Algorithmic Finance 2 (2013) 3–43
DOI 10.3233/AF-13015
IOS Press

Cluster formation and evolution in networks of financial market indices

Leonidas Sandoval Junior
Insper, Instituto de Ensino e Pesquisa Rua Quatá, 300, São Paulo, SP CEP 04546-042, Brazil
E-mail: leonidassj@insper.edu.br

Abstract. Using data from world stock exchange indices prior to and during periods of global financial crises, clusters and networks of indices are built for asset graphs based on distance thresholds and diverse periods of time, so that it is then possible to analyze how clusters are formed according to correlations among indices and how they evolve in time, particularly during times of financial crises. Further analysis is made on the eigenvectors corresponding to the second highest eigenvalues of the correlation matrices, revealing a structure peculiar to markets that operate in different time zones. We also study the survivability of connections and of clusters through time and the influence of noise in centrality measures applied to the networks of financial indices. The results show how the world's main stock market indices evolved in the last few decades with respect to their clustering structure, how their connections survive in time, and which indices are more central, according to different criteria. In particular, we witness the early formation and evolution of two main clusters, an American and an European one, the formation of a Pacific Asian cluster, and later on, of an Arab cluster. This analysis complements previous studies of the interdependencies of stock markets worldwide.

Keywords: networks, financial markets, cluster, evolution.

1. Introduction

The study of the dynamics of clusters has been occupying the minds of many researchers in the past years. In particular, one would like to study how clusters form and evolve in time. Networks that evolve in time are commonly known as *adaptive networks*, and they have been studied in a large number of fields. See Gross and Sayama (2009), Gross (2010), and Miller and Page (2007) for reviews on the subject.

In particular, some adaptive networks based on financial data have been the subject of attention of some research. As examples, Sieczka and Hołyst (2009) studied the correlations in commodity markets, investigating correlations of future contracts for commodities, and concluding that the market was constantly getting more correlated within the investigated period. Reimann and Tupak (2010) developed and analyzed a model for the dynamics of prices using statistical properties, and showed that the coupling of processes in time is essential for understanding the behavior of a financial market. In Qiu et al. (2010), based on the daily data of American and Chinese stock markets, the dynamic behavior of a financial network with static and dynamic thresholds was studied, and it was shown that the dynamic threshold suppresses the large fluctuation induced by the cross-correlation of individual stock prices, and leads to a stable topological structure in the dynamic evolution.

Fenn et al. (2010) explored the dynamical clustering of exchange rates, investigated the roles of exchange rates in the network and showed that they depend on their position within a community, which changes in time. The article also shows major structural changes that occurred within the foreign exchange market and how they relate to crises in that market. In Song et al. (2011), the evolution of clusters based on the PMFG was studied using 57 stock market indices worldwide, showing that the network presents both a slow dynamics, related with the growing integration of financial markets worldwide, and a fast one, resulting from particular events that end up affecting the whole market.

The evolution of 418 of the constituting stocks of the S&P 500 was studied by Kenett et al. (2011). The authors found a rapid market transition at the end of 2001 that could lead to market systemic collapses. In Hołyst (2012), the author analyzes the network of the correlation strength between the 19 most developed countries, with respect to their GDP (Gross Domestic Product) per capita and its evolution in time.

In the present work, we build and analize networks based on a variety of indices of stock exchanges around the globe. The indices cover the five habitable continents, spanning a diversity of markets and economies so as to represent both evolved and developing economies. The series were taken both prior to and during some of the severest financial crises of the past decades, namely the 1987 Black Monday, the 1997 Asian Financial Crisis, the 1998 Russian Crisis, the burst of the dot-com bubble in 2001, the crisis after September, 11, 2001, and the recent Subprime Mortgage Crisis, that reached its peak in 2008, so that different regimes of volatility are represented. The idea is to be able to study how clusters behave when going from periods of low volatility to more turbulent times.

The correlation matrix of the time series of financial data encodes a large amount of information, and an even greater amount of noise. That information and noise must be filtered if one is to try to understand how the elements (in our case, indices) relate to each other and how that relation evolves in time. One of the most common filtering procedures is to represent those relations using a *Minimum Spanning Tree* (MST) (Mantegna, 1999; Bonanno et al., 2001, 2004; Micchichè et al., 2003; Coelho et al., 2007; Borghesi et al., 2007; Brida, 2008; Kwapień et al., 2009a, 2009b; Drożdż et al., 2010; Zhanga et al., 2011; Keskin et al., 2011; Sandoval, 2012), which is a graph containing all indices, connected by at least one edge, so that the sum of the edges is minimum, and which presents no loops. Another type of representation is that of a *Planar Maximally Filtered Graph* (PMFG) (Tumminello et al., 2005; Coronnello et al., 2005, 2007, 2010; Aste et al., 2005; Tumminello et al., 2007, Kenett et al., 2010), which admits loops but must be representable in two dimensional graphs without crossings.

Yet another type of representation is obtained by establishing a number which defines how many connections (edges) are to be represented in a graph of the correlations between nodes. There is no limitation with respect to the crossing of edges or to the formation of loops, and if the number is high enough, then one has a graph where all nodes are connected to one another. These are usualy called *Asset Trees*, or *Asset Graphs* (Onnela et al., 2002, 2003a,b,c, 2004; Sinha and Pan, 2007; Ausloos and Lambiotte, 2007; Sandoval, 2011), since they are not trees in the network sense. Another way to build asset graphs is to establish a value (threshold) such that distances above it are not considered. This eliminates connections (edges) as well as indices (nodes), but also makes the diagrams more understandable by filtering both information and noise. Some previous works using graphic representations of correlations between international assets (indices or otherwise) can be found in Coelho et al. (2007), Coronnello et al. (2007), Ausloos and Lambiotte (2007), and Eryiğit and R. Eryiğit (2009).

Many of our results corroborate the ones obtained in Eryiğit and R. Eryiğit (2009), where the authors used daily and weekly index changes of 143 stock market indices from 59 different countries. Their analysis of the asset graphs, minimum spanning trees (MST) and planar maximally filtered graphs (PMFG) of those networks leads to the results that globalization has been increasing in recent years, that North American and European markets are much more strongly connected among themselves compared to the integration with the other geographical regions, that the integration of East Asian markets among themselves as well as to the Western markets is rather weak, and that the clustering of the indices is mostly geographical.

Here, we analyze some of the dynamics and evolution of networks based on many of the indices of the world's stock exchanges. The analysis is based both on the hierarchical structure of the networks derived from the correlation of their nodes and on their evolution in time. The data used in this article consist on some of the benchmark indices of a diversity of stock exchanges around the world. The number goes from 16 indices (1986) to 92 indices (2007 to 2011). The choice of indices was mainly based on availability of data, since some of the stock exchanges being studied did not develop or had no recorded indices until quite recently. As there are differences between some of the days certain stock markets operate, some of the operation days had to be deleted and others duplicated. The rule was the following: when more than 30% of the markets did not operate on a certain day, that day was deleted. When that number was bellow 30%, we repeated the value of the index from the previous day for the markets that did not open. Special care had to be given to markets whose weekends did not correspond to the usual ocidental days, like some Arab countries. For

those markets, we adjusted the weekends in order to match those of the majority of markets. All that was done in order to certify that we minimized the number of days with missing data, so that we could measure the log-returns of two consecutive days of a market's operation whenever that was possible.

In Section 2, asset graphs for 92 benchmark indices of a variety of stock exchanges worldwide are built using a correlation based distance measure. Section 3 considers the survivability of clusters through time. Section 4 shows a study of some centrality measures of the networks obtained from the asset graphs and of how they change with the value of the threshold used in order to build each graph and also how the node degrees evolve in time. Section 5 consists on a general conclusion.

One contribution of this article is the study of the evolution in time of a large number of stock market indices, involving both influent indices and marginal ones, focusing on times of crises. The cluster structure of the networks studied is analyzed by using a hierarchical structure based on the correlation between indices, thus showing which market indices are closer among themselves. The survivability of the connections is also studied using three different measures, including a new one, *survivability*, and the centrality of nodes is analyzed using a variety of centrality measures. So, another contribution of this article is the study of this set of data from many aspects. Also, this article helps clarifying the differences between clusters obtained from different threshold values, showing there is stability of results when choosing thresholds close to the noise level.

2. Asset graphs

In this section, correlation matrices based on the time series of benchmark indices of the variety of stock exchanges explained before are built using the Spearman rank correlation. The reason why we are using the Spearman correlation and not the usual Pearson correlation is because the former is more useful when measuring linear relations between variables, and the latter is better at measuring nonlinear relations between them (there is a comparison between both coefficients in Subsection 2.2). Each correlation matrix thus obtained is then used in order to build a measure of distance between nodes (indices), and then threshold values are chosen so that all distances bellow the threshold are used in order to create an asset graph, and the ones above it are discarded. By using a variety of thresholds, up to a limit where random noise becomes very intense, one is then able to witness the formation of clusters, first isolated and then unified, as the threshold goes up. The upper limit for the threshold is obtained by using randomized data based on the original data so as to keep the original frequency distributions but destroy any temporal relation between them. In order for one to contemplate the true distances between indices, two dimensional maps are built using Classical Multidimensional Scaling (Borg & Groenen, 2005).

Here we build asset graphs by assigning different threshold values for a distance measure obtained from the correlation between financial market indices. The data consists on the indices of stock markets worldwide (for a list of the indices, and their symbols in this article, see Appendix A), collected from the years 1986–1987, 1997–1998, 2000–2001, and 2007–2011, each period pertaining to a different crisis. In the main body of the article, we shall concentrate on the analysis of 2007–2011, and the results for the remaining crises shall be placed in Appendix B. Clusters are defined using a hierarchical criterium based on the distances between nodes: for a given threshold, nodes that are connected bellow the threshold are considered as forming a cluster.

The indices' values are computed daily, and here represented by P_t for day t. Then, we calculate their log-returns, defined by

$$R_t = \ln(P_t) - \ln(P_{t-1}), \tag{1}$$

where P_{t-1} is the return on day $t-1$.

The log-returns are then used to calculate the correlations between pairs of indices. From the many correlation measures, we chose Spearman's rank correlation, since it is better at identifying nonlinear correlations between variables. The most common choice would be Pearson's correlation, and we make a comparison between results obtained based on both types of correlations later in this same section.

Given a correlation coeficient c_{ij} in a correlation matrix of all indices, a distance measure (Mantegna, 1999) is defined as

$$d_{ij} = \sqrt{2\,(1 - c_{ij})}. \tag{2}$$

As correlations between indices vary from -1 (anticorrelated) to 1 (completely correlated), the distance between them vary from 0 (totally correlated)

to 2 (completely anticorrelated). Totally uncorrelated indices would have distance 1 between them.

Based on the distance measures, two dimensional coordinates are assigned to each index using an algorithm called Classical Multidimensional Scaling (Borg and Groenen, 2005), which is based on minimizing the stress function

$$S = \left[\frac{\sum_{i=1}^{n} \sum_{j>i}^{n} \left(\delta_{ij} - \bar{d}_{ij}\right)^2}{\sum_{i=1}^{n} \sum_{j>i}^{n} d_{ij}^2} \right]^{1/2}, \quad (3)$$

where δ_{ij} is 1 for $i=j$ and zero otherwise, n is the number of rows of the correlation matrix, and $\bar{d}_{ij}$ is an m-dimensional Euclidean distance (which may be another type of distance for other types of multidimensional scaling) given by

$$\bar{d}_{ij} = \left[\sum_{a=1}^{m} \left(x_{ia} - x_{ja}\right)^2 \right]^{1/2}. \quad (4)$$

The outputs of this optimization problem are the coordinates x_{ij} of each of the nodes, where $i=1,\ldots,n$ is the number of nodes and $j=1,\ldots,m$ is the number of each dimension, in an m-dimensional space. It is customary for the representation of the network to be well represented in smaller dimensions than m, and in the case of this article we shall consider $m=2$ for a 2-dimensional visualization of the network. The choice is a compromise between fidelity to the original distances and the easiness of representing the networks. In Shapira et al. (2009), the authors use *Principal Component Analysis* (Chou, 1975) in order to represent multidimensional data based on financial time series. Their three-dimensional representation of the results account for more than 75% of the original information.

As some correlations may be the result of random noise, we ran some simulations based on randomized data, consisting on randomly reordering the time series of each index so as to destroy any true correlations between them but maintain each frequency distribution intact. The result of such simulations (1000 for each period being studied) is a distance value above which correlations are probably due to random noise (this value is usually close to $d=1.3$). Then, asset graphs are built using distance measures as thresholds. As an example, for threshold $T=0.5$, one builds an asset graph where all distances bellow 0.5 are represented as edges (connections) between nodes (indices). All distances above this threshold are removed, and all indices that do not connect to any other index bellow this threshold are also removed.

In what follows, we show the resulting asset graphs concerning each year from 2007–2011, for some values of distance thresholds (for the other years, please see Apeendix B). Those values were chosen so as to help visualization, which becomes harder for higher thresholds and many more connections.

2.1. 2007 to present day - subprime mortgage crisis and credit crisis

The latest large financial crisis began in 2007, reached its peak in 2008, and is happening until now. This crisis was triggered by the default of a large number of mortgages in the USA. Subprimes are loans to borrowers who have low credit scores. Most of them had a small initial interest rate, adjustable for future payments, which led to many home foreclosures after the rates climbed substantially. Meanwhile, the loans were transformed into pools that were then resold to interested investors. Since the returns of such investments were high, a financial bubble was created, inflating the subprime mortgage market until the defaults started to pop up.

Because of their underestimation of risk, financial institutions worldwide lost trillions of dollars, and many of them declared bankruptcy. Because of that, credit lines tightened around the world, taking the financial crisis to the so called real economy, affecting many European countries and placing the Eurozone in danger. The world is yet to recover from this crisis, and many institutions are still to lose a good part of their assets in the following years.

In our research, we use a total of 92 indices (listed in Appendix A), 4 from North America, 2 from Central America, 2 from the islands of the Atlantic, 6 from South America, 35 from Europe, 2 from Eurasia, 28 from Asia, 2 from Oceania, and 11 from Africa.

2.1.1. Graphs for 2007

For the year 2007, whe have 252 days of data for 92 indices. The average minimum distance for 1000 simulations with randomized data is 1.24 ± 0.02, and the average maximum distance is 1.57 ± 0.01. About

14% of the connections are bellow the minimum threshold and none is above it. Figure 1 shows the two dimensional view of the asset graphs for the second semester of 2007, for thresholds $T = 0.4$, $T = 0.6$, $T = 0.9$, $T = 1.0$, and $T = 1.2$.

At $T = 0.4$, we already have connections between the S&P 500 and the Nasdaq, and between the UK and France, what is a consequence of greater unification of markets in time. For $T = 0.5$, new connections are made in Europe, now also with the indices of Germany, Italy, Belgium, and the Netherlands. We also have a connection between South Africa and Namibia, what is to be expected, since Namibia is very much dependent on South Africa, both politically and economically. Few modifications occur for $T = 0.6$; the European cluster is joined by Switzerland and Spain, and few other connections are made. Brazil joins the American network at $T = 0.7$. Austria, Sweden, and Finland join the European network at this threshold. For $T = 0.8$, Canada, Mexico, and Argentina join the American cluster, and the European cluster densifies, and receives the addition of Ireland, Luxembourg, Denmark, Norway, the Czech Republic, and South Africa. Greece and Turkey establish a connection, and a Pacific Asian cluster forms, with Japan, Hong Kong, South Korea, Malaysia, Singapore, Indonesia, and Australia. At $T = 0.9$, the American and European clusters are dense, but still not connected among themselves. Chile joins the American network through Mexico, Portugal, Greece, Poland, and Turkey join the European cluster, and the Pacific Asian cluster, still disconnected from the others, grows denser and receives the addition of Taiwan. The joining of the three clusters occur for $T = 1.0$, at which the European cluster also receives the addition of the indices from Hungary, Russia, Cyprus, and Israel. The connection between the Pacific Asian cluster to the European one is made mainly by Singapore. At $T = 1.1$, many more connections are made, but few indices join the global cluster, namely Peru, Iceland, India, and New Zealand. Colombia and Romania join the global cluster at $T = 1.2$, which is also joined by Serbia, Croatia, Montenegro, Estonia, Latvia, Lithuania, Ukraine, Thailand, Philippines, and Egypt. Noise starts to take over at $T = 1.3$ and above, and many more connections are made until $T = 1.6$.

2.1.2. Graphs for 2008

For the year 2008, whe have 254 days of data for 92 indices. The average minimum distance for 1000 simulations with randomized data is 1.24 ± 0.02, and the average maximum distance is 1.57 ± 0.01. About 18% of the connections are bellow the minimum threshold and none is above it. Figure 2 shows the two dimensional view of the asset graphs for the second semester of 2008, for thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

Once again, the first connections start at $T = 0.4$, between the S&P 500 and the Nasdaq, and between the European indices from the UK, France, and the Netherlands. For $T = 0.5$, the European cluster grows with Germany, Switzerland, Italy, Sweden, and Spain, and the connection between South Africa and Namibia appears. For $T = 0.5$, the European cluster grows denser, and receives the addition of Belgium and Finland. For $T = 0.7$, the European cluster is even denser, and with the addition of Austria, Luxembourg, Denmark, Portugal, and the Czech Republic. There also appears a connection between Greece and Cyprus, but both indices are still disconnected from the European cluster. For $T = 0.8$, Mexico joins the North American cluster, the European cluster connects with the indices from Ireland, Norway, and Poland, and a link forms between Hong Kong and Singapore. At $T = 0.9$, Canada, Brazil, and Argentina join the American cluster, and Chile connects with the European cluster. Greece, Hungary, Romania, Turkey, Cyprus, and South Africa also join the European cluster. Japan, Taiwan, South Korea, and Australia join Hong Kong and Singapore in a Pacific Asian cluster. For the first time at this threshold, a connection appears between the United Arab Emirates and Qatar. The American, European, and Pacific Asian clusters join at $T = 1.0$, Colombia connects with the European cluster, and Peru with both the American and the European clusters. Croatia, Ukraine, Russia, and Israel connect with the European cluster, and an isolated link forms between Estonia and Lithuania. The emerging Arab cluster is now joined by Oman, and India, Taiwan, and New Zealand join the Pacific Asian cluster. The merging of clusters become stronger for $T = 1.1$, Iceland, Slovenia, Bulgaria, Estonia, Latvia, and Lithuania join the European network, and Malaysia and Indonesia join the Pacific Asian cluster. Jordan, Kuwait, and Bahrein join the Arab cluster, which is still disconnected from the others. Connections start to multiply at $T = 1.2$, but few new indices are added to the global cluster, with which the Arab cluster is now connected. From $T = 1.3$ onwards, noise starts to dominate, up to the last connections, at $T = 1.6$.

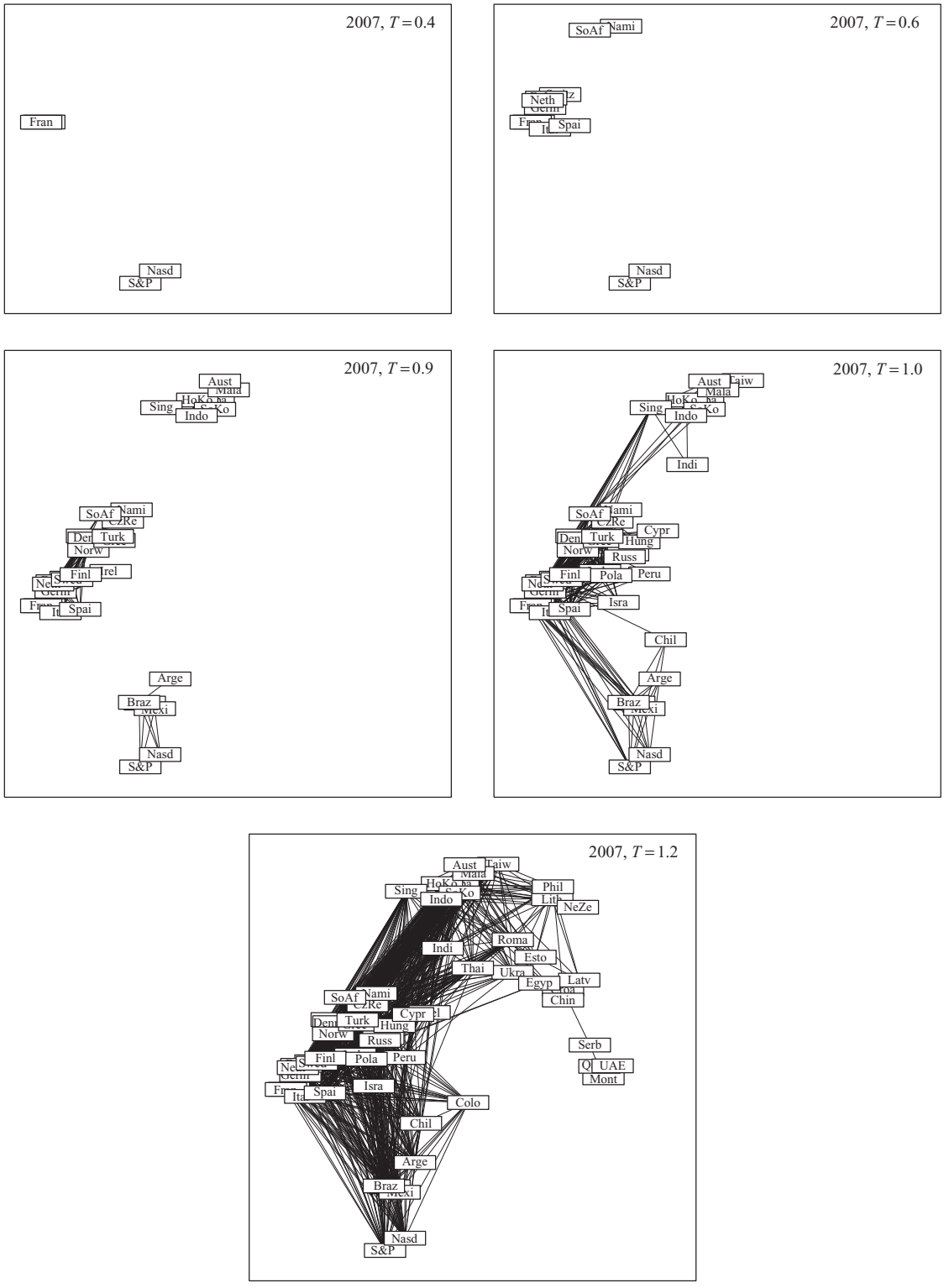

Fig. 1. Asset graphs for the year 2007, with thresholds $T = 0.4$, $T = 0.6$, $T = 0.9$, $T = 1.0$, and $T = 1.2$.

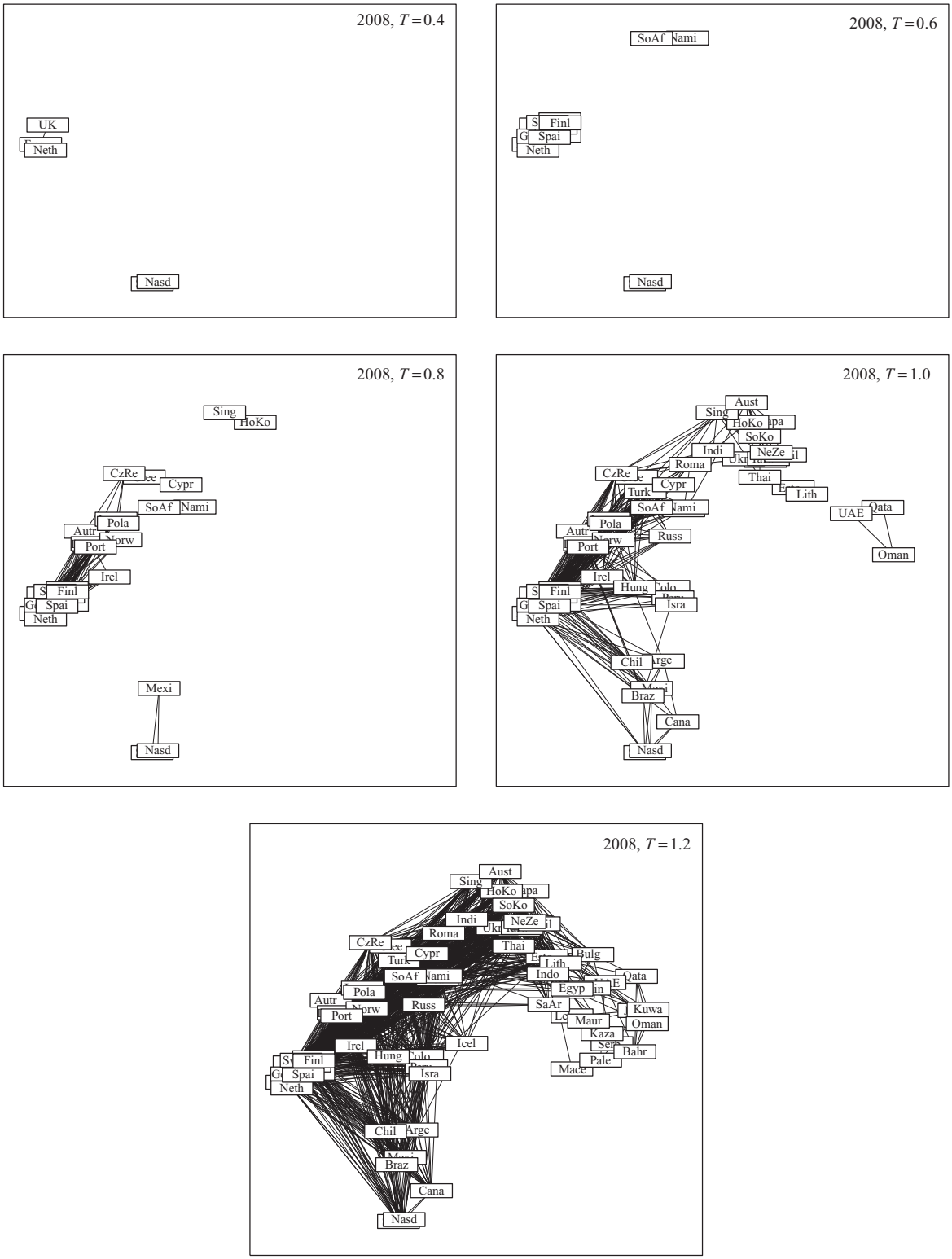

Fig. 2. Asset graphs for the year 2008, with thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

2.1.3. Graphs for 2009

For the year 2009, whe have 253 days of data for 92 indices. The average minimum distance for 1000 simulations with randomized data is 1.24 ± 0.01, and the average maximum distance is 1.57 ± 0.01. About 18% of the connections are bellow the minimum threshold and none is above it. Figure 3 shows the two dimensional view of the asset graphs for the second semester of 2009, for thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

At $T = 0.4$, the American and European clusters begin to form, the American cluster comprised of the S&P 500 and the Nasdaq indices, and the European cluster comprised of the indices from France, Germany, and the Netherlands. At $T = 0.5$, the indices from the UK, Italy, and Spain join the European cluster, and the connection between South Africa and Namibia is formed. For $T = 0.6$, the European cluster receives the index from Switzerland and makes more connections among its members. At

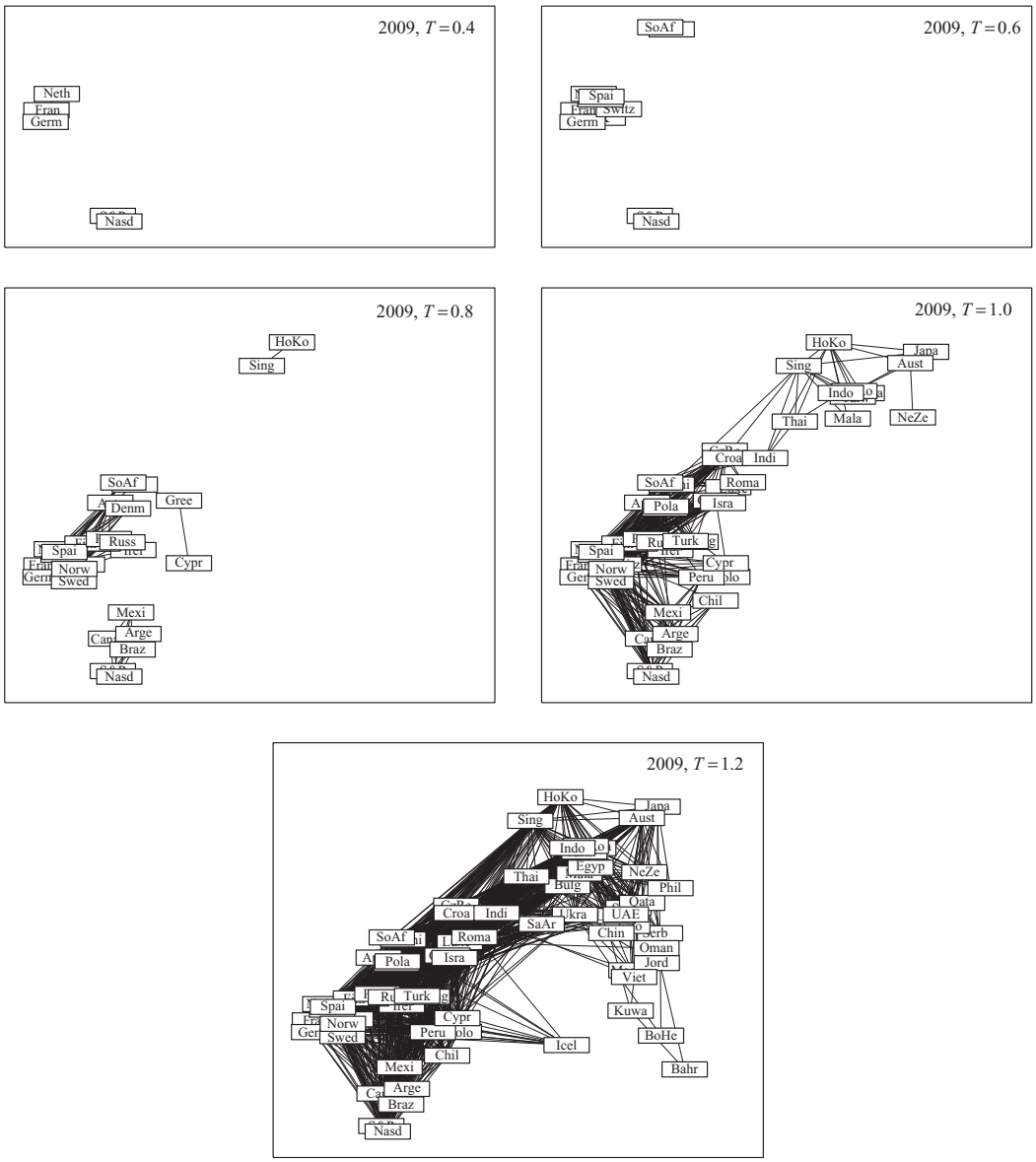

Fig. 3. Asset graphs for the year 2009, with thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

$T = 0.7$, Canada joins the American cluster and the European one densifies, and adds to itself Belgium, Sweden, Finland, Norway, and Portugal. Also, Greece and Cyprus form a connection. At $T = 0.8$, the American cluster densifies, and receives the indices from Mexico, Brazil, and Argentina. The European cluster adds Ireland, Austria, and Denmark to its ranks, and Hong Kong and Singapore make a connection with each other. For $T = 0.9$, the American and European clusters form many connections, and Luxembourg, Greece, the Czech Republic, Hungary, Poland, Russia, Turkey, Cyprus, and the pair South Africa - Namibia join the European cluster. Meanwhile, a Pacific Asian cluster is truly formed, with the addition of India, Japan, South Korea, Thailand, Indonesia, and Australia to the existing pair Hong Kong - Singapore. Integration between the American and the European clusters increases at $T = 1.0$. Chile integrates with the American cluster and forms one connection with the European one, Colombia forms connections with the European cluster, and Peru forms connections with both. Croatia, Romania, and Israel form connections with the European cluster, Kazakhstan links with Hong Kong, and the Pacific Asian cluster adds Malaysia and the Philippines to its ranks. All clusters densify for $T = 1.1$, Bulgaria, Estonia, and Lithuania join together the European cluster, which is also joined by Ukraine, and Saudi Arabia makes one connection with Europe and one with Pacific Asia. The indices from Jordan, Qatar, the United Arab Emirates, Oman, and Egypt form an Arab cluster. China, Taiwan, and the Philippines join the Pacific Asian cluster, which also makes many connections with Egypt. The threshold $T = 1.2$ brings even more densification to the now global cluster, but the only new indices to join are those from Serbia, Slovenia, Bosnia and Herzegovina, and Macedonia. Kuwait and Bahrein connect with the Arab cluster, which is now linked to the global cluster, and Vietnam makes a connection with the Pacific Asian cluster. Many more connections occur for $T = 1.3$ and beyond, and from $T = 1.4$ up to $T = 1.6$, where the last connections are made, noise dominates the picture.

2.1.4. Graphs for 2010

For the year 2010, whe have 256 days of data for 92 indices. The average minimum distance for 1000 simulations with randomized data is 1.24 ± 0.02, and the average maximum distance is 1.56 ± 0.01. About 17% of the connections are bellow the minimum threshold and none is above it. Figure 4 shows the two dimensional view of the asset graphs for 2011 for thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

At $T = 0.4$, we have the two usual small clusters, the American one formed by the S&P 500 and the Nasdaq, and the European cluster formed by France, Germany, and the Netherlands. For $T = 0.5$, the European cluster is joined by the indices of the UK, Italy, Belgium, and Spain. Again, at this threshold we have the connection between South Africa and Namibia. For $T = 0.6$, Switzerland joins the European cluster. For $T = 0.7$, this same cluster becomes denser and receives the addition of the indices from Austria, Sweden, Finland, Norway, and Portugal. Also, Greece and Cyprus make a connection at this threshold. For $T = 0.8$, Canada and Argentina join the American cluster, Ireland, the Czech Republic, and South Africa join the European cluster, and Russia makes a feeble connection with the same cluster. For $T = 0.9$, Mexico and Brazil join the American cluster, and the S&P 500 connect with the Netherlands. Luxembourg, Denmark, Hungary, Poland, and Turkey join the European cluster. A Pacific Asian cluster is formed by Japan, Hong Kong, Taiwan, South Korea, Singapore, Indonesia, and Australia. The merging of the American and the European clusters again happens at $T = 1.0$, but the Pacific Asian cluster is very lightly connected with the European cluster by a link through India, which now joins the Pacific Asian cluster, together with China and New Zealand. Connections between all three clusters increase for $T = 1.1$, and the global cluster is joined by Chile, Peru, Greece, Croatia, Romania, Estonia, Lithuania, Ukraine, Kazakhstan, Cyprus, Israel, Saudi Arabia, Thailand, Malaysia, the Philippines, and Egypt. Qatar connects with both Japan and Australia, and one cannot detect an Arab cluster. The global cluster becomes denser at $T = 1.2$, but the only additions to it are the indices from Colombia, Bulgaria, Qatar, United Arab Emirates, and Oman. Many more connections are formed for $T = 1.2$, and noise dominates from $T = 1.3$, up to $T = 1.6$.

2.1.5. Graphs for 2011

For the year 2011, whe have 256 days of data for 92 indices. The average minimum distance for 1000 simulations with randomized data is 1.25 ± 0.01, and the average maximum distance is 1.56 ± 0.01. About 17% of the connections are bellow the minimum threshold and none is above it. Figure 5 shows the two dimensional view of the asset graphs for 2011

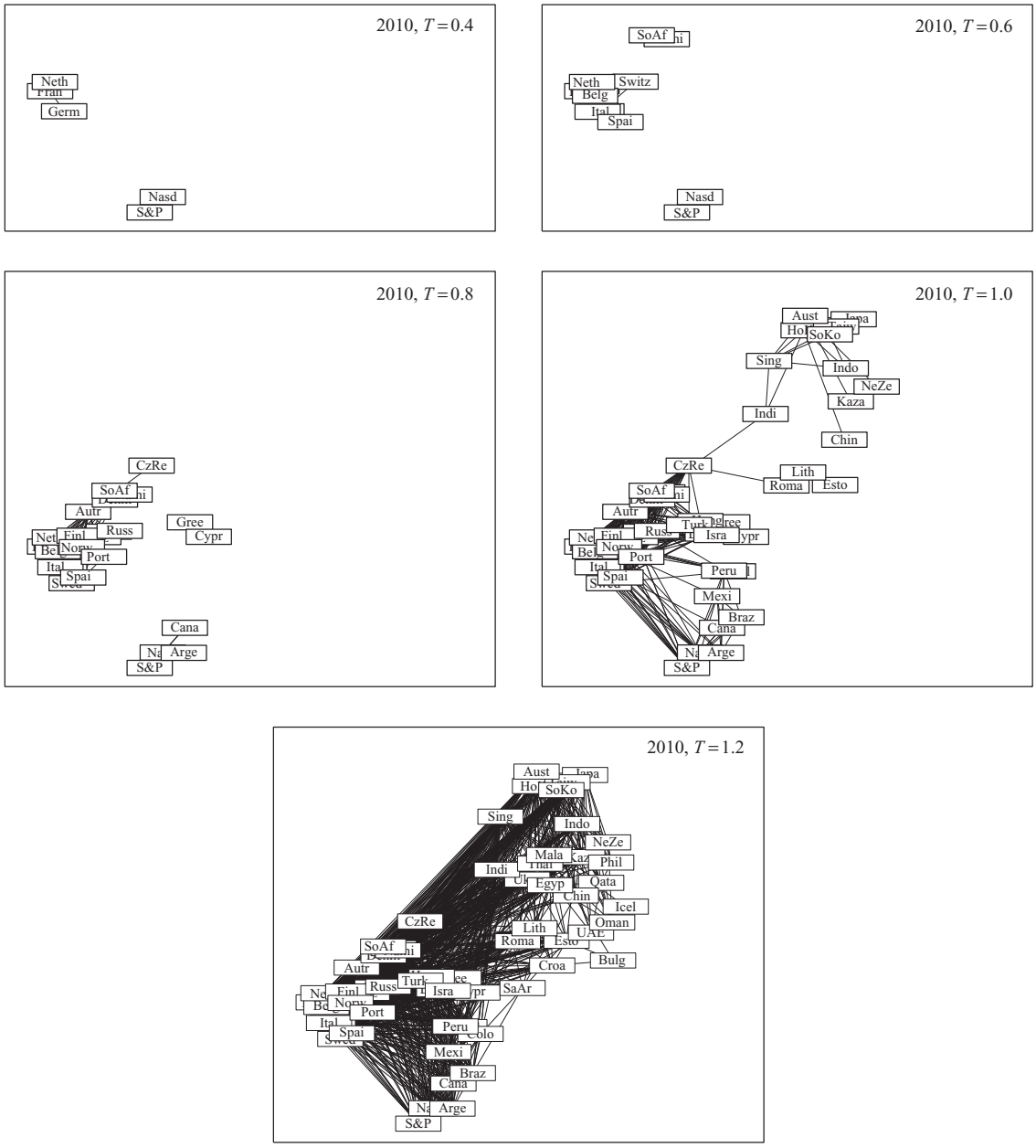

Fig. 4. Asset graphs for the year 2010, with thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

for thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

At $T = 0.4$, there is a connection between the S&P 500 and the Nasdaq, connections between France, Germany, Belgium, and the Netherlands, and also a connection between South Africa and Namibia. For $T = 0.5$, the UK, Italy, and Spain join the European cluster. For $T = 0.6$, the European cluster receives the addition of Ireland, Sweden, Finland, and Norway. At $T = 0.7$, Switzerland, Austria, and Portugal join the European cluster, and a link forms between Hong Kong and Singapore. At $T = 0.8$, the American and the European clusters already start to merge, Canada, Mexico, and Argentina join the American cluster, and Luxembourg, Denmark, the Czech Republic, Poland, Russia, and the pair South Africa - Namibia join the European cluster. Greece and Cyrpus form a connection, and the Pacific Asian cluster receives the addition

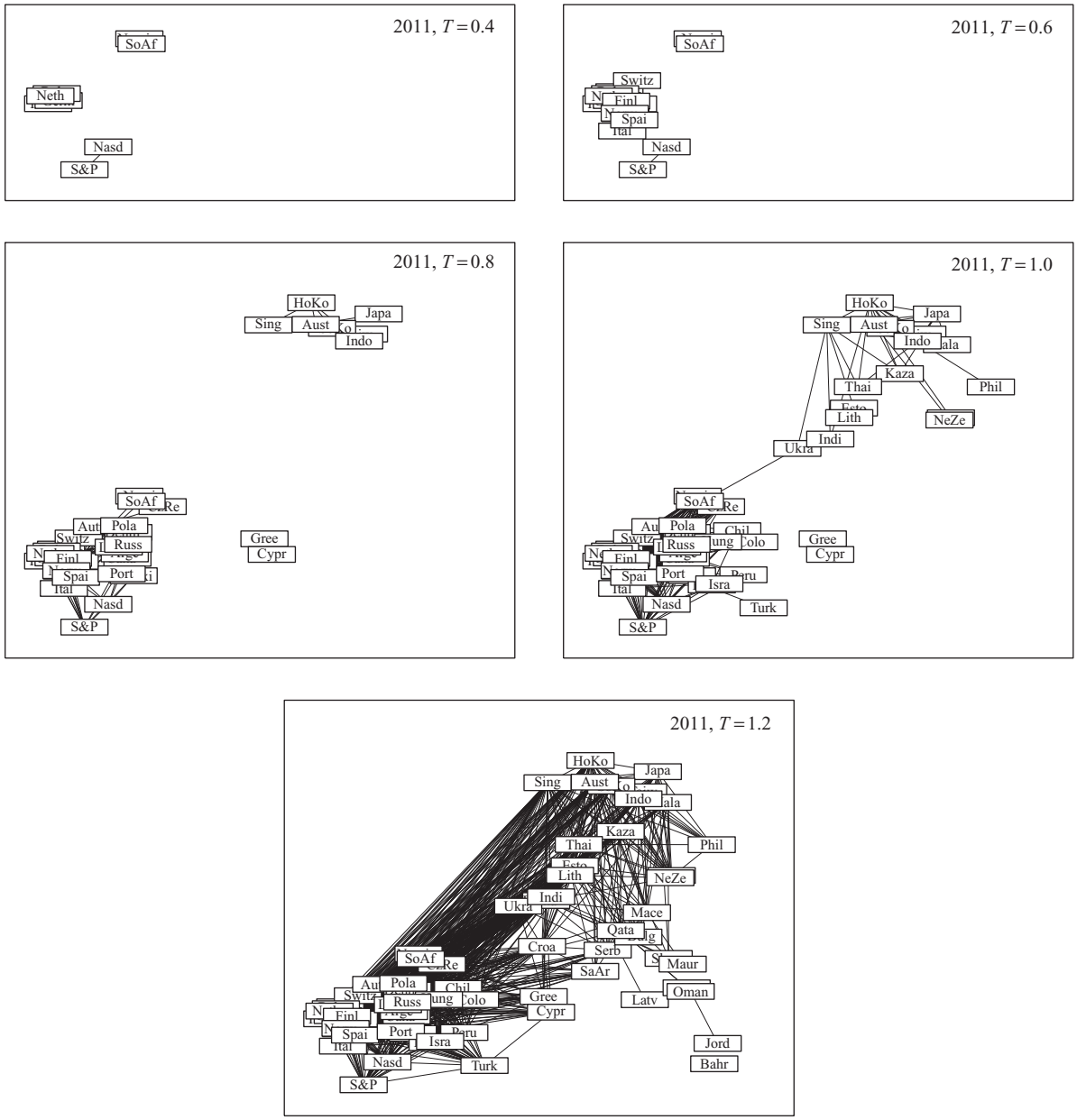

Fig. 5. Asset graphs for the year 2011, with thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

of the indices from Japan, Taiwan, South Korea, Indonesia, and Australia. For $T = 0.9$, Brazil joins the American cluster, Hungary joins the European cluster, and Thailand and Malaysia join the Pacific Asian cluster. At $T = 1.0$, Chile, Colombia, and Peru join the mergerd American and European clusters, and Ukraine, Turkey, and Israel make connections with nodes from the European cluster. A link is formed between Estonia and Lithuania, India connects with the Pacific Asian cluster, which is also joined by Kazakhstan, the Philippines and New Zealand. For $T = 1.1$, Greece, Croatia, Romania, Estonia, Lithuania, and Cyprus join the American-European cluster, which is now connected to the Pacific Asian cluster. India and China now join the Pacific Asian cluster. Many more connections are formed at $T = 1.2$, but most of them involve the desnsification of the already existing clusters. The additions now are Servia, Slovenia,

Macedonia, Bulgaria, Latvia, Saudi Arabia, Jordan, and Qatar. More connections are formed at $T = 1.3$, and noise dominates from $T = 1.4$ to $T = 1.6$, when all connections have been made.

2.2. Whole data set and Pearson correlation

I shall now analyze two types of criticism the previous results might face. The first one is the use of possibly too few data in order to generate the networks, which might introduce much statistical noise to the results obtained. The second is the use of the Spearman rank correlation instead of the Pearson correlation.

In order to investigate the effects of the amount of data used and the choice of correlation, we calculated correlation matrices based on the whole data (1271 days) from 2007 to 2011, based on the Spearman rank correlation and on the Pearson correlation. The reason for choosing data from 2007 to 2011, only, is because for these years we have a stable number of indices, 92, for a sufficiently large number of years. Since the two correlation measures do not have necessarily the same values, we performed 1000 simulations with randomized data for each type of correlation and computed the average minimum value of the correlations for the randomized data. The results were 1.340 ± 0.006 for the Spearman correlation and 1.30 ± 0.05 for the Pearson correlation.

We then computed the distance matrices for each case and adjacency matrices for threshold values equal to 100%, 75%, and 50% of the minimum noise level computed from the simulations with randomized data (we did not perform the test for 25% of the minimum values because there are no connections formed in either case for this threshold value). For 100% of the minimum simulated values, we have 3926 common connections, 88% of the total of the connections obtained from the Spearman correlation and 98% of the total of the connections obtained from the Pearson correlation. For 75% of the threshold values, we have 708 common connections, 96% of the total of the connections obtained from the Spearman correlation and 86% of the total of the connections obtained from the Pearson correlation. For 50% of the threshold values, we have 82 common connections, 100% of the total of the connections obtained from the Spearman correlation and 74% of the total of the connections obtained from the Pearson correlation.

So, we confirmed that both correlation measures yield very similar results, what also indicates that the correlations between the indices have a linear character.

Figure 6 shows the asset graphs for distances obtained from the Spearman and the Pearson correlations for a threshold $T = 1.0$, equivalent to roughly 75% of the average minimum values obtained from the simulations with randomized data.

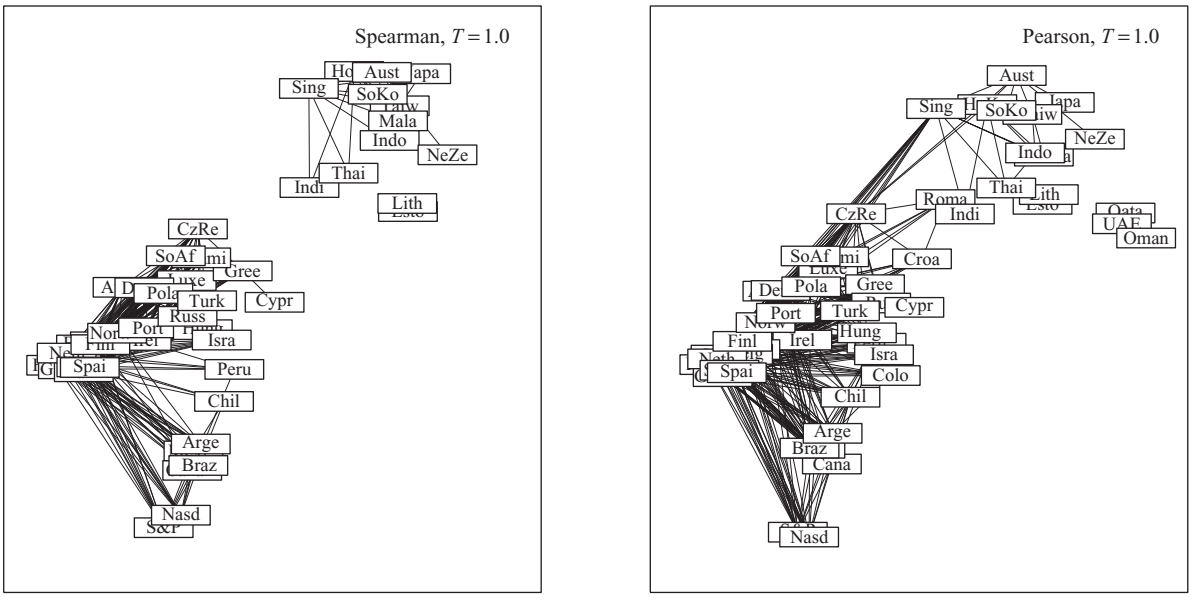

Fig. 6. Asset graphs for the whole period of data (2007 to 2010). The figure on the left is based on the Spearman rank correlation, and the figure on the right is based on the Pearson correlation, both with threshold $T = 1.0$.

Also, if one compares Figure 6 (left) with the asset graphs obtained for the separate years from 2007 to 2011, one can see the connections are quite stable in time, which is a discussion that shall be pursued further in Section 3.

2.3. Lagging indices

A third issue that could be raised is the use of same-day data for calculating correlations, when in truth markets do not operate at the same time. As an example, there is no intersection between the opening hours of the Tokyo Stock Exchange and the New York Stock Exchange. The issue of to lag or to not lag data is an important one, which shall be resolved in a separate article (Sandoval, 2012). What can be done for now is to lag some indices so that we will compare them to the non-lagged indices of the previous day. In order to decide which indices to lag, we shall take a look at the eigenvector of the second highest eigenvalue of the correlation matrix between the 92 indices analyzed for the years 2007–2011.

The eigenvector of the first highest eigenvalue ($\lambda_1 = 24.28$ in this case) is commonly associated with a market mode which explains the common movement of the indices (Laloux et al., 1999, 2000, Plerou et al., 1999; 2000, 2002; Rosenow et al., 2002; Ormerod, 2005; Kulkarni & Deo, 2007; Kwapień et al., 2006a,b; Potters et al., 2005; Wilcox & Gebbie, 2007; Pan & Sinha, 2007; Dimov et al., 2009; Bouchaud & Potters, 2011; Rak et al., 2008; Conlon et al., 2009; Namaki et al., 2010; Oh et al., 2011). Such eigenvalue is set far apart from all others, as shown in Figure 7, being much higher than it would be expected from those of a correlation matrix based on noise, and its corresponding eigenvector shows equal contributions from nearly every index being studied. As an example, the correlation between the log returns of the MSCI (Morgan and Stanley Capital International) World Index, based on a portfolio of indices from a variety of developed economies worldwide, and of the log returns of an index built using the eigenvector of the largest eigenvalue as a vector of weights for the indices used, is 0.85.

In Figure 7, the noise region is drawn as a shaded area. This area was calculated by randomizing each time series of each index so as to eliminate any correlation between them, but keeping their frequency distributions intact. The result was obtained from 1000 simulations using randomized data, but is very similar to the theoretical result obtained using Random Matrix Theory (Mehta, 2004), which considers the eigenvalue spectra due to matrices whose elements are essentially random.

The second highest eigenvalue ($\lambda_2 = 5.97$ in this case) is set not as far apart from the others in the region considered as noise as the highest one. Its corresponding eigenvector, e_2, (Figure 8) reveals a part of the internal structure of the correlations that is peculiar to indices of stock exchanges worldwide,

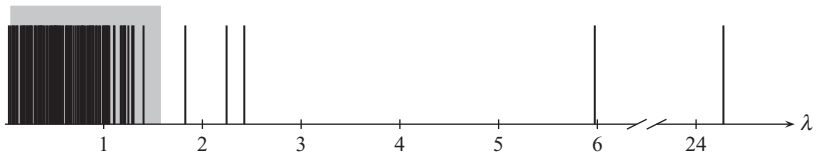

Fig. 7. Eigenvalues of the correlation between stocj market indices for data from 2007 to 2010, in order of magnitude. The shaded area corresponds to the eigenvalues obtained with randomized data.

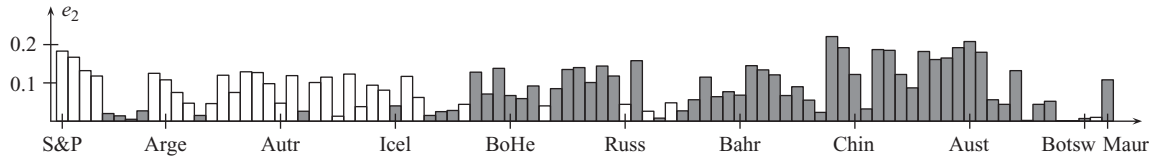

Fig. 8. Contributions of the stock market indices to eigenvector e_2, corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, Pana, CoRi, Berm, Jama, Braz, **Arge**, Chil, Colo, Vene, Peru, UK, Irel, Fran, Germ, Swit, **Autr**, Ital, Malt, Belg, Neth, Luxe, Swed, Denm, Finl, Norw, **Icel**, Spai, Port, Gree, CzRe, Slok, Hung, Serb, Croa, Slov, **BoHe**, Mont, Mace, Pola, Roma, Bulg, Esto, Latv, Lith, Ukra, **Russ**, Kaza, Turk, Cypr, Isra, Pale, Leba, Jord, SaAr, Kuwa, **Bahr**, Qata, UAE, Oman, Paki, Indi, SrLa, Bang, Japa, HoKo, **Chin**, Mong, Taiw, SoKo, Thai, Viet, Mala, Sing, Indo, Phil, **Aust**, NeZe, Moro, Tuni, Egyp, Ghan, Nige, Keny, Tanz, Nami, **Bots**, SoAf, and **Maur**.

which operate at different intervals of time. In the figure, white bars correspond to positive signs in the eigenvector, and gray bars correspond to negative signs in the eigenvector.

All indices of countries that are from Greece to the East, in terms of latitude, with the exceptions of Russia, Turkey, Israel, and South Africa, appear with negative values in eigenvector e_2. All these exceptions are indices from stock exchanges that have a non-negative intersection with the opening hours of the New York Stock Exchange. Also, from Greece to the West, the indices from Panama, Costa Rica, Bermuda, Jamaica, Venezuela, Malta, and Iceland, all of them indices with very low correlation with all the others, also appear with small negative values in the same eigenvector. Similar evidences may be found if one analyzes the data for 1986–1987, 1997–1998, and 2000–2001.

If one laggs all the indices that have negative values in eigenvector e_2, what means that one compares their returns at day $t+1$ with day t of all the other indices, one may then create a new matrix of returns and thus a new correlation matrix. The eigenvalues of such a matrix are plotted in Figure 9, together with the region associated with noise, the result of yet more 1000 simulations with randomized data.

The first largest eigenvalue is still associated with a market mode, but the second largest eigenvalue, more detached now from the others than in the non-lagged case, means something else. Figure 10 shows the values of the indices in the new eigenvalue e_2, associated with the second largest eigenvalue. Once more, white bars correspond to positive signs in the eigenvector, and gray bars correspond to negative signs in the eigenvector.

Eigenvector e_2 detects the structure of the European block, which consists on indices from countries from Central Europe, and also of indices from Hungary, Poland, Russia, Turkey, Israel, and South Africa, which basically are part of the European cluster. Also appearing with negative but diminute values in eigenvector e_2 are the indices from Panama, Chile, Colombia, Peru, and Mongolia, most of them with marginal connections in the correlation network.

Now, let us analyze the network formed using lagged data. Figure 11 shows the resulting asset graph for lagged data at threshold $T = 1.0$ using Spearman correlation in comparison to the network based on the original (non-lagged) data. What one can readily see is that the new coordinates for the lagged indices are more stretched, with the American cluster placed in between the European and the Pacific Asian clusters, and with some of the outlier indices from Europe placed farther than the Pacific Asian ones.

In order to evaluate the possible differences between the networks obtained from the original data and with

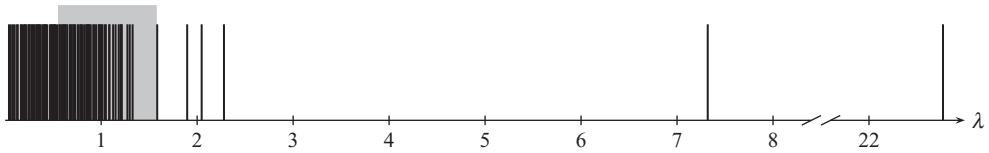

Fig. 9. Eigenvalues of the correlation for data from 2007 to 2010, with some indices lagged by one day, in order of magnitude. The shaded area corresponds to the eigenvalues predicted for randomized data.

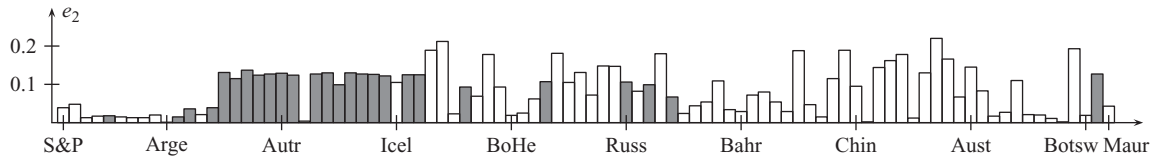

Fig. 10. Contributions of the stock market indices to eigenvector e_2, corresponding to the second largest eigenvalue of the lagged correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 2001. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, Pana, CoRi, Berm, Jama, Braz, **Arge**, Chil, Colo, Vene, Peru, UK, Irel, Fran, Germ, Swit, **Autr**, Ital, Malt, Belg, Neth, Luxe, Swed, Denm, Finl, Norw, **Icel**, Spai, Port, Gree, CzRe, Slok, Hung, Serb, Croa, Slov, **BoHe**, Mont, Mace, Pola, Roma, Bulg, Esto, Latv, Lith, Ukra, **Russ**, Kaza, Turk, Cypr, Isra, Pale, Leba, Jord, SaAr, Kuwa, **Bahr**, Qata, UAE, Oman, Paki, Indi, SrLa, Bang, Japa, HoKo, **Chin**, Mong, Taiw, SoKo, Thai, Viet, Mala, Sing, Indo, Phil, **Aust**, NeZe, Moro, Tuni, Egyp, Ghan, Nige, Keny, Tanz, Nami, **Bots**, SoAf, and **Maur**.

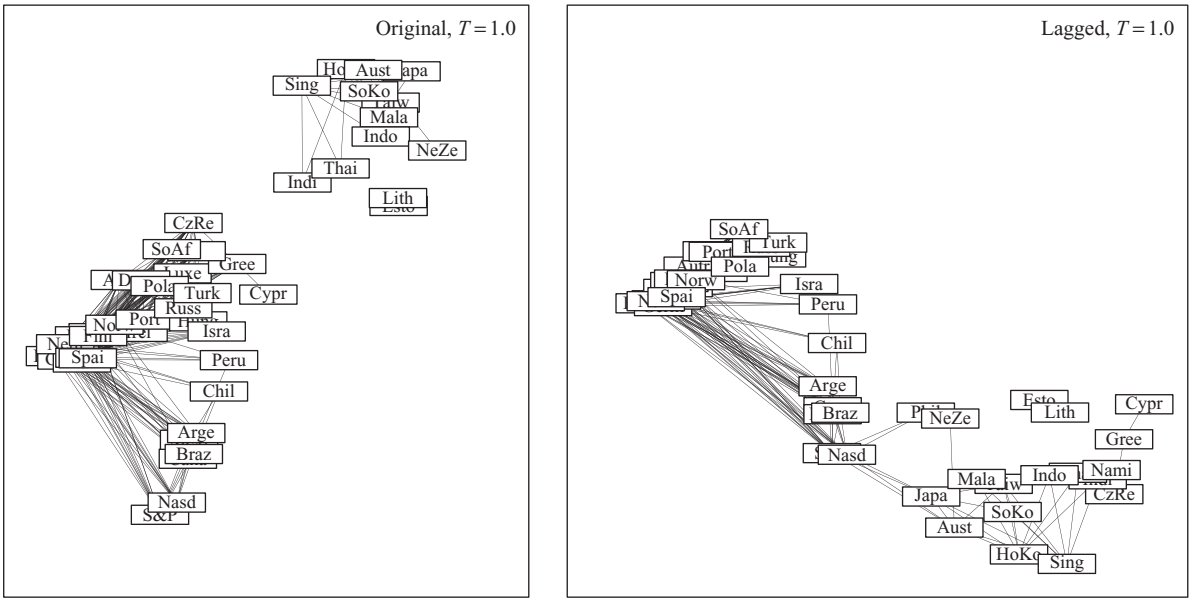

Fig. 11. Asset graphs for the whole period of data (2007 to 2011). The figure on the left is based on the original data, and the figure on the right, on lagged data, both with threshold $T = 1.0$.

lagged indices, we once again computed the distance matrices for each case and adjacency matrices for threshold values equal to 100%, 75%, and 50% of the minimum noise level computed from the simulations with randomized data (which was, for both sets of data, $T = 1.3$). For 100% of the minimum simulated values, we have 3114 common connections, 86% of the total of the connections obtained from the original data and 87% of the total of the connections obtained from the lagged data. For 75% of the threshold values, we have 618 common connections, 86% of the total of the connections obtained from the original data and 98% of the total of the connections obtained from the lagged data. For 50% of the threshold values, we have 54 common connections, 75% of the total of the connections obtained from the original data and 100% of the total of the connections obtained from the lagged data. This confirms the robustness of the asset graphs that have been built here so far also with respect to lagging or not some indices.

2.4. Community detection

Up to now, we have declared some combinations of nodes are clusters based on hierarchies derived from threshold values of a distance measure allied with visual inspection. It would be interesting to check for clusters using a community detection method. For a review on community detection methods, see Fortunato, 2010. The method that is used most extensively is the so called *modularity maximization* (for details, please reffer to Fortunato, 2010, and Newman, 2006). Here, we shall use an algorithm developed in Martelot and Hankin, 2012, for fast community detecton using the modularity maximization method. The data are the log-returns of the 92 indices from 2007 to 2011. The clusters detected depend on the threshold value, and we made calculations for thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$. The results are displayed in Figure 12, where communities are represented with different symbols, and links are withdrawn for $T = 1.2$ (for clarity of vision).

For $T = 0.4$, there are two communities: the first one made of the two nodes, S&P 500 and Nasdaq, and the other made of the nodes from the UK, France, Germany, and the Netherlands. For $T = 0.6$, four communities are detected: 1) S&P and Nasdaq; 2) France, Italy, Belgium, and Spain; 3) the UK, Germany, Switzerland, the Netherlands, Sweden, and Finland; 4) South Africa and Namibia. For $T = 0.8$, there are six communities: 1) S&P 500, Nasdaq, Canada, and Mexico; 2) Austria and the Czech Republic; 3) the UK, Ireland, France, Germany, Switzerland, Italy, Belgium, the Netherlands, Luxembourg, Sweden, Denmark, Finland, Norway, Spain, and Portugal; 4) Greece and

Fig. 12. Communities detected by modularity maximization in asset graphs for the whole period of data (2007 to 2011) for thresholds $T = 0.4$, $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

Cyprus; 5) Hong Kong and Singapore; 6) Namibia and South Africa. For $T = 1.0$, the number of communities is four: 1) S&P 500, Nasdaq, Canada, Mexico, Brazil, Argentina, Chile, Peru, the UK, France, Germany, Switzerland, Italy, Belgium, the Netherlands, Sweden, Finland, Spain, and Israel; 2) Estonia and Lithuania; 3) Ireland, Austria, Luxembourg, Denmark, Norway, Portugal, Greece, the Czech Republic, Hungary, Poland, Russia, Turkey, Cyprus, Namibia, and South Africa; 4) India, Japan, Hong Kong, Taiwan, South Korea, Thailand, Malaysia, Singapore, Indonesia, Australia, and New Zealand. Finally, for $T = 1.2$, there are just two large communities and an isolated index: 1) S&P 500, Nasdaq, Canada, Mexico, Brazil, Argentina, Chile, Colombia, Peru, the UK, Ireland, France, Germany, Switzerland, Austria, Italy, Belgium, the Netherlands, Luxembourg, Sweden, Denmark, Finland, Norway, Spain, Portugal, Greece, Hungary, Croatia, Poland, Russia, Turkey, Cyprus, Israel, India, Namibia, and South Africa. 2) the Czech Republic, Slovenia, Romania, Bulgaria, Estonia, Lithuania, Ukraine, Kazakhstan, Jordan, Saudi Arabia, Qatar, United Arab Emirates, Oman, Japan, Hong Kong, China, Taiwan, South Korea, Thailand, Malaysia, Singapore, Indonesia, the Philippines, Australia, New Zealand, and Egypt. 3) Iceland, as an isolated index.

One can notice that the communities that are formed are basically the same as those determined before. Most are based on geographical regions and on cultural ties. For lower threshold values, the only communities are the American and the European clusters, and the pair South Africa - Namibia. For intermediate values of the threshold, a Pacific Asian cluster appears, and the European cluster is divided in two: one of highly connected indices, all of them from Central and Western Europe, and other of marginally connected indices, mainly from Eastern Europe and Eurasia. For the highest threshold value represented,

the community detection algorithm does not separate the American from the European cluster, and what we have is, basically, a separation between Western and Eastern indices, what may also be a consequence of the difference of operating hours.

2.5. Overview

By considering all the time periods being studied, the first fact that stands out is the presence of two clusters throughout all periods: the American and the European clusters. The connection between the S&P 500 and the Nasdaq is present at all times and at the smallest thresholds. The European cluster also appears early on, but its core varies with time, beginning with the pair Germany-Netherlands and then evolving to a new core, consisting on the UK, France, Germany, Switzerland, Italy, the Netherlands, Sweden, and Spain. It is interesting to notice that the UK, which was more connected to the American cluster in the 80's, moves to the European cluster later on (see Appendix B).

At higher thresholds, the American cluster is joined by Canada and some Latin American indices, and the European cluster grows with the addition of some more indices of Scandinavian countries and of other Western European indices. As the thresholds grow, Europe is joined by Eastern European countries and only then we witness the formation of a Pacific Asian cluster, which starts to solidify after 1997, which coincides with the Asian Financial Crisis. South Africa and, in a lesser proportion, Israel, connect with Europe rather than with their neighbouring countries. Australia and New Zealand are more connected with Pacific Asia than with Europe. Indices from the Caribbean, of most islands, of the majority of Africa, and of the Arab countries, connect only at much higher values of the threshold, where noise reigns.

At later stages (2007–2011), one can also identify two other persistent connections: one between South Africa and Namibia, and other between Greece and Cyprus. These are not surprising, since Namibia, whose economy is in early stages of development, is highly dependent both economically and politically on South Africa. Now Cyprus is divided between Greek and Turkish populations, and the Stock Exchange to which the index from Cyprus is assigned is in Nicosia, in the Greek part of the island. We also have a similar connection, at higher thresholds, between Estonia and Lithuania, both former members of the extinct Soviet Union. At higher thresholds, and at the latest years of data, one also witnesses the formation of an Arab cluster.

What is also clear from the graphics is that the financial indices tend to group according to the geographic proximity of their countries, and also according to cultural similarities, what corroborates results previoulsy found by using nonlinear regression (gravity) models (Flavin et al., 2002; Gormus et al., 2011). Examples are the connections of South Africa and Israel with Europe, or between Namibia and South Africa, or Greece and Cyprus. Other striking feature is that the American and European clusters are fairly independent, and only connect at higher thresholds, when both clusters have already grown larger and denser. These remarks are corroborated by using a community detection method. It is good to remind ourselves that networks built from correlation matrices are not directed networks, so that we can deduct no effects of causality between the indices using just the correlation matrix.

So, what we see are three major clusters, based on geographical and sometimes cultural ties, which maintain some indepedence, with the addition of more indices as the threshold values grow. From the data concerning the periods of crisis, we can also see that networks shrink in times of high volatility, what means that there is higher correlation between the indices, as already stated in Sandoval and Franca, 2012. All connections are stable both in what concerns the choice of correlation measure and the lagging of some of the indices so as to acommodate differences in operating hours of the stock exchanges. The communities detected by using the hierarchical method are similar to the ones obtained using a modularity maximization community detection method.

3. Survivability of connections and clusters

This section considers the topic of survivability of clusters through time, which is essential, for instance, if one wishes to build risk-minimizing portfolios that are useful for future investments when based on past data. Single and multi step survival ratios (Onnela et al., 2003a) and a survivability measure developed in Sandoval, 2012, are used in order to make this analysis, and the possibility of building asset graphs based on survivability is contemplated.

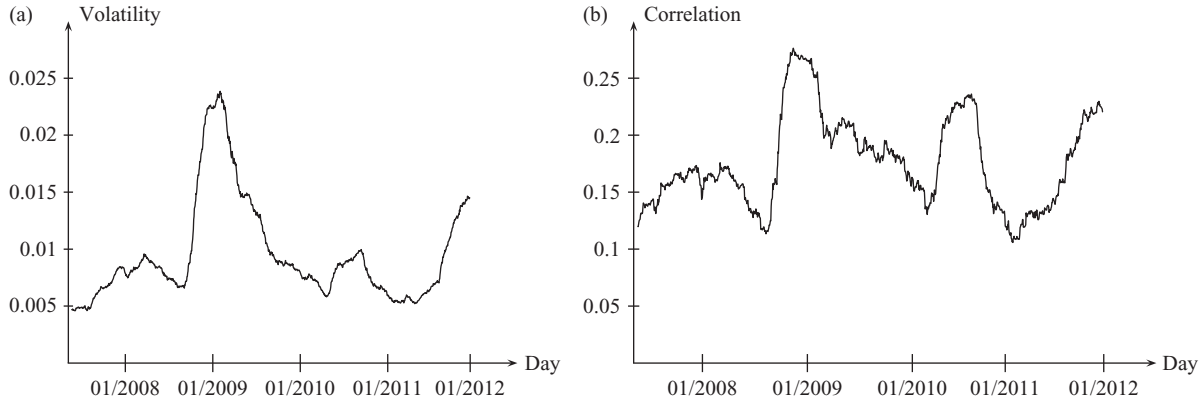

Fig. 13. a) Average volatility of the MSCI World Index in time. b) Average correlation of the 92 indices in time.

We shall then turn our attention to the survivability of connections between indices, and specifically of clusters of indices, through time. By survivability we mean the endurance of connections in time given a certain threshold. The survival of connections is heavily dependent on the threshold value chosen for the network. For $T = 1.6$, as an example, all connections survive 100% of the time, since all indices are fully connected at this threshold. Now, for lower threshold values, survivability becomes scarce.

The timeline that is being studied is varied in terms of volatility and correlation of indices, and it only covers the period from 2007 to 2011, since it is the only one for which we have a good amount of continuous data for a large enough number of indices. Figure 13a shows a graph of the average volatility of the MSCI (Morgan and Stanley Capital International) World Index, based on a portfolio of indices from a variety of developed economies worldwide. The average volatility of this index is calculated by taking the standard deviations of the log-returns of the index for running windows of 100 days each with the windows shifted by one day at a time. Plotted in Figure 13b is the average correlation of the 92 indices we are using in this research from correlation matrices calculated in overlapping windows of 100 days of the complete time series from 2007 to 2011. Since we are considering windows of 100 days, we assign the result of each window to the last day of that window. So, the curves in both graphs start 100 working days after the beginning of 2007 (05/22/2007).

What may be clearly noticed is that high volatility is associated with high correlation of financial markets (as stated in Sandoval and Franca, 2012, and references therein), what means that markets tend to behave similarly in times of crisis. In the case of the data based on 2007–2011, the correlation between the average volatility of the MSCI World index and the average correlation of the correlation matrix is 0.85.

3.1. Single step survival ratio

Measuring survivability of connections in a network is not a difficult task, but there are many ways to do that and they are not all equivalent. The simplest measure is to consider the *single step survival ratio* (Onnela et al., 2003a), here adapted to asset graphs, defined as the fraction of edges (connections) common to two consecutive asset graphs, divided by the number of edges of the asset graph. This can be easily calculated by defining the adjacency matrix A (Newman, 2010), which is a matrix whose elements a_{ij} are one if there is a connection between indices i and j and zero otherwise. In the present case, an asset graph may be represented as the distance matrix between all indices, but limited by a threshold value T. So, the adjacency matrix associated with this distance matrix with a cut-off has ones for distances $d_{ij} \leq T$ and zeros for $d_{ij} > T$. Now, it is fairly easy to define the single step survival ratio: it is the sum of all elements of the product of an adjacency matrix at time t with an adjacency matrix at time $t+1$, divided by the number of edges at time t:

$$S(t) = \frac{1}{M(t)} \sum_{i,j=1}^{N} A_{ij}(t) A_{ij}(t+1), \tag{5}$$

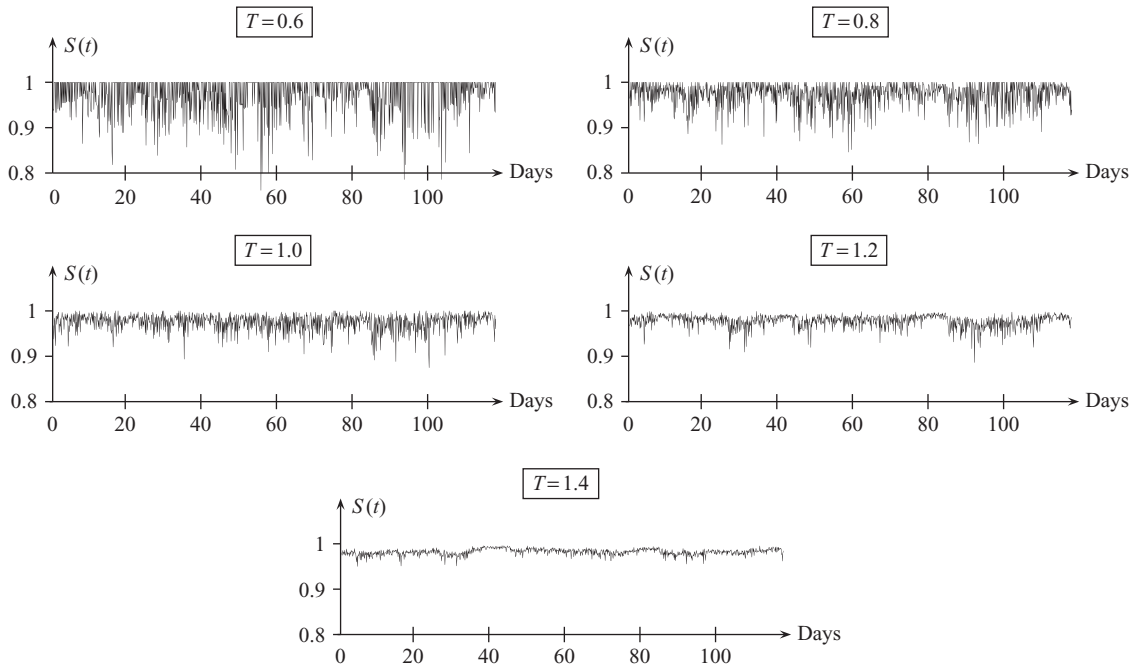

Fig. 14. Single step survival ratio for various asset graphs based on a diversity of thresholds for running windows of 100 days with steps of one day at a time.

where $A_{ij}(t)$ is the element ij of matrix A(t), $A_{ij}(t+1)$ is the element ij of matrix $A(t+1)$, and N is the total number of nodes in the correlation matrix (this is the same as the scalar product of both matrices divided by $M(t)$). Since we are using a distance threshold and not the number of edges to establish an asset graph, the number of edges is not constant in time. So, we are using $M(t)$ as the number of connections in the asset graph related with $A(t)$.

The results depend on the threshold T, and Figure 14 shows the results for thresholds $T = 0.6$, $T = 0.8$, $T = 1.0$, $T = 1.2$, and $T = 1.4$, just below and above the noise threshold. For high threshold values up to $T = 1.7$, $S(t)$ assumes values that increasingly lead to $S(t) = 1$ for all values of t, since $T = 1.7$ is the distance at which all nodes are connected and so all connections survive all the time.

As expected, the single step survival ratio varies much from $T = 0.6$ to $T = 0.8$, since one has, for these thresholds, very few connections, although strong ones. As the value of the threshold grows, one has many more connections, and then it is expected that the single step survival ratio goes slowly towards the unit value. Strong variations of the single step survival ratio do not necessarily coincide with any particular important event. When analyzing this, one should bear in mind that the single step survival ratio is being calculated over windows of 100 days, and so it is not very sensitive when pinpointing events that happen in one or few days.

If one now wishes to compare the single step survival ratio of real data with the one of randomized data, the first striking difference between them is that, for randomized data, there are virtually no connections bellow $T = 1.2$. So, one can only compare the results for connections above this distance value. Figure 15 shows the comparison of the single step survival ratio for real (graph on the left) and randomized (graph on the right) data for a typical simulation where all time series were randomly shuffled, both for $T = 1.3$. One thing to notice is that the single step survival ratio for randomized data oscillates around 0.89. The other is that it does not come close to the value 1, so that the single step survival ratio of randomized data is never large, and it is also much more volatile (high standard deviation). The fact that the survival ratio is so high for randomized data is that it is an effect of my use of overlapping windows of 100 days with shifts of one day at a time. This also applies to the real data, but does not explain by itself the high survival ratio in that case.

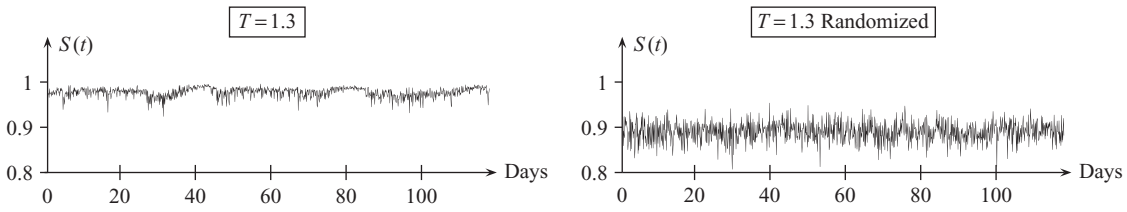

Fig. 15. Single step survival ratio for $T = 1.3$ for running windows of 100 days with steps of one day at a time. The graph on the left is based on real data, and the graph on the right is based on randomized data.

3.2. Multi step survival ratio

A natural generalization of the single step survival ratio is the multi step survival ratio (Onnela et al., 2003a), which is the percentage of edges that exist in a certain number of windows, divided by the total number of edges. Here we shall adapt this definition to the case where asset graphs are built based on threshold values for the distance, in which case the number of edges may vary in each window. The definition here of a multi step survival ratio for p windows is

$$S_p(t) = \frac{1}{M(t)} \sum_{i,j=1}^{N} A_{ij}(t) A_{ij}(t+1) \cdots A_{ij}(t+p), \tag{6}$$

where $A_{ij}(t)$ is the element ij of matrix A(t), N is the total number of nodes in the correlation matrix, and $M(t)$ is the total number of edges in time t.

It is expected that the multi step survival ratio drops as the number of windows p grow, since it is increasingly difficult for a connection (edge) to survive at all intervals being considered. It is also expected that for a large enough value of the thresholds, where all nodes are connected to all others, this survival ratio will always have the value 1. Figure 16 (left graph) shows the multi step survival ratios for thresholds ranging from $T = 0.1$ to $T = 1.5$ (fully connected network) as functions of the number of windows.

Note the high concentration from thresholds $T = 0.7$ to $T = 1.4$. It is at these values that one has the most information about connections yet with a bearable amount of noise. For values bellow $T = 0.6$, there are too few connections, although strong ones, which makes it easier for the multi step survival ratio to drop faster, and for very high thresholds, there are so many connections that it is easy for a particular edge to survive all the time.

The graph on the left of Figure 16 shows the multi step survival ratio for a typical simulation with randomized data, obtained by shufling randomly the time series of the indices. One difference is that there are no asset graphs for distance thresholds bellow $T = 1.3$, since there are no connections formed bellow this threshold for randomized data (only very rarely). Other difference is that the multi step survival ratio for randomized data decays nearly exponentially, as it was to be expected from data that should not be related.

It is also interesting to compare this multi step survival ratio with the one obtained from a minimum spanning tree (Sandoval, 2012). Since edges built using a minimum spanning tree algorithm are much less stable than for asset graphs, it is much easier for those connections to change over time, leading to a near exponential decay of the multi step survival ratio, what does not occur for connections based on asset graphs.

3.3. Survivability

The multi step survival ratio depends on the starting point of the data being considered and also on the time shift between windows. So, it is a measure that, like the single step survival ratio, depends on time. An alternative measure derived in Sandoval, 2012, considers each day as a starting point and then tests the multi step survival for a number D of days (5, for example) of an asset graph built with a threshold T. Then the percentage of tests in which a particular edge survives is used in order to define a survivability measure. As an example, the connection between nodes i and j for five consecutive days is tested with many different starting points in running windows of length Δt. If the connection survives in 30% of the tests, then its survivability is $s = 0.3$. So, survivability is a global (in terms of time) measure of how stable a connection between two nodes is through a certain period of time. Dividing the number of surviving edges by the total

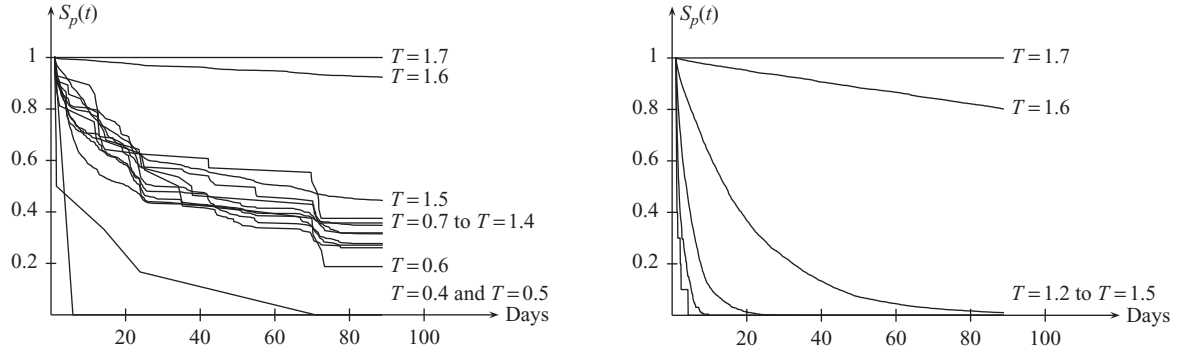

Fig. 16. Left: multi step survival ratios for thresholds ranging from $T = 0.4$ to $T = 1.7$ as functions of the number of windows. Right: the same for randomized data, from $T = 1.2$ to $T = 1.7$.

number of edges of the asset graph, one obtains the survivability ratio, S_r, that goes from 0 to 1.

Performing an analysis of survivability for the data concerning the 92 stock market indices from 2007 to 2011, one obtains the results in Figure 17. In this figure, the survivability ratio is plotted against the percentage of tests in which that number of connections survived.

Once again, there is a concentration of curves from $T = 0.7$ to $T = 1.4$, which is the region around the noise limit. For these threshold values, the percentage of connections that survive 100% of the time is around 22% which, compared with the survivability of connections in a Minimum Spanning Tree (Sandoval, 2012) is, not surprisingly, very high. For $T = 0.4$, there is just one connection (from the original 27) that survive 100% of the time: the one between France and Germany. For $T = 0.6$, there are three connections that survive 100% of the time, those between France and Germany, S&P 500 and Nasdaq, and between France and the Netherlands. For $T = 0.8$, this number goes to 21, and for $T = 1.0$, there are 36 connections that survive 100% of the time, forming an American cluster (S&P 500 and Nasdaq), and an European cluster involving the UK, France, Germany, Switzerland, Austria, Italy, Belgium, the Netherlands, Sweden, Finland, and Spain, and an African cluster comprised of Namibia and South Africa. By fixing the threshold value and varying the survivability ratio, one may obtain a sequence of networks that grow as the survivability ratio drops. This procedure may be used in order to build "survivability networks", which may bring some information on the stability of connections.

Also in Figure 17, to the right, we plot the survivability ratio for a network obtained by randomly shuffling the time series of the 92 indices being studied. The curves are results of single simulations, since the computations are quite long, but one does not lose too much information here by doing this. From $T = 0.1$ to $T = 0.7$, there are no connections in the network and so one cannot even define a survivability ratio. From $T = 0.8$ to $T = 1.1$, the survivability curves drop exponentially, and for values higher than $T = 1.2$, it quickly approaches the line $S = 1$, since all connections are active all the time for large enough values of the threshold. The difference between the results obtained with randomized data and with real data are striking: survivability is often much higher for real data, showing there must be information on the connections formed by their correlations.

Figure 18 (graph on the left) analyzes the relationship between distance and survivability for $T = 1.3$. The horizontal axis is the distance between nodes as calculated by (2), and the vertical axis is the percentage of time a connection survives. What one can readily see is that, like in the case of Mininum Spanning Trees, there is no linear relation between the two variables. Looking at the data, one could see that around 68% of the connections have survivability zero, specially those that represent distances higher than 1.2. Also, around 11% of the connections have survivability 1 (100%), none of them with a distance higher than 1.6. The remaining 21% of connections survive in clumps around survivabilities 0.2, 0.5, and 0.8, the majority with distances ranging from 1.1 to 1.4.

The graph on the right in Figure 18 shows the relationship between distance and survavibility ratio based on randomized data and a threshold $T = 1.3$ (thresholds bellow $T = 1.3$ have no connections for randomized data). The difference is clear, as distances concentrate around a certain value and connections tend to survive little. About 96% of the

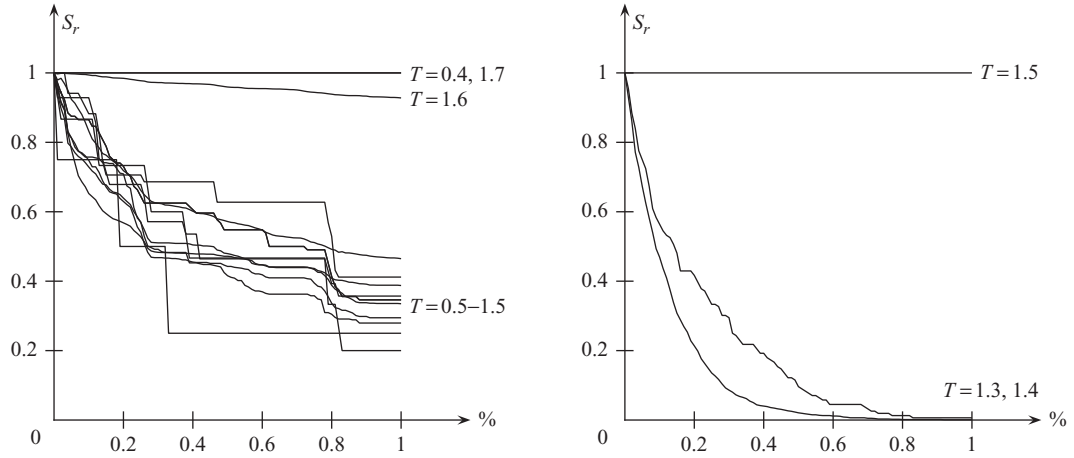

Fig. 17. Survivability ratios for thresholds ranging from $T = 0.1$ to $T = 1.7$ as functions of the percentage of surviving windows. On the left, it is plotted the survivability for the network obtained from the stock market indices, which is zero for $T < 0.4$; on the right, the survivability for a network built with randomized data, which is zero for $T < 1.3$.

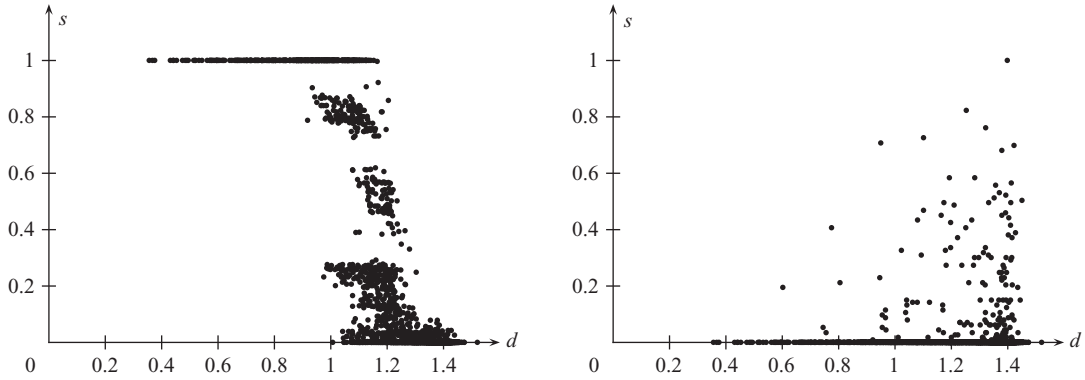

Fig. 18. Left: scatter plot of the distance d between nodes and the percentage of time a connection survives, s, for $T = 1.3$. Right: scatter plot of the distance d between nodes and the percentage of time a connection survives, s, for $T = 1.3$ and randomized data.

connections have survivability zero and only one has survivability 1.

So, concluding this section, we saw that there are remarkable differences between real data and randomized data in terms of survivability, be it measured by the single step survival ratio, the multi step survival ratio or the survivability ratio. For real data, the curves for thresholds ranging from $T = 0.7$ to $T = 1.4$ tend to agglomerate, being that the region where one can get the most information about the network with an acceptable level of noise.

We could also see that some connections, and by extension, some clusters remain constant in time. Those clusters are heavily associated with geographic position and on cultural and ethnical ties. Those connections that were most intense also tended to last longer, although there are many exceptions to this rule. Next, the effects of the choice of threshold and of time in centrality measures of network theory are studied.

4. Centrality measures

Centrality measures are of key importance in network theory (Newman, 2010). Whereas a node has many connections, or it is on the shortest path between many pairs of nodes, or if it lies within a heavily connected region, are of great importance for models of propagation of information, of diseases, and on the vulnerability of a network to attacks, as examples. So, many measures of how central a node is in a network were developed, and each one associates a different

concept to the word "centrality". Here, we shall see how one of those measures change over time, and what is the influence of the choice of threshold when building an asset graph on those centrality measures.

4.1. Node degree

The node degree of a node is the number of connections it has in the network. Most of the stocks have low node degree, and some of them, called hubs, have a large node degree and are generally nodes that are more important in the dynamics of a network. Considering the adjacency matrix A of a network, the node degree of node i can be defined mathematically as

$$Nd_i = \sum_{j=1}^{N} A_{ij}. \tag{7}$$

Node degrees vary considerably according to the threshold value of an asset graph. It is expected that, for small threshold values, where very few connections are formed, most node degrees are zero, and, for high threshold values, where all connections are formed, all nodes will have the highest node degree that is possible. As the maximum node degree varies from 1 to 92 as thresholds grow, we divided all node degrees by the maximum node degree for better comparison of the asset graphs.

Considering that the asset graphs are formed only by those nodes which are connected to any other node, then the asset graphs for low threshold values are much smaller. Taking this into account, one must eliminate from the network all nodes which have node degree zero. Figure 19 illustrates frequency distributions of the normalized node degrees for thresholds going from $T = 0.4$ to $T = 1.5$, and for randomized data for $T = 1.4$.

Extreme values for the threshold lead to distributions that are heavily concentrated. Again, the distributions close to the noise threshold seem to be the ones that give the most information. From thresholds $T = 0.4$ to $T = 1.1$, Central European indices occupy the top positions, particularly France, the UK, the Netherlands, and Germany. For higher thresholds, the top positions are occupied in an increasingly random way.

From Figure 19, one can also see that the resulting histograms are definitely not exponentially decreasing, so they cannot be represented by power laws of the type $p_k = ck^{-\alpha}$, where p_k is the frequency distribution for the value k, and c and α are constants. Power laws are the hallmark of a diversity of complex systems, as in the study of earthquakes, the world wide web, networks of scientific citations, of film actors, of social interactions, protein interactions, and many other topics (Newman, 2006). Networks whose centrality measures follow this type of distribution are often called scale-free networks, and this behavior can best be visualized if one plots a graph of the cumulative frequency distribution of a centrality in terms of the centrality values, both in logarithmic representation (Newman, 2006). Figure 20 shows the logarithm of the cumulative frequency distributions as functions of the logarithm for $T = 0.4$ to $T = 1.5$ (graph on the left), and for randomized data (graph on the right), for $T = 1.4$ (for $T = 1.5$, the curve is on the vertical axis).

So, although the cumulative frequency distribution of the node degrees for the asset graphs is not the one of a scale-free network, they are definitely not the same as the cumulative frequency distribution of randomized data. A cautionary note must be added here. As the networks that are being considered in this analysis are very small compared with scale-free networks like the Internet, or pure scale-free networks, which must be infinite in size conclusions about these asset graphs networks, particularly the ones with smaller thresholds and a very small number of nodes, are not reliable.

In Figure 21, we show the evolution of the average node degree as a function of time, $\overline{Nd}(t)$, which is the mean of the node degrees of the nodes belonging to each asset graph (the number of nodes may change in time). In the calculation of the graph, we have used windows of 100 days, shifted one day at a time. It does make data overlap, but gives a better number of points for the plot. The curves are, as expected, very similar to the graph of the average correlation in time and, as a consequence, to the graph of the volatility of the world market as measured by the MSCI World Index. Figure 22 shows the time series of the average node degree for $T = 1.2$ (in black) in terms of the volatility of the MSCI World Index (in gray), both normalized so as to have average zero and standard deviation one, for a better comparison.

4.2. Node strength

The strength of a node is the sum of the correlations of the node with all other nodes to which it is connected. If C is the matrix that stores the correlations between n nodes that are linked in the asset graph, then the node

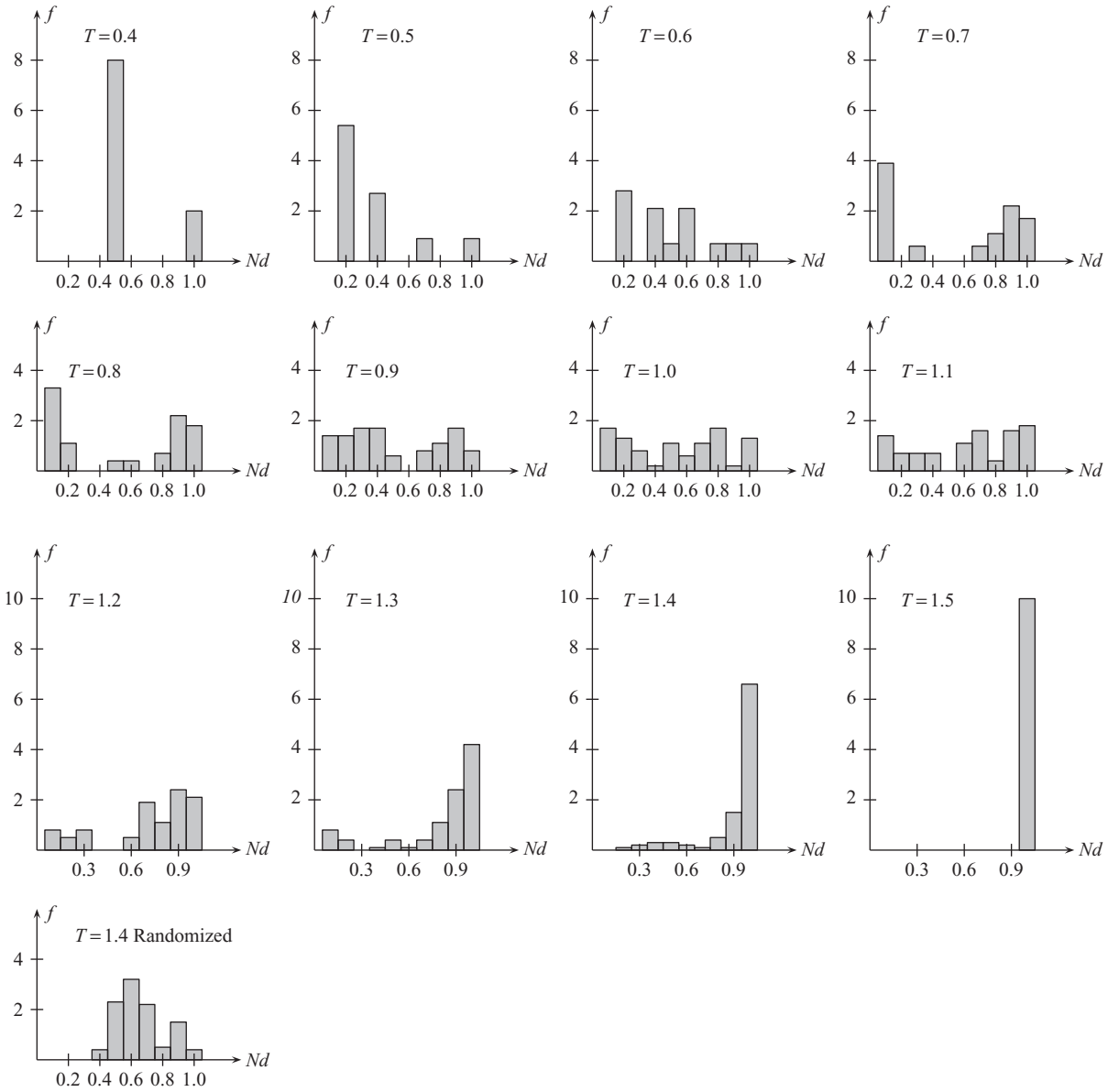

Fig. 19. Histograms of the normalized node degrees of asset graphs based on thresholds varying from $T = 0.4$ to $T = 1.5$, plus histogram for randomized data at $T = 1.4$.

strength of node i is given by

$$Ns_i = \sum_{j=1}^{n} C_{ij}, \tag{8}$$

where C_{ij} is an element of matrix C.

For the asset graphs of the world stock indices, the Central European indices appear as the most central according to node strength for all threshold values. The topmost indices are those from France the Netherlands, Germany, Belgium, and the UK.

Figure 23 displays the cumulative frequency distribution of the node strength in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data (from $T = 0.4$ to $T = 1.5$), and the figure on the right represents the curves for randomized data (for $T = 1.4$ and $T = 1.5$). Again, the results are remarkably different

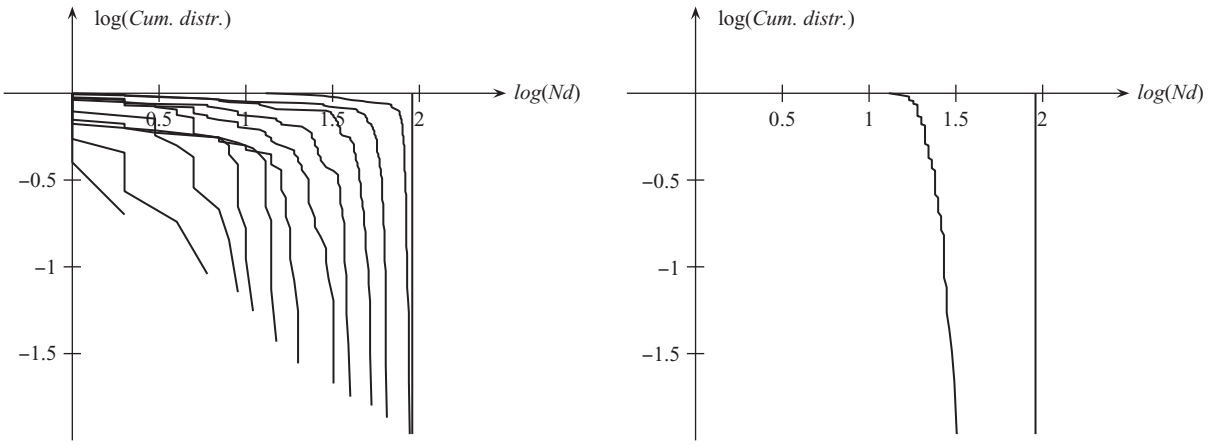

Fig. 20. Cumulative frequency distribution of the node degree in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data, and the figure on the right represents the curve for randomized data.

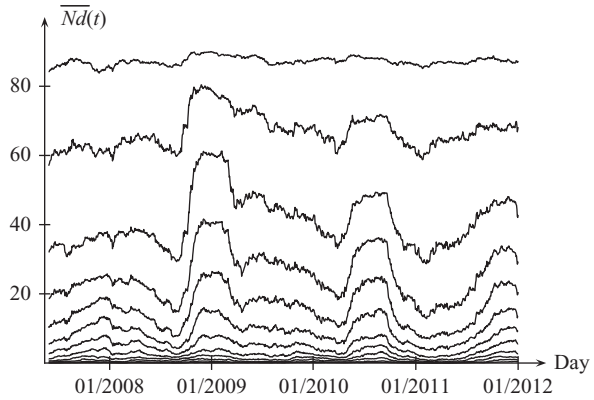

Fig. 21. Evolution of the average node degree, $\overline{Nd}(t)$, in time for different values of T.

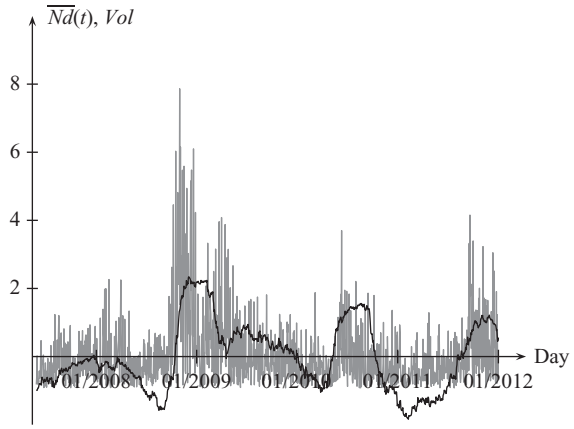

Fig. 22. Average node degree $\overline{Nd}(t)$ for $T = 1.2$ (black) and volatility of the MSCI World Index (*Vol*) (gray), both normalized to average zero and standard deviation one.

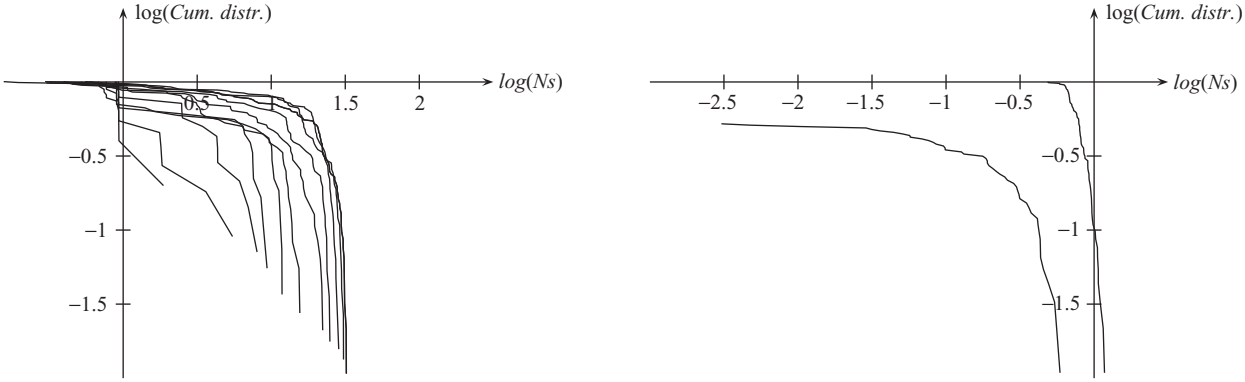

Fig. 23. Cumulative frequency distribution of the node strength in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data, and the figure on the right represents the curve for randomized data.

from randomized data, but they do not seem to correspond to scale-free networks.

4.3. Eigenvector centrality

One node may have a low dregree, but it may be connected with other nodes with very high degrees, so it is, in some way, influent. A measure that takes into account the degree of neighboring nodes when calculating the importance of a node is called *eigenvector centrality*. If one considers the eigenvectors of the adjacency matrix of the distance matrix for the asset graph, and choosing its largest value, one may then define the eigenvector with largest eigenvalue by the equation

$$AX = \lambda X, \tag{9}$$

where X is the eigenvector with the largest eigenvalue λ. The eigenvector centrality of a node i is then defined as the ith element of eigenvalue X:

$$Ec_i = x_i, \tag{10}$$

where x_i is the element of X in row i.

Here, there is a clearer set of nodes which exhibit more eigenvector centrality values than the others in a consistent way for the thresholds closest to the limit of the noise region. France, Germany, the Netherlands, the UK, Belgium, and Austria appear as the top nodes from $T = 0.4$ up to $T = 1.0$, what is not surprising, since they inhabit a densely connected region of the asset graphs.

Figure 24 displays the cumulative frequency distribution of the eigenvector centrality in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data, and the figure on the right represents the curves for randomized data. All curves are quite close to one another, and the curve for the extreme value $T = 1.5$ is the same as the one obtained for randomized data (everyone is connected with everyone).

4.4. Betweenness

The betweenness centrality measures how often a node lies on the paths between other vertices. It is an important measure of how much a node is important as an intermediate between other nodes. It may be defined as

$$Bc^k = \sum_{i,j=1}^{n} \frac{n_{ij}^k}{m_{ij}}, \tag{11}$$

where n_{ij} is the number of shortest paths (geodesic paths) between nodes i and j that pass through node k and m_{ij} is the total number of shortest paths between nodes i and j. Our network is fully connected, so we need not worry about m_{ij} being zero.

From $T = 0.4$ to $T = 0.7$, the nodes with larger values of betweenness centrality are those of Central Europe, mainly France, the Netherlands, and Germany. For $T = 0.8$, the S&P 500 also exhibits some centrality, and, for $T = 1.0$ and $T = 1.1$, Singapore, Australia, Hong Kong, and Qatar assume the highest values for betweenness centrality.

Figure 25 displays the cumulative frequency distribution of the betweenness centrality in terms of the centrality values, both in logarithmic representation. Like before, the figure on the left displays the curves

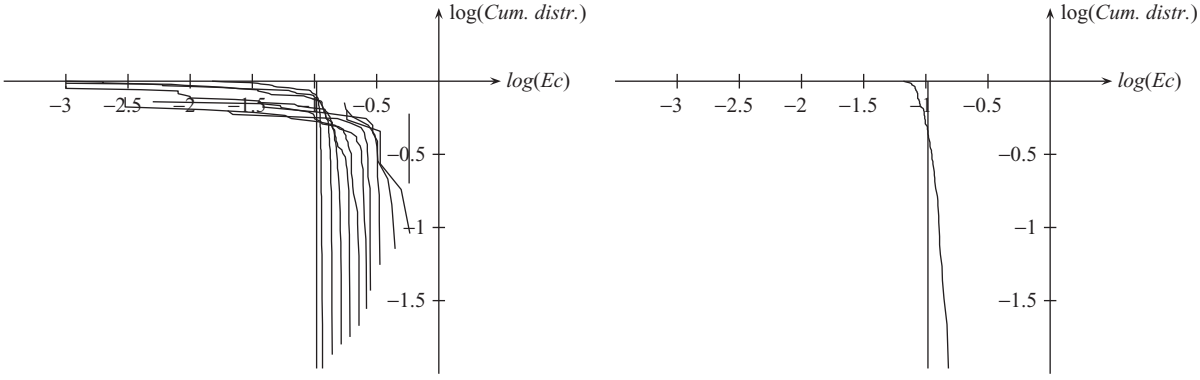

Fig. 24. Cumulative frequency distribution of the eigenvector centrality in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data, and the figure on the right represents the curves for randomized data.

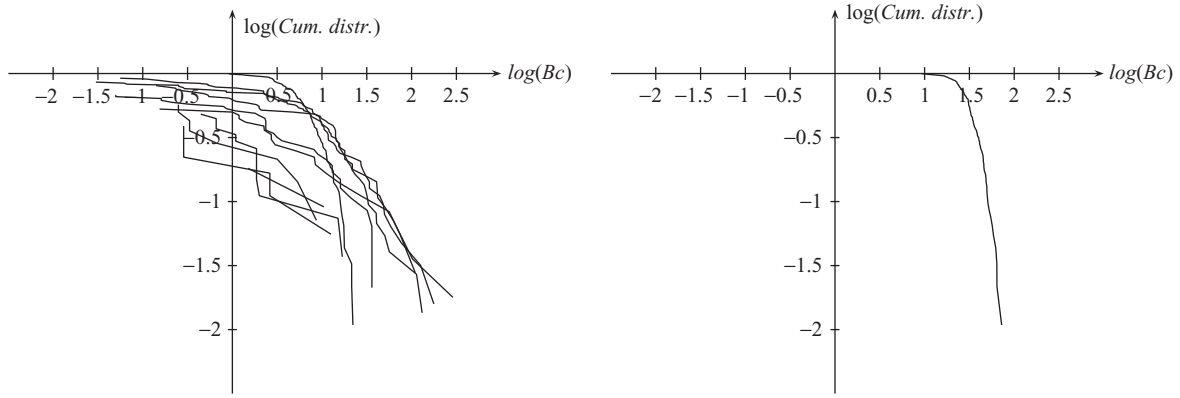

Fig. 25. Cumulative frequency distribution of the betweenness centrality in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data, and the figure on the right represents the curves for randomized data.

for real data, and the figure on the right represents the curves for randomized data. The curves for real and randomized data are all dissimilar, and there are no curves for $T = 1.5$ in both cases.

4.5. Harmonic closeness

Another measure of centrality is the *closeness centrality*, which measures the average distance between one node and all the others. It is defined as the measure of the mean geodesic distance for a given node i, which is given by

$$\ell_i = \frac{1}{n} \sum_{j=1}^{n} d_{ij}, \tag{12}$$

where n is the number of vertices and d_{ij} is a geodesic (minimum path) distance from node i to node j. This measure is small for highly connected vertices and large for distant or poorly connected ones. In order to obtain a measure that is large for highly connected nodes and small for poorly connected ones, one then defines the *harmonic closeness centrality* of node i as

$$Hc_i = \frac{1}{n-1} \sum_{i=1, i\neq j}^{n} \frac{1}{d_{ij}}. \tag{13}$$

There is again a set of nodes which consistently have the highest harmonic closeness centralities for asset graphs close to the noise threshold. These are Central European indices, like France, Germany, the Netherlands, and the UK. For higher threshold values, Singapore and Hong Kong also appear with high harmonic closeness centralities.

Figure 26 displays the cumulative frequency distribution of the harmonic closeness centrality in terms

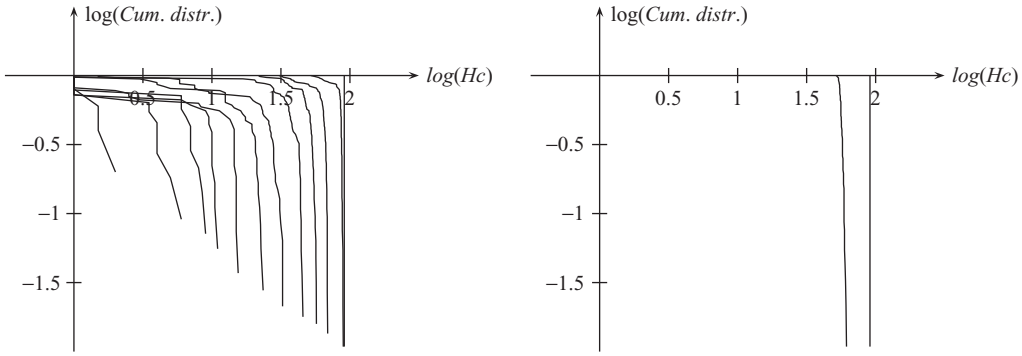

Fig. 26. Cumulative frequency distribution of the harmonic closeness centrality in terms of the centrality values, both in logarithmic representation. The figure on the left displays the curves for real data, and the figure on the right represents the curves for randomized data.

of the centrality values, both in logarithmic representation. Once more, the figure on the left displays the curves for real data, and the figure on the right represents the curves for randomized data. The curves for thresholds $T = 1.4$ and $T = 1.5$ are similar to the ones obtained based on randomized data.

4.6. Overview

Our analysis of the centrality measures showed that there is a strong dependence of centralities on the threshold chosen to define a particular asset graph, and that, although the probability frequency distributions and the cummulative frequency distributions of the four centralities we have researched do not resemble those of scale-free networks, they are remarkably different from their counterparts based on randomized data, except for high values of the threshold. The values of the threshold between $T = 1.0$ and $T = 1.1$, close to the noise threshold, are believed to be the ones that generate asset graphs with more information also in terms of centralities. Central European indices, like the UK, France, the Netherlands, Germany, Belgium, and Austria, carry high centrality values, in general, and Singapore and Hong Kong present high betweenness centrality values, mostly because they connect most Pacific Asian indices with European ones. The final figure of this article (Figure 27) shows the five centrality measures discussed in this section as functions of indices.

From the figure, one can see that Central Europe and part of Eastern Europe present a consistent large number of centralities, except for betweenness at higher threshold values. North America and part of Pacific Asia also behave, in a lesser degree, in the same way. For betweenness centrality, we have peaks in Singapore and Hong Kong for higher values of the threshold, since they connect nodes from Pacific Asia to European nodes. The eigenvector centrality and betweenness centrality are particularly useful in determining high centrality, in its two main aspects: nodes that are well connected in regions of densely connected nodes, and nodes that link clusters of nodes.

5. Conclusion

The aims of this article may be divided into three parts: the first one was to use asset graphs based on thresholds for a distance measure obtained from the correlation matrices of time series of world stock market indices in order to study their behavior due to changes in the threshold value and also their evolution in time; the second was to analyze the survivability of the connections between indices in asset graphs and, as a consequence, of clusters based on regional, economical, and cultural ties (Flavin et al., 2002; Gormus et al., 2011), and their dependence on the threshold values; the third was to analyze some centrality measures of the nodes in the previous asset graphs, and their dependence on threshold values.

What we have shown is that financial markets tend to group themselves according to geographical region, and also according to economic and cultural ties. Two main clusters were detected, an American one, and an European one, both existing already at low threshold values (small distances) and very stable in time (high survivability). Other clusters form at higher levels of the threshold, like a Pacific Asian cluster and, at higher threshold values, an Arab one. All clusters tend to integrate at higher levels of the threshold, beginning by an absorption of some Eastern European nodes, plus

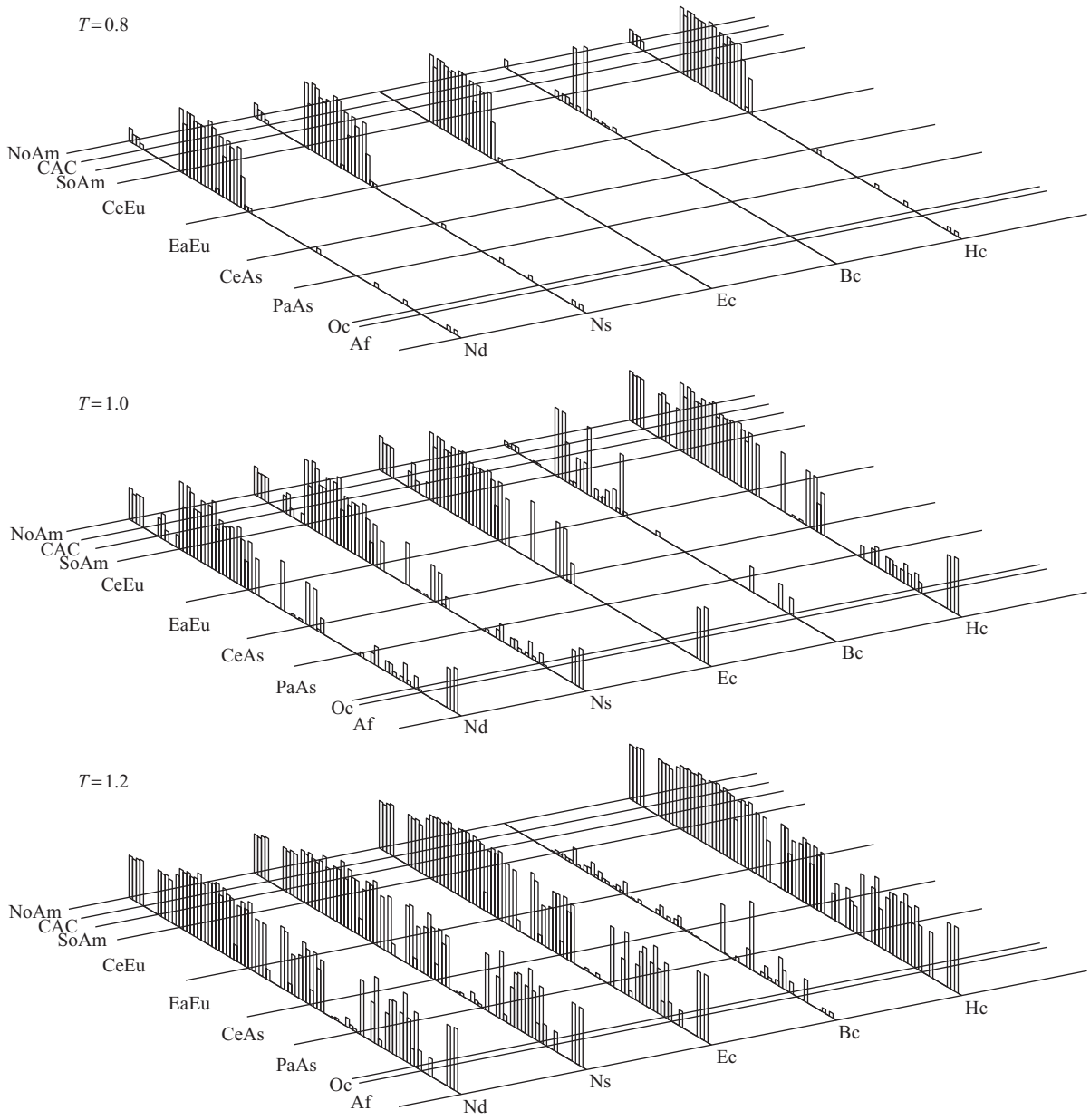

Fig. 27. Centrality measures for asset graphs defined by $T = 0.7$, $T = 0, 6$, and $T = 0.5$ in terms of node. The horizontal lines indicate divisions between regions: NoAm for North America, SoAm for South America, CeEu for Central Europe, EaEu for Eastern Europe, CeAs for Central Asia, Pa As for Pacific Asia, Oc for Oceania, and Af for Africa.

Israel and South Africa, by the European cluster, and an enlargement of the North American cluster by the addition of Latin American nodes. An unification of the American, European, and Pacific Asian clusters occurs at intermediate values of the threshold. We could also see, looking also in Apendix B, that, in times of crisis, like 1987, 1997, 1998, 2001, and 2008, those asset graphs tend to shrink in size and augment in number of nodes, both consequences of the higher correlations between indices. These results are robust with respect to both the choice of correlation coefficient (Spearman or Pearson, specifically), and the lagging of some of the indices in order to take into account the difference in operating hours. We also used simulations

with randomized data in order to establish threshold values above which random noise has a strong influence on the results.

Another result was that the second highest eigenvalue of the correlation matrix for the time series of stock exchange indices seems to be connected with the difference in operating hours, separating the indices in two groups, Western and Eastern ones. By lagging the second group of indices in one day, we then obtain a new structure for the second eigenvalue, separating the European cluster from the others.

We could also see that Central Europe exhibits high values of centrality according to five centrality measures, two of them (node degree and node strength) based on the number and strength of connections of a node, one of them (eigenvector centrality) which also takes into account how connected each nodes' neighbors are, one (harmonic closeness centrality) measuring the average (inverse) distance of a node to all the others, and betweenness centrality, which measures how often a node is in between shortest paths connecting other nodes. We also saw that, for higher threshold values, Hong Kong and Singapore surpass the others as highly central with respect to betweeness centrality. We also made a study of the evolution of the average node degree in time and showed that it is highly correlated with the average correlation of nodes in time.

All results showed there is a lot of information that can be obtained from asset graphs based on distance thresholds, and that the most significant ones seem to be those around the noise threshold. Most results have been compared with the ones obtained by simulations with randomized data, and they are remarkably different from those, except for high values of the threshold, which leads to the belief that the connections between international stock exchange indices are not formed at random, but have a complex structure underlying their behavior. The stability of some of the main connections also reinforce the idea that those ties are not formed at random. Finally, even though we have few points of data, asset graphs seem not to exhibit characteristics of scale-free networks when analyzed in their centralities, although they are still quite different from random ones in terms of these same centrality measures.

The same techniques used here may be applied to a variety of networks, including but not limited to other networks based on financial data. Knowing where to look in terms of threshold values for obtaining asset graphs with good informative value may be also useful for future research of other networks.

Acknowledgments

The author acknowledges the support of this work by a grant from Insper, Instituto de Ensino e Pesquisa. This article was written using LaTeX, all figures were made using PSTricks, and the calculations were made using Matlab, Excel, and Ucinet. All data and algorithms used are freely available upon request to the author at leonidassj@insper.edu.br.

Appendix

A. Stock market indices

The next table (Table 1) shows the stock market indices we used, their original countries, the symbols we used for them in the main text, and their codes in Bloomberg. In the tables, we use "SX" as short for "Stock Exchange". Some of the indices changed names and/or method of calculation and are designated by the two names, prior to and after the changing date.

B. Asset graphs for 1986–1987, 1997–1998, and 2000–2001

In this appendix, we show the asset graphs for the crisis of 1987 (Black Monday), 1997 (Asian Crisis), 1998 (Russian Crisis), 2000 (Internet Bubble) and 2001 (September, 11). The analysis is made by choosing appropriate threshold values and presenting the results in 2-dimensional pictures, following the same rules as in the main text.

B.1. 1986 and 1987 - Black Monday

The first asset graphs to be built are based on data for the years 1986 and 1987, the year that preceded and the one that witnessed the crisis known as Black Monday, whose peak occurred on Monday, October 19, 1987. Not yet completely explained, the crisis made stock markets worldwide drop up to 45% in less than two weeks, and represented the worst financial crisis since 1929.

The networks for 1986 were built using the indices of 16 markets, here named after the countries they belong to (except for the ones from the USA), all of them detailed in the Appendix: S&P 500 and Nasdaq, both from the USA, Canada, Brazil, the United Kingdom, Germany (West Germany at the time), Austria, the Netherlands, India, Sri Lanka,

Table 1
Names, Codes, and Abbreviations of the Stock Market Indices Used in this Article

Index	Country	Symbol	Code in Bloomberg
North America			
S&P 500	United States of America	S&P	SPX
Nasdaq Composite	United States of America	Nasd	CCMP
S&P/TSX Composite	Canada	Cana	SPTSX
IPC	Mexico	Mexi	MEXBOL
Central America			
Bolsa de Panama General	Panama	Pana	BVPSBVPS
BCT Corp Costa Rica	Costa Rica	CoRi	CRSMBCT
Caribbean			
Jamaica SX Market Index	Jamaica	Jama	JMSMX
British overseas territories			
Bermuda SX Index	Bermuda	Berm	BSX
South America			
Ibovespa	Brazil	Braz	IBOV
Merval	Argentina	Arge	MERVAL
IPSA	Chile	Chil	IPSA
IGBC	Colombia	Colo	IGBC
IBC	Venezuela	Vene	IBVC
IGBVL	Peru	Peru	IGBVL
Western and Central Europe			
FTSE 100	United Kingdom	UK	UKX
ISEQ	Ireland	Irel	ISEQ
CAC 40	France	Fran	CAC
DAX	Germany	Germ	DAX
SMI	Switzerland	Swit	SMI
ATX	Austria	Autr	ATX
FTSE MIB or MIB-30	Italy	Ital	FTSEMIB
Malta SX Index	Malta	Malt	MALTEX
BEL 20	Belgium	Belg	BEL20
AEX	Netherlands	Neth	AEX
Luxembourg LuxX	Luxembourg	Luxe	LUXXX
OMX Stockholm 30	Sweden	Swed	OMX
OMX Copenhagen 20	Denmark	Denm	KFX
OMX Helsinki	Finland	Finl	HEX
OBX	Norway	Norw	OBX
OMX Iceland All-Share Index	Iceland	Icel	ICEXI
IBEX 35	Spain	Spai	IBEX
PSI 20	Portugal	Port	PSI20
Athens SX General Index	Greece	Gree	ASE
Eastern Europe			
PX or PX50	Czech Republic	CzRe	PX
SAX	Slovakia	Slok	SKSM
Budapest SX Index	Hungary	Hung	BUX
BELEX 15	Serbia	Serb	BELEX15

Continued

Table 1
Continued

Index	Country	Symbol	Code in Bloomberg
CROBEX	Croatia	Croa	CRO
SBI TOP	Slovenia	Slov	SBITOP
SASE 10	Bosnia and Herzegovina	BoHe	SASX10
MOSTE	Montenegro	Mont	MOSTE
MBI 10	Macedonia	Mace	MBI
WIG	Poland	Pola	WIG
BET	Romania	Roma	BET
SOFIX	Bulgaria	Bulg	SOFIX
OMXT	Estonia	Esto	TALSE
OMXR	Latvia	Latv	RIGSE
OMXV	Lithuania	Lith	VILSE
PFTS	Ukraine	Ukra	PFTS
Eurasia			
MICEX	Russia	Russ	INDEXCF
ISE National 100	Turkey	Turk	XU100
Western and Central Asia			
KASE	Kazakhstan	Kaza	KZKAK
CSE	Cyprus	Cypr	CYSMMAPA
Tel Aviv 25	Israel	Isra	TA-25
Al Quds	Palestine	Pale	PASISI
BLOM	Lebanon	Leba	BLOM
ASE General Index	Jordan	Jord	JOSMGNFF
TASI	Saudi Arabia	SaAr	SASEIDX
Kuwait SE Weighted Index	Kuwait	Kuwa	SECTMIND
Bahrain All Share Index	Bahrain	Bahr	BHSEASI
QE or DSM 20	Qatar	Qata	DSM
ADX General Index	United Arab Emirates	UAE	ADSMI
MSM 30	Oman	Oman	MSM30
South Asia			
Karachi 100	Pakistan	Paki	KSE100
SENSEX 30	India	Indi	SENSEX
Colombo All-Share Index	Sri Lanka	SrLa	CSEALL
DSE General Index	Bangladesh	Bang	DHAKA
Asia-Pacific			
Nikkei 25	Japan	Japa	NKY
Hang Seng	Hong Kong	HoKo	HSI
Shangai SE Composite	China	Chin	SHCOMP
MSE TOP 20	Mongolia	Mong	MSETOP
TAIEX	Taiwan	Taiw	TWSE
KOSPI	South Korea	SoKo	KOSPI
SET	Thailand	Thai	SET
VN-Index	Vietnam	Viet	VNINDEX
KLCI	Malaysia	Mala	FBMKLCI
Straits Times	Singapore	Sing	FSSTI
Jakarta Composite Index	Indonesia	Indo	JCI
PSEi	Philippines	Phil	PCOMP

Continued

Table 1
Continued

Index	Country	Symbol	Code in Bloomberg
Oceania			
S&P/ASX 200	Australia	Aust	AS51
NZX 50	New Zealand	NeZe	NZSE50FG
Northern Africa			
CFG 25	Morocco	Moro	MCSINDEX
TUNINDEX	Tunisia	Tuni	TUSISE
EGX 30	Egypt	Egyp	CASE
Central and Southern Africa			
Ghana All Share Index	Ghana	Ghan	GGSEGSE
Nigeria SX All Share Index	Nigeria	Nige	NGSEINDX
NSE 20	Kenya	Keny	KNSMIDX
DSEI	Tanzania	Tanz	DARSDSEI
FTSE/Namibia Overall	Namibia	Nami	FTN098
Gaborone	Botswana	Bots	BGSMDC
FTSE/JSE Africa All Share	South Africa	SoAf	JALSH
SEMDEX	Mauritius	Maur	SEMDEX

Japan, Hong Kong, Taiwan, South Korea, Malaysia, and Indonesia. The number of countries is small, mainly due to lack of data, but offers a relatively wide variety of nations and cultures in three continents. The networks for 1987 were built using 23 indices, adding to the ones of 1986 indices from Ireland, Sweden, Finland, Spain, Greece, and the Philippines.

B.1.1. Graphs for 1986

For the year 1986, whe have 257 days of data for 16 indices. The average minimum distance for 1000 simulations with randomized data is 1.30 ± 0.02, and the average maximum distance is 1.52 ± 0.02. About 6% of the connections are bellow the minimum threshold and only one is above it. Figure 28 shows the two dimensional view of the asset graphs for 1986, for $T = 1.0$ and $T = 1.2$.

The first connection is formed at $T = 0.8$, between the S&P 500 and the Nasdaq. This cluster grows only at $T = 1.0$, with the addition of Canada. At $T = 1.0$, we also have a connection between West Germany and the Netherlands. At $T = 1.2$, the UK joins the North American cluster, which also connects with the Netherlands and, through it, with Germany. For $T = 1.3$, the Netherlands establishes itself as a hub, and Japan and India join the North American cluster. Random noise starts to have more effects at $T = 1.4$, so the many connections that are formed above $T = 1.4$ and $T = 1.5$ cannot be trusted. The last connections occur at $T = 1.6$.

B.1.2. Graphs for 1987

For the year 1987, whe have 256 days of data for 23 indices. The average minimum distance for 1000 simulations with randomized data is 1.28 ± 0.02, and the average maximum distance is 1.54 ± 0.02. About 9% of the connections are bellow the minimum threshold and only one is above it. Figure 29 shows the two dimensional view of the asset graphs for the second semester of 1987, for $T = 1.0$ and $T = 1.2$.

The first connections are again established between the S&P 500 and the Nasdaq, but now at $T = 0.7$. At $T = 0.9$, Canada joins the North American cluster. A connection between West Germany and the Netherlands only occur at $T = 1.0$, and at $T = 1.1$, the UK connects with both the North American cluster and the other European indices, with Ireland attached to it. The Netherlands is again a hub, connecting with most of the indices. For $T = 1.2$, Sweden joins the European cluster, and Japan connects with both Germany and Sweden. For $T = 1.3$, many more connections are made, mainly by the newcomers Hong Kong and Malaysia. Connections above $T = 1.4$ are already dominated by noise, and the last connections occur for $T = 1.5$.

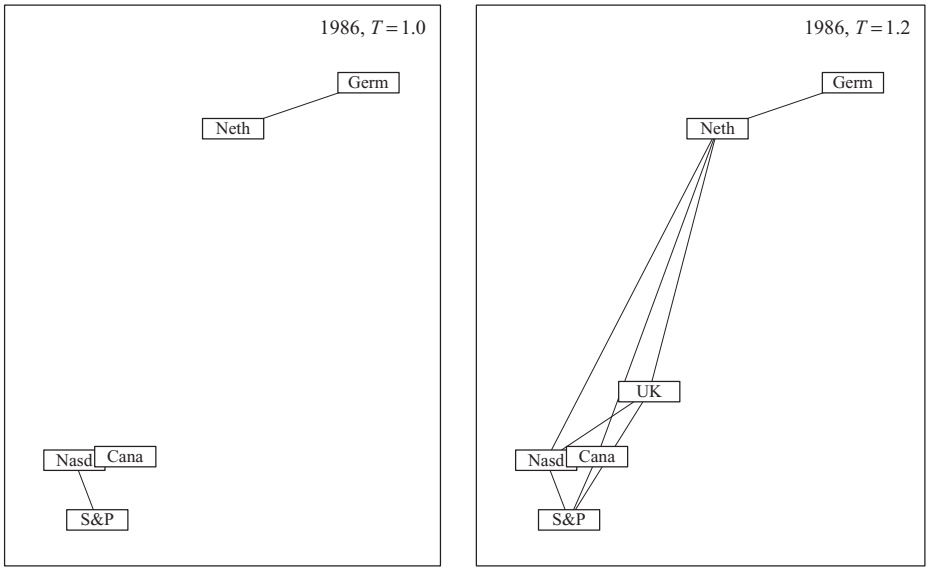

Fig. 28. Asset graphs for the year 1986, with thresholds $T = 1.0$ and $T = 1.2$.

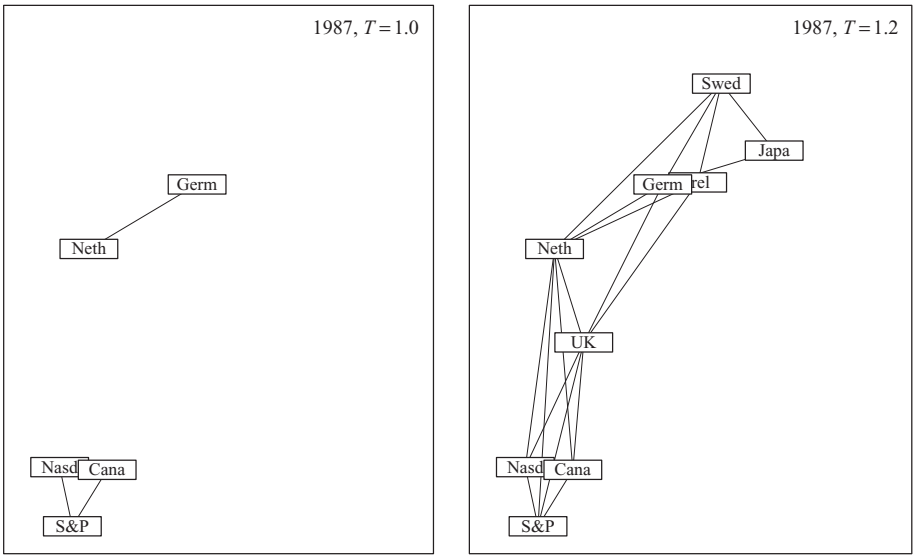

Fig. 29. Asset graphs for the year 1987, with thresholds $T = 1.0$ and $T = 1.2$.

B.2. 1997 and 1998 - Asian Financial Crisis and Russian Crisis

We now jump some years ahead to the next two crises we are going to analyze. The first one is the Asian Financial Crisis of 1997, which began with the devaluation of the Thailandese currency and spread to other Pacific Asian markets. The second was almost a consequence of the first one, as prices of commodities fell worldwide, affecting the Russian economy with particular acuteness. The networks for 1997 are built using 57 indices, adding the indices from Mexico, Costa Rica, Bermuda, Jamaica, Argentina, Chile, Venezuela, Peru, France, Switzerland, Belgium, Denmark, Norway, Iceland, Portugal, the Czech Republic, Slovakia, Hungary, Poland, Estonia, Turkey, Israel, Lebanon, Saudi Arabia, Oman, Pakistan, China, Thailand, Australia, Morocco, Ghana, Kenya, South

Africa, and Mauritius. For 1997, we have no data about Russia, which is added to the data used for building networks for 1998 (which then has 58 indices).

B.2.1. Graphs for 1997

For the year 1997, we have 256 days of data for 57 indices. The average minimum distance for 1000 simulations with randomized data is 1.25 ± 0.01, and the average maximum distance is 1.55 ± 0.02. About 12% of the connections are bellow the minimum threshold and only one is above it. Figure 30 shows the two dimensional view of the asset graphs for the second semester of 1997, for thresholds $T = 0.7$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

The first connections occur for $T = 0.7$, between the S&P 500 and the Nasdaq, and between the Netherlands and Switzerland. For $T = 0.8$, we can talk about a North American and an European clusters, with Canada joining the two indices of the USA and the UK, France, Germany, Belgium, Sweden, Finland, and Spain joining the Netherlands and Switzerland. For $T = 0.9$, Argentina joins the now American cluster, and the European cluster becomes denser and is joined by Austria and Denmark. Mexico and Brazil, through Argentina, join the American cluster at $T = 1.0$, and at this same threshold, Europe becomes a very dense cluster, receiving the addition of the indices from Ireland and Portugal. At $T = 1.1$, both clusters join, mainly through Canada, Chile connects with the European and not with the American cluster, and Australia and South Africa connect with the European cluster. At $T = 1.2$, the American and European clusters are intertwined; Chile, Peru, Ireland, Norway, Portugal, Hungary, Poland, Australia, and

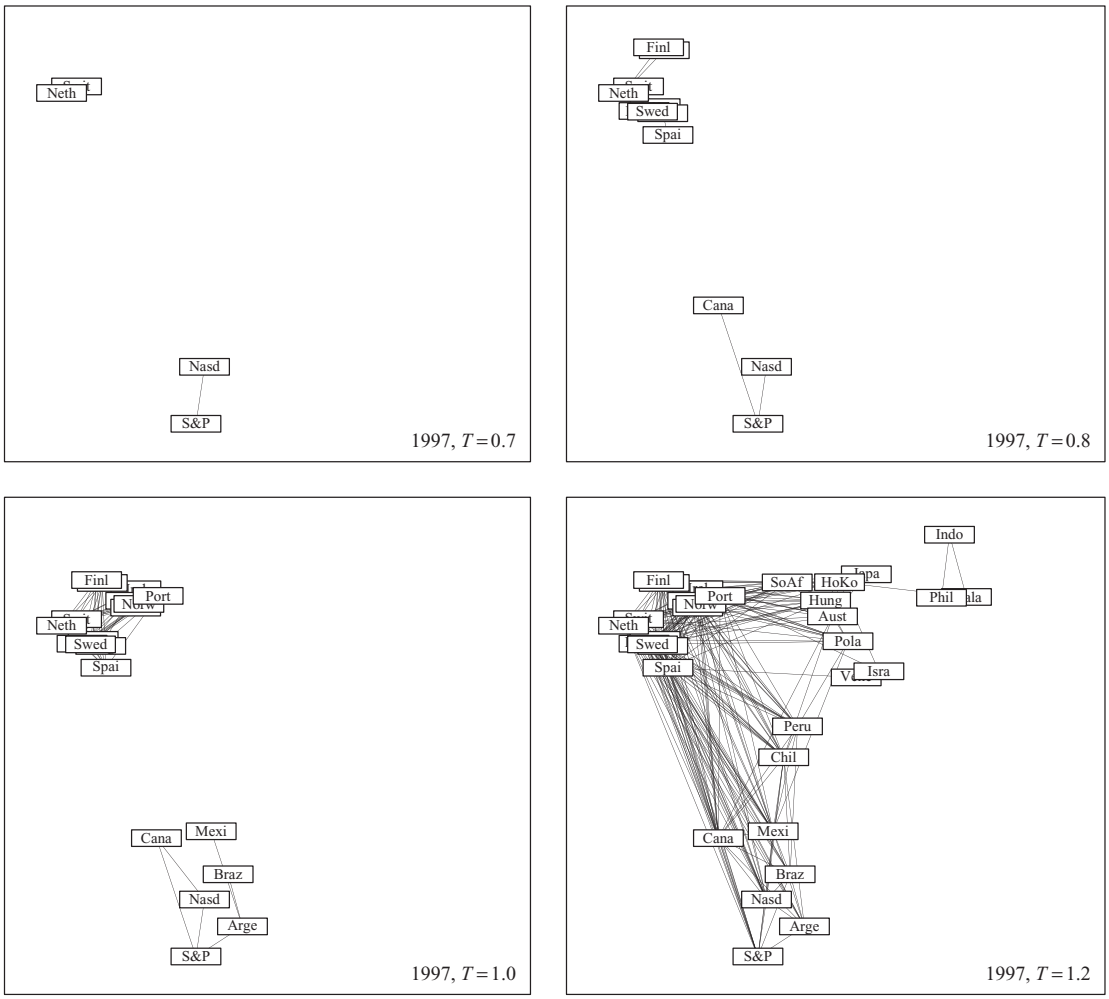

Fig. 30. Asset graphs for the year 1997, with thresholds $T = 0.7$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

South Africa also join this large cluster. A Pacific Asian cluster is formed by the indices from Japan, Hong Kong, Malaysia, Philippines, and Australia. This cluster is strongly linked with the American-European cluster, mainly by Hong Kong and Australia. From $T = 1.3$ onwards, many connections are formed, probably mostly due to random noise, and the last connections occur for $T = 1.6$.

B.2.2. Graphs for 1998

For the year 1998, whe have 256 days of data for 58 indices. The average minimum distance for 1000 simulations with randomized data is 1.25 ± 0.02, and the average maximum distance is 1.55 ± 0.01. About 19% of the connections are bellow the minimum threshold and none is above it. Figure 31 shows the two dimensional view of the asset graphs for the second semester of 1998, for thresholds $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

Again, the first connection is between the S&P 500 and the Nasdaq, at $T = 0.6$. For $T = 0.7$, an European cluster forms with the connections between the indices from the UK, France, Germany, Switzerland, Belgium, the Netherlands, Sweden, and Spain. The most connected nodes are those of the Netherlands and of France. For $T = 0.8$, the European cluster gets denser and receives Denmark and Portugal, but weakly connected. For $T = 0.9$, Canada joins the North American cluster and a Latin American cluster forms between Mexico, Brazil, and Argentina. Ireland, Austria, and Norway join the European cluster, now even denser, and South Africa connects with Germany. At $T = 1.0$, the American clusters merge, and Canada establishes connections with four European indices. The Czech Republic and Hungary join the European cluster, and there appears a strange correlation between Poland and Australia. Hong Kong connects with South Africa, which is now fully integrated with the European cluster. At $T = 1.1$, many more connections are formed between American and European indices, wich now include Chile, Greece, Poland, Turkey, and Israel. Hong Kong is also highly connected to European markets, and a Pacific Asian cluster is formed with the indices from Japan, Hong Kong, Thailand, Malaysia, Indonesia, Philippines, and Australia. Many more connections are formed for $T = 1.2$, and by then the

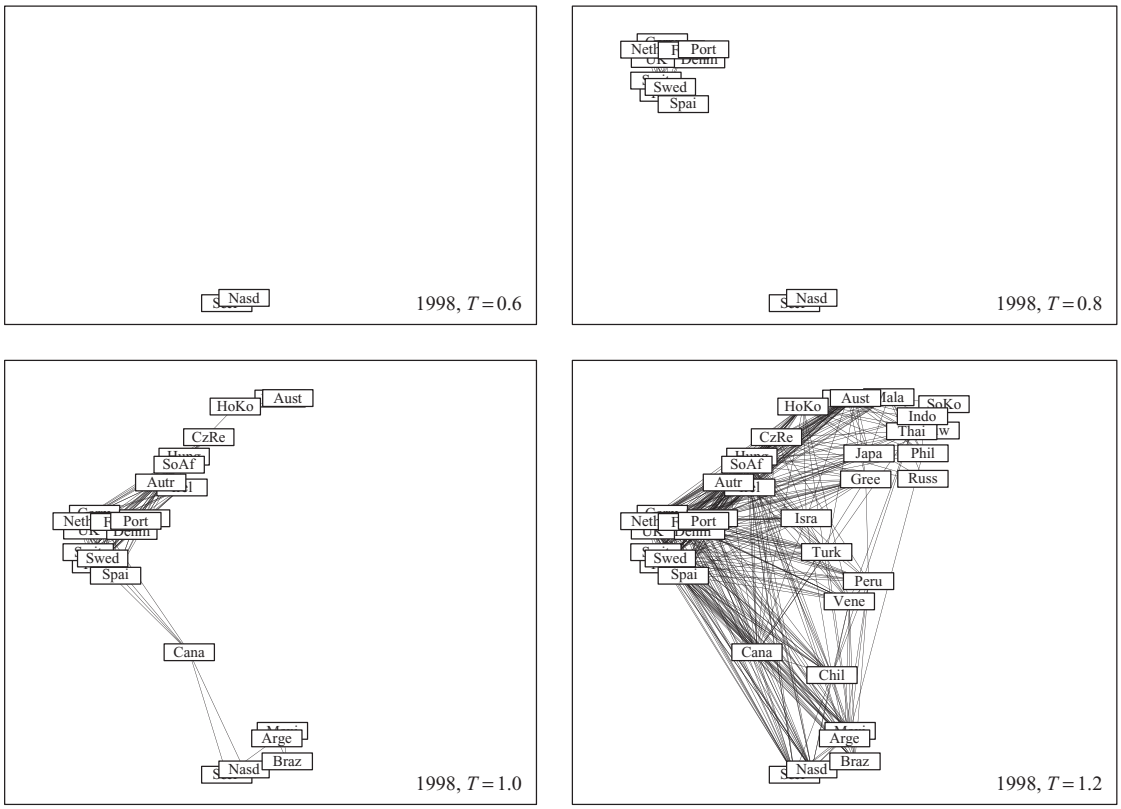

Fig. 31. Asset graphs for the year 1998, with thresholds $T = 0.7$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

world is a large single cluster. Noise rapidly dominates for thresholds above $T = 1.3$, and the last connections occur for $T = 1.6$.

B.3. 2001 - Burst of the dot-com bubble and September 11

Two very distinct crises arose in 2001: the first was the burst of a bubble due to the overvaluing of the new internet-based companies, and the second was the fear caused by the severest terrorist attack a country has ever faced, when the USA was targeted by commercial airplanes overtaken by terrorists. Both crises were very different in origin and in duration, and their analysis might shed some light on how crises may develop.

For the networks concerning the year 2000, we add now the indices from Panama, Italy, Malta, Luxembourg, Romania, Latvia, Lithuania, Ukraine, Palestine, Jordan, Qatar, Mongolia, Singapore, Tunisia, Egypt, and Nigeria, for a total of 74 indices. For the networks of 2001, we also add the indices from Bulgaria, Kazakhstan, Vietnam, New Zealand, and Botswana, for a total of 79 indices.

B.3.1. Graphs for 2000

For the year 2000, whe have 254 days of data for 74 indices. The average minimum distance for 1000 simulations with randomized data is 1.24 ± 0.02, and the average maximum distance is 1.56 ± 0.01. About 9% of the connections are bellow the minimum threshold and none is above it. Figure 32 shows the two dimensional view of the asset graphs for the second semester of 2000, for thresholds $T = 0.7$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

At $T = 0.7$, a connection is formed between the S&P 500 and the Nasdaq, and connections are formed

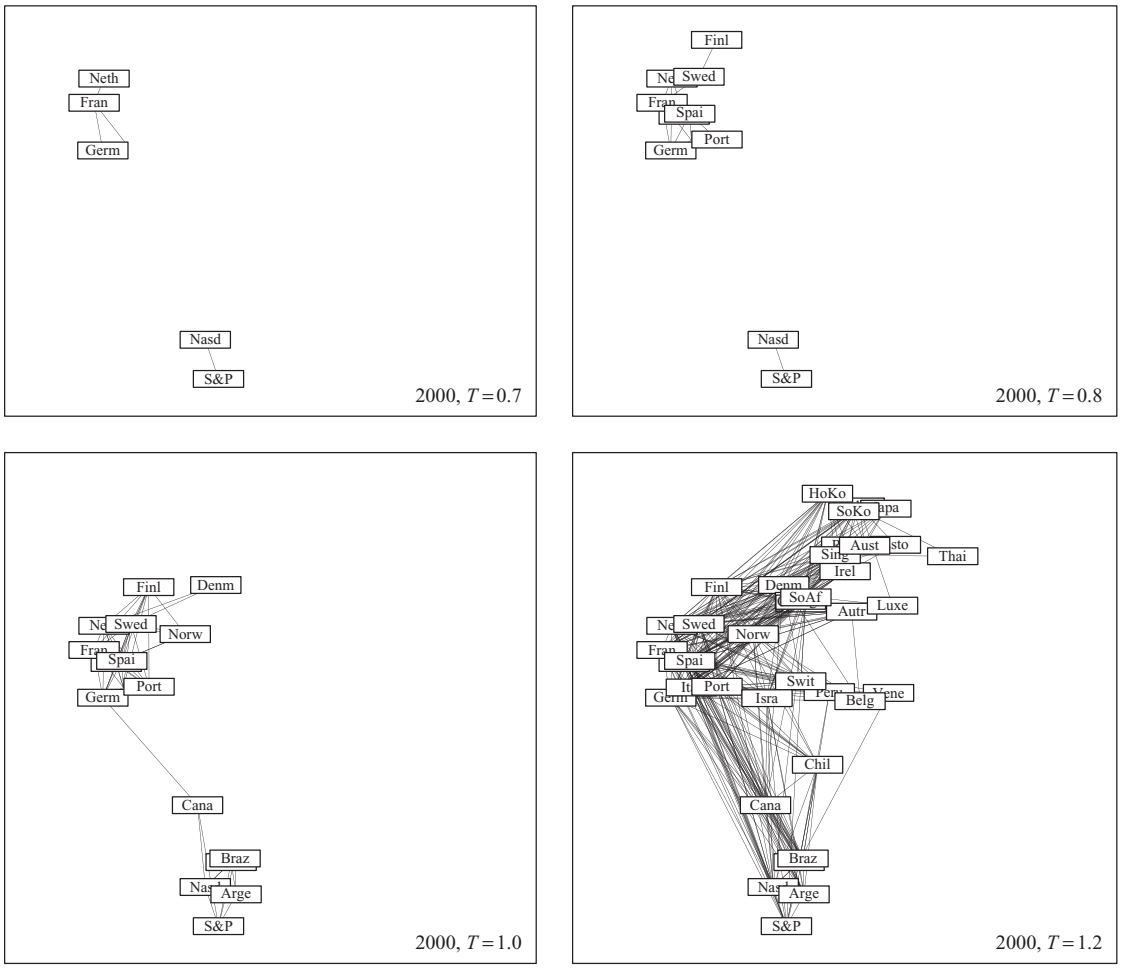

Fig. 32. Asset graphs for the year 2000, with thresholds $T = 0.7$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

between the indices from France, Germany, Italy, and the Netherlands. For $T = 0.8$, the European cluster grows with the addition of the UK, Sweden, Finland, Spain, and Portugal. At $T = 0.9$, Canada and Mexico join the North American cluster. Brazil and Argentina join the now American cluster at $T = 1.0$. At this same threshold, the European cluster receives the indices from Denmark and Norway, and a single connection is made between Canada and Germany. For $T = 1.1$, both clusters merge, and Chile, Ireland, Switzerland, the Czech Republic, Hungary, Russia, Israel, and South Africa join this merged cluster. A Pacific Asian cluster is formed around the indices of Japan and Hong Kong, involving the indices of South Korea, Singapore, and Australia, being some of the indices of the Pacific Asian cluster also connected with the European cluster. A myriad of new connections is formed at $T = 1.2$ and above it, but without the formation of any new cluster. The last connections occur, once more, for $T = 1.6$.

B.3.2. Graphs for 2001

For the year 2001, whe have 253 days of data for 79 indices. The average minimum distance for 1000 simulations with randomized data is 1.24 ± 0.02, and the average maximum distance is 1.57 ± 0.01. About 14% of the connections are bellow the minimum threshold and none is above it. Figure 33 shows the two dimensional view of the asset graphs for the second semester of 2001, for thresholds $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

Once more, for $T = 0.6$, there is a connection between the S&P 500 and the Nasdaq, and connections between the UK, France, Germany, Italy, the Netherlands, and Spain, being France the central node of the cluster. For $T = 0.7$, the European cluster grows denser, and receives the indices from Switzerland and Sweden. At $T = 0.8$, Canada joins the North American cluster, and the European cluster receives the addition of Belgium and Finland. For $T = 0.9$, the European cluster grows denser and receives the indices from Norway and Portugal. Mexico joins the North American cluster at $T = 1.0$, at the same moment in which the North American cluster establishes connections with the Eurpean one, which now has Ireland and Denmark among its ranks. At $T = 1.1$, Brazil and Argentina join the American cluster, wich is now even more connected with the European cluster, which

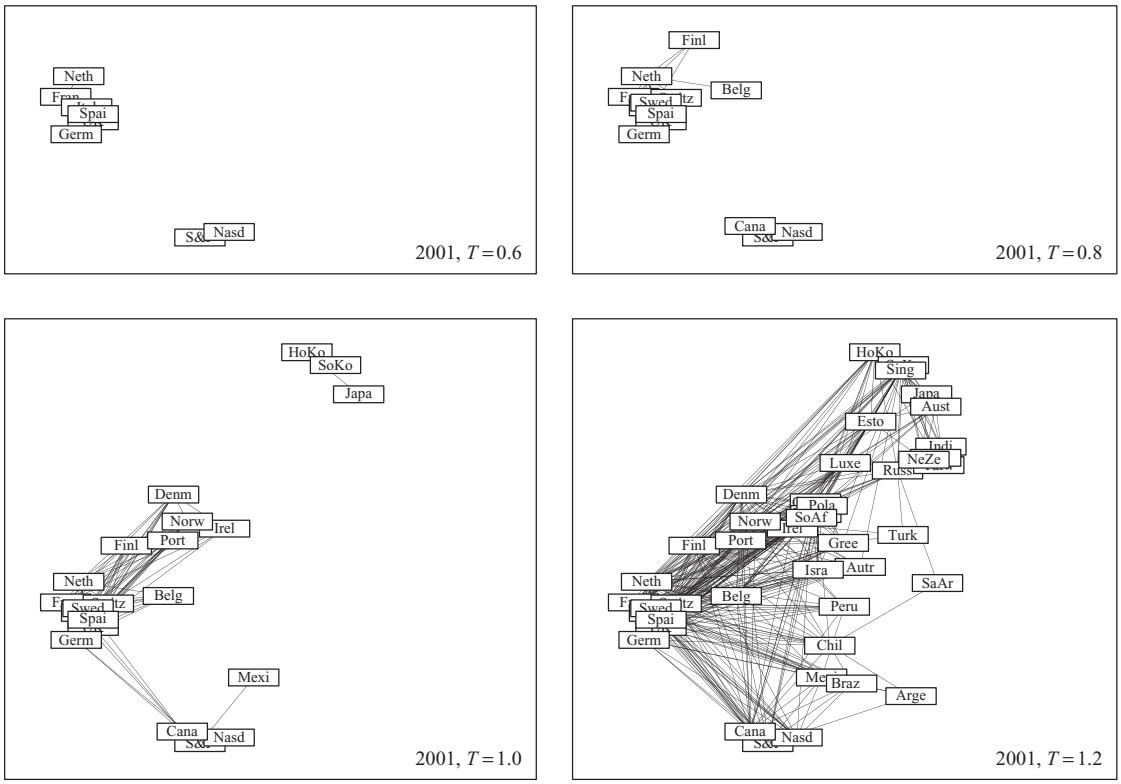

Fig. 33. Asset graphs for the year 2001, with thresholds $T = 0.6$, $T = 0.8$, $T = 1.0$, and $T = 1.2$.

is joined by Chile, Austria, Luxembourg, the Czech Republic, Hungary, Poland, Russia, Israel, and South Africa. A relatively isolated Pacific Asian cluster is formed between Japan, Hong Kong, South Korea, Singapore, and Australia. At $T = 1.2$, further integration occurs, and many more connections are formed. Noise starts having more effect from $T = 1.3$, and above it, and the last connections occur for $T = 1.6$.

References

Aste, T., Di Matteo, T., 2005. Correlation filtering in financial time series, Noise and Fluctuations in Econophysics and Finance. Proc. SPIE 5848, 100–109.

Ausloos, M., Lambiotte, R., 2007. Clusters or networks of economies? A macroeconomy study through gross domestic product. Physica A 382, 16–21.

Bonanno, G., Lillo, F., Mantegna, R.N., 2001. High-frequency cross-correlation in a set of stocks. Quant. Financ. 1, 96–104.

Bonanno, G., Caldarelli, G., Lillo, F., Micciché, S., Vandewalle, N., Mantegna, R.N., 2004. Networks of equities in financial markets. Eur. Phys. J. B 38, 363–371.

Borg, I., Groenen, P., 2005. Modern Multidimensional Scaling: Theory and Applications, second ed. Springer-Verlag.

Borghesi, C., Marsili, M., Micciché, S., 2007. Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode. Phys. Rev. E 76, 026104.

Bouchaud, J.-P., Potters, M., 2011. Financial applications of random matrix theory: A short review. In: Akemann, G., Baik, J., Di Francesco, P. (Eds.), The Oxford Handbook of Random Matrix Theory. Oxford University Press.

Brida, J.G., Risso, W.A., 2008. Multidimensional minimal spanning tree: The Dow Jones case. Physica A 387, 5205–5210.

Chou, Y., 1975. Statistical Analysis. Holt International, ISBN 0030894220, Sect. 17.9.

Coelho, R., Gilmore, C.G., Lucey, B., Richmond, P., Hutzler, S., 2007. The evolution of interdependence in world equity markets - evidence from minimum spanning trees. Physica A 376, 455–466.

Conlon, T., Ruskin, H.J., Crane, M., 2009. Cross-correlations dynamics in financial time series. Physica A 388, 705–714.

Coronnello, C., Tumminello, M., Lillo, F., Micchichè, S., Mantegna, R.N., 2005. Sector identification in a set of stock return time series traded at the London Stock Exchange. Acta Phys. Pol. B 36, 2653–2679.

Coronnello, C., Tumminello, M., Lillo, F., Micchichè, S., Mantegna, R.N., 2007. Economic sector identification in a set of stocks traded at the New York Stock Exchange: A comparative analysis. Proc. SPIE 6601, 66010T.

Dimov, I.I., Kolm, P.N., Maclin, L., Shiber, D.Y.C., 2009. Hidden noise structure and random matrix models of stock correlations. Working Paper 2009-4, New York University, Courant Institute of Mathematical Sciences.

Drożdż, S., Kwapień, J., Speth, J., 2010. Coherent patterns in nuclei and in financial markets. AIP Conf. Proc. 1261, 256–264.

Eryiğit, M., Eryiğit, R., 2009. Network structure of cross-correlations among the world market indices. Physica A 388, 3551–3562.

Fenn, D.J., Porter, M.A., Mucha, P.J., McDonald, M., Williams, S., Johnson, N.F., et al., 2010. Dynamical Clustering of Exchange Rates. arXiv:0905.4912v2.

Flavin, T., Hurley, M., Rousseau, F., 2002. Explaining stock market correlation: A gravity model approach. Manchester Sch. 70, 87–106.

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486, 75–174.

Gormus, S., Guloglu, B., Gunes, S., 2011. Determinant of stock market return correlation: An extended gravity model approach. Intl. J. Contem. Econ. Admin. Sci. 1, 298–312.

Gross, T., Sayama, H., 2009. Adaptive Networks: Theory. Models and Applications (Understanding Complex Systems). Springer.

Gross, C., 2010. Complex and Adaptive Dynamical Systems: A Primer, second ed. (Springer Complexity). Springer.

Hołyst, J.A., 2012. Network Analysis of Correlation Strength between the Most Developed Countries. arXiv:1211.3599v1.

Laloux, L., Cizeau, P., Bouchaud, J.-P., Potters, M., 1999. Noise dressing of financial correlation matrices. Phys. Rev. Lett. 83, 1467–1470.

Laloux, L., Cizeau, P., Potters, M., Bouchaud, J.-P., 2000. Random matrix theory and financial correlations. Math. Models Methods Appl. Sci.

Kenett, D.Y., Shapira, Y., Madi, A., Bransburg-Zabary, S., Gur-Gershgoren, G., Ben-Jacob, E., 2011. Index cohesive force analysis reveals that the US market

became prone to systemic collapses since 2002. PLoS One 6 (4), e19378.

Kenett, D.Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R.N., Ben-Jacob, E., 2010. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. PLoS One 5 (12), e15032.

Keskin, M., Deviren, B., Kocalkaplan, Y., 2011. Topology of the correlation networks among major currencies using hierarchical structure methods. Physica A 390, 719–730.

Kulkarni, V., Deo, N., 2007. Correlation and volatility of an Indian stock market: A random matrix approach. Eur. Phys. J. B 60, 101–109.

Kwapień, J., Gworek, S., Drożdż, S., Górski, A. (2009). Analysis of a network structure of the foreign currency exchange market. J. Econ. Interact. Coord. 4, 55–72.

Kwapień, J., Gworek, S., Drożdż, S., 2006a. Structure and evolution of the foreign exchange networks. Acta Physica Polonica B 40, 175–194.

Kwapień, J., Drożdż, S., Oświęcimka, P., 2006b. The bulk of the stock market correlation matrix is not pure noise. Physica A 359, 589–606.

Kwapień, J., Drożdż, S., Górski, A.Z., Oświęcimka, P., 2006. Asymmetric matrices in an analysis of financial correlations. Acta Physica Polonica B 37, 3039–3048.

Mantegna, R.N., 1999. Hierarchical structure in financial markets. Eur. Phys. J. B 11, 193.

Le Martelot, E., Hankin, C., 2012. Fast Multi-Scale Detection of Relevant Communities. arXiv:1204.1002.

Mehta, M.L., 2004. Random Matrices. Academic Press.

Micchichè, S., Bonanno, G., Lillo, F., Mantegna, R.N., 2003. Degree stability of a minimum spanning tree of price return and volatility. Physica A 324, 66–73.

Miller, J.H., Page, S.E., 2007. Complex Adaptive Systems: An Introduction to Computational Models of Social Life (Princeton Studies in Complexity). Princeton University Press.

Namaki, A., Jafari, G.R., Raei, R., 2010. Comparing TEPIX as an emerging market with efficient market by Random Matrix Theory. Preprint.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA. 103, 8577–8582.

Newman, M.E.J., 2010. Networks, and Introduction. Oxford University Press.

Oh, G., Eom, C., Wang, F., Jung, W.-S., Stanley, H.E., Kim, S., 2011. Statistical properties of cross-correlation in the Korean stock market. Eur. Phys. J. B 79, 55–60.

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., 2002. Dynamic asset trees and portfolio analysis. Eur. Phys. J. B 30, 285–288.

Onnela, J.-P., Chakraborti, A., Kaski, K., 2003a. Dynamics of market correlations: Taxonomy and portfolio analysis. Phys. Rev. E 68, 1–12.

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., 2003b. Dynamic asset trees and black monday. Physica A 324, 247–252.

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., Kanto, A., 2003c. Asset trees and asset graphs in financial markets. Phys. Scripta T 106, 48–54.

Onnela, J.-P., Kaski, K., Kertész, J., 2004. Clustering and information in correlation based financial networks. Eur. Phys. J. B 38, 353–362.

Ormerod, P. (2005). The convergence of European business cycles 1980-2004. Acta Physica Polonica B 36, 2747–2756.

Pan, R.K., Sinha, S., 2007. Collective behavior of stock price movements in an emerging market. Phys. Rev. E 76, 1–9.

Plerou, V., Gopikrishman, P., Rosenow, B., Amaral, L.A.N., Stanley, H.E., 1999. Universal and non-universal properties of cross-correlations in financial time series. Phys. Lett. 83, 1471–1474.

Plerou, V., Gopikrishman, P., Rosenow, B., Amaral, L.A.N., Stanley, H.E., 2000. A Random Matrix Theory approach to financial cross-correlations. Physica A 287, 374–382.

Plerou, V., Gopikrishman, P., Rosenow, B., Amaral, L.A.N., Guhr, T., Stanley, H.E., 2002. Random matrix approach to cross-correlations in financial data. Phys. Rev. E 65, 066126.

Potters, M., Bouchaud, J.-P., Laloux, L., 2005. Financial applications of random matrix theory: Old laces and new pieces. Proceedings of the Cracow conference on "Applications of Random Matrix Theory to economy and other complex systems".

Qiu, T., Zheng, B., Chen, G., 2010. Adaptive Financial Networks with Static and Dynamic Thresholds. arXiv:1002.3432v1.

Rak, R., Kwapień, J., Drożdż, S., Oświęcimka, P., 2008. Cross-correlations in Warsaw Stock Exchange. Acta Physica Polonica A 114, 561–568.

Reimann, S., Tupak, A., 2010. Dynamics on/in Financial Markets: Dynamical Decoupling and Stylized Facts. arXiv:1004.1522v1.

Rosenow, B., Plerou, V., Gopikrishnan, P., Amaral, L.A.N., Stanley, H.E., 2002. Application of random matrix theory to study cross-correlations of stock prices. Int. J. Theoret. Appl. Financ. 3, 399–403.

Sandoval Jr., L., 2012. Pruning a minimum spanning tree. Physica A 391, 2678–2711. doi:10.1016/j.physa.2011.12.052.

Sandoval Jr., L., 2012. A Map of the Brazilian Stock Market. Adv. Complex Syst. 15, 1250042-1–1250042-40.

Sandoval Jr., L., 2012. To Lag or Not to Lag?. arXiv:1201.4586.

Sandoval Jr., L., Franca, I.D.P., 2012. Correlation of financial markets in times of crisis. Physica A 391, 187–208.

Shapira, Y., Kenett, D.Y., Ben-Jacob, E., 2009. The index cohesive effect on stock market correlations. Eur. Phys. J. B 72, 657–669.

Sieczka, P., Hołyst, J.A., 2009. Correlations in commodity markets. Physica A 388, 1621–1630.

Sinha, S., Pan, R.K., 2007. Uncovering the internal structure of the Indian financial market: Cross-correlation behavior in the NSE. In: Econophysics of Markets and Business Networks. Springer, pp. 215–226.

Song, D.-M., Tumminello, M., Zhou, W.-X., Mantegna, R.N., 2011. Evolution of worldwide stock markets, correlation structure and correlation based graphs. Phys. Rev. E 84, 026108.

Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N., 2005. A tool for filtering information in complex systems. Proc. Natl. Acad. Sci. USA. 102, 10421–10426.

Tumminello, M., Di Matteo, T., Aste, T., Mantegna, R.N., 2007. Correlation based networks of equity returns sampled at different time horizons. Eur. Phys. J. B 55, 209–217.

Tumminello, M., Lillo, F., Mantegna, R.N., 2010. Correlation, hierarchies, and networks in financial markets. J. Econ. Behav. Organ. 75, 40–58.

Wilcox, D., Gebbie, T., 2007. An analysis of cross-correlations in an emerging market. Physica A 375, 584–598.

Zhanga, Y., Leea, G.H.T., Wonga, J.C., Kokb, J.L., Prustyb, M., Cheong, S.A., 2011. Will the US Economy Recover in 2010? A minimal spanning tree study. Physica A 390, 2020–2050.