sri Lanka, Malaysia, Pakistan, Indonesia, Israel, Hong Kong, Japan, Jordan, Philippines, Taiwan, South Korea, and Singapore. We pick these fourteen nations driven by two logics; the first logic is that in the Asian region, these set of countries belong to the most emerging economies class, and the second logic is that the equity markets in these countries have transformed to a relatively large extent because of the financial liberalization in the real economy started long back. We used fourteen years (1st January 2002 to 31st January 2016) data for the current study, and the study dataset had 3511 data points. The data was sourced in local currency from the Thomson Reuters Eikon database, the web portals of Yahoo Finance, and Investing.com. There were instances where prices of market indices for some days were missing. We replaced the missing values with the previous day's closing price of the index with an assumption that no trading activity was conducted on the previous day.

3.2. The rolling window approach

One of the key objectives of this study is the examination of the temporal patterns of cross-market connectedness in MST and cross-market clusters formation. In accordance with this goal, the rolling window approach was utilized for exploring this time-variant nature of the Asian indices network. We implement this rolling window approach on the fourteen years study dataset (3510 log returns) of the 14 Asian indices. From implementing this line of approach, we generate 151 temporally synchronous observations. Each of this observation has a size of 500 data points, and each subsequent observation is separated by a time-scale of 20 data points. For each of these 151 temporally synchronous observations, cross-correlation matrices are generated, and subsequently, from these matrices both the representative MST diagrams and the representative Hierarchical clustering trees are created. We also compute the network measures for the 151 observations.

3.3. Minimum spanning tree

We follow the methodological approach of construction of MST as postulated in the seminal study by Mantegna (1999). The log returns of the Asian market are first computed using the following equation:

$$S_i = \ln\left(\frac{CP_t}{CP_{t-1}}\right) \tag{1}$$

Where, CP is the daily closing price of the Asian market index in the respective currencies. The daily price returns are log transformed so as to generate a normalized data wherein there can be measurement of returns in a comparable metric irrespective of the origin of price series.

The cross-correlation coefficient is next computed by utilizing the following equation:

$$\rho_{i,j} \equiv \frac{\langle S_i S_j\rangle - \langle S_i\rangle\langle S_j\rangle}{\sqrt{\left(\langle S_i^2\rangle - \langle S_i\rangle^2\right)\left(\langle S_j^2\rangle - \langle S_j\rangle^2\right)}} \tag{2}$$

In Eq. (2), the variable $< S_i>$ represents the mean log return for a duration t. Here t is the size of the rolling window observation.

To construct an MST, the distance metric should fulfil the following three pre-requisite conditions, which are

(i) $d_{i,j} = 0$ *if and only if* $i = j$
(ii) $d_{i,j} = d_{j,i}$
(iii) $d_{i,j} \leqslant d_{i,k} + d_{k,j}$

The study by Mantegna (1999) has used the following distance measure for construction of the equity networks as it fulfills all the three axioms mentioned above.

$$d_{i,j} = \sqrt{2(1-p_{i,j})} 0 \leqslant d_{i,j} \leqslant 2 \tag{3}$$

We also use the same measure (Eq. (3)) for the Asian indices network. In our study, we construct the MST using the Prim's algorithm.

3.4. Hierarchical clustering

The Hierarchical Clustering is a data mining approach frequently used in financial networks to retrieve clusters possessing information of economic value. Using this tree method, the data can be ranked according to the sub-dominant ultrametric distances, and an ordered set of clusters can be so obtained.

Financial markets especially the equity markets have been considered as a type of complex systems (Johnson et al., 2003). In several disciplines of biological, physical and social sciences (especially economic sciences), it has observed that there is an existence of a nested hierarchical organizational structure. In this organizational framework, the elements comprising the system can be made segregated into distinct clusters, which may further be segregated into smaller sub-clusters (Simon, 1962). The dynamical nature of the complex systems is impacted by this hierarchy in the organizational structure; whereby multi-scale level of interactions plays an essential role to generate the structural makeup. Hence, a significant step in modelling and understanding such systems is to perform a mathematical depiction of the system level hierarchies (Tumminello et al., 2010). It is generally preferred to employ hierarchical clustering approaches when the underlying dataset has inherent hierarchical structures, which are true in the case of current study dataset consisting of correlations of financial time series. In addition to this, we can readily associate a clustering tree obtained from hierarchical clustering procedure to a correlation-based network acquired from a cross-correlation matrix. For instance, one can associate the minimum spanning tree of market indices (which reflects the shortest connecting distances) to either the single linkage or average linkage cluster analysis trees (Tumminello et al., 2010). The advantages that hierarchical clustering hold over non-hierarchical methods are as follows: a) in case of hierarchical clustering any valid measure of distance (here in the case ultrametric distance $d_{i,j}$) can be used as input which is not the case in non-hierarchical clustering methods like k-means, b) no previous information about the number of clusters is required for performing the hierarchical clustering steps, c) the dendrogram obtained from the hierarchical clustering approaches can provide a lot of insightful information into the data structure and can also facilitate identification of outliers in the study dataset, d) the final outcome of the k-means clustering of market indices returns may be in the form of segments. These segments may range in between the roughest one (wherein the cluster occupies all the data points) to the optimum ones (wherein the cluster is the form of a single data point); however, the segments would not possess any nested architecture and hence would not disclose any proper hierarchical patterns present in the data. Thus in this study, we employ a hierarchical clustering approach for deciphering the hierarchies underlying the interdependency structure of cross-market interactions in Asia.

We utilized the subdominant ultrametric distance $d_{i,j}$ as an input distance measure for generation of the hierarchical clustering trees. The values of the distance measure $d_{i,j}$ are already in a scaled form, so it can directly be taken as an input for generation of the hierarchical clustering trees without further transformation or scaling steps. The Average Linkage Clustering method was further used taking this distance as an input (Tumminello et al., 2010) to decipher the cluster formation among Asian indices.

We also perform the clustering of the Asian indices data using non-hierarchical K-means clustering approach, and perform a comparative analysis of the findings with that of hierarchical clustering. The K-means clustering analysis is performed on a sample of three study datasets, i.e. the datasets belonging to the pre-crisis, the crisis and the post-crisis periods. We employ the elbow method (Kassambara, 2017), the silhouette method (Rousseeuw, 1987) and the gap statistic method (Tibshirani et al., 2001) to obtain the optimum number of clusters prior to performing K-means clustering step. The clusters identified from K-means clustering for the three time-frames are compared with the ones derived from hierarchical clustering method.

3.5. Network topological properties

We embrace a few evaluative quantitative measures to grasp the nature of topology and network statistics of the Asian indices networks. We compute the mean correlation coefficients, the average MST length and its inverse, the maximum Eigen value, along with the normalized tree length to comprehend the regional interdependencies and effects of dynamic connectedness in the current studied Asian indices networks.

3.5.1. Mean correlation coefficients

The mean correlation coefficients in this study is computed using the following equation,

$$\bar{p} = \frac{1}{N(N-1)}\sum_{i\neq j} p_{ij}^{t} = \frac{2}{N(N-1)}\sum_{i<j} p_{ij}^{t} \tag{4}$$

3.5.2. Average MST length

The MST length referred by L is computed as the summation of all the sub-dominant ultrametric distances between nodes i and j that is having edge linkage in a given MST.

$$L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij} \tag{5}$$

Hence, we can state the inverse form of the MST length (L^{-1}) as an indirect means of measuring the integration levels implicit in the cross-market index correlations. We employ the inverse form of the average MST length as a measure for studying the evolutionary patterns of the cross-market indices of the Asian region.

3.5.3. Normalized tree length

The normalized tree length $L(t)$ was postulated by Onnela et al. (2002). The equation for computing the normalized tree length is given below,

$$L(t) = \frac{1}{N(N-1)}\sum_{d_{ij}\in T^{t}} d_{ij} \tag{6}$$

We use this equation for computation of the normalized tree lengths of the 151 MST plots representative of the 151 temporally synchronous observations.

3.5.4. Largest Eigen value

We can test the random behaviour in the stock returns by utilizing the concepts of random matrix theory (RMT). In general, for a correlation matrix of security returns in any equity market, the largest eigenvalue has a significant large magnitude relative to the others in the range specified by this equation above. A lot of past studies have postulated that this largest eigenvalue is capable of capturing the economic information encompassing the entire market-wide effect of the totality of securities in that particular market. And, this measure has a high level of robustness to changes in the market conditions impacting this collective movement. The magnitude of this parameter will provide us with a reflection of the collective evolutionary trend of the Asian capital markets, and thereby it also serves as a pin-pointer to inclining trends of cross-market correlations.

3.6. Hop count computation for India specific connectivity in MST

The hop count in the context of cross-market indices network refers to the number of intermediate indices through which data must pass between a source index and a destination index. This can give us a quantifiable measure of the contagion from a hub node of Asian indices network to India capital market. We compute the hop count between Indian capital market index and the other indices in all the 151 temporally synchronous observations. Hop count is the number of edges between two nodes i and j of the MST. We employ a weighted approach for computing a score of the closeness of an index to the BSE. The weight assigned according to the number of hop counts from BSE to the reference index is provided in Table 1. In our study, we assumed that indices having a hop count less than or equal to three is considered to be near BSE and those greater that three are considered distant. This weighted approach is executed for all the 151 observations, and

Table 1
Weight assigned as per the number of hop counts from BSE to a given Asian index.

Number of hope counts	Weight assigned
1	1
2	0.75
3	0.5
4	0.3
5	0.2
>6	0.1

based on the average weighted hop count scores the indices are ranked.

4. Computational analytical results and discussions

We have utilized dual data analytics strategies in this paper to track and enquire into the statistical properties, hierarchical features and dynamical cross-market connectivity structures among Asian markets for the fourteen year study period. The Asian indices MST graph obtained by this approach is an edge subset which forms a tree-like structure, and this sub-graph so constructed has a minimum summed up edge weight total. After retrieving the MST of the Asian indices that are representative of the 151 temporally synchronous observations, we move further for mapping the subdominant ultrametric space on to these respective hierarchical trees. And, in this manner, we attempt to decipher the distribution of cross-market hierarchies and intra-cluster and inter-cluster connectedness. In the hierarchical clustering trees, the vertical ax is an indicator of the ultrametric distance at which the attachment occurs between any two Asian indices.

4.1. Temporal dynamics of cross-market cluster formation

We observe that over the incrementing time scales, several strong clusters form and several of these clusters are persisting over long periods. Some weak clusters are also formed whose permanency is for few observations and subsequently, they break-up and re-join to some other groups. The first cluster to form with the lowest intra-cluster distance is between SEHK and SGX (average distance at which the cluster forms is 0.85 for observation 1–46). We observe that in a majority of the observations in the range of 1–118, this (SEHK and SGX) is the closest cluster. From observation 119 to the last observation (observation number 151), the cluster of two indices namely KRX and TWSE is the closest cluster (average length of cluster formation was 0.7 initially, but later on it rose to 0.9); followed by SEHK and KRX which comes at the second position in this range of observations. From observation 1 to 15, the 2nd closest cluster is formed by KRX and TWSE (average distance at which the cluster forms is 0.9). From the observation 16 to 35, the 2nd closest cluster is formed by KRX and TSE (average distance at which the cluster forms is 0.91). From observation 36 to 40, the closest cluster is the group of KRX and TWSE and the group of SEHK and SGX is the 2nd closest cluster. From observation 41 onwards the cluster of SEHK and SGX again regains the 1st position as the closest cluster. Again from observation 41 to 63, the cluster of KRX and TSE is the 2nd closest cluster whose average lengths during the initial phase is 0.85, and it later drops to 0.75. From observation 133 again the average length of all the clusters relatively increases, and this incline in intra-cluster length is observed till the last observation. From the observation 61 to 87, we observe that the most distant clade is KSE, followed by CSE and JSE. In all other observations prior to 61, JSE was the most distant clade among all the indices. From 119 to the last observation 151, the three group cluster of JSE, CSE and KSE are the most distantly visible group, and has an average cluster formation length at 1.25. From observation 47 onwards there is a significant reduction in the distances at which the clusters are formed. The average distance of the cluster formation further drops down from the starting of observation 50. From observation 1 to 49, JSE and CSE are the most distinctly identified distant clusters having an average distance of cluster formation at 1.38. However, from observation 50 to 56, JSE and CSE are distinctly different class as all clusters form under the same hierarchy in these observations. Though, during these observations we notice that the JSE and CSE are the last clades to be attached to the entire hierarchical tree, and have an average distance of cluster formation at 1.35, which is a significant reduction from the normal states. From observation 32 to 46, another distinctly visible cluster is formed by BSE and IDX having an average distance of cluster formation at 1.3. A big cluster formed of TWSE, TSE, KRX, SEHK and SGX having an average distance of cluster formation at 1.0 is visible throughout all the observations. Some of the indicative cluster formation patterns explained above is provided in Fig. 1.

4.2. Analysis of cluster formation and MST connectivity structure during 2008 financial crisis

We make a comparative analysis of the connectivity structure in the MST plots for the three time-periods: the pre-crisis period, the crisis period, and the post-crisis period. In the plot of pre-crisis period (Fig. 2(a)), we can notice that the hub nodes are the SEHK and SGX. The Asian indices of MYX and KRX are also branching nodes. The most distant nodes from the hub nodes are CSE and KSE.

In the plot of the crisis period (Fig. 2(b)), we can notice that the hub nodes are SGX and SEHK. The MST has shrunk relatively to previous time scale, yet distant nodes such as CSE and KSE are still

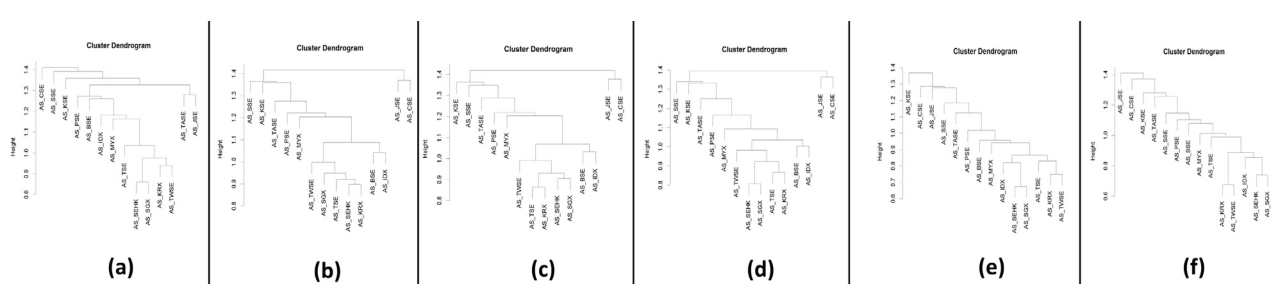

Fig. 1. Some of the sample Hierarchical clusters for the Asian indices at different periods of time (a) Hierarchical cluster for the observation No. 2 for the time period between 29 January 2002 and 21 January 2004; (b) Hierarchical cluster for the observation No. 33 for the time period between 15 July 2004 and 10 July 2006; (c) Hierarchical cluster for the observation No. 36 for the time period between 7 October 2004 and 4 October 2006; (d) Hierarchical cluster for the observation 43 for the time period between 2 May 2005 and 30 April 2007; (e) Hierarchical cluster for the observation 64 for the time period between 2 January 2007 and 7 January 2009; (f) Hierarchical cluster for the observation 119 for the time period between 13 June 2011 and 11 June 2013.

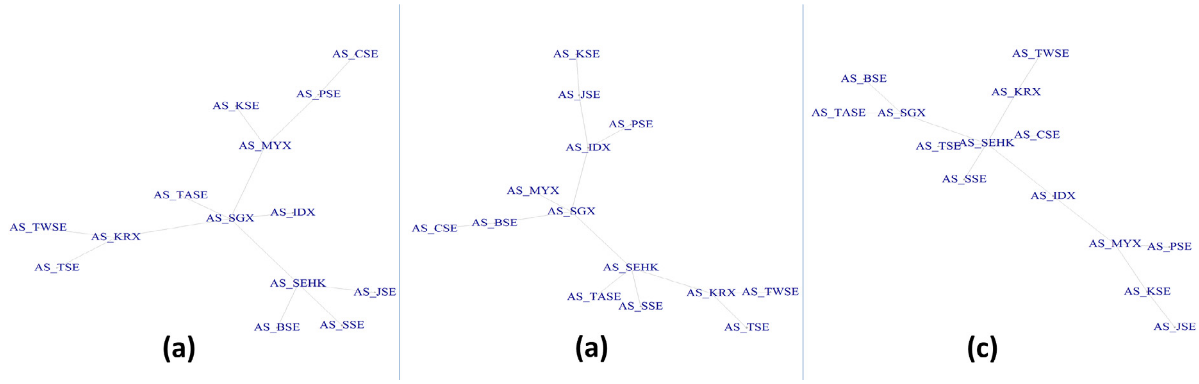

Fig. 2. (a) MST of pre-crisis period for the Asian indices; (b) MST of crisis period for the Asian indices; and (c) MST of post-crisis period for the Asian indices.

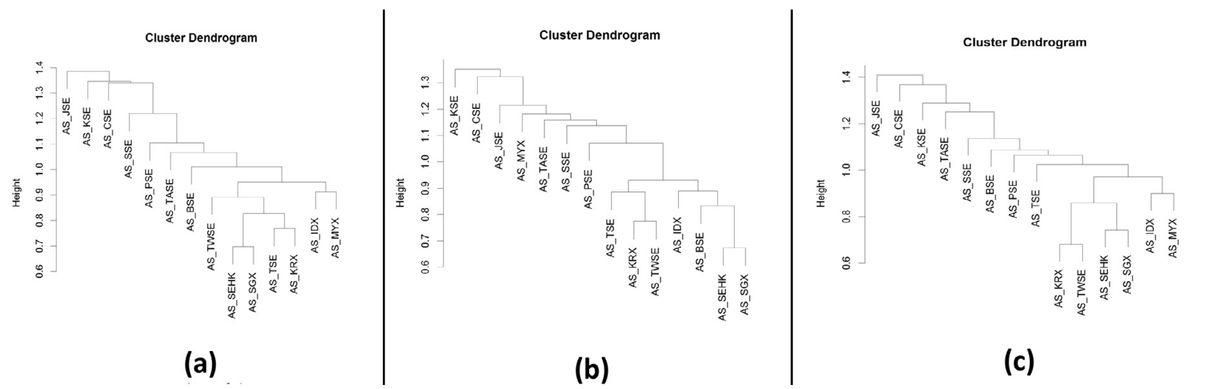

Fig. 3. (a) Hierarchical clustering dendrogram of pre-crisis period for the Asian indices; (b) Hierarchical clustering dendrogram of crisis period for the Asian indices; and (c) Hierarchical clustering dendrogram of post-crisis period for the Asian indices.

relevant and display weak correlation. This implies that by investing at this distant terminal links, diversification benefits can still be accrued during the crisis period.

In the plot of the post-crisis period (Fig. 2(c)), we can notice that the hub nodes remain identical (namely SGX and SEHK) to the ones observed during pre-crisis and post-crisis phases. The distant nodes in the case are JSE and KSE.

We make comparative analysis of the clustering structure in Hierarchical Clustering trees for the three time-periods: the pre-crisis period, the crisis period, and the post-crisis period. We find that during the 2008 financial crisis the height of the dendrogram increases (Fig. 3(b)). Also, we notice in Fig. 3(b) that the average distance of matrix also decreases during the 2008 financial crisis. The average distance of matrices in the pre-crisis period, the crisis period and the post-crisis period distances are 0.537735, 0.532507, and 0.548411 respectively. Although there is a noticeable shrinkage in the MST distance during the financial crisis, but it not at higher side relative to that of studies on European and American markets or on other developed markets. This gives us an indication that there is an increase in correlation but still there is a relative weakness in linkage structure compared to other developed markets.

In the Hierarchical tree plot of pre-crisis period (Fig. 3(a)), we can notice that the closest clusters are formed by SEHK and SGX, having a distance of cluster formation at 0.7. The next closest cluster is formed by TSE and KRX, materializing at a distance of 0.78. Another distinct cluster formation is observed between IDX and MYX. The JSE, KSE and CSE are the most distant clades. They form a cluster at a distance of 1.34. The JSE is the most distant clades among all.

In the Hierarchical tree plot of the crisis period (Fig. 3(b)), we can notice that the closest clusters are SEHK and SGX having a distance of cluster formation at 0.67, which is relatively lesser than during the pre-crisis period. The next closest cluster is TWSE and KRX, forming at a distance of 0.78. Another distinct observation is the height of the tree is lesser than that of pre-crisis periods, revealing that there is shrinkage of intra-cluster and inter-indices distances respectively. The KSE, JSE, and CSE are the most distant clades. They form a cluster at a distance 1.22. The KSE is the most distant clades among all.

In the Hierarchical tree plot of post-crisis period (Fig. 3(c)), we can notice that the closest clusters are KRX and TWSE having a distance of cluster formation at 0.69. The next closest cluster is SEHK and SGX, formed at a distance of 0.78. Another distinctly observable cluster is formed between IDX and MYX. Another distinct observation is that the height of the tree is relatively increased in comparison to the crisis periods, and has recovered back to pre-crisis state. It reveals that the shrinkage of intra-cluster and inter-indices distances have reduced considerably during this period. The JSE, CSE, and KSE are the most distant clades. They form a cluster at a distance 1.28. The JSE is the most distant clades among all.

We next discuss the results obtained from the K-means clustering on the study dataset of pre-crisis, crisis and post-crisis. The number of clusters obtained for the pre-crisis period using the elbow method, the silhouette method and the gap statistic method is exhibited in the Supplementary Fig. 1. Similar natures of plots are also obtained for the crisis and the post-crisis periods. The optimum number of clusters is obtained is three for all the study datasets (pre-crisis, crisis and post-crisis periods). The Asian indices that form the clusters obtained for the study dataset of pre-crisis, crisis periods and post-crisis periods are presented in Table 3. On comparison of the clustering results of hierarchical and k-means, we can note that in case of all the three periods (pre-crisis, crisis and post-crisis) the Asian indices of AS_JSE, AS_KSE, and AS_CSE are the most distant nodes in the dendrograms (Fig. 3(a)–(c)); and the same has been reflected in the cluster 1 obtained by the K-means clustering approach (Table 3). Another visible aspect in the dendrogram derived from the hierarchical clustering process (Fig. 3(a)–(c)) is that the cluster formed by AS_SEHK and AS_SGX is the closest one in all the observations; however, in the case of clusters obtained from the K-means clustering, this information is not truly reflected. The third visible aspect that is to be noted is that the cluster 3 (Table 3) obtained by the K-means clustering has AS_BSE, AS_IDX, AS_TSE, AS_PSE, AS_KRX and AS_TWSE as a separated cluster; however, the hierarchical organization cannot be truly deciphered from this cluster information. On the other hand, the dendrogram obtained from the hierarchical clustering on all the three observation has clearly exhibited the nested hierarchy of these market indices (AS_BSE, AS_IDX, AS_TSE, AS_PSE, AS_KRX and AS_TWSE).

4.3. Network metrics analysis

It is well established in several past studies that node distance measure of a given Minimum Spanning Tree is a monotonically decreasing function of the inter-node correlation. In several complex network studies on equity markets, it is well observed that there is a subsequent increment in the correlation structure between stocks (and indices) with the initiation of global (or regional) financial turmoil and this successively results in shrinkage of overall MST length. We notice the similar type of pattern in the Asian indices network also. In the Fig. 4, we can notice that during two periods (2008 Sub-prime mortgage financial crisis, and early 2011 to mid-2012 European debt crisis) of the financial crisis there is steady incline in the inverse of MST length, which indicates that the MST has relatively shrunk during this phases. Once the first crisis (2008 Sub-prime mortgage financial crisis) phases passes out, the inverse MST length returns back to its normal state and again rises during the 2nd crisis (debt crisis in the Euro zone during periods of initial months of 2011 to middle phase of 2012). However, the rise during the 1st crisis phase is relatively higher than in the case of the 2nd crisis.

In Fig. 5, we can notice that the trend line of the temporally variation plot of the maximum Eigen value (generally accepted as carrying the collective information of all the interacting stocks or indices) of the market indices network is similar to the plot of the mean correlation coefficient (Fig. 6). We can observe similar peaks during two crisis periods (2008 Sub-prime mortgage financial crisis and debt crisis in the Euro zone during periods of initial months of 2011 to middle phase of 2012).

It is well established in the several studies of equity network in past that there is a subsequent increment in the correlation structure between the stocks (and the indices) with the initiation of global (or regional) financial turmoil. The similar patterns are observed in the time variant plot of the Asian indices correlation structure. In the Fig. 6, we can notice that during two periods i.e. the 2008 Sub-prime mortgage financial crisis (late months of 2008 to middle of 2009), and the debt crisis in the Euro zone (initial months of 2011 to middle periods of 2012), there is a steady

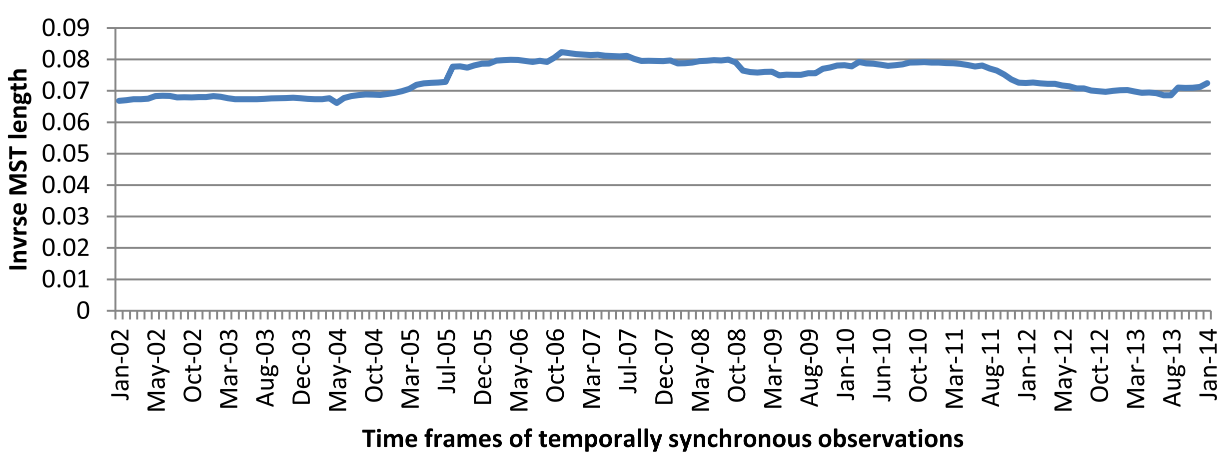

Fig. 4. Time variant plot of inverse MST length for the Asian indices network.

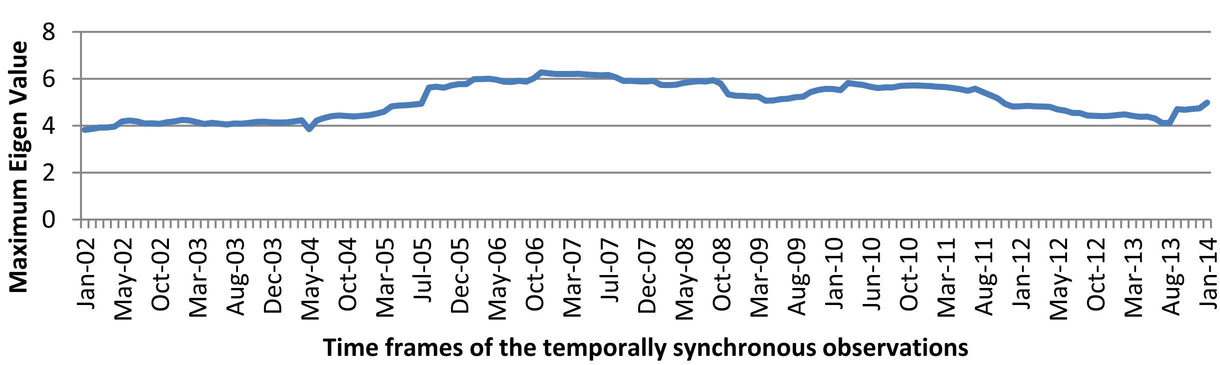

Fig. 5. Time variant plot of the Maximum Eigen value for the 151 correlation matrices.

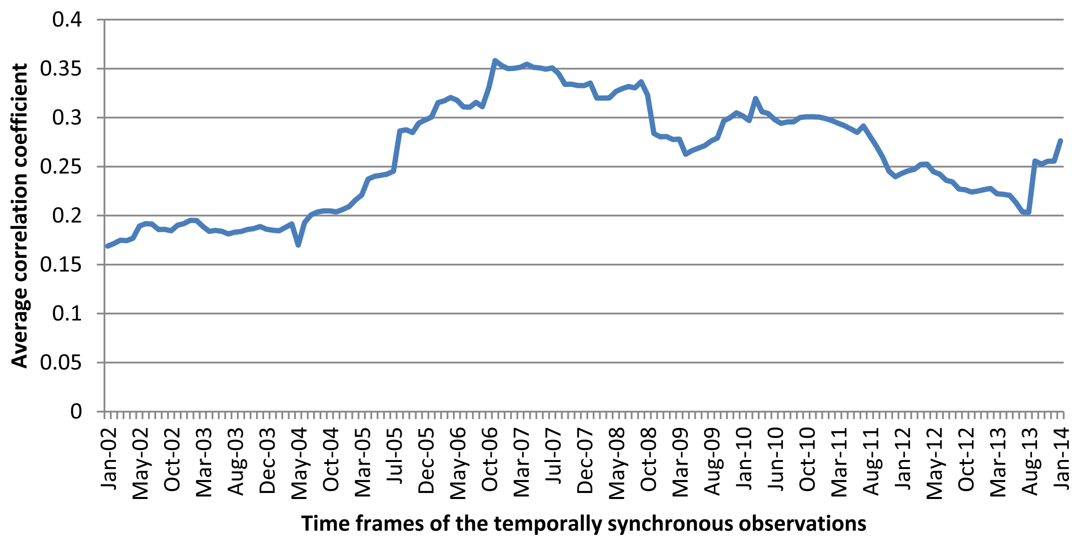

Fig. 6. Time variant plot of the average correlation coefficient for the 151 correlation matrices.

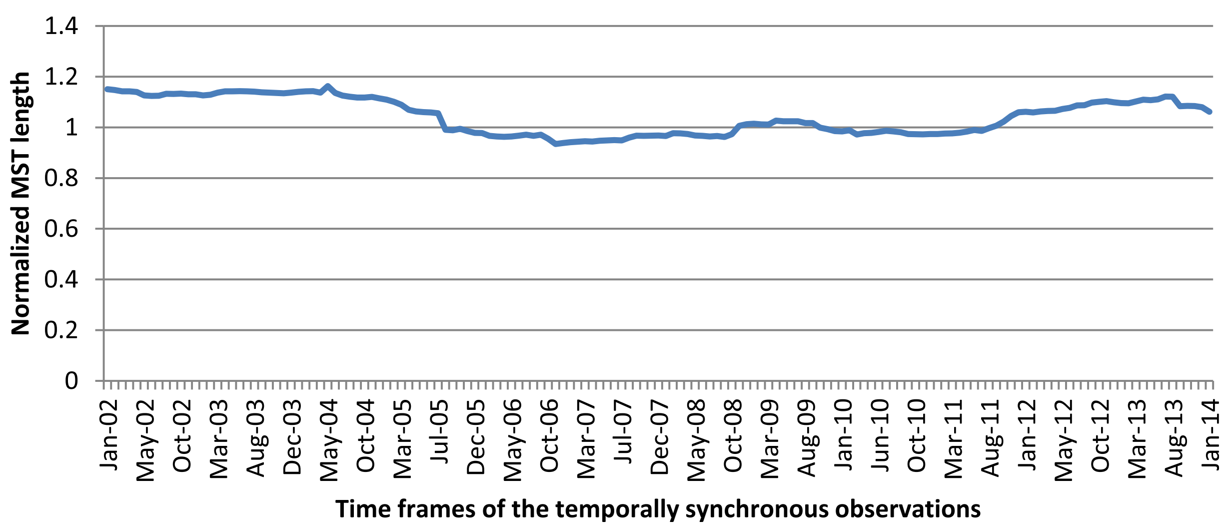

Fig. 7. Time variant plot of the Normalized MST lengths for the Asian indices networks.

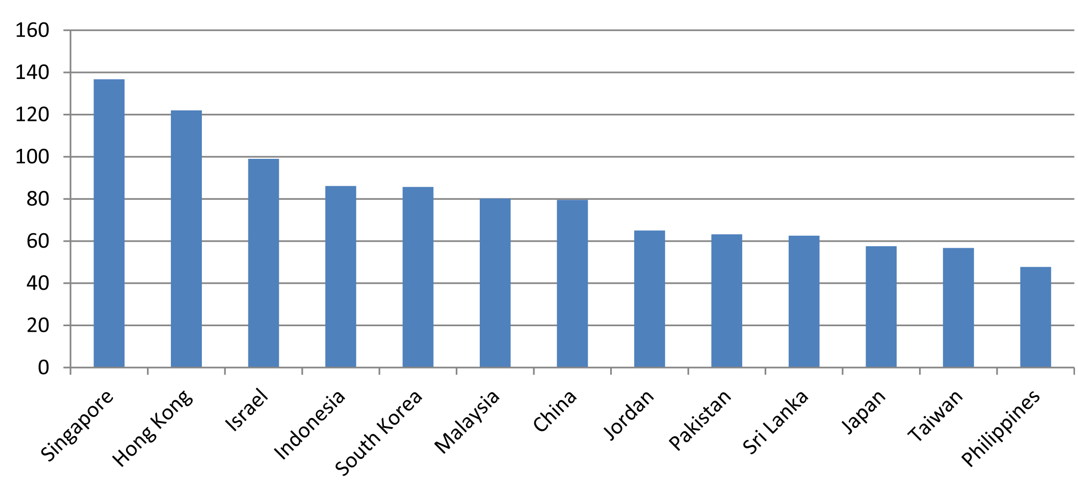

Fig. 8. Summed up weighted hop measure of Asian indices (indicative of distance from Indian market index in MST).

Table 2
Ranking based on least number of hops from BSE.

Index name	Country	Weighted hop count	Rank
AS_SGX	Singapore	136.75	1
AS_SEHK	Hong Kong	122	2
AS_TASE	Israel	99	3
AS_IDX	Indonesia	86.1	4
AS_KRX	South Korea	85.7	5
AS_MYX	Malaysia	80.2	6
AS_SSE	China	79.55	7
AS_JSE	Jordan	65	8
AS_KSE	Pakistan	63.2	9
AS_CSE	Sri Lanka	62.55	10
AS_TSE	Japan	57.55	11
AS_TWSE	Taiwan	56.7	12
AS_PSE	Philippines	47.75	13

incline in the average correlation values of the correlation matrices belonging to 14 Asian indices. However, we can note that the rise of the correlation values for the 1st crisis phase is relatively higher than in the 2nd crisis phase.

In past empirical studies on equity networks, it is well established that there exists a strong correlation between the normalized tree length measure and the existent diversification prospective. A relatively greater magnitude of this metric speaks of a scenario in which there is a high degree of inter-indices linkage, and the reverse is also true. We note that the normalized tree length was varying in the range between 0.949 and 1.150.

We also notice that there are exactly opposing fluctuation patterns in the trend lines of the inverse MST length and the normalized tree length. In particular, both inverse MST length and normalized tree length exhibits a special kind of fluctuations during the time-frames of the 2008 subprime crisis in the US and the 2011–12 debt crisis in the Euro zone. The same kind of movement is also noticeable in the case of the mean correlation coefficient. The normalized tree length fluctuated in the interval of 0.949 and 1.150.

In this regard, we sensibly infer that this network metrics are able to capture the trend of financial turmoil in spite of them being a form of lagging indicator. This is in line with the past findings, which contends that the inter-nodal cross-market co-movements in the equity market are suggestively impacted by global or regional financial events of influencing nature in the real economic space (Flavin et al., 2008).

4.4. Analysis of the India-specific connectivity with Asian peers in MST plot

The weighted sum based on the number of hops from BSE (Bombay Stock Exchange) is given in Fig. 7. We can observe from the weighted sum values, that the market indices of Singapore and Hong Kong are the most important link in the Asian indices network in the context of their impact on the BSE. It is followed by Israel, Indonesia and South Korea. The Asian indices of Philippines, Taiwan, Japan, Srilanka, Pakistan and Jordan have the least weighted sum values of hop count, which indicates their very limited impact on the BSE (Fig. 8). Table 2 presents the ranks of the Asian indices based on the weighted sum values of hop counts.

The rankings of each of the Asian indices based on the weighted sum measure are given in Table 1. We also plot the number of hops of the Asian indices ranked in the top four positions (in the ranking list in Table 1) in Fig. 9.

Table 3
Clusters obtained using K-means clustering method.

Time-frames	Asian market indices in Cluster 1	Asian market indices in Cluster 2	Asian market indices in Cluster 3
Pre-crisis period	AS_SSE, AS_JSE, AS_KSE, and AS_CSE	AS_SGX	AS_BSE, AS_SEHK, AS_IDX, AS_TASE, AS_TSE, AS_MYX, AS_PSE, AS_KRX, and AS_TWSE
Crisis period	AS_SSE, AS_TASE, AS_JSE, AS_MYX, AS_KSE, and AS_CSE	AS_SEHK	AS_BSE, AS_IDX, AS_TSE, AS_PSE, AS_SGX, AS_KRX, and AS_TWSE
Post-crisis period	AS_TASE, AS_JSE, AS_KSE, and AS_CSE	AS_SSE and AS_SEHK	AS_BSE, AS_IDX, AS_TSE, AS_MYX, AS_PSE, AS_SGX, AS_KRX, and AS_TWSE

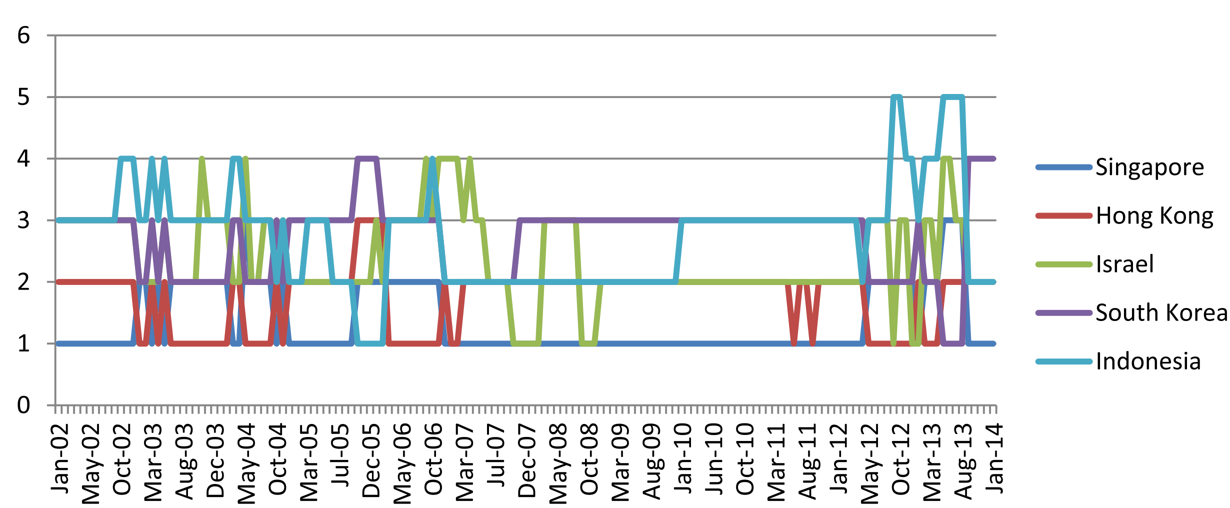

Fig. 9. Number of hops jumps from Indian market index to the top four (based on rankings in Table 2) Asian Indices in MST plots (of the 151 temporally synchronous observations).

5. Conclusion

We begin our study with some interesting questions pertaining to cross-market connectivity structure and inter-market clusters. The first set of questions was about the presence of close clusters in the Asian indices networks. From our analysis, we could observe that there is presence of three to four close clusters possessing very low ultrametric distance throughout the 151 observations. For a long period of time, some close linked clusters are existent and their decoupling happens very rarely. The second set of questions was about hub nodes present in the Asian indices network. From our analysis, we could observe that Hong Kong, Singapore and South Korea serve as hub nodes in several of the MST diagrams reflecting their central positions in the Asian indices network. Indian equity market is very well coupled with these hub nodes as reflected by weighted hop count analysis, where these hub nodes (Hong Kong, Singapore and South Korea) figure as the top-ranked indices. This also indicates that Indian market has a high exposure to risks originating out of any of these hub nodes (hub nodes country's markets) or risks of global nature directed at these hub nodes. The third set of questions was regarding the structural change in MST and Hierarchical clustering trees during the 2008 financial crisis. In the context of the average MST length, during the 2008 financial crisis there was noticeable shrinkage in the MST length; however, the shrinkage is relatively lesser to other indices network studied by previous researchers. Before and during 2008 financial crisis, SGX was the central node through which most information between any two indices would pass, and SEHK seems to be the hub node where the major block of Asian indices are connected. However, in the post-crisis period phase, we notice that SEHK has become the central node through which all cross-indices information should pass, and SGX has become a small hub node connecting some indices. In the hierarchical clusters of the pre-, during-, and the post- crisis periods respectively, we can observe that there is a change in the height of the clusters. The average height of the clusters has relatively decreased during the crisis periods and during the post-crisis period, the height has inclined. The comparative analysis of the hierarchical clustering results (obtained from pre-crisis, crisis and post-crisis periods) with that of K-means clustering approach revealed that both the methods are able to identify the distant Asian indices correctly; however, the nested hierarchical organization of other Asian indices cannot be truly deciphered from the K-means clustering unlike that of hierarchical clustering wherein the nested hierarchy is distinctly visible. The fourth set of questions was about the association of financial crisis with the change in key network measures. From our analysis, we notice that all the network measures seem to be a lagging indicator of the increase in correlation structure during a financial crisis.

There are five key findings of this current study: (i) the first one is that the Asian countries of Hong Kong and Singapore form the strongest clusters with minimum intra-cluster distance, which is followed by cluster formed by South Korea with Taiwan and the cluster formed by South Korea with Japan. The most distant clusters are formed by Jordan, Sri Lanka and Pakistan; (ii) the second key finding is that Hong Kong forms the major hub node in a predominant number of the MST plots, followed by Singapore and South Korea which all forms hub nodes. This finding is in line with the finding of the study by Şensoy & Tabak (Şensoy and Tabak, 2014); (iii) the third key finding is that the distance at which the cluster formation occurs relatively decreases during times of financial crisis and again rebounds to the original state in normal conditions; (iv) In terms of closeness to the Indian market index, the Asian market indices of Singapore, Hong Kong, Israel and Indonesia are the closest; and (v) There has been noticeable decline in the inverse MST length and observable incline in the average correlation coefficients, the maximum Eigen values and the normalized MST lengths during the both the phases of 2008 sub-prime financial crisis in US and the debt crisis in Eurozone.

We can use the identified indices from MST which are the terminal ends of MST for making a diversified regional Asian portfolio. The paper by Esfahanipour and Zamanzadeh (Esfahanipour and Zamanzadeh, 2013) has demonstrated that it is possible to generate an efficient frontier for the entire market by choosing a selected set of terminal stocks of the MST plot, and by doing so the risk-return outcome of the whole market can be matched with that of selected set of stocks. In future studies, the same kind of approach can be performed to generate a cross-market Markowitz's (Markowitz, 1952) optimized portfolio for Asian region.

Acknowledgements

The authors acknowledge Mr. Manu and Mr. Deepak from School of Management Studies, National Institute of Technology, Calicut for their kind assistance during the course of the research work. The author acknowledges Ms. Meena Panicker, Faculty, SOMS, NITC for the language editions.

Funding

The authors are grateful for the NIT Calicut institutional scholarship for the purpose of the research study.

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jksuci.2017.11.002.

References

Bekaert, G., Erb, C.B., Harvey, C.R., Viskanta, T.E., 1998. Distributional characteristics of emerging market returns and asset allocation. J. Portfolio Manage. 24, 102–116. https://doi.org/10.3905/jpm.24.2.102.

Bhattacharjee, B., Shafi, M., Acharjee, A., 2016. Investigating the influence relationship models for stocks in indian equity market: a weighted network modelling study. PLOS One 11. https://doi.org/10.1371/journal.pone.0166087.

Bonanno, G., Vandewalle, N., Mantegna, R.N., 2000. Taxonomy of stock market indices. Phys. Rev. E 62, R7615. https://doi.org/10.1103/physreve.62.r7615.

Claus, E., Lucey, B.M., 2012. Equity market integration in the Asia Pacific region: Evidence from discount factors. Res. Int. Bus. Finance 26, 137–163. https://doi.org/10.1016/j.ribaf.2011.09.002.

Coelho, R., Hutzler, S., Repetowicz, P., Richmond, P., 2007. Sector analysis for a FTSE portfolio of stocks. Phys. A: Stat. Mech. Appl. 373, 615–626. https://doi.org/10.1016/j.physa.2006.02.050.

Cupal, M., Deev, O., Linnertová, D., 2012. Network structures of the european stock markets. In: Proceedings of 30th International Conference Mathematical Methods in Economics. Silesian University in Opava, School of Business Administration Karviná, Karviná, pp. 79–84.

Dekker, A., Sen, K., Young, M.R., 2001. Equity market linkages in the Asia Pacific region. Global Finance J. 12, 1–33. https://doi.org/10.1016/s1044-0283(01)00025-4.

Eryiğit, M., Eryiğit, R., 2009. Network structure of cross-correlations among the world market indices. Phys. A: Stat. Mech. Appl. 388, 3551–3562. https://doi.org/10.1016/j.physa.2009.04.028.

A. Esfahanipour S.E. Zamanzadeh 2013. A Stock Market Filtering Model Based on Minimum Spanning Tree in Financial Networks. Amirkabir International Journal of Science & Research (Modeling, Identification, Simulation & Control) 45, 67–75.

Flavin, T.J., Panopoulou, E., Unalmis, D., 2008. On the stability of domestic financial market linkages in the presence of time-varying volatility. Emerging Markets Rev. 9, 280–301. https://doi.org/10.1016/j.ememar.2008.10.002.

Gao, Y.-C., Zeng, Y., Cai, S.-M., 2015. Influence network in the Chinese stock market. J. Stat. Mech: Theory Exp. 2015, P03017. https://doi.org/10.1088/1742-5468/2015/03/p03017.

Georg, C.-P., 2013. The effect of the interbank network structure on contagion and common shocks. J. Bank. Finance 37, 2216–2228. https://doi.org/10.1016/j.jbankfin.2013.02.032.

Gérard, B., Thanyalakpark, K., Batten, J.A., 2003. Are the East Asian markets integrated? Evidence from the ICAPM. J. Econ. Bus. 55, 585–607. https://doi.org/10.1016/s0148-6195(03)00055-9.

Gilmore, C.G., Lucey, B.M., Boscia, M., 2008. An ever-closer union? Examining the evolution of linkages of European equity markets via minimum spanning trees. Phys. A: Stat. Mech. Appl. 387, 6319–6329. https://doi.org/10.1016/j.physa.2008.07.012.

Harvey, C., 1994. Predictable risk and returns in emerging markets. Rev. Financial Studies 8, 773–816. https://doi.org/10.3386/w4621.

Jang, W., Lee, J., Chang, W., 2011. Currency crises and the evolution of foreign exchange market: Evidence from minimum spanning tree. Phys. A: Stat. Mech. Appl. 390, 707–718. https://doi.org/10.1016/j.physa.2010.10.028.

Jung, W.-S., Chae, S., Yang, J.-S., Moon, H.-T., 2006. Characteristics of the Korean stock market correlations. Phys. A: Stat. Mech. Appl. 361, 263–271. https://doi.org/10.1016/j.physa.2005.06.081.

Jung, W.-S., Kwon, O., Wang, F., Kaizoji, T., Moon, H.-T., Stanley, H.E., 2008. Group dynamics of the Japanese market. Phys. A: Stat. Mech. Appl. 387, 537–542. https://doi.org/10.1016/j.physa.2007.09.022.

Kassambara, A., 2017. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning, Multivariate Analysis. CreateSpace Independent Publishing Platform.

Lautier, D., Raynaud, F., 2012. Systemic risk in energy derivative markets: a graph-theory analysis. Energy J. 33, 217–242. https://doi.org/10.5547/01956574.33.3.8.

Mantegna, R., 1999. Hierarchical structure in financial markets. Eur. Phys. J. B 11, 193–197. https://doi.org/10.1007/s100510050929.

Markowitz, H., 1952. Portfolio Selection. J. Finance 7, 77. https://doi.org/10.2307/2975974.

Ng, T.H., 2002. Stock market linkages in South-East Asia. Asian Econ. J. 16, 353–377. https://doi.org/10.1111/1467-8381.00157.

Johnson, N.F., Jefferies, P., Hui, P.K., 2003. Financial Market Complexity: What Physics Can Tell Us About Market Behaviour. Economics & Finance,. Oxford University Press.

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertiész, J., 2002. Dynamic asset trees and portfolio analysis. Eur. Phys. J. B – Condens. Matter 30, 285–288. https://doi.org/10.1140/epjb/e2002-00380-9.

Pan, R.K., Sinha, S., 2007. Collective behavior of stock price movements in an emerging market. Phys. Rev. E 76, 046116. https://doi.org/10.1103/physreve.76.046116.

Peter, H., Kannan, P., 2007. Growth and returns in emerging markets. In: Takatoshi, I., Andrew, R. (Eds.), International Financial Issues in The Pacific Rim: Global Imbalances, Financial Liberalization, And Exchange Rate Policy. University of Chicago Press.

Reyes, J.A., Minoiu, C., 2011. A network analysis of global banking: 1978-2009. IMF Working Papers 11/74, 1–41. https://doi.org/10.5089/9781455227051.001.

Roy, R.B., Sarkar, U.K., 2011. Identifying influential stock indices from global stock markets: a social network analysis approach. Proc. Comput. Sci. 5, 442–449. https://doi.org/10.1016/j.procs.2011.07.057.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Setiawan, K., 2011. An alternative perspective on the stock markets integration: multilateral measure of the degree of integration. In: Proceedings of the 3rd EMG Conference on Emerging Markets Finance. CASS Business School – City University London, London, pp. 1–33.

Sensoy, A., Tabak, B.M., 2014. Dynamic spanning tree approach – the case of Asia-pacific stock markets. Working Paper Series (Research Department of Borsa İstanbul) 15, 1–26.

Simon, H.A., 1962. The Architecture of Complexity. Proc. Am. Philos. Soc. 106, 467–482.

Tabak, B.M., Serra, T.R., Cajueiro, D.O., 2010. Topological properties of stock market networks: the case of Brazil. Phys. A: Stat. Mech. Appl. 389, 3240–3249. https://doi.org/10.1016/j.physa.2010.04.002.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc.: Series B (Stat. Methodol.) 63, 411–423. https://doi.org/10.1111/1467-9868.00293.

Tumminello, M., Lillo, F., Mantegna, R.N., 2010. Correlation, hierarchies, and networks in financial markets. J. Econ. Behav. Organ. 75, 40–58. https://doi.org/10.1016/j.jebo.2010.01.004.

Unteroberdoerster, O., Pongsaparn, R., 2011. Financial Integration and Rebalancing in Asia. IMF Working Papers 11, 1–32. https://doi.org/10.5089/9781463922641.001.

Wang, G.-J., Xie, C., Stanley, H.E., 2016. Correlation structure and evolution of world stock markets: evidence from pearson and partial correlation-based networks. Comput. Econ. 1–19. https://doi.org/10.1007/s10614-016-9627-7.

Worthington, A.C., Katsuura, M., Higgs, H., 2003. Price linkages in asian equity markets: evidence bordering the asian economic, currency and financial crises. Asia-Pacific Finance Markets 10, 29–44. https://doi.org/10.1023/b:fejm.0000039873.53973.3c.