Eur. Phys. J. B **74**, 243–249 (2010)
DOI: 10.1140/epjb/e2010-00079-4

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Topological properties of commodities networks

B.M. Tabak[1,2,a], T.R. Serra[3], and D.O. Cajueiro[3,4]

[1] Departamento de Pesquisa, Banco Central do Brasil, SBS Quadra 3, Bloco B, ED. Sede, 13 andar, Brasília, DF, 70074-900 Brasil
[2] Department of Economics, Universidade Católica de Brasília, Brasília, DF, 70790-160, Brazil
[3] Department of Economics, Universidade de Brasília, Campus Darcy Ribeiro, Prédio da FACE, Asa Norte, 70910-900 Brasília, Brazil
[4] National Institute of Science and Technology for Complex Systems, Rio de Janeiro, RJ, 22290-180, Brazil

Received 14 April 2009 / Received in final form 4 September 2009
Published online 2 March 2010 –

Abstract. This paper investigates the topological properties of the commodities networks. We have found that commodities form strong clusters and are homogeneous with relation to sector (metals, agriculture and energy). We also develop a dynamic approach suggesting that agriculture commodities are very important in the network, followed by metals and energy. Furthermore, the parameters that characterize the network seem to be changing over time.

1 Introduction

Recent literature has employed network theory to unveil some characteristics of important economic networks [1–8]. However, most the recent literature has studied networks generated by correlations of stock prices. In this paper we focus on commodity prices and the main objective is to characterize the topology and taxonomy of the commodity network. The commodity network is an interesting case of study for a number of reasons. Before the existence of money, commodities were used to buy and to trade all kind of goods. Nowadays, the commodities are not only used to exchange goods but it is used as primary raw materials in all production stages. In crisis and turbulence periods they can be seen as a measure of value.

Many developing countries, and in particular most of those with weak growth performance, remain highly dependent on commodities for their trade, production income and employment. Large structural shifts in the global economy have been steadily reflected in commodities prices increases, especially of food, metals and energy. For example, recently we have witnessed the increase of crude oil's price, reaching historical prices over 140 dollars. For countries (net exporters) this increase in prices is positive for the Balance of Payments. However, for other countries these shocks may propagate within the economy increasing overall inflation. For this reason, the study of commodities is of great relevance. In this paper we will study how the different commodities form a network and analyze their relationship using graph theory tools. This a topic of utmost importance as it may reveal patterns within the commodities network.

There is a vast literature about the study of topological properties of networks, which has associated a meaningful economic taxonomy. Mantegna [2] presents the minimum spanning tree (MST) and the associated subdominant ultrametric tree in which he selects a topological space of stocks and is useful to describe the financial markets.

Recent research has found interesting patterns in stock market networks using concepts from Graph theory [3–8]. Overall this literature has suggested that it is possible to uncover the formation of clusters within stock market networks and that it is possible to define a taxonomy for the stocks within the network. These results are important and have many implications for the design of portfolio and risk management models.

Lien and Yang [9] examine the asymmetric effect of basis on the time-varying variance and correlation of spot and futures returns and its consequence in dynamic futures hedging strategies in commodity markets. They propose an alternative specification of the Bivariate GARCH model in which the effect is incorporated for estimating MVRHs (dynamic minimum variance hedge ratios). They applied it to the commodity market and have found that the basis effect is asymmetric. They also found that the model with the asymmetric effect provides greater risk reduction than the conventional models, illustrating importance of the asymmetric effect when modeling the joint dynamics of spot and futures returns and hence estimating hedging strategies.

Matia et al. [10] found that the price fluctuations for commodities have a significantly broader multifractal spectrum than for stocks. They observe the clustering of commodities, shuffling the returns by randomly exchanging pairs. Therefore, they found that the commodity series loses its clustering and the stock series resembles the original one. It permits to identify that commodities and stocks are similar to stocks for large fluctuations and they differ for small fluctuations. They also find that, for

[a] e-mail: benjamin.tabak@bcb.gov.br

commodities, stronger higher-order correlations in price fluctuations result in broader multifractal spectra.

Some authors [11] have applied network-based approach to analyze the Brazilian interbank network structure. They found that the network has high heterogeneity, a weak evidence of community structure and that the market is characterized by money centers having exposures to many banks. Tabak et al. [12] apply the network theory to investigate the topological properties of the Brazilian term structure of interest rates. They found that the short-term interest rate is the most important within the interest rates network, in accordance with the Expectation Hypothesis of interest rates. They also found that the Brazilian interest rates network forms clusters by maturity.

Sieczka and Holyst [13] investigated correlations of future contracts for commodities traded at different markets over the period of 1998 to 2007. They created a MST. They also studied dynamic properties of correlations and found that the market was constantly getting more correlated within the studied period.

Recent contributions have shown that entropy measures can be used to assess the topology of complex networks ([14]) and that tools from complex networks can be useful in evaluating different data-sets such as trades in the world trade web ([15]). Furthermore, Arianos and Carbone [16] have developed a method to estimate cross-correlation of long-range dependent series.

Overall, little is known regarding the topology of commodities networks and its dynamics properties. This paper seeks to fill this gap with a dynamic analysis of commodities networks.

The remainder of the paper is structured as follows. Section 2 introduces the methodology and the sampling procedures, whereas Section 3 shows the data and Section 4 presents empirical results. Finally, Section 5 provides some final considerations.

2 Commodity networks

We use the cross-correlation in commodities prices changes from January 1, 1991 to February 8, 2008. The dataset is daily closing prices denominated in US dollars. From the log-return of the commodities prices, which is defined as $S_i(t) = \ln(Y_i(t)) - \ln(Y_i(t-1))$, where $Y_i(t)$ is the price of commodity i at time t we could calculate the cross-correlation function as:

$$\rho_{i,j} = \frac{\langle S_i \cdot S_j\rangle - \langle S_i\rangle\langle S_j\rangle}{\sqrt{(\langle S_i^2\rangle - \langle S_i\rangle^2)(\langle S_j^2\rangle - \langle S_j\rangle^2)}}, \quad (1)$$

where $\langle S_i\rangle$ represents the statistical average of $S_{i,t}$ for a given time period. All cross-correlations range from −1 to 1 where a value of $\rho_{i,j} = 1$ implies that commodities have a perfect correlation between them, and a value of $\rho_{i,j} = -1$ suggests that these commodities are perfectly anti-correlated. The matrix $\rho_{i,j}$ has $n \times n$ order if there are n commodities in the sample and is symmetric as $\rho_{i,j} = \rho_{j,i}$.

2.1 Minimum spanning tree

To study the topology of the network the MST requires the use of a variable that can be interpreted as distance, satisfying the three axioms of Euclidean distance. Therefore, we transform this matrix in order to build a distance matrix. To build the commodity network we employ the metric distance proposed by Mantegna [2], $d_{i,j} = \sqrt{2(1-\rho_{i,j})}$, where $\rho_{i,j}$ is the correlation between changes in commodities i and j[1].

The MST is a graph that connects all the n nodes of the graph with $n-1$ edges, such that the sum of all edge weights is a minimum. The MST extracts significant information from the distance matrix and it reduces the information space from $\frac{n\times(n-1)}{2}$ correlations to $n-1$ tree edges. It is the spanning tree of the shortest length using the Prim algorithm of the $d_{i,j}$. Prim's algorithm is an algorithm in graph theory that finds a minimum spanning tree for a connected weighted graph. This means it finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. If the graph is not connected, then it will only find a minimum spanning tree for one of the connected components[2].

Define the maximal distance $d^*_{i,j}$ between two successive commodities when moving from commodity i to commodity j over the shortest path of the MST connecting these two commodities. The distance $d^*_{i,j}$ is called subdominant ultrametric distance and a space connected by these distances provides a topological space that has associated a unique indexed hierarchy. This distance satisfies the above axioms of Euclidean distance and also the following ultrametric inequality:

$$d_{i,j} \le \max[d_{i,k}, d_{k,j}]. \quad (2)$$

To construct the hierarchical tree and a better interpretation of it we used the subdominant ultrametric distance $d^*_{i,j}$ and employ the complete linkage clustering method.

In the complete linkage method distance between nodes is defined as the distance between the most distant pair of nodes. The distance D_c can be defined as:

$$D_c = \mathrm{Max}(d_{i,j}). \quad (3)$$

2.2 Weighted networks measures

In order to study the evolution of the networks parameters over time we estimate a variety of network measures[3]

[1] This metric satisfies the three axioms of Euclidean distance: (i) $d_{i,j} = 0$ if and only if $i = j$, (ii) $d_{i,j} = d_{j,i}$, and (iii) $d_{i,j} \le d_{i,k} + d_{k,j}$.

[2] The Prim's algorithm has the following steps: (1) choose a pair of commodities with the nearest distance and connect with a line proportional to this distance; (2) connect a pair with second nearest distance; (3) connect the nearest pair that is not connected by the same tree; and (4) repeat step three until all commodities are connected in one tree.

[3] See [17] for a discussion on network measures.

using a moving window of fixed size (1008 observations). Therefore, we estimate the parameters for the sample that comprises observation 1 to 1008, 2 to 1009, and so forth until we use all the sample. This dynamic approach allows studying the evolution of the network over time.

Domination power measures [18], also called β-measure of individual nodes, are able to find the centrality in a network that takes the direction and the weight of the relations into account and could be described as below:

$$\beta(i) = \sum_{j=1}^{n} \frac{w(i,j)}{\lambda(j)}, \tag{4}$$

where $w(i,j)$ is the weight of links connected to vertex i and λ_j is the dominance weight of node j given by

$$\lambda(j) = \sum_{i=1}^{n} w(i,j). \tag{5}$$

We also study closeness centrality [19], which proxies for the proximity to the rest of vertices in the network. The higher its value, the closer that vertex is to the others (on average). Given a vertex k and a graph G, it can be defined as:

$$C(k) = \frac{1}{\sum_{h \in G} d_G(k,h)}, \tag{6}$$

where $d_G(k,h)$ is the minimum distance from vertex k to vertex h. This measures the influence of a vertex in a graph.

For a given vertex i with connectivity k_i and strength s_i all weights $w_{i,j}$ can be of the same order if $s_i = k_i$ and if one weight dominates (or a small number of vertexes) over the others, we may have an heterogenous case. This measure is the disparity [17] of vertex i and can be measured as:

$$D(i) = \sum_{j \in v(i)} \frac{w_{ij}}{s_i}^2, \tag{7}$$

where $v(i)$ is the set of neighbors of i.

The weighted clustering coefficient [20] can be measured as:

$$c_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{i,j} + w_{i,h})}{2} a_{ij} a_{ih} a_{jh}, \tag{8}$$

where s_i is the vertex strength, k_i is the degree of vertex i. This coefficient is within the $(0, 1)$ interval. This measures considers not just the number of closed triplets in the neighborhood of a node but also their total relative weight with respect to the strength of the node.

We calculate the entropy [21] of the network, which is given by:

$$H = -\sum_{k} P(k) \log(P(k)). \tag{9}$$

The maximum value of entropy is given for a uniform degree distribution, whereas the minimum is zero (all vertices have the same degree)[4]. This measure provides the average heterogeneity of the network.

[4] The relative weights of the distance matrix are employed to calculate the probabilities.

3 Data

We evaluate 20 commodities using daily recorded database from January 2, 1991 to February 8, 2008. We have obtained 4463 observations for commodities spot prices, denominated in US dollars[5]. We build the networks using correlation distances between the spot prices of commodities in the sample following Mantegna [2]. We study the commodities in the networks built using the previous MST, based on the concept of ultrametricity.

We also study the dynamics of the evolution of network measures over time. We employ a recursive approach in which we use the first 1008 observations to estimate closeness centrality, disparity, entropy, clustering and dominance. In order to analyze the data we average these measures by sector: metals, agriculture and energy to compare the sectors relative importance within the network.

4 Empirical results

Networks analysis are useful to provide a representation of a broad variety of complex systems. The networks property of hierarchy is useful to observe that the networks often have structure in which vertices cluster together into groups that then join to form groups of groups, from the lowest levels of organization up to the level of the entire network.

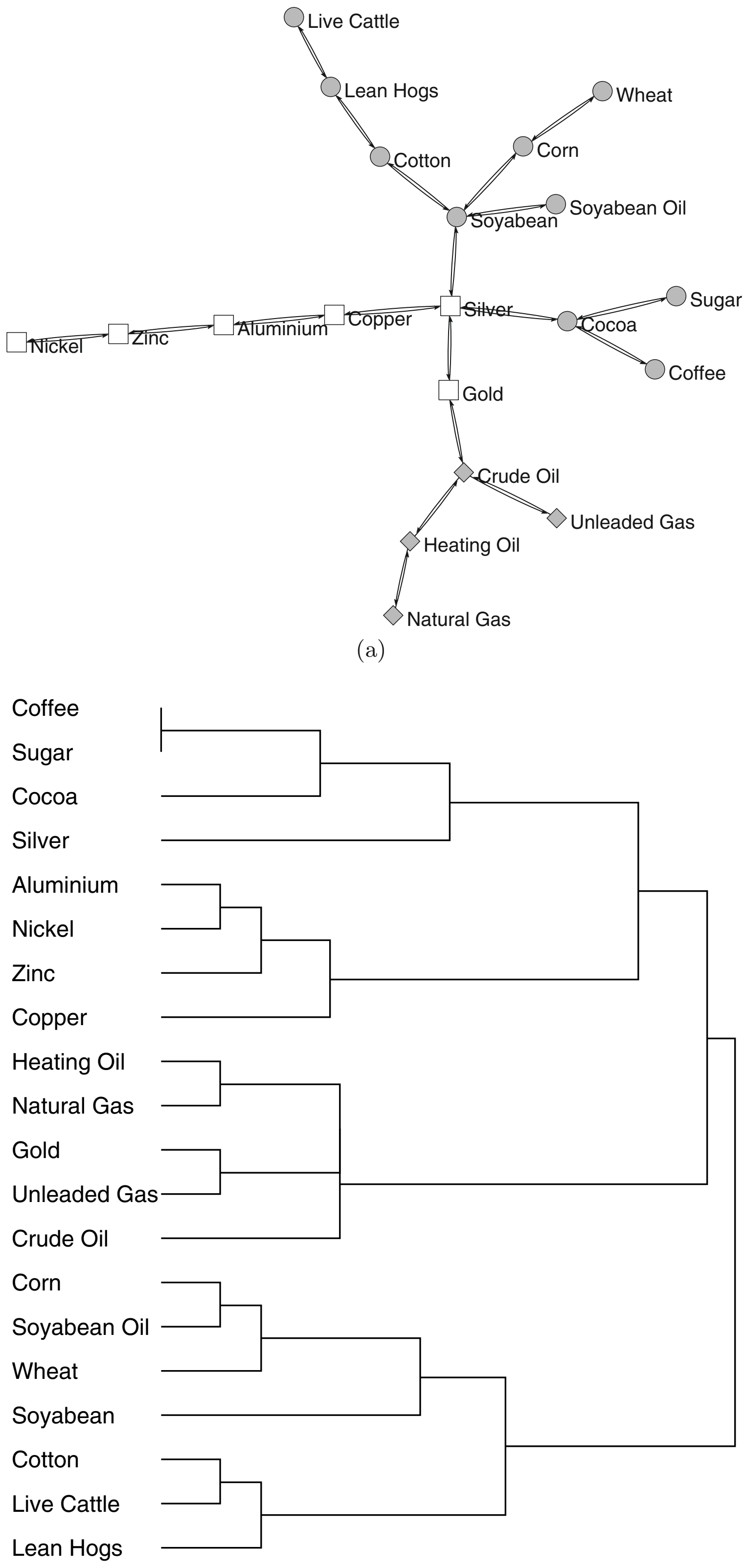
The commodities or groups of commodities with high values of the distances $d_{i,j}$ are influenced by factors, which are specific to these commodities. For low values of $d_{i,j}$, the commodities are affected either by factors which are common to all commodities and by factors which are specific to the considered group of commodities. The length of the segments observed for each group, quantifies the relative relevance of these factors.

The patterns that we can observe are that, if the distances of $d_{i,j}$ are low, the groups of commodities tend to form strong clusters and for high values of $d_{i,j}$ these links tend to be weak. The commodities sector has groups of commodities strongly clustered. This occurs because commodities with tendency to suffer more influence of external risks tend to create clusters more complete and robust.

For our full sample we identify the values of $d_{i,j}$ between 0 and 1.41 indicating that the commodities sector forms strong clusters and is homogeneous for all characteristics. The division of the full sample in two parts of equal size, does not affect the analysis of the network. The division give us the values of $d_{i,j}$ varying from 0 to 1.43, for the first set, and $d_{i,j}$ varying from 0 to 1.42, for the second set of data. We observe that the values does not change significantly, indicating that the patterns and characteristics of the full sample continue for the divided sample.

Figure 1 presents the MST and the taxonomy Hierarchical tree of the subdominant ultrametric associated to the MST, for a network based on spot prices of

[5] In order to check whether results depend on the specific sample the 4463 observations were divided in two samples, with 2231 for the first set and 2232 for the second.



(b)

Fig. 1. Plot of the MST of a network based on spot prices of commodities distances topology, for the full sample. The node shape is based on commodities sector. The correspondence is: ellipse for Agriculture, Square for Metals and Diamond for Energy, (b) plot of the Taxonomy Hierarchical tree of the subdominant ultrametric associated to the MST of spot prices of commodities distances. The correspondence is based on commodities names.

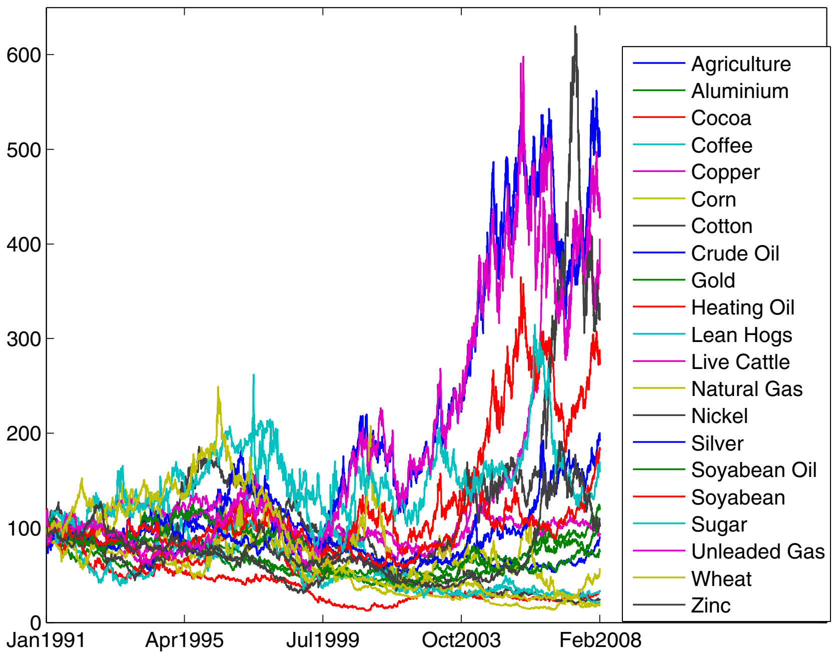


Fig. 2. (Color online) Plot of the evolution of commodity indices for the period from January 2, 1991 to February 8, 2008.

commodities distances of the full sample (January 2, 1991 to February 8, 2008). We present the network based on commodities names[6].

Figure 2 presents the evolution of commodity indices for the period from January 1991 to February 2008.

The MST and taxonomy tree presented in Figure 1 suggest the emergence of clusters and we can see that these commodities are clustered by sector and independent of the time width. The metal commodities form homogeneous cluster, such as the agriculture and energy commodities.

Figures 3 and 4 presents the evolution of dominance, clustering, disparity, closeness centrality and entropy for the commodities network. Results for the dominance (β) measure suggests that the commodities from the agriculture sector are the most important in the network followed by metals and energy, respectively. Dominance measures fall within the $(0.8404-1.1932)$, $(0.9526-1.4763)$, and $(0.4991-0.8542)$ range for the metals, agriculture and energy sectors, respectively. The dynamics of the series are different for each sector.

The average weighted clustering coefficients fall within the $(0.1180-0.3291)$, $(0.1190-0.3632)$, and $(0.1350-0.3501)$ range for the metals, agriculture and energy sectors, respectively. In the beginning of the sample the difference between sectors was small, but is has increased substantially in the recent years. An interesting feature is that these sector share similar dynamics over time.

The average disparity measures fall within the $(0.0559-0.0873)$, $(0.0560-0.0701)$, and $(0.0626-0.1025)$ range for the metals, agriculture and energy sectors, respectively. These sector display very different dynamics with the energy sector showing the highest disparity over most of the sample. In the end of the sample the metals sector shows a substantial increase in disparity.

The average closeness centrality fall within the $(0.0367-0.0582)$, $(0.0369-0.0657)$, and $(0.0376-0.0642)$ range for the metals, agriculture and energy sectors, respectively. These sectors display different dynamics over time and the energy sector has the highest centrality in the beginning of the sample, whereas the agriculture sector has the highest centrality for the middle of the sample. In the end of the sample the metals sector shows the highest closeness centrality.

Finally, the entropy measures imply an increase in network heterogeneity in the 1999–2001 period, with a subsequent fall in the 2002–2004 period. These changes were followed by similar changes, which seem to be cyclical in the last part of the sample. Therefore, the degree of heterogeneity seems to be fluctuating over time.

Since 1973 oil crisis, oil and energy have become more volatile. Regnier [22] points out that oil prices are highly volatile (93%) compared with all products manufactured in the US since 1986. Specifically, crude oil prices are 65% more volatile than other products. However, recently, commodities prices in general have become more volatile. Prices have been through a sudden rose since the beginning of 2004 and have lasted for several years. Disparity figures for the energy sector seems to be fluctuating widely if compared to other sectors, which is in line with these changes in volatility[7].

[6] We also split the sample in two and plot both the MST and the taxonomy Hierarchical tree of the subdominant ultrametric associated to the MST and find qualitatively the same results.

[7] Several authors have documented these fluctuations (see Radetzki [23], Akram [24],Wei and Zhu [25]).

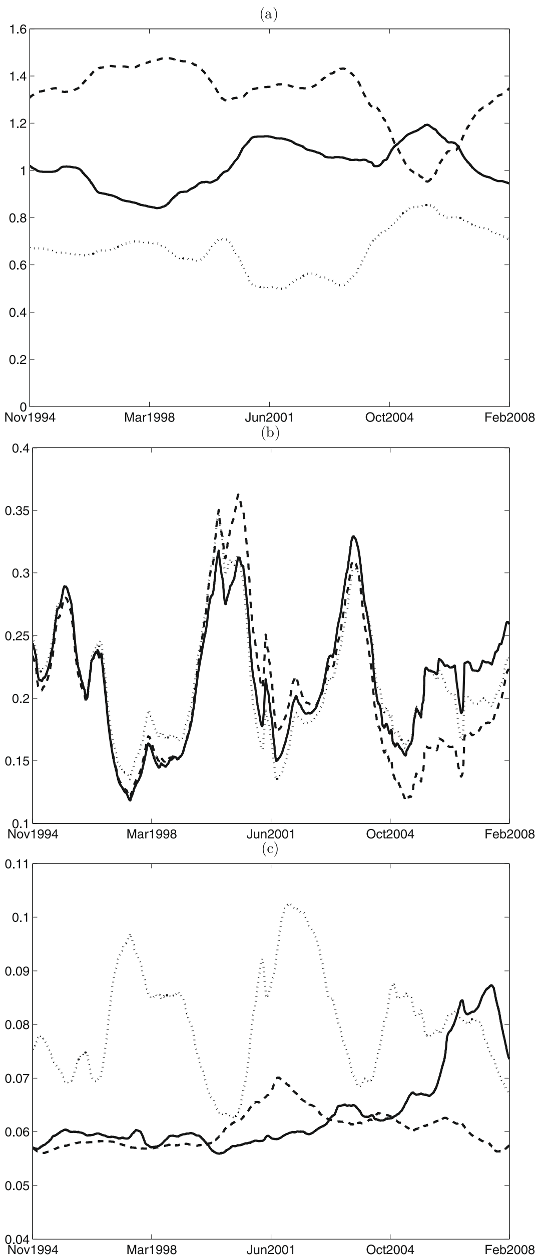


Fig. 3. Plot of the average (a) dominance, (b) clustering coefficient and (c) disparity for the Metals (solid line), Agriculture (dashed line) and Energy (dotted line) sectors. We employ 1008 observations (approximately four years of data) for each calculation and estimate these measures recursively on a moving average basis.

In general spikes in the dynamics of clustering coefficients, disparity, closeness centrality and entropy occur approximately in 2001, 2004 and 2008, which have been years in which commodities have suffered substantial shocks due to the war in Iraq and Afghanistan and the recent subprime crisis. In the latter, the failure of major banks in the US and Europe have provoked a worldwide confidence crisis, which have had important implications on commodity prices.

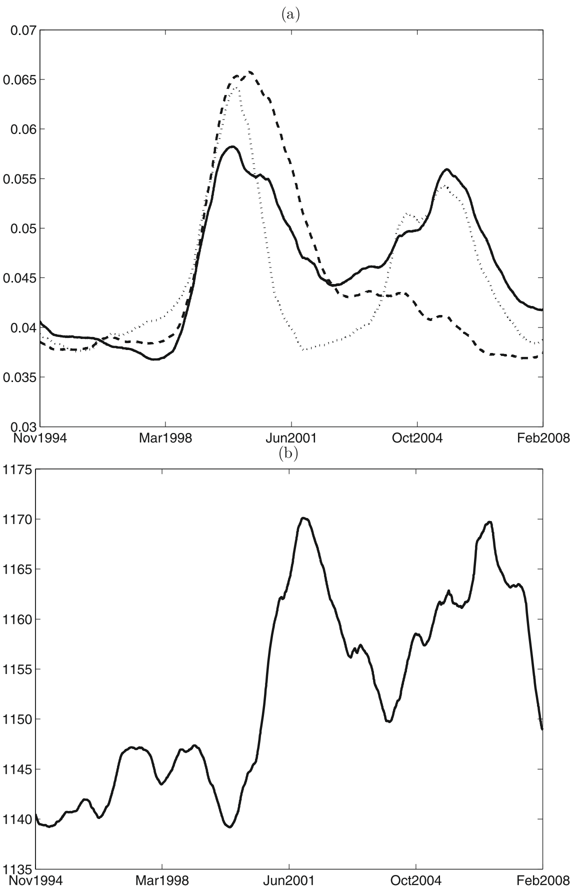


Fig. 4. (a) Plot of the average closeness centrality for the Metals (solid line), Agriculture (dashed line) and Energy (dotted line) sectors and (b) plot of the entropy for the entire network over time. We employ 1008 observations (approximately four years of data) for each calculation and estimate these measures recursively on a moving average basis.

We also compare the measures associated with the commodities (correlation) network to the measures associated to the randomized versions of this network. Since the metric distance d_{ij} of this network belongs to the interval $[0, 2]$, we have built random networks that satisfy this characteristic. In one version of the randomized network, d_{ij} has uniform distribution in the interval $[0, 2]$. In the second version version of the randomized network, d_{ij} has a normal distribution with the same mean and the same variance of the original network truncated in 0 and 2. In both versions of the randomized network, we found that the disparity, entropy, strength and domination power measures are higher than the original network.

5 Conclusions

This paper shows that the MST and the ultrametric hierarchical tree can be used to analyze the commodities sector. The evidence suggests that the commodities sector forms clusters and is homogeneous with respect to the sector. In other words, if the price of a commodity raise, the group of this commodity tends to raise too. The results also provide an indication that the sector is vulnerable to the influence of external risks and tends to create complete and robust clusters. We found that the time width does not influence the analysis of the clusterization of the commodities sector and as well as the full sample, the sets of commodities forms clusters and are homogeneous by sector.

We also develop a dynamic analysis of the commodities network using weighted network measures: closeness centrality, disparity, entropy, dominance and clustering. Empirical results suggests that agriculture commodities are very important in the network, followed by metals and energy. Furthermore, the parameters that characterize the network seem to be changing over time. Our results suggest that the analysis of networks should analyze also the temporal dimension.

The authors wish to thank anonymous reviewers that have helped improve the paper. Benjamin M. Tabak and Daniel O. Cajueiro are grateful to the Brazilian agency CNPQ for financial support. The opinions expressed in this paper are those of the authors and do not necessarily represent those of the Central Bank of Brazil.

References

1. R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999)
2. R.N. Mantegna, Eur. Phys. J. B **11**, 193 (1999)
3. G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Phys. Rev. E **68**, 046130 (2003)
4. G. Bonanno, N. Vandewalle, R.N. Mantegna, Phys. Rev. E **62**, R7615 (2000)
5. J.P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Phys. Rev. E **68**, 056110 (2003)
6. J.P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, Physica A **324**, 247 (2003)
7. W.S. Jung, S. Chae, J.S. Yang, H.T. Moon, Physica A **361**, 263 (2006)
8. W.S. Jung, O. Kwon, F. Wang, T. Kaizoji, H.T. Moon, H.E. Stanley, Physica A **387**, 537 (2008)
9. D. Lien, L. Yang, Journal of Banking & Finance **32**, 187198 (2008)
10. K. Matia, Y. Ashkenazy, H.E. Stanley, Europhysics Letters **61**, 422 (2003)
11. D.O. Cajueiro, B.M. Tabak, Physica A **387**, 6825 (2008)
12. B.M. Tabak, T.R. Serra, D.O. Cajueiro, Physica A **388**, 1137 (2009)
13. P. Sieczka, J.A. Holyst, Physica A: Statistical Mechanics and its Applications **388**, 1621 (2009)
14. J. Gomez-Gardenes, V. Latora, Phys. Rev. E **78**, 065102 (2008)
15. D. Garlaschelli, T.D. Matteo, T. Aste, G. Caldarelli, M. Loffredo, Eur. Phys. J. B **57**, 159 (2007)
16. S. Arianos, A. Carbone, Journal of Statistics Mechanics – Theory and Experiment **03037** (2009)
17. M. Barthlemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Physica A **346**, 34 (2005)
18. R. van den Brink, R.P. Gilles, Social Networks **22**, 141 (2000)
19. G. Sabidussi, Psychometrika **31**, 581 (1966)
20. J. Saramki, M. Kivel, J.P. Onnela, K. Kaski, J. Kertsz, Phys. Rev. E **75**, 027105 (2007)
21. L.D. Costa, F.A. Rodrigues, G. Travieso, P.R.V. Boas, `[arXiv:cond-mat 0505185]`, 1 (2005)
22. E. Regnier, Energy Economics **29**, 405 (2007)
23. M. Radetzki, Resources Policy **31**, 56 (2006)
24. Q. Akram, Working paper, Norges Bank (2008)
25. S. Wei, Z. Zhu, Energy Economics **28**, 523 (2006)