MAPPING THE BRAZILIAN STOCK MARKET: INSIGHTS FROM COMPLEX NETWORKS ANALYSIS

Leandro Coghi Bernardelli[1] Carlos Enrique Carrasco-Gutierrez[2] Danilo Gomes Vieira[3]

Abstract: This study analyzed the putative partial correlation structure and clustering patterns of the Brazilian stock market, represented by the stocks composing the Ibovespa index, through Dynamic Bayesian Networks (DBNs) over the period from January 2, 2007, to December 31, 2024. This network was used to calculate distance measures between stocks and to investigate the network topology through mixed and global complex network metrics. The results show that the Brazilian stock market predominantly exhibits positive partial correlations, suggesting joint movements in the long term, although this dynamic may vary in the short term. The network core is relatively dense and composed of several stocks, including those with the largest capitalization and liquidity in the index. These nodes occupy strategic positions and have a higher potential to propagate financial shocks. Authority points are concentrated in companies focused on the domestic market, favored by the growth of high-end consumption over the last decade. On the other hand, hubs mostly consist of smaller capitalization and less liquid companies, also focused on the domestic market.

Keywords: Complex Networks, Stock Markets, Dynamic Bayesian Networks, Topology, Complex Network Measures.
JEL classification: C11; C45; E44; G1.

1. INTRODUCTION

Complex networks are a recent analytical approach through which complex systems can be examined more thoroughly, enabling the study of diverse and intricate systems in ways that classical graph theory cannot accommodate.

The stock market is a typical complex system in which stock price fluctuations are not independent but highly interconnected, exhibiting strong correlations with the economic sectors and industries to which the stocks belong. Therefore, cross-correlations among stocks are particularly important for understanding interactive mechanisms and for optimizing portfolio choices that investors must make in response to the prevailing economic environment. Wang, Xie, and Stanley (2018) categorized correlation-based networks into three main types: i) Pearson correlation networks (Boginski et al., 2005; Bonanno et al., 2003; Mantegna, 1999; J.-

[1] Graduate Program in Economics at the Catholic University of Brasília (PPGE/UCB) and Caixa Econômica Federal, Brasília, DF, Brazil. E-mail address: bernardellileandro@gmail.com.
[2] Graduate Program in Economics at the Catholic University of Brasília (PPGE/UCB) and Graduate Program in Public Policy at the Catholic University of Brasília (MPPP/UCB), Brasília, DF, Brazil. E-mail address: carlos.carrasco.gutierrez@gmail.com.
[3] Graduate Program in Economics at the Brazilian Institute of Education, Development and Research (MPE /IDP), Brasília and Caixa Econômica Federal, Brasília, DF, Brazil. E-mail address: vieira.danilo.gomes@gmail.com.

P. Onnela et al., 2003; Tabak et al., 2010; Tumminello et al., 2005; Yang et al., 2013); ii) networks constructed with other correlation methods (Brida & Risso, 2010; Cao et al., 2017; Korkusuz et al., 2022; Sensoy & Tabak, 2014; G. J. Wang et al., 2012); and iii) networks based on partial correlation (Ballester et al., 2023; Kenett et al., 2010, 2015; Millington & Niranjan, 2020; So et al., 2021). Additionally, stochastic processes and random matrix theory have also been introduced in other studies on complex networks in stock markets (Heimo et al., 2008, 2009; Ma et al., 2004; Ren & Zhou, 2014).

Brazil hosts the largest stock market in Latin America and, as the region's leading economy, attracts the attention of global investors seeking portfolio diversification. In addition, the country's relative economic and political stability is a significant factor in the selection of emerging markets with investment potential. The financial, agribusiness, and mineral/energy sectors stand out for their high economic efficiency. With a consumer market of over 215 million people, Brazil combines attributes that make it particularly attractive both to international investors aiming to expand their exposure to equities on a global scale and to domestic investors seeking to broaden their portfolio diversification.

In this study, we analyze the structure of putative partial correlations and clustering patterns of the Brazilian stock market, represented by the stocks comprising the Ibovespa[4] index, through Dynamic Bayesian Networks (DBN) over the period from 01/01/2007 to 12/31/2024. This network was used to compute distance measures between stocks and to investigate the network's topology by means of mixed and global complex network metrics. We adopt the approach of Opgen-Rhein and Strimmer (2007b), which enhances James–Stein shrinkage in the VAR process, and construct the networks based on the inclusion of edges (partial correlations) ordered by decreasing coefficients. To control the false positive rate, we apply the multiple testing correction method (q-value) proposed by Schäfer and Strimmer (2005).

The main empirical findings are as follows: 1) The results show that the Brazilian stock market predominantly exhibits positive partial correlations, suggesting joint long-term movements, although this dynamic may vary in the short term. Additionally, some negative partial correlations were identified, which may provide hedging opportunities associated with differences in firms' exposure to domestic or international markets and sectoral sensitivity to domestic interest rates. 2) The network analysis revealed that authority nodes are concentrated in companies focused on the domestic market, which have benefited from the growth of high-

[4] The Ibovespa is the main performance indicator of stocks traded on B3, comprising the most important companies in the Brazilian capital market. Created in 1968, it has, over the past five decades, established itself as a benchmark for investors around the world.

end consumption over the past decade. In contrast, the hubs are mostly smaller-cap and less liquid companies, also focused on the domestic market, which can be tactically used by investors seeking gains during periods of heightened volatility. 3) The core of the financial network, identified through the coreness measure, is relatively dense and comprises several stocks, including those with larger market capitalization and liquidity within the index. These nodes occupy strategic positions and have greater potential to transmit financial shocks. This subset of assets forms a more cohesive structure — a feature that contrasts with the overall pattern of the network, which remains sparse.

This study offers several relevant contributions to the literature on financial networks. First, it advances the frontier of knowledge by employing putative partial correlations to model interactions among assets, enabling the identification of direct relationships between stocks in the Brazilian equity market over an extensive period of 17 years (01/01/2007 to 12/31/2024). Second, we propose the use of a distance metric based on partial correlations, a tool that proves effective for representing and interpreting structures within complex networks. Third, we analyze the topology of the stock network through 11 metrics drawn from the complex networks literature, covering both mixed and global measures. Finally, we apply community detection algorithms to understand the dynamics of clustering among Brazilian stocks, revealing potential latent sectoral or behavioral structures in the market.

In the literature, the seminal work by Mantegna (1999) introduced the study of network topology using the Minimum Spanning Tree (MST), obtaining a hierarchical arrangement of stocks in the U.S. equity market. Bonanno, Vandewalle, and Mantegna (2000) applied the MST to the global financial market, identifying relevant taxonomic structures given appropriate time horizons and reference currencies. J. P. Onnela, Kaski, and Kertész (2004) and J.-P. Onnela et al. (2003) applied the MST method to 477 NYSE stocks and reported that the asset tree exhibited a pronounced clustering feature, in addition to the degree distribution following a power law. Thus, we observe that these seminal studies applied the Minimum Spanning Tree (MST) methodology.

Other authors have applied alternative correlation-based methods. Pozzi, Di Matteo, and Aste (2013) found that allocating investments to stocks from peripheral regions, characterized by sparse connections in filtered financial networks based on MST and PMFG (Planar Maximally Filtered Graphs), effectively reduces allocation risk. MST and PMFG networks include few edges, which can reduce complexity and facilitate analysis, but inevitably lead to the loss of structural information about internal correlations. Sensoy and Tabak (2014) used Dynamic Spanning Trees (DST) based on ARMA-FIEGARCH-cDCC models to study Asian assets, whereas Cao, Shi, and Li (2017) applied the Spearman correlation coefficient to rank connections within moving time windows.

Partial correlation networks are also used in the literature. Kenett et al. (2010) introduced the maximally filtered planar graph of partial correlation (PCPG), while Kenett et al. (2015) statistically validated significant links, revealing important influence relationships. So, Chan, and Chu (2021) constructed dynamic networks based on partial correlations above 0.5. Ballester, López, and Pavía (2023) applied Dynamic Bayesian Networks (DBN) to model systemic credit risk in Europe.

Finally, when we examine the literature on stock market networks in Brazil, we find the seminal work of Tabak, Serra, and Cajueiro (2010), who studied the topology of a network constructed using the Minimum Spanning Tree (MST) method, based on the concept of ultrametricity. Their analysis used the correlation matrix of weekly closing prices for 47 stocks over the period from January 7, 2000, to February 29, 2008. Magno et al. (2019) analyzed the topology of the stock network of the Ibovespa index from different perspectives, such as vertex centrality, community organization, robustness, and distribution of influence, covering the period from August to December 2017. More recently, de Pontes and Rêgo (2022) investigated the relationship between the stock market and macroeconomic variables by employing the concept of complex networks, in which they constructed 43 monthly networks based on Pearson correlations of the logarithmic returns of individual assets.

In Section 2, we detail the theoretical framework used for the estimation of Dynamic Bayesian Networks, as well as the complex network measures employed. In Section 3, we examine the estimated networks and the calculated complex network measures. Finally, in Section 4, we present the conclusions found throughout this study.

2. METHODOLOGY

2.1. Primary Concept of Bayesian Network (BN)

A Bayesian Network (BN) is a graphical structure that allows us to represent an uncertain domain and reason about it. A BN consists of nodes and edges (arcs or links). The nodes represent a set of random variables within the domain. The set of edges (arcs or links) connects pairs of nodes, representing direct dependencies between the variables. Assuming the variables are discrete, the strength of the relationship between variables is quantified by conditional probability distributions associated with each node. The only restriction on allowed edges in a BN is that there should be no directed cycle, meaning you cannot return to a node simply by following directed edges. Therefore, these networks are called Directed Acyclic Graphs (DAGs) (Korb & Nicholson, 2004).

Let $G = (V, E)$ be a directed acyclic graph (DAG), where V is a finite set of nodes, and E is a finite set of directed edges between the nodes. Each node $v \in V$ in this graph represents a random variable X_v and comprises the set of variables in G. Given any pair of nodes X and $Y \in V$, if there is a directed edge from X to Y, X is called the parent node of Y. For each parent node of v, the notation $pa(v)$ is adopted. Additionally, a conditional probability function between v and $pa(v)$ is defined as $p(x_v|x_{pa(v)})$. The set of relational probability functions in the network is denoted as P. A BN for a given set of random variables is the pair (G, P).

It is important to highlight that the theory of Bayesian Networks (BN) includes the concept of *d-separation*, which guarantees the conditional and directional independence of a set of random variables (Carvalho & Chiann, 2013). The Markov condition implies that all *d-separations* are conditional independences, and all conditional independences implied by the Markov condition are identified by *d-separation*. In other words, if $(\mathbb{G}, P)$ satisfies the Markov condition, every *d-separation* in $\mathbb{G}$ is a conditional independence in P. Furthermore, every conditional independence that is common to all probability distributions satisfying the Markov condition with DAG $\mathbb{G}$ is identified by *d-separation* (Korb & Nicholson, 2004; Neapolitan, 2004).

2.2. Bayesian Network Learning

Analyzing from a statistical perspective, learning a network represents the estimation of model parameters based on a certain criterion and having knowledge of a specific dataset (Carvalho & Chiann, 2013).

The Bayesian approach is used to estimate parameters within the network, aiming to encode uncertainty about the parameters θ into a $prior$ distribution $p(\theta)$ and applying the data $\bar{d}$ (through the likelihood function). By applying Bayes' theorem, the uncertainty is updated with the $posterior$ distribution $p(\theta|\bar{d})$, as per Equation 1:

$$p(\theta|\bar{d}) = \frac{f(\bar{d}|\theta)h(\theta)}{\int_{\theta\in\Theta} f(\bar{d}|\theta)h(\theta)d\theta}, \theta \in \Theta, \tag{1}$$

given Θ as the parameter space, $\bar{d}$ as the random sample from the probability distribution $p(x|\theta)$, and $p(\bar{d}|\theta)$ as the likelihood function.

2.3. Concept of Dynamic Bayesian Network (DBN)

Dynamic Bayesian Network (DBN) is a temporal extension extensively used to model temporal relationships among variables over different time periods (Murphy, 2002). The unfolding of an interaction graph over time is highly beneficial for accommodating potential loops and feedbacks in the network's topology, as required by definition for a Bayesian Network (Nagarajan et al., 2013).

In a DBN, it is assumed that the dependency relationships are represented by a Vector Autoregressive Process (VAR), where it is important that the time series are stationary. In a VAR(p) process of order p, the observed variables at any time $t \geq p$ satisfy Equation 2:

$$X(t) = A_1 X(t-1) + \cdots + A_i X(t-i) + \cdots + A_p X(t-p) + B + \varepsilon(t). \tag{2}$$

Where:

$X(t) = \big(X_i(t)\big), i = 1, \ldots, k,$ is the vector of k observed variables at time t;

$A_i, i = 1, \ldots, p$ are coefficient matrices of size $k\ x\ k$;

B is a vector of size k that represents the baseline measurement for each variable, and is invariant as a function of time;

$\varepsilon(t)$ is a vector of white noise of size k, with zero mean $E\big(\varepsilon(t)\big) = 0$ and time-invariant positive definite covariance matrix $Cov\big(\varepsilon(t)\big) = \sum$.

Thus, if we assume a VAR(p) of order 1, as in Equation 3:

$$X(t) = AX(t-1) + B + \varepsilon(t), \quad com \quad \varepsilon(t) \sim N(0, \sum), \tag{3}$$

the edges are defined between two successive time periods, and this set of edges is determined by all nonzero coefficients in matrix A (see Equation 4). If an element $a_{ij}, i \neq j$ is non-zero, then the network includes an edge from $X_i(t-1)$ to $X_j(t)$. Additionally, we assume that the error term for each variable X_i is independent of the other variables and their respective error terms, meaning the elements off the diagonal in $\sum$ can be set to 0.

A VAR(1) process, with a diagonal error covariance matrix $\sum$, can be represented as a Dynamic Bayesian Network, where the edges of the network correspond to the non-zero elements of the coefficient matrix A. Specifically, the elements a_{ij} of matrix A indicate the non-zero coefficients, which define the dependencies between the variables. This structure helps visualize the direct relationships between the variables over time, with the diagonal elements of $\sum$ representing the variances of the error terms for each variable:

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{bmatrix}. \tag{4}$$

Thus, we can represent the Dynamic Bayesian Network as demonstrated in Appendix A, Figure 6.

2.4. Dynamic Bayesian Network Learning Algorithms

Classical ordinary least squares (OLS) estimates of the regression coefficients A and B from Equation 2 or 3, in the case of multiple individuals, can only be calculated when $n \gg k$, ensuring that the sample covariance matrix has full rank. For real-world data, regularized estimators are necessary in almost all cases.

According to Nagarajan, Scutari, and Lèbre (2013), regularized estimators can be divided into four approaches:

The Least Absolute Shrinkage and Selection Operator or LASSO (Tibshirani, 1996): the estimation procedure tends to produce some coefficients exactly equal to zero by applying an L_1-norm penalty to the sum of the coefficients. Variable selection thus becomes straightforward: only the non-zero coefficients define significant dependency relationships.

James–Stein Shrinkage Approach: An efficient estimator of the covariance matrix can be obtained by "shrinking" the empirical correlation coefficients toward zero and the empirical variances toward their median. Stein (1956) and James and Stein (1961) demonstrated that the resulting correlation matrix dominates the empirical one, meaning its mean squared error is never worse than that of the empirical correlation matrix. Opgen-Rhein and Strimmer (2007b) proposed an application of the James–Stein shrinkage approach to the VAR process, achieving better results than many classical approaches. These improved estimates of the regression coefficients, which are essentially a function of the covariance matrix of X, are attributed to this method. The network structure is determined by the inclusion of edges in descending order of the coefficients. To control the False Discovery Rate (FDR), the local FDR approach suggested by Schäfer and Strimmer (2005) can be employed for multiple testing correction (see Appendix, B). Finally, in static Bayesian networks, James–Stein shrinkage is used to compute regularized partial correlations and conditional probabilities, which are employed in conditional independence tests.

First-Order Conditional Dependencies Approximation: This approach is based on first-order conditional dependencies. The central concept of this method is the low-order conditional dependency graph, which originated within the context of graphical modeling theory with directed acyclic graphs. The directed acyclic graph defining the dynamic Bayesian network is approximated by first-order conditional dependencies. Lèbre (2009) proposes the G1DBN algorithm, which implements dynamic Bayesian network learning through a two-step procedure. First, it learns a directed acyclic graph that encodes first-order partial dependency relationships. Then, it infers the actual structure of the dynamic Bayesian network using the graph from the previous step.

Modular Networks: Chiquet et al. (2009) developed learning algorithms based on LASSO[5], adding a clustering effect for multiple data sets. The algorithm was named SIMoNe (Statistical Inference for MOdular Networks) and applies a score-based approach, seeking a latent clustering of the network to guide edge selection through an adaptive L_1 likelihood penalty, particularly for VAR(1) processes.

In our work, we apply the approach of Opgen-Rhein and Strimmer (2007b), which achieved better results for the VAR process by reducing empirical correlation coefficients compared to classical models. For determining the inclusion of network edges, we apply the local FDR approach suggested by Schäfer and Strimmer (2005) for multiple testing correction. Thus, all estimations of the Dynamic Bayesian Networks were carried out using the GeneNet package in RStudio. In the next subsection, we present the theoretical rationale of the GeneNet package for network estimation.

2.4.1. GeneNet Package in R

The GeneNet estimation package in R implements a linear shrinkage estimator for a covariance matrix and the selection of a Gaussian Graphical Model (GGM) based on partial correlation obtained from the shrinkage estimator suggested by Opgen-Rhein and Strimmer (2007b). The package applies multiple testing using the local false discovery rate (fdr) approach through the method of Schäfer and Strimmer (2005), where the GGM selection controls the false discovery rate at a predetermined α level. Specifically, one of the most commonly used linear shrinkage estimators, S^*, for the covariance matrix $\sum$ is:

$$S^* = \lambda^* T + (1 - \lambda^*)S, \tag{5}$$

where $S = \left(s_{ij}\right)_{1 \le ij \le p}$ is the sample covariance matrix, $T = diag\left(s_{11}, s_{22}, \dots, s_{pp}\right)$ is the target shrinkage matrix, and $\lambda^* = \frac{\sum_{i \neq j} \widehat{Var}(s_{ij})}{\left(\sum_{i \neq j} s_{ij}^2\right)}$ is the optimal shrinkage intensity.

This estimator S^*, the partial correlation matrix $P = \left(\hat{\rho}^{ij}\right)_{1 \le i,j \le p}$, is defined as:

$$\tilde{r} = \hat{\rho}^{ij} = -\frac{\widehat{\omega}_{ij}}{\sqrt{\widehat{\omega}_{ii}\widehat{\omega}_{jj}}}, \tag{6}$$

where $\widehat{\Omega} = \left(\widehat{\omega}_{ij}\right)_{1 \le i,j \le p} = (S^*)^{-1}$.

[5]Least Absolute Shrinkage and Selection Operator.

To identify the significant edges, we apply the approach of Schäfer and Strimmer (2005), which assumes the distribution of partial correlations as a mixture:

$$f(\tilde{r}) \ = \ \eta_0 f_0(\tilde{r}; k) + (1 \ - \ \eta_0) f_A(\tilde{r}), \tag{7}$$

where f_0 is the null distribution, f_A is the alternative distribution corresponding to the true edges, and n_0 is the unknown mixture parameter. Using the algorithm developed by Opgen-Rhein and Strimmer (2007a), the GeneNet package identifies the significant edges that have the local false positive rate:

$$\mathrm{Prob}(\text{null edge}|\tilde{r}) = \mathrm{fdr}(\tilde{r}) = \frac{\hat{\eta}_0 f_0\left(\tilde{r};\ \hat{k}\right)}{\hat{f}(\tilde{r})}, \tag{8}$$

smaller than the pre-determined level α, where $|\tilde{r}|\mathrm{Be}\left(\tilde{r}^2; \frac{1}{2}, \frac{k-1}{2}\right)$, and $\mathrm{Be}(x; a, b)$ is the Beta distribution density, with k being the reciprocal variance of the null $\tilde{r}$.

For further details on the approach used, including the local false discovery rate (FDR) and the significance measure q-value (α), see Appendix B and C.

2.5. Application of Complex Network Measures

Once the estimations of Dynamic Bayesian Networks are performed, we use the statistically significant partial correlations to calculate the distance measures between the financial assets. These distance measures based on partial correlation can be useful for complex networks where the focus is on the direct relationships between financial assets, eliminating the effect of other financial assets. Thus, it can provide a more accurate view of the 'direct' connections between variables or elements. In Equation 9, we present the calculation of the distance:

$$d_{ij}^{partial} \ = \ \sqrt{2\left(1 - \tilde{r}_{ij}\right)}\ , \tag{9}$$

where $\tilde{r}_{ij}$ is the partial correlation calculated by the GeneNet package (see section 2.4.1.).

We apply the toolbox of complex network measures to investigate how large sets of dynamic systems, interacting through complex wiring topology, can behave collectively (Silva & Zhao, 2016).

In this work, we will categorize our complex network measures following the classification proposed by Silva and Zhao (2016) , which categorizes these measures into three categories:

1st- <u>Strictly Local Measures</u>: they use only information from the vertex itself to be computed, meaning they are always vertex-level measures;

2nd- <u>Mixed Measures</u>: they utilize strictly local information and topological information from their direct and indirect neighborhoods. The additional information can range from a topology that is simply quasi-local, such as the number of triangles in the neighborhood, to long-range information like the shortest path between the two farthest pairs of vertices. These mixed measures are always vertex-level measures;

3rd- <u>Global Measures</u>: they utilize the entire structure of the network to be computed. Global measures are always network-level measures.

The measures considered in this study include both mixed and global metrics, which will serve as tools to understand the dynamics of the structure of the Brazilian stock market network over the past 17 years.

2.5.1. Mixed Measures

i. *Authority Score*: It is the main eigenvector of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, where **A** is the adjacency matrix of the graph. The measure is known to be the first approach to centrality proposed by Jon Kleinberg. An authority is a vertex pointed to by many important hubs - such as a significant research article or authoritative figure on a particular subject.

ii. *Hub Score*: It is the main eigenvector of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$, where **A** is the adjacency matrix of the graph. The measure is known to be the second approach to centrality proposed by Jon Kleinberg. A hub is a vertex that points to many important authorities - such as a review article in a network of scientific citations.

iii. *Coreness (K-core)*: It is the maximum subgraph in which each vertex has at least a degree of k. The coreness of a vertex is k if it belongs to the k-core but not to the (k + 1)-core.

iv. *Transitivity (Clustering Coefficient)*: It measures the probability that adjacent vertices of a vertex are connected. We apply the Global Transitivity, which is the ratio of the count of triangles to the count of connected triples in the graph. We use the definition by A. Barrat:

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{ih} a_{jh}, \tag{10}$$

where s_i is the strength of vertex i, a_{ij} are elements of the adjacency matrix, k_i is the degree of vertex i, and w_{ij} are the weights.

2.5.2. Global Measures

i. *Diameter*: The diameter of $\mathcal{G}$, T, is the length of the largest pairwise distance in $\mathcal{G}$. Formally, it is given by:

$$T = \max_{u,v \in \mathcal{V}} d_{uv}. \tag{11}$$

The diameter can be interpreted as the longest chain of intermediation in the network.

ii. *Mean Distance*: It calculates the average path length in a graph by computing the shortest paths between all pairs of vertices (in both directions for directed graphs, as in our case).

iii. *Modularity*: measures how good the division is, or how separated different types of vertices are from each other. In other words, the main idea of modularity is to calculate the fraction of edges that fall within given groups minus the expected value if the edges were randomly distributed. For a given division of the network's vertices into some modules, modularity reflects the concentration of vertices within the modules compared to a random distribution of links among all vertices, regardless of the modules. In Equation 12, we present the formula for calculating modularity for directed graphs as presented in this work:

$$Q = \frac{1}{m} \sum_{i,j} \left(\mathbf{A}_{ij} - \gamma \frac{k_i^{out} \, k_j^{in}}{m} \right) \delta(c_i, c_j), \tag{12}$$

where m is the number of edges, $\mathbf{A}_{ij}$ is the element of the adjacency matrix in row i and column j, $k_i^{out}$is the out-degree of i, k_j^{in} is the in-degree of j, c_i is the type (or component) of i, c_j is the type of j, the sum covers all pairs i and j of vertices, and $\delta(x, y)$ is 1 if $x = y$ and 0 otherwise. The resolution parameter γ allows weighting the null random model, which can be useful when finding partitions with high modularity. Maximizing modularity with higher values of the resolution parameter often results in more smaller clusters when locating partitions with high modularity. Lower values often result in fewer larger clusters. The original definition of modularity is recovered by setting γ to 1.

iv. *Assortativity*: captures, in a structural sense, the preference of vertices to attach to others that are similar or dissimilar in terms of degree. In our study, we apply three approaches to this measure.

The first one is nominal assortativity, calculated as follows:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}, \tag{13}$$

where e_{ij} is the fraction of edges connecting vertices of type i and j, $a_i = \sum_j e_{ij}$, and $b_j = \sum_i e_{ij}$.

The second variant is assortativity, which is based on values assigned to the vertices, defined as:

$$r = \frac{1}{\sigma_o \sigma_i} \sum_{jk} jk\left(e_{jk} - q_j^o q_k^i\right), \tag{14}$$

where $q_i^o = \sum_j e_{ij}$, $q_i^i = \sum_j e_{ji}$, and, additionally, σ_q, σ_o, and σ_i are the standard deviations of q, q^o, and q^i, respectively.

The third variant is assortativity degree, which uses the vertex degree (minus one) as the vertex value and is called assortativity.

v. *Edge Density*: The density of a graph is the ratio between the number of edges and the number of possible edges.

3. EMPIRICAL RESULTS

3.1. Data

The daily closing price series of the stocks were obtained from Yahoo Finance for the period from January 2, 2007, to December 31, 2024. The selected stocks are those that were part of the Ibovespa index in the portfolio composition for the period September–December 2024, which included 86 stocks. For descriptive statistics as well as the names and business sectors of the companies, see Appendix D and E.

The calculation of daily returns was based on the closing price, applying the natural logarithm, as per Equation 15:

$$R_{i,t} = \left[\ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)\right] \times 100. \tag{15}$$

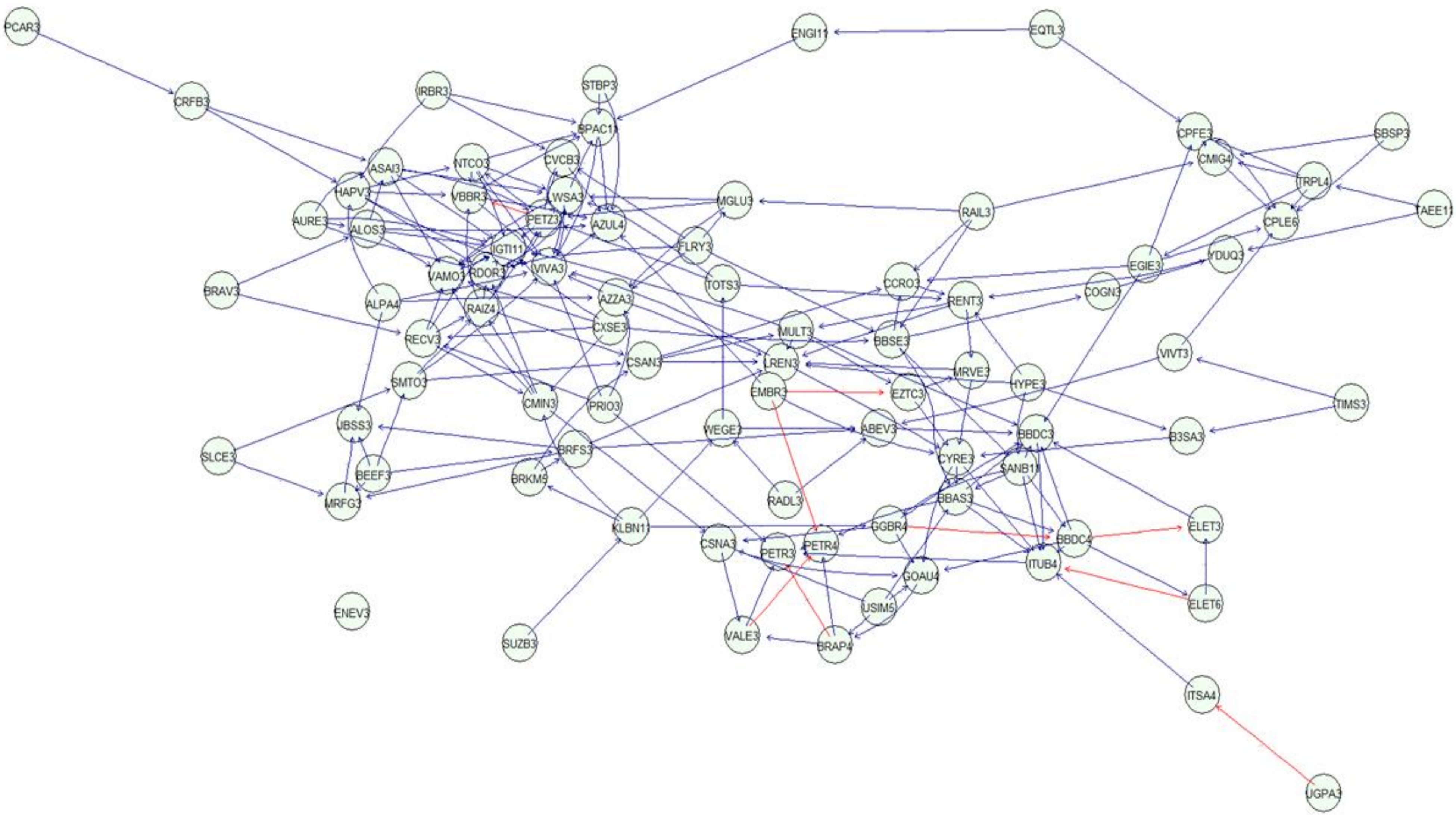
3.2. Estimation of Dynamic Bayesian Networks

We use the logarithm of daily stock returns to estimate Dynamic Bayesian Networks. We conducted Augmented Dickey-Fuller and Phillips-Perron stationarity tests at 1% significance level. All the time series of log stock returns from the studied financial assets are stationary, which is an important requirement as pointed out by Nagarajan, Scutari, and Lèbre (2013) for estimating Dynamic Bayesian Networks.

Efron (2007) suggests q-values between 0.05 and 0.15 as feasible choices for estimating networks, with these q-value limits interpreted as a conservative Bayesian factor for interpreting the FDR. To assess result sensitivity, all networks were generated with q-values equal to 0.05, 0.1, and 0.15, and the networks remained unchanged, demonstrating estimation robustness. Finally, the edges represent the direction of partial correlation, with blue indicating positive and red indicating negative associations.

We estimated the network and the aim was to infer if there is any network formation pattern among the stocks. In Figure 1, we present the estimated network over the analyzed period. Appendix F presents the values of the partial correlations found in the estimated dynamic Bayesian network.

Figure 1 - Dynamic Bayesian Network[6]



The blue edges represent positive partial correlations, while the red edges represent negative partial correlations. | To correlate the stock ticker traded on the Brazilian stock exchange with its respective company name and sector of activity, see Appendix E | To check the partial correlation values, see Appendix F.

[6] The Dynamic Bayesian Network estimations were conducted in RStudio using the GeneNet package version 1.2.16.

In Figure 1, we observe that the estimated dynamic Bayesian network presents 217 partial correlations, predominantly positive. This pattern suggests a unidirectional behavior of the Brazilian stock market over the 17 years analyzed, reflecting strong co-movement among assets. On the other hand, we identified nine negative partial correlations, indicating the existence of potential hedging opportunities between specific pairs of stocks. These negative partial correlations may be partly attributed to the heterogeneity in firms' exposure to domestic and international markets, as well as to the differentiated sectoral sensitivity to the level of the domestic yield curve.

In the analysis of the clustering structure of the estimated network, we observe that stocks tend to cluster with others operating in similar or related sectors of the economy. This finding is consistent with the study by Tabak, Serra, and Cajueiro (2010), who identified, for the Brazilian stock market network, groups of stocks that are homogeneous with respect to their sector, often forming strong clusters. In other words, such clustering among companies from related sectors may amplify the sensitivity of these stock prices to adverse economic and political events.

When examining the stocks of companies operating in the power generation and/or transmission, sanitation, and telecommunications sectors, we find that these assets exhibit a lower number of associations with other stocks in the network. This behavior can be attributed to the fact that these companies operate in highly regulated sectors with greater predictability of cash flows, which results in lower price volatility. Consequently, these stocks tend to be less correlated with the other assets that make up the Ibovespa index, suggesting a more defensive profile within the investment portfolio.

Given the associations identified among the analyzed assets, we further deepen the analysis and interpretation of the financial network by applying a set of metrics derived from complex network theory. This set of measures enables us to capture additional and relevant information about the structure and behavior of the network, offering a more comprehensive understanding of the interactions among assets over time.

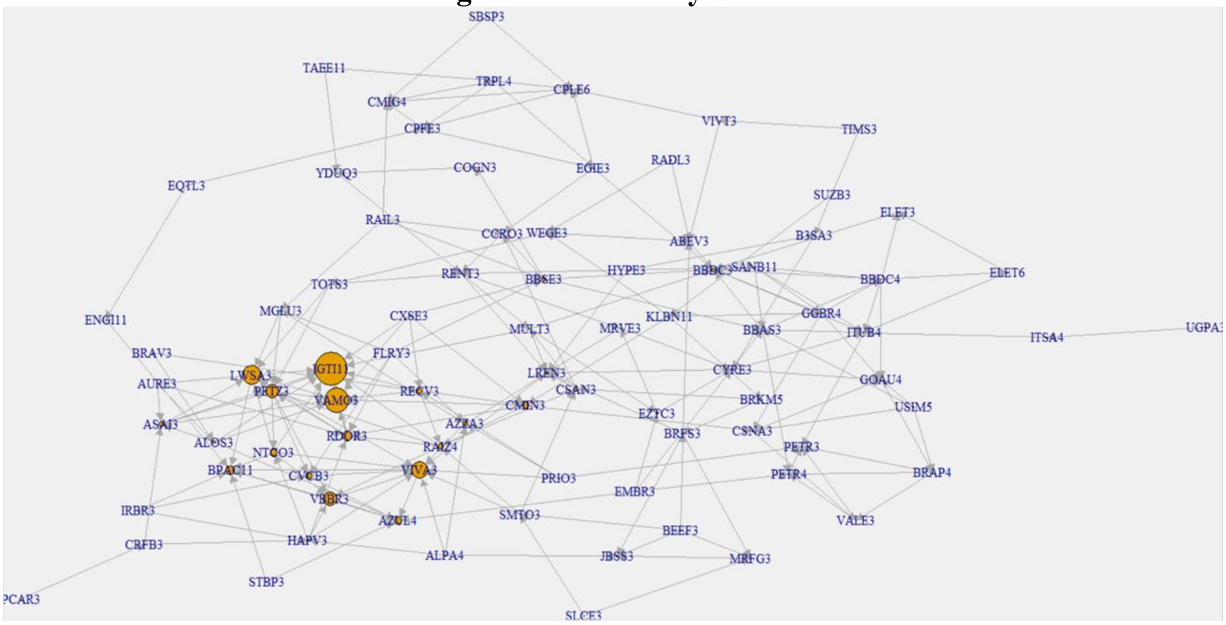
3.3. Applied Complex Network Measures

We applied 11 complex network measures, encompassing both asset-specific characteristics and metrics that capture the overall structure of the financial networks studied, as proposed by Silva and Zhao (2016). This set of measures provides a comprehensive understanding of the dynamics of associations among stocks over time, enabling the identification of association patterns that dynamic Bayesian network estimates alone may not reveal.

To calculate the distance measure between the studied stocks, we used the partial correlations estimated by the dynamic Bayesian network and computed the distance measure as defined in Equation 9.

The first two measures employed to analyze the estimated financial network were the mixed metrics Authority Score and Hub Score, which are complementary approaches used to understand the centrality of financial assets within the estimated networks. These measures are effective for acyclic networks, as is the case with our network. Figure 2 and Figure 3 present the authority points and hubs, respectively.

Figure 2 - Authority Score



To correlate the stock ticker traded on the Brazilian stock exchange with its respective company name and sector of activity, see Appendix E.

When analyzing the authority points, we observe that they correspond to companies focused on the domestic market with strong growth margins and consumers with high purchasing power. The authorities IGT11 (Iguatemi), VIVA3 (Vivara), and RDOR3 (Rede D'Or São Luiz) operate targeting high-income consumers in the shopping centers, jewelry, and hospital sectors, respectively. Therefore, the expansion of the high-end market in Brazil[7] benefits companies serving this clientele, in addition to offering higher margins and a more resilient market.

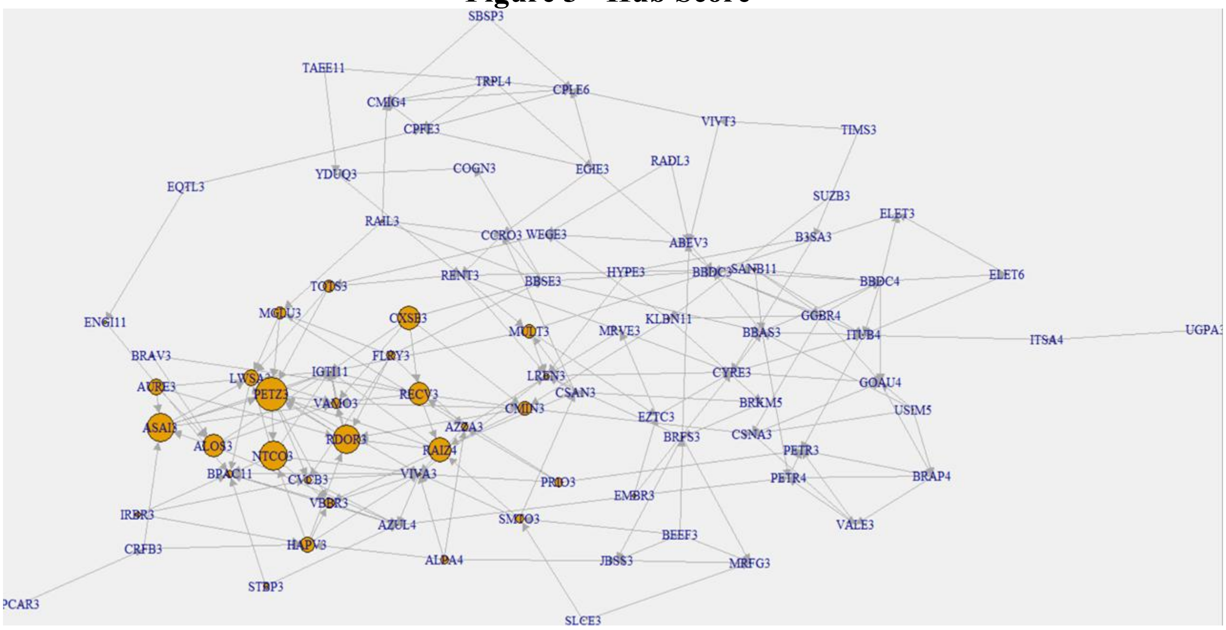
[7] The *Luxury Lab Global* 2024 report by Euromonitor International indicates that the luxury market — including retail and services such as travel — is expected to generate approximately USD 1.4 trillion in global revenue in 2024, reaching USD 1.7 trillion by 2030. Brazil is highlighted in the study as the ninth-fastest growing market among the top ten, alongside China, Malaysia, Thailand, Indonesia, India, the Philippines, Romania, Mexico, and Hong Kong. The Brazilian luxury segment also stands out in terms of its size within the region: the country, together with Mexico, accounts for 25% of the luxury market in Latin America. The report further projects growth of 7% in 2025 and a cumulative increase of 22% over the next five years in Brazil.

Other stocks that stand out as authorities in the network are VAMO3, LWSA3, and PETZ3, belonging respectively to the truck rental, cloud services, and pet products sectors. These assets reflect the dynamics of segments with high growth potential, driven by the increase in household income in Brazil. This scenario favors the expansion of demand for goods distribution — which is heavily reliant on road transportation in the country — more intensive use of cloud-based technological solutions, and rising spending on pet care and products.

In the literature, Magno et al. (2019) also identified authority points in the Brazilian stock network concentrated in companies focused on the domestic market, consistent with our findings. In the Pakistani stock market, Memon and Yao (2019) found that central nodes were companies from the cement (construction) and textile sectors, which have significant growth potential in a developing country context. In China, Z. Wang et al. (2024) identified that the industrial sector accounts for 30% of key nodes based on degree and centrality, with media information services representing 10%. In other words, the international literature shows a certain convergence regarding the sectors that hold authority within stock market networks, while also reflecting the specific characteristics of each country and its companies.

Analyzing the other mixed centrality measure, the Hub Score, we find that the hubs are companies that are not among the largest market capitalizations of the Ibovespa index and are mostly focused on the domestic market. It is noteworthy that these stocks are not the most liquid in the market, serving instead as tactical positions for investors seeking returns during periods of volatility. Magno et al. (2019) also found that the hubs in the Brazilian stock market network are companies that are not among the largest capitalizations of the Ibovespa index, but rather investment firms — that is, firms that hold stakes in other companies.

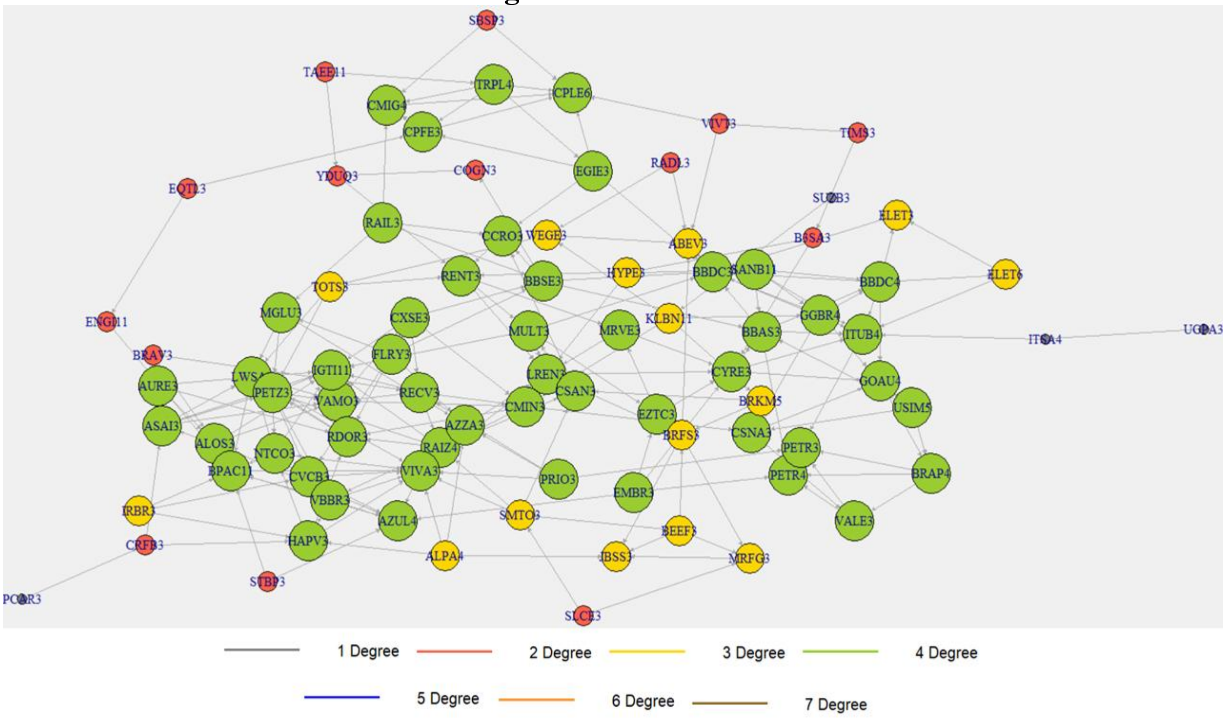
Figure 3 - Hub Score



To correlate the stock ticker traded on the Brazilian stock exchange with its respective company name and sector of activity, see Appendix E.

The third mixed network measure employed in our study is Coreness, which identifies the “core” of the network — that is, the nodes most deeply connected to other important nodes. In Figure 4, we present the Coreness of the stock market financial network.

Figure 4 - Coreness



To correlate the stock ticker traded on the Brazilian stock exchange with its respective company name and sector of activity, see Appendix E.

The analysis of the network coreness (Figure 4) reveals that its core is composed of stocks with at least four connections, indicating a dense structure of interactions. It is observed that a significant share of assets is part of this core, regardless of their market capitalization level or sector. This result suggests that broad upward or downward movements in the stock market are likely to strongly impact the prices of these stocks, given their high degree of interdependence. In the short term, investors able to identify asymmetries in risk and return within this core may benefit through arbitrage strategies, such as selling overvalued assets and buying undervalued ones. In the long term, however, the density of the core may reduce the effect of diversification, as stocks tend to move more synchronously within this central structure.

The last mixed measure used to analyze the financial network is Global Transitivity (or Clustering Coefficient), which assesses the degree of connectivity among a node’s neighbors — that is, the tendency for connections to form triangles. Over the 17 years analyzed, this metric presented a value of 0.2818 (see Table 1), indicating a moderate level of local cohesion among the assets. By comparison, Tabak, Serra, and Cajueiro (2010), when analyzing the main sectors of the Brazilian stock market, observed a downward trend in this measure, with values decreasing from 0.35 to 0.06 between 2000 and 2008. In the Chinese context, Huang, Zhuang, and Yao (2009) and Z. Wang et al. (2024) found values ranging between 0.1 and 0.7 during the period from 2003 to 2007, while in the United States, Rakib, Nobi, and Lee (2021) reported a range between 0.39 and 0.47 from 1998 to 2012, and Tse, Liu, and Lau (2010) found values between 0.1 and 0.43 from 2005 to 2009. These results indicate that the formation of triangles — and thus the degree of local cohesion in the network — can vary significantly over time and across markets. Even so, clustering remains a useful tool for understanding the mechanisms of shock propagation in financial networks.

Table 1 - Complex Network Measures Employed

Diameter	Mean Distance	Modularity	Assortativity	Nominal Assortativity	Degree Assortativity	Global Transitivity	Edge Density
10.4242	4.5047	0.5465	0.0301	-0.0146	0.255	0.2818	0.0304

We used seven global measures to understand the overall dynamics of our estimated partial correlation network. The first measure is the network diameter, which was 10.4242, indicating a more dispersed or poorly connected network. This can be explained by economic dynamics and the high volatility of agents' expectations, which directly affect stock prices and alter the connections among assets on a daily basis. In their study, Memon and Yao (2019) found diameters ranging from 3 to 9 across different periods for the Pakistani stock market. Meanwhile, Tse, Liu, and Lau (2010) reported values between 7 and 30 for the U.S. stock market — the largest in the world in terms of capitalization and number of listed stocks — which is one of the main factors contributing to higher network diameter values.

The second global measure calculated was the Mean Distance, which was 4.5047, showing that the relatively high diameter influences the Mean Distance. In other words, connections in the financial market are more volatile, resulting in less efficient linkages. Similar values were found in other studies (Caraiani, 2012; Memon & Yao, 2019; Rakib et al., 2021; Tse et al., 2010).

The calculated modularity was 0.5465, indicating a good division of the Brazilian stock market network into groups. The result shows that companies from similar sectors tend to associate with greater cohesion, while firms from very different sectors tend to be associated to a lesser extent. This finding suggests that unsystematic risk mitigation can be achieved through diversification among stocks within a portfolio. In Brazil, Magno et al. (2019) found a modularity of 0.8556 for the period from August to December 2017. Meanwhile, Memon and Yao (2019) reported values ranging from 0.044 to 0.417 for the Pakistani stock market.

The fourth global measure analyzed is Assortativity, which measures whether nodes with similar characteristics (typically the number of connections, or degree) tend to connect with each other. In addition to assortativity, we employed two variations of this measure: Nominal and Degree. When we examine the network using the Assortativity and Nominal Assortativity metrics, we observe that the network is neutral, indicating the association of nodes with similar values and similar categories, respectively. In contrast, Degree Assortativity shows a moderately positive value, suggesting that stocks with many neighbors tend to connect with other well-connected stocks. In other words, there is a cluster or community structure among hubs (high-degree nodes), as can be seen in Figure 3.

The last global measure analyzed is Edge Density, which assesses the overall connectivity of the network. We obtained a value of 0.0304, indicating a very sparse network (few connections relative to the total possible). This is expected for financial networks, which exhibit more volatile associations due to the heterogeneous expectations of financial agents and the high volatility of the macroeconomic and political environment. Additionally, Brazil being an emerging market amplifies this volatility, as foreign investors evaluate global risk–return asymmetries when allocating their portfolios. In the United States, Rakib, Nobi, and (Lee 2021) found values ranging from 0.38 to 0.42, showing that mature stock markets have lower volatility and more stable connections.

Overall, the complex network measures used to analyze the Brazilian stock market network reveal the formation of certain clusters among companies from related sectors. This observation motivated an analysis through community detection. In the next section, we present the construction of these communities.

3.3.1. Identification of Communities in the Brazilian Stock Market

We constructed the stock communities using the algorithm proposed by Clauset, Newman, and Moore (2004) to identify the highest modularity within the Brazilian stock market network. In Figure 5, we present the resulting communities.

The analysis of the identified communities within the network reveals the formation of six distinct groups, among which there are numerous inter-community connections (represented by the red lines), indicating that these communities exert mutual influence on one another. A joint interpretation of Figure 1 and Figure 5 shows that the predominance of positive partial correlations reinforces unidirectional movements among assets. In this context, even diversification strategies based on allocating assets from different communities may see reduced effectiveness, since the presence of multiple interconnections among communities facilitates the propagation of shocks and the convergence of returns.

The red community (Figure 5, upper right side) is composed of companies that are capital-intensive and traditional banks; in other words, it consists of members operating in monopolistic or oligopolistic sectors with significant barriers to entry. We infer that this community can be classified as an allocation in defensive stocks, characterized by less volatile returns in both dividends and capital appreciation.

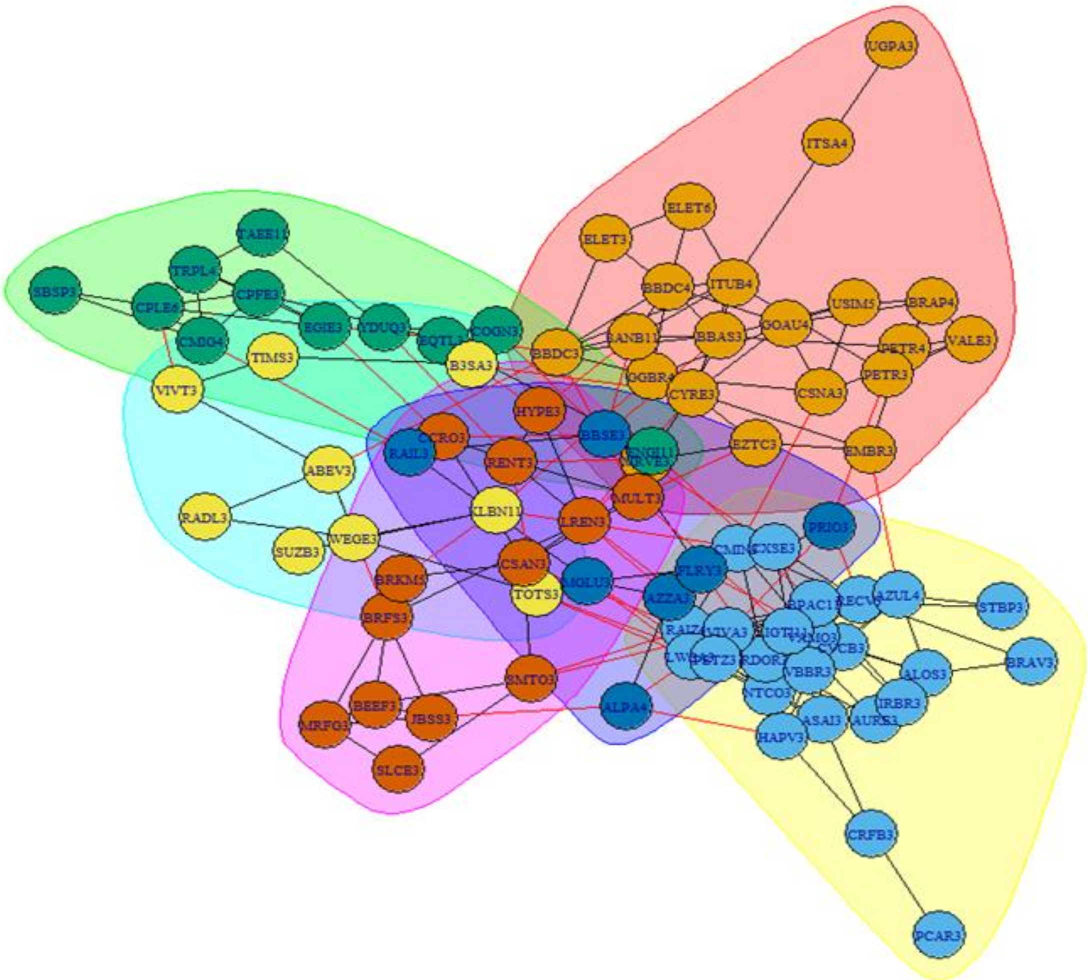
The yellow community (Figure 5, lower right side) consists predominantly of stocks from companies focused on the domestic market and end consumers (Business to Consumer). It is observed that these companies generally are not among the largest capitalizations of the Ibovespa index and tend to exhibit greater volatility in their returns. Thus, such assets can be used as strategic allocations to capture short- and medium-term gains, especially when investors identify risk–return asymmetries that enable more opportunistic arbitrage or stock-picking strategies.

The dark blue and light blue communities (Figure 5) are the smallest among all identified communities. The companies within these communities are efficient market leaders in their respective sectors, and their stock prices are relatively high given the quality of the assets — in other words, these stocks trade at a premium. Allocating to this type of stock is suitable for investors with a long-term approach, as the main gains are expected to come from capital appreciation.

The dark pink community is primarily composed of meat processing companies and exporters of agricultural commodities, with a main focus on the international market. Thus, this community offers investors an opportunity for diversification through assets directly exposed to international demand. On the other hand, the presence of stocks from companies oriented toward the domestic market within the same community highlights sectoral heterogeneity, allowing investors to diversify their allocation between firms more dependent on the external sector and others more linked to domestic consumption.

The last community, the green one, is predominantly composed of companies from the electric sector, with the exception of two firms from the education sector. Thus, this community exhibits low sectoral heterogeneity, indicating limited internal diversification. Consequently, excessive allocation across multiple stocks within this community may offer limited diversification benefits, making it more advantageous to concentrate investments in a smaller number of representative assets from this group.

Figure 5 - Community Detection in the Brazilian Stock Market Network



The red edges represent connections between individuals from different communities, while the black edges represent connections between individuals within the same community. | To correlate the stock ticker traded on the Brazilian stock exchange with its respective company name and sector of activity, see Appendix E.

4. CONCLUSION

The main objective of this study was to investigate the partial correlation structure and clustering patterns of the Brazilian stock market over the period from January 2, 2007, to December 31, 2024. To this end, Dynamic Bayesian Networks were employed to estimate putative partial correlations, following the methodology proposed by Opgen-Rhein and Strimmer (2007b) and Schäfer and Strimmer (2005). Based on the estimated network, we calculated distance measures between stocks and analyzed its topology using complex network metrics, encompassing both mixed and global measures. This approach enabled a more detailed understanding of the interdependence patterns and the underlying structure of the Brazilian stock network over the past 17 years.

The results reveal significant patterns of partial correlations, with predominantly positive connections, suggesting that, in the long term, the Brazilian stock market tends to move in the same direction, although this interdependence dynamic may vary in the short term. On the other hand, we identified nine negative partial correlations, indicating the existence of potential hedging opportunities between certain pairs of stocks. These negative partial correlations can be partly attributed to heterogeneity in companies’ exposure to domestic and international markets, as well as sector-specific sensitivities to the domestic interest rate curve.

The analysis of the network topology, through the employed complex network measures, revealed that authority points are concentrated in companies focused on the domestic market, characterized by strong growth margins and consumers with higher purchasing power. It is observed that the expansion of the high-end consumer segment over the last decade has particularly benefited these firms. Meanwhile, the identified hubs largely correspond to companies that are not among the largest market capitalizations and mostly operate in the domestic market as well. Notably, these assets tend to exhibit relatively lower liquidity, thereby serving as tactical positioning instruments for investors seeking to capture additional returns in periods of heightened volatility.

The core of the financial network, represented by coreness, is relatively dense, with many stocks belonging to this core, including the largest capitalizations and the most liquid stocks in the index. These nodes occupy strategic positions and can transmit financial shocks with greater intensity. This indicates the existence of a subset of assets with multiple strong connections among themselves, which is uncommon throughout the entire network that tends to be sparse in financial networks, as is the case here, where we found a diameter of 10.4242.

We identified six distinct communities, yet numerous inter-community connections (represented by red lines) indicate that these communities exert mutual influence on each other. Thus, even diversification strategies based on allocating assets across different communities may see reduced effectiveness, as the presence of multiple interconnections among communities facilitates the propagation of shocks and the convergence of returns.

For investors, the main takeaway from this study is the identification of a central core within the network, composed of the leading companies included in the Ibovespa index. Potential shocks affecting this core have the capacity to propagate to other assets, even within a network that exhibits dispersed characteristics. Furthermore, we find that partial correlations predominantly lie in the positive domain, favoring unidirectional price movements over the long term. However, the presence of some negative partial correlations indicates opportunities for diversification strategies, allowing risk mitigation by combining assets with distinct return dynamics.

Finally, for future research, we suggest the application of nonlinear models and machine learning techniques in the estimation of stock networks, simulating shock propagation within the estimated networks.

REFERENCES

Ballester, L., López, J., & Pavía, J. M. (2023). European systemic credit risk transmission using Bayesian networks. *Research in International Business and Finance*, *65*, 101914. https://doi.org/10.1016/J.RIBAF.2023.101914

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, *48*(2), 431–443. https://doi.org/10.1016/J.CSDA.2004.02.004

Bonanno, G., Caldarelli, G., Lillo, F., & Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, *68*(4), 046130. https://doi.org/10.1103/PhysRevE.68.046130

Bonanno, G., Vandewalle, N., & Mantegna, R. N. (2000). Taxonomy of stock market indices. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *62*(6 Pt A). https://doi.org/10.1103/PHYSREVE.62.R7615

Brida, J. G., & Risso, W. A. (2010). Dynamics and structure of the 30 largest North American companies. *Computational Economics*, *35*(1), 85–99. https://doi.org/10.1007/s10614-009-9187-1

Cao, G., Shi, Y., & Li, Q. (2017). Structure Characteristics of the International Stock Market Complex Network in the Perspective of Whole and Part. *Discrete Dynamics in Nature and Society*, *2017*. https://doi.org/10.1155/2017/9731219

Caraiani, P. (2012). Characterizing emerging European stock markets through complex networks: From local properties to self-similar characteristics. *Physica A: Statistical Mechanics and Its Applications*, *391*(13), 3629–3637. https://doi.org/10.1016/J.PHYSA.2012.02.008

Carvalho, J. V. de F., & Chiann, C. (2013). *Redes Bayesianas: Um Método para Avaliação de Interdependência e Contágio em Séries Temporais Multivariadas*. https://www.scielo.br/j/rbe/a/wqXSrKBQRpSnnFr73XjfhZN/?format=pdf&lang=pt

Chiquet, J., Smith, A., Grasseau, G., Matias, C., & Ambroise, C. (2009). SIMoNe: Statistical Inference for MOdular NEtworks. *Bioinformatics*, *25*(3), 417–418. https://doi.org/10.1093/BIOINFORMATICS/BTN637

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *70*(6), 6. https://doi.org/10.1103/PHYSREVE.70.066111/FIGURES/3/MEDIUM

de Pontes, L. S., & Rêgo, L. C. (2022). Impact of macroeconomic variables on the topological structure of the Brazilian stock market: A complex network approach. *Physica A: Statistical Mechanics and Its Applications*, *604*, 127660. https://doi.org/10.1016/J.PHYSA.2022.127660

Efron, B. (2005, July). *Correlation and Large-Scale Simultaneous Significance Testing | Stanford Digital Repository*. Technical Report Nº 2005-21-B/235 . https://purl.stanford.edu/vj518hf1106

Efron, B. (2007). Size, power and false discovery rates. *The Annals of Statistics*, *35*(4), 1351–1377. https://doi.org/10.1214/009053606000001460

Heimo, T., Kaski, K., & Saramäki, J. (2009). Maximal spanning trees, asset graphs and random matrix denoising in the analysis of dynamics of financial networks. *Physica A: Statistical Mechanics and Its Applications*, *388*(2–3), 145–156. https://doi.org/10.1016/J.PHYSA.2008.10.007

Heimo, T., Tibély, G., Saramäki, J., Kaski, K., & Kertész, J. (2008). Spectral methods and cluster structure in correlation-based networks. *Physica A: Statistical Mechanics and Its Applications*, *387*(23), 5930–5945. https://doi.org/10.1016/J.PHYSA.2008.06.028

Huang, W. Q., Zhuang, X. T., & Yao, S. (2009). A network analysis of the Chinese stock market. *Physica A: Statistical Mechanics and Its Applications*, *388*(14), 2956–2964. https://doi.org/10.1016/J.PHYSA.2009.03.028

James, W., & Stein, C. (1961). Estimation with quadratic loss. *In: Neyman J (Ed) Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 361–379. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.stat.yale.edu/~hz68/619/Stein-1961.pdf

Kenett, D. Y., Huang, X., Vodenska, I., Havlin, S., & Stanley, H. E. (2015). Partial correlation analysis: applications for financial markets. *Quantitative Finance*, *15*(4), 569–578. https://doi.org/10.1080/14697688.2014.946660/ASSET//CMS/ASSET/F8571C1A-2432-469B-863C-948D897F4E25/14697688.2014.946660.FP.PNG

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. https://doi.org/10.1371/JOURNAL.PONE.0015032

Korb, K. B., & Nicholson, A. E. (2004). *Bayesian artificial intelligence*. Chapman & Hall/CRC.

Korkusuz, B., McMillan, D. G., & Kambouroudis, D. (2022). Complex network analysis of volatility spillovers between global financial indicators and G20 stock markets. *Empirical Economics*, *64*(4), 1517–1537. https://doi.org/10.1007/S00181-022-02290-W/TABLES/3

Lèbre, S. (2009). Inferring dynamic genetic networks with low order independencies. *Statistical Applications in Genetics and Molecular Biology*, *8*(1). https://doi.org/10.2202/1544-6115.1294

Ma, W. J., Hu, C. K., & Amritkar, R. E. (2004). Stochastic dynamical model for stock-stock correlations. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *70*(2), 9. https://doi.org/10.1103/PHYSREVE.70.026101/FIGURES/4/THUMBNAIL

Magno, C., Barbosa, G., Gabriel, L., Felix, S., Xavier, C. R., Da, V., & Vieira, F. (2019). Investigating the Relation Between Companies with Topological Analysis of a Network of Stock Exchange in Brazil. *Journal of Information and Data Management*, *10*(3), 130-145–130 – 145. https://periodicos.ufmg.br/index.php/jidm/article/view/13591

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *European Physical Journal B*, *11*(1), 193–197. https://doi.org/10.1007/S100510050929/METRICS

Memon, B. A., & Yao, H. (2019). Structural Change and Dynamics of Pakistan Stock Market during Crisis: A Complex Network Perspective. *Entropy 2019, Vol. 21, Page 248*, *21*(3), 248. https://doi.org/10.3390/E21030248

Millington, T., & Niranjan, M. (2020). Partial correlation financial networks. *Applied Network Science*, *5*(1), 1–19. https://doi.org/10.1007/S41109-020-0251-Z/FIGURES/13

Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning* [Thesis, University of California]. https://ibug.doc.ic.ac.uk/media/uploads/documents/courses/DBN-PhDthesis-LongTutorail-Murphy.pdf

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). *Bayesian Networks in R with Applications in Systems Biology* (1st ed., Vol. 48). Springer. https://doi.org/doi.org/10.1007/978-1-4614-6446-4.

Neapolitan, R. E. (2004). *Learning Bayesian Networks* (1st ed.). Pearson.

Onnela, J. P., Kaski, K., & Kertész, J. (2004). Clustering and information in correlation based financial networks. *The European Physical Journal B 2004 38:2*, *38*(2), 353–362. https://doi.org/10.1140/EPJB/E2004-00128-7

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *68*(5). https://doi.org/10.1103/PhysRevE.68.056110

Opgen-Rhein, R., & Strimmer, K. (2007a). From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology*, *1*(1), 1–10. https://doi.org/10.1186/1752-0509-1-37/FIGURES/3

Opgen-Rhein, R., & Strimmer, K. (2007b). Learning causal networks from systems biology time course data: An effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics*, *8*(SUPPL. 2), 1–8. https://doi.org/10.1186/1471-2105-8-S2-S3/FIGURES/3

Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific Reports 2013 3:1*, *3*(1), 1–7. https://doi.org/10.1038/srep01665

Rakib, M. I., Nobi, A., & Lee, J. W. (2021). Structure and dynamics of financial networks by feature ranking method. *Scientific Reports 2021 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-97100-1

Ren, F., & Zhou, W. X. (2014). Dynamic Evolution of Cross-Correlations in the Chinese Stock Market. *PLOS ONE*, *9*(5), e97711. https://doi.org/10.1371/JOURNAL.PONE.0097711

Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*(1), 1–30. https://doi.org/10.2202/1544-6115.1175/MACHINEREADABLECITATION/RIS

Sensoy, A., & Tabak, B. M. (2014). *Dynamic spanning trees in stock market networks: The case of Asia-Pacific*. www.bcb.gov.br/?CONTACTUS

Silva, T. C., & Zhao, L. (2016). Machine learning in complex networks. In *Machine Learning in Complex Networks* (1st ed.). Springer International Publishing. https://doi.org/10.1007/978-3-319-17290-3/COVER

So, M. K. P., Chan, L. S. H., & Chu, A. M. Y. (2021). Financial Network Connectedness and Systemic Risk During the COVID-19 Pandemic. *Asia-Pacific Financial Markets*, *28*(4), 649–665. https://doi.org/10.1007/s10690-021-09340-w

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivaruate normal distribution. *Neyman J (Ed) Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, 197–206. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.stat.yale.edu/~hz68/619/Stein-1956.pdf

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *64*(3), 479–498. https://doi.org/10.1111/1467-9868.00346

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Https://Doi.Org/10.1214/Aos/1074290335*, *31*(6), 2013–2035. https://doi.org/10.1214/AOS/1074290335

Tabak, B. M., Serra, T. R., & Cajueiro, D. O. (2010). Topological properties of stock market networks: The case of Brazil. *Physica A: Statistical Mechanics and Its Applications*, *389*(16), 3240–3249. https://doi.org/10.1016/J.PHYSA.2010.04.002

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. In *Source: Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 58, Issue 1).

Tse, C. K., Liu, J., & Lau, F. C. M. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, *17*(4), 659–667. https://doi.org/10.1016/J.JEMPFIN.2010.04.008

Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10421–10426. https://doi.org/10.1073/PNAS.0500298102/SUPPL_FILE/00298TABLE7.XLS

Wang, G. J., Xie, C., Han, F., & Sun, B. (2012). Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. *Physica A: Statistical Mechanics and Its Applications*, *391*(16), 4136–4146. https://doi.org/10.1016/j.physa.2012.03.036

Wang, G. J., Xie, C., & Stanley, H. E. (2018). Correlation Structure and Evolution of World Stock Markets: Evidence from Pearson and Partial Correlation-Based Networks. *Computational Economics*, *51*(3), 607–635. https://doi.org/10.1007/S10614-016-9627-7/FIGURES/15

Wang, Z., Zhang, G., Ma, X., & Wang, R. (2024). Study on the Stability of Complex Networks in the Stock Markets of Key Industries in China. *Entropy 2024, Vol. 26, Page 569*, *26*(7), 569. https://doi.org/10.3390/E26070569

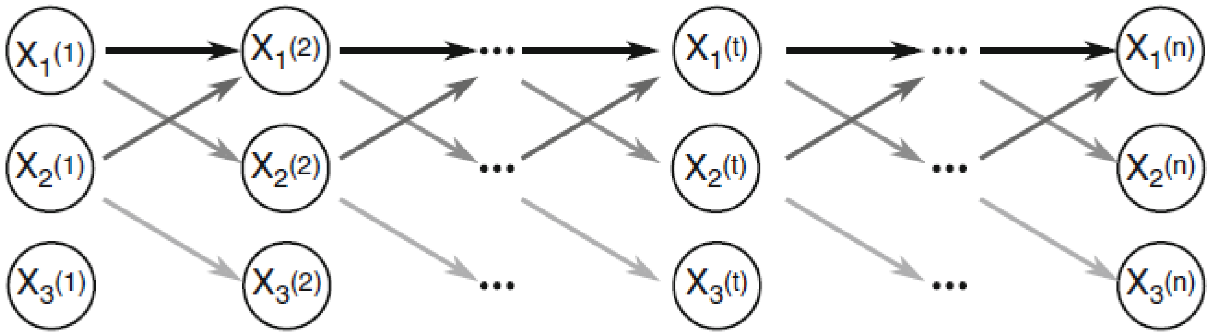
Yang, C., Shen, Y., & Xia, B. (2013). Evolution of shanghai stock market based on maximal spanning trees. *Modern Physics Letters B*, *27*(3). https://doi.org/10.1142/S021798491350022X;PAGE:STRING:ARTICLE/CHAPTER

APPENDIX

A. Graphical Structure of a Dynamic Bayesian Network

We present the structure of a dynamic Bayesian network over multiple time periods:

Figure 6 - Graphical Representation of a Dynamic Bayesian Network with Temporal Variation



If the information of the variables is available over time, it is possible to map the feedback cycles and loops across temporal points. Therefore, we can define these interactions as a time-homogeneous dynamic Bayesian network, assuming that, at each time period t, all the parents of each node are measured at the previous time point $t-1$. It is important to emphasize that the dynamic Bayesian network at a given time period cannot have acyclic forward-backward movements.

Source: Nagarajan, Scutari, and Lèbre (2013).

B. Significance Measures: local false discovery rate (fdr) and q-value

An efficient estimator of the covariance matrix can be obtained by setting the empirical correlation coefficients to zero and the empirical variances to their medians (Nagarajan et al., 2013).

In this context, the work conducted by Schäfer and Strimmer (2005) and Opgen-Rhein and Strimmer (2007) present an algorithm that enables robust estimation of VAR(1) coefficients for Dynamic Bayesian Networks. Gaussian Graphical Models (GGM) are estimated for each Directed Acyclic Graph (DAG) based on applying shrinkage estimators to estimated covariances and partial correlation matrices, which represent the interactions between variables. The network structure is determined by including edges in descending order of coefficients and employing multiple tests of the local false discovery rate (local fdr), which tests the existence of false positives (edges with null probability) and eliminates them. Thus, only significant edges remain.

In Equation 21, we define the observed partial correlations $\tilde{r}$ in the edges:

$$f(\tilde{r}) = \eta_0 f_0(\tilde{r}; k) + (1 - \eta_0) f_A(\tilde{r}), \tag{16}$$

where f_0 is the null distribution, η_0 is the (unknown) proportion of "null edges", and f_A is the distribution of observed partial correlations assigned to actually existing edges. The null density f_0 is given by:

$$f_0(\tilde{r}; k) = (1 - \tilde{r}^2)^{(k-3)/2} \frac{\Gamma\left(\frac{k}{2}\right)}{\pi^{1/2}\Gamma\left(\frac{(k-1)}{2}\right)} = |\tilde{r}|\mathrm{Be}\left(\tilde{r}^2; \frac{1}{2}, \frac{k-1}{2}\right), \tag{17}$$

where $\mathrm{Be}(x; a, b)$ is the Beta distribution and k is the degrees of freedom, equal to the reciprocal variance of the null $\tilde{r}$. Fitting this mixture density allows k, η_0, and f_A to be determined. Subsequently, we can calculate the specific edge's false discovery rate (fdr) using Equation 18:

$$\mathrm{Prob}(\text{null edge}|\tilde{r}) = \mathrm{fdr}(\tilde{r}) = \frac{\hat{\eta}_0 f_0(\tilde{r};\, \hat{k})}{\hat{f}(\tilde{r})}. \tag{18}$$

The Equation 18 highlights the local fdr as the posterior probability of an edge being null given $\tilde{r}$. In the estimation process, the q-values associated with each edge and the probabilities of an edge being non-null $(1 - \mathrm{fdr})$ are calculated, and these quantities can be used to define the significance level of the edges.

The q-value is a measure of statistical significance adjusted by the FDR (False Discovery Rate) and is used to address multiple testing statistical problems, where q-values represent the expected percentage of false positives among significance tests. On the other hand, p-values only indicate the significance level considering the overall number of performed tests (Storey, 2002).

Storey (2002) defines the FDR as the expected proportion of false positives among all rejected hypotheses multiplied by the probability of at least one rejection occurring. Storey (2003) establishes the pFDR (positive false discovery rate) to demonstrate that we are conditioned on at least one positive finding occurring:

$$\mathrm{pFDR} = \mathbb{E}\left(\frac{V}{R} | R > 0\right), \tag{19}$$

where V is the number of type I errors (or false positive results), and R is the number of rejected hypotheses. For a set of rejection regions $\{\Gamma\}$, these could be all sets of the form $[c, \infty)$ for $-\infty \leq c \leq \infty)$, where the p-value of an observed statistic $T = t$ is defined as:

$$p - \text{value}(t) = \min_{\{\Gamma: t \in \Gamma\}} \{\Pr(T \in \Gamma | H = 0)\}. \tag{20}$$

According to Storey (2002), the p-value provides a measure of the strength of the observed statistic in terms of committing a type I error, i.e., it is the minimum rate of type I error that can occur when rejecting a statistic with value t for the set of rejection regions. On the other hand, the q-value is a measure of the strength of an observed statistic in relation to pFDR, being the minimum pFDR that can occur when rejecting a statistic with value t for the set of significance regions (Storey, 2003). Thus, for an observed statistic $T = t$, the q-value of t is defined as:

$$q(t) = \inf_{\{\Gamma: r \in \Gamma\}} \{\text{pFDR}(\Gamma)\}. \tag{21}$$

When statistics are independent p-values, the definition is simplified, and the set of rejection regions takes the form $[0;\, \gamma]$ and pFDR can be written more simply. Therefore, for a set of hypothesis tests conducted with independent p-values, the q-value of the observed p-value p is:

$$q(p) = \inf_{\gamma \geq p} \{\text{pFDR}(\gamma)\} = \inf_{\gamma \geq P} \left\{ \frac{\pi_0 \gamma}{\Pr(P \leq \gamma)} \right\}. \tag{22}$$

According to Schäfer and Strimmer (2005) , the q-value is intrinsically related to local Bayesian Fdr statistics. Efron (2005) asserts that using local Fdr is more appropriate because it naturally fits the mixture model setup and considers dependencies among estimated correlation coefficients. Therefore, we choose to use it as our measure of significance.

Efron (2007) asserts that the Bayesian posterior probability that a case is null given z, by definition, is the local false discovery rate:

$$\text{Fdr}(z) \equiv \Pr\{\text{null}|z\} = \frac{p_0 f_0(z)}{f(z)} = \frac{f_0^+(z)}{f(z)}, \tag{23}$$

where $p_0 = \Pr\{\text{null}\}, f_0(z) = \text{density if null},$

$p_1 = \Pr\{\text{non} - \text{null}\}, f_1(z) = \text{density if non} - \text{null}.$

Benjamini and Hochberg (1995) theorized about the discovery rate that relies on tail areas rather than densities. Assuming $F_0(z)$ and $F_1(z)$ as the cumulative distribution functions corresponding to $f_0(z)$ and $f_1(z)$, they define $F_0^+(z) = p_0 F_0(z)$ and $F_0(z) = p_0 F_0(z) + p_1 F_1(z)$. Consequently, we can infer that the posterior probability of a case being null given its z-value, $\mathcal{Z}$, is less than some value z is:

$$\text{Fdr}(z) \equiv \Pr\{\text{null}|\mathcal{Z} \leq z\} = \frac{F_0^+(z)}{F(z)}. \tag{24}$$

Therefore, $\mathrm{Fdr}(z)$ corresponds to the q-value defined by Storey (2002) and to the false discovery rate value of the tail area achieved at a specific observed value $Z = z$. Analytically, Fdr is a conditional expectation of fdr:

$$\mathrm{Fdr}(z) = \frac{\int_{-\infty}^{z} \mathrm{fdr}(Z) f(Z)\, dZ}{\int_{-\infty}^{z} f(Z)\, dZ} = E_f\{\mathrm{fdr}(Z) | Z \le z\}, \tag{25}$$

E_f indicates the expectation concerning $f(z)$. In other words, $\mathrm{Fdr}(z)$ is the average of $\mathrm{fdr}(Z)$ for $Z \le z$. $\mathrm{Fdr}(z)$ will be lower than $\mathrm{fdr}(z)$ in the usual situation where $\mathrm{fdr}(z)$ decreases as $|z|$ increases. In the works of Schäfer and Strimmer (2005) and Opgen-Rhein and Strimmer (2007), the authors detail the construction of the algorithm.

Finally, we can express that each new DAG ($posterior$ distribution) is composed of the previous DAG's data ($prior$ distribution) plus the data from that period (likelihood). Therefore, the parameters are sequentially updated over time, so that the past parameters of dependency influence the estimation of future dependency. It is important to mention that the initial distribution ($prior$) is uniform. By analyzing the networks formed over the periods, we will observe changes in interactions between assets through the direction of edges.

C. How GeneNet Uses the VAR Model

The dynamic procedure in the GeneNet package (invoked with *ggm.estimate.pcor(method = "dynamic")*) performs an internal pre-processing step based on a VAR(1) model. This VAR is not used for forecasting or for economic interpretation; its sole purpose is statistical filtering to remove serial autocorrelation before estimating the network.

1. Preliminary temporal modeling

For each variable (gene or, in this context, financial asset), GeneNet internally estimates:

$$X(t) = AX(t-1) + \varepsilon(t), \tag{26}$$

where:

A is the VAR(1) coefficient matrix; and

$\varepsilon(t)$ denotes the residuals, representing the component of the series not explained by its immediate past.

This step is executed automatically by the method, requiring no manual VAR estimation from the user.

2. Why GeneNet applies a VAR before network estimation

Time series — especially financial ones — exhibit autocorrelation, meaning each variable is influenced by its own past.

If GeneNet were to compute partial correlations directly from the raw series, many of the detected connections would be spurious, driven only by the temporal structure of the individual series.

The internal VAR step serves to:

- Remove autocorrelation;
- Disentangle the endogenous temporal component from the cross-sectional relational component; and
- "Clean" the series by producing residuals free from their own past dependence.

3. The VAR residuals are used to estimate the network

After the VAR step, GeneNet does not use the original series. It relies exclusively on:

$$\varepsilon(t) \tag{27}$$

These residuals capture what remains after removing the temporal effect, isolating:

- Contemporaneous interactions;
- Pure conditional dependencies among assets;
- Variability attributable to other variables rather than to time.

4. Estimation of dynamic partial correlations

Using the residuals, GeneNet estimates:

$$\mathrm{PCOR_{ij}} = \text{partial correlation between } \varepsilon_\mathrm{i}(\mathrm{t}) \text{ and } \varepsilon_\mathrm{j}(\mathrm{t}) \text{ conditional on all other variables.} \tag{28}$$

These partial correlations undergo shrinkage and are subsequently subjected to significance testing.

5. Determination of putative direction

The dynamic method assumes a temporal structure in which:

- effects at $t-1$→ outcomes at t.

Thus, when the dynamic partial correlations suggest dependence between filtered series, GeneNet infers a putative direction:

$$X_i(t-1) \rightarrow X_j(t) \tag{29}$$

This direction is putative–i.e., based on temporal ordering rather than confirmed causal relations.

6. Summary

GeneNet uses a VAR(1) solely as a filtering step to remove serial autocorrelation; the network itself is estimated from the partial correlations of the VAR residuals, which represent the pure conditional dependencies among the variables.

Table 2 provides an overview of the network estimation procedures implemented in RStudio.

Table 2 - Procedures Performed by GeneNet to Estimate Dynamic Bayesian Networks in RStudio

Order	What Happens	R Function	Input	What Is Generated	Why Is It Necessary?	Output Format
1	Load and organize time-ordered data	Import and structuring	Original expression matrix	Time × financial asset matrix	Preserve temporal ordering	Observational matrix
2	Fit a VAR(1) from financial asset to financial asset	Internal step within *ggm.estimate.pcor(method = "dynamic")*	Original expression matrix	VAR residuals	Remove temporal autocorrelation	Residual matrix and
2.1	Estimate dynamic partial correlations (shrinkage)	*ggm.estimate.pcor(..., "dynamic")*	Residual matrix	Dynamic partial-correlation matrix	Capture conditional dependence among financial assets	Matrix (p × p)
3	Test the significance of edges	*network.test.edges()*	Partial-correlation matrix	Test statistics, p-values, local FDR	Verify that an edge is not noise	Table of financial-asset pairs
4	Edge selection	*extract.network(cutoff = α)* Observation: Efron (2007) suggests q-values between 0.05 and 0.15 as feasible choices for estimating networks, with these q-value limits interpreted as a conservative Bayesian factor for interpreting the FDR.	local FDR	Selected bipartite network	Control false positives	List or adjacency matrix
5	Determination of (putative) direction	Implicit in the dynamic method	Selected network	Arrows A → B	Represent temporal influence	Directed network
6	Visualization and export	*network.make.graph() / edge.info()* Additionally, the packages *Rgraphviz* and *Igraph* are employed.	Final network	Graph	Interconnections	DOT file / image

D. Descriptive Statistics of the Analyzed Stocks

Table 3 - Descriptive Statistics of Daily Stock Returns

Stocks (Ticker)	Nº	Mean (%)	Standard Deviation (%)	Min (%)	Pctl(25) (%)	Median (%)	Pctl(75) (%)	Max (%)
SANB11	3,780	0.002446	2.134232	-14.47255	-1.191562	0	1.206289	14.64963
TAEE11	4,250	0.024738	1.733931	-19.88508	-0.632915	0	0.708485	15.6638
CYRE3	4,465	0.011265	3.168307	-28.3029	-1.596107	0	1.637468	54.83727
LREN3	4,465	0.023018	2.663833	-23.72442	-1.425953	0	1.459879	19.23719
PCAR3	4,465	-0.016351	3.538475	-109.7087	0	0	0	87.1839
MRFG3	4,343	-0.001145	3.066896	-28.54623	-1.606892	0	1.550425	22.49198
RAIZ4	849	-0.142463	2.628552	-10.00835	-1.632688	0	1.384793	8.76096
CXSE3	916	0.038556	1.876431	-7.786509	-0.955238	0.080133	0.976988	11.46918
HAPV3	1,658	-0.055189	3.463617	-40.88956	-1.738454	0	1.589141	24.15302
ABEV3	4,465	0.025406	1.719111	-17.17594	-0.829352	0	0.93701	11.14413
CPFE3	4,465	0.017791	1.845412	-18.59675	-0.876647	0	0.952104	11.60436
CSAN3	4,465	-0.005563	2.544767	-20.37942	-1.34567	0	1.368981	20.6614
ALPA4	4,220	0.012119	2.56826	-30.24723	-1.226899	0	1.311487	25.40254
FLRY3	3,732	0.009068	1.976428	-16.55144	-1.10499	0	1.07433	14.92547
EZTC3	4,348	0.008167	2.748915	-22.66709	-1.437011	0	1.47429	22.60296
EGIE3	4,465	0.020724	1.711398	-12.08226	-0.871889	0	0.917437	15.00872
RADL3	4,465	0.094712	2.343202	-22.87967	-1.086351	0	1.22157	33.64715
TIMS3	4,465	0.003347	3.174406	-30.81368	-1.184226	0	1.18872	33.24802
TOTS3	4,465	0.04803	2.306991	-17.7996	-1.207036	0	1.328605	18.06502
TRPL4	4,465	0.025416	1.80299	-28.52039	-0.80187	0	0.896873	15.76289
CVCB3	2,748	-0.0811	3.801039	-42.73101	-1.869207	-0.130316	1.657431	28.0706
PETR3	4,465	0.007756	2.905908	-35.20543	-1.399351	0.022654	1.468356	20.50237
EQTL3	4,156	0.011033	3.15947	-157.6719	-0.906389	0	1.059276	12.68012
ELET3	4,465	0.006251	3.033406	-33.33303	-1.540987	0	1.449995	40.07593
PETR4	4,465	0.008142	2.875742	-35.23667	-1.359144	0.029527	1.412683	20.06706
KLBN11	2,719	0.028544	1.83312	-16.51433	-1.04197	0	1.110528	9.638604
ASAI3	958	-0.097156	2.460966	-9.672961	-1.628397	-0.155589	1.33612	13.71092
PETZ3	1,067	-0.125048	3.655702	-15.70037	-2.333338	-0.212987	1.852959	31.5853
MGLU3	3,396	0.008878	3.923384	-25.91731	-2.058871	-0.057362	1.965052	31.69127
ELET6	4,220	0.011777	2.750708	-22.41635	-1.266031	0	1.293886	27.82434
BRAV3	81	-0.165313	4.012167	-9.872132	-2.354158	-0.735085	1.854506	10.10867
CRFB3	1,854	-0.054446	2.327223	-14.00354	-1.420983	-0.058377	1.258949	12.01443
AURE3	690	-0.090717	1.500577	-11.56086	-0.768457	0	0.706617	8.908363
BBSE3	2,903	0.026838	1.840067	-10.78646	-0.949392	0.060866	1.018079	10.35229
WEGE3	4,465	0.074177	2.161062	-23.09206	-0.996492	0	1.147273	25.13144
RDOR3	1,005	-0.095739	2.420395	-8.830301	-1.608299	-0.185011	1.325305	8.924461
SMTO3	4,437	0.024376	2.575271	-39.09458	-1.187146	0	1.275031	22.52917
COGN3	3,179	-0.030165	3.483838	-39.93862	-1.674384	0	1.549376	59.78374
AZUL4	1,923	-0.09594	4.356725	-45.9937	-2.004418	-0.03773	1.861639	32.19634
RAIL3	2,424	0.001507	2.958543	-29.76324	-1.293628	0	1.330003	19.97797
ITSA4	4,465	0.016246	2.483462	-16.25189	-1.087524	0	1.107817	22.43192
VIVT3	4,465	-0.000347	1.815743	-13.19197	-0.939651	0	0.987666	11.61958
PRIO3	3,522	0.015402	4.445319	-45.47703	-1.951285	0	1.870111	60.79894
UGPA3	4,465	-0.287872	21.91624	-1.455.22	-0.65251	0	0.67484	78.54298
GGBR4	4,465	0.005102	2.750531	-66.05394	-1.282398	0	1.255248	21.1529

Stocks (Ticker)	Nº	Mean (%)	Standard Deviation (%)	Min (%)	Pctl(25) (%)	Median (%)	Pctl(75) (%)	Max (%)
VAMO3	975	-0.053985	3.080169	-12.3379	-1.702753	0	1.747798	13.54212
CMIN3	963	-0.061261	2.638523	-12.72022	-1.614402	-0.188497	1.337832	10.2362
CPLE6	4,465	0.028842	2.246195	-16.84706	-1.174203	0	1.264797	19.94002
MRVE3	4,328	-0.012957	3.26644	-24.478	-1.741011	0	1.759417	28.76823
EMBR3	4,465	0.021103	2.612044	-30.7091	-1.228514	0	1.264699	20.29285
VBBR3	1,688	-0.01409	2.551863	-16.59729	-1.399568	0	1.296018	15.45941
ALOS3	297	-0.070124	1.473575	-5.721433	-0.965783	-0.086732	0.714117	4.45088
HYPE3	4,147	0.018642	2.300227	-16.49877	-1.156871	0	1.181622	19.18003
SUZB3	4,465	0.052465	5.62	-200.4179	0	0	0	299.5732
CSNA3	4,465	-0.004805	3.442707	-33.14569	-1.822177	0	1.721211	19.62756
BPAC11	1,953	0.091556	2.830497	-31.28402	-1.27678	0	1.41345	24.60323
CCRO3	4,465	0.006821	2.350308	-19.76103	-1.239382	0	1.207253	20.01345
STBP3	3,950	0.066486	2.747385	-26.84588	0	0	0.098075	106.6524
B3SA3	4,257	-0.02865	3.204633	-108.2746	-1.42988	0	1.462226	17.60072
NTCO3	1,250	-0.084641	3.417355	-27.76317	-1.722508	-0.103785	1.585247	15.85109
IRBR3	1,847	-0.090196	3.808685	-38.51375	-1.548569	0	1.439135	26.74135
VIVA3	1,296	-0.016144	3.241153	-38.47292	-1.505681	0	1.501217	26.31463
LWSA3	1,218	-0.036045	4.25391	-22.31435	-2.374514	-0.124555	2.407972	25.78291
RECV3	913	0.006512	2.913423	-21.4007	-1.780254	-0.052783	1.811743	11.06887
BRKM5	4,465	-0.006239	3.009904	-28.06809	-1.527329	-0.039834	1.376308	27.57693
IGTI11	776	-0.021843	2.082705	-8.810727	-1.264688	-0.015117	1.239513	7.16725
MULT3	4,324	0.022768	2.306704	-25.34731	-1.159012	0	1.15521	26.98049
CMIG4	4,465	0.010735	2.484889	-23.63889	-1.197856	0	1.288427	16.38476
SLCE3	4,353	0.041026	2.603356	-24.7241	-1.232041	0	1.307215	32.56833
BRAP4	4,465	0.007593	2.701958	-22.72463	-1.31095	0	1.347226	16.76439
ENEV3	4,229	-0.06782	3.277604	-44.18327	-1.336321	0	1.227014	28.76821
BEEF3	4,329	-0.030037	2.665003	-20.5485	-1.405186	0	1.31871	16.70541
BBAS3	4,465	0.017723	2.595258	-23.78914	-1.306729	0	1.315807	18.82563
YDUQ3	4,091	0.001922	3.124676	-29.00104	-1.596585	0	1.619471	25.09278
GOAU4	4,465	-0.017782	2.947906	-23.91844	-1.572582	0	1.557662	18.76267
USIM5	4,465	-0.027244	3.409367	-26.85427	-1.765025	-0.049864	1.626051	30.08923
ITUB4	4,465	0.019334	2.170086	-19.80153	-1.109953	0	1.128585	21.00393
SBSP3	4,465	0.046746	2.38747	-19.63795	-1.21963	0.048925	1.304429	16.09304
BRFS3	4,465	0.012872	2.568193	-21.99871	-1.226999	0	1.255249	16.90601
RENT3	4,465	0.036862	2.697434	-26.9333	-1.390164	0	1.417846	24.09552
VALE3	4,465	0.011461	2.667381	-28.18222	-1.377065	0	1.34682	19.35735
ENGI11	3,763	0.036326	2.049572	-18.74254	-0.596306	0	0.699351	26.26767
BBDC3	4,465	0.005865	2.161448	-17.45041	-1.125995	0	1.114515	15.4641
AZZA3	3,455	0.009573	2.499479	-18.23216	-1.377795	0	1.366678	15.37447
BBDC4	4,220	-0.015282	2.412985	-56.65282	-1.098732	0	1.105343	19.98888
JBSS3	4,406	0.037968	3.089754	-37.60512	-1.609214	0	1.621653	24.09157

E. Listed Companies and Their Respective Sectors

Table 4 - Names of Listed Companies and Their Business Sectors

Nº	Stocks (Ticker)	Company	Business Sector
1	ABEV3	Ambev S/A	Beverages / Consumption
2	ALOS3	Allos	Shopping Centers
3	ALPA4	Alpargatas	Footwear / Apparel
4	ASAI3	Assaí	Food Retail (Cash & Carry)
5	AURE3	Auren	Energy / Generation
6	AZUL4	Azul	Aviation
7	AZZA3	Azzas 2154	Apparel / Footwear
8	B3SA3	B3	Stock Exchange / Financial Services
9	BBAS3	Banco do Brasil	Banks
10	BBDC3	Bradesco	Banks
11	BBDC4	Bradesco	Banks
12	BBSE3	BB Seguridade	Insurance / Pensions
13	BEEF3	Minerva	Food / Animal Protein
14	BPAC11	BTG Pactual	Banks / Financial Services
15	BRAP4	Bradespar	Holdings (Investment Companies)
16	BRFS3	BRF S.A.	Food / Animal Protein
17	BRKM5	Braskem	Petrochemicals / Chemicals
18	CCRO3	CCR S.A.	Concessions / Highways / Logistics
19	CMIG4	Cemig	Energy / Generation and Distribution
20	CMIN3	CSN Mineração	Mining
21	COGN3	Cogna	Education
22	CPFE3	CPFL Energia	Energy / Generation and Distribution
23	CPLE6	Copel	Energy / Generation and Distribution
24	CRFB3	Carrefour Brasil	Food Retail
25	CSAN3	Cosan	Energy / Logistics / Sugar-Ethanol
26	CSNA3	Companhia Siderúrgica Nacional (CSN)	Steelmaking / Mining
27	CVCB3	CVC Brasil	Tourism / Travel
28	CXSE3	Caixa Seguridade	Insurance
29	CYRE3	Cyrela	Construction / Real Estate
30	EGIE3	Engie Brasil	Energy / Generation
31	ELET3	Eletrobras	Energy / Generation and Transmission
32	ELET6	Eletrobras	Energy / Generation and Transmission
33	EMBR3	Embraer	Aerospace / Defense
34	ENEV3	Eneva	Energy / Generation and Commercialization
35	ENGI11	Energisa	Energy / Distribution
36	EQTL3	Equatorial Energia	Energy / Distribution
37	EZTC3	Eztec	Construction / Real Estate
38	FLRY3	Fleury	Healthcare / Diagnostics
39	GGBR4	Gerdau	Steelmaking
40	GOAU4	Metalúrgica Gerdau	Steelmaking
41	HAPV3	Hapvida	Healthcare / Health Insurance
42	HYPE3	Hypera	Pharmaceuticals
43	IGTI11	Iguatemi	Shopping Centers
44	IRBR3	IRB Brasil RE	Reinsurance
45	ITSA4	Itaúsa	Holding Company (main asset: Itaú Unibanco)
46	ITUB4	Itaú Unibanco	Banks
47	JBSS3	JBS	Food / Animal Protein
48	KLBN11	Klabin	Pulp and Paper
49	LREN3	Lojas Renner	Apparel Retail
50	LWSA3	LWSA (antiga Locaweb)	Technology / Digital Services
51	MGLU3	Magazine Luiza	Retail / E-commerce
52	MRFG3	Marfrig	Food / Animal Protein
53	MRVE3	MRV	Construction / Real Estate
54	MULT3	Multiplan	Shopping Centers
55	NTCO3	Grupo Natura	Cosmetics / Personal Care

Nº	Stocks (Ticker)	Company	Business Sector
56	PCAR3	Pão de Açúcar (GPA)	Food Retail
57	PETR3	Petrobras	Oil / Gas
58	PETR4	Petrobras	Oil / Gas
59	PETZ3	Petz	Retail / Pet Products and Services
60	PRIO3	PetroRio (PRIO)	Oil / Oil & Gas
61	RADL3	Raia Drogasil	Pharmaceutical Retail
62	RAIL3	Rumo	Logistics / Rail Transport
63	RAIZ4	Raízen	Energy / Biofuels
64	RDOR3	Rede D'Or	Healthcare / Hospitals
65	RECV3	PetroReconcavo	Oil / Oil & Gas
66	RENT3	Localiza	Vehicle Rental / Fleets
67	RRRP3	3R Petroleum	Oil / Oil & Gas
68	SANB11	Santander Brasil	Banks
69	SBSP3	Sabesp	Sanitation
70	SLCE3	SLC Agrícola	Agribusiness / Agriculture
71	SMTO3	São Martinho	Agribusiness / Sugar-Ethanol
72	STBP3	Santos Brasil	Logistics / Ports
73	SUZB3	Suzano	Pulp and Paper
74	TAEE11	Taesa	Energy / Transmission
75	TIMS3	TIM	Telecommunications
76	TOTS3	Totvs	Technology / Software
77	TRPL4	Transmissão Paulista (CTEEP)	Energy / Transmission
78	UGPA3	Ultrapar	Fuel Distribution
79	USIM5	Usiminas	Steelmaking
80	VALE3	Vale	Mining / Metals
81	VAMO3	Vamos	Truck & Machinery Leasing
82	VBBR3	Vibra Energia (antiga BR Distribuidora)	Fuel Distribution
83	VIVA3	Vivara	Jewelry / Retail
84	VIVT3	Telefônica Brasil (Vivo)	Telecommunications
85	WEGE3	WEG	Electrical Equipment / Automation
86	YDUQ3	Yduqs	Education

F. Partial Correlations of Dynamic Bayesian Networks

Table 5 - Partial Correlations Calculated

SANB11→BBAS3	SANB11→BBDC4	SANB11→BBDC3	SANB11→ITUB4	SANB11→GGBR4	TAEE11→YDUQ3	TAEE11→TRPL4
0.06074	0.08419	0.08549	0.11996	0.12517	0.06953	0.09466
CYRE3→ITUB4	CYRE3→BBAS3	CYRE3→GOAU4	LREN3→VIVA3	LREN3→CYRE3	PCAR3→CRFB3	MRFG3→JBSS3
0.06937	0.07618	0.08228	0.07328	0.09588	0.08584	0.20793
RAIZ4→CSAN3	RAIZ4→RDOR3	RAIZ4→PETZ3	RAIZ4→VBBR3	RAIZ4→IGTI11	CXSE3→VAMO3	CXSE3→RECV3
0.08231	0.09064	0.10037	0.1048	0.16899	0.07119	0.09756
CXSE3→BBSE3	CXSE3→CMIN3	CXSE3→IGTI11	HAPV3→NTCO3	HAPV3→VBBR3	HAPV3→VIVA3	HAPV3→RDOR3
0.09856	0.09891	0.11939	0.06551	0.08992	0.10593	0.12523
ABEV3→BBDC3	CPFE3→CMIG4	CPFE3→CPLE6	CSAN3→MULT3	CSAN3→LREN3	CSAN3→CCRO3	ALPA4→JBSS3
0.06121	0.09494	0.14794	0.06374	0.06401	0.07004	0.05888
ALPA4→HAPV3	ALPA4→AZZA3	ALPA4→VIVA3	FLRY3→BBSE3	FLRY3→CVCB3	FLRY3→MGLU3	FLRY3→RDOR3
0.06423	0.06517	0.07089	0.0606	0.06307	0.06345	0.06604
FLRY3→AZZA3	EZTC3→MRVE3	EZTC3→CYRE3	EGIE3→BBDC3	EGIE3→CCRO3	EGIE3→CPLE6	EGIE3→CPFE3
0.06969	0.1795	0.19594	0.06046	0.08386	0.08434	0.15382
RADL3→WEGE3	RADL3→ABEV3	TIMS3→B3SA3	TIMS3→VIVT3	TOTS3→RENT3	TOTS3→PETZ3	TOTS3→LWSA3
0.06764	0.07636	0.08359	0.13759	0.06093	0.06768	0.08397
TRPL4→CPFE3	TRPL4→CPLE6	TRPL4→CMIG4	TRPL4→EGIE3	CVCB3→VIVA3	CVCB3→AZUL4	EQTL3→ENGI11

0.06477	0.07723	0.11187	0.11591	0.10645	0.32277	0.06195
EQTL3→CPFE3	ELET3→BBDC3	PETR4→PETR3	KLBN11→GGBR4	KLBN11→CMIN3	KLBN11→BRKM5	KLBN11→WEGE3
0.07896	0.06091	0.88336	0.06576	0.07877	0.08387	0.1155
ASAI3→PETZ3	ASAI3→LWSA3	ASAI3→IGTI11	ASAI3→VAMO3	PETZ3→VBBR3	PETZ3→CVCB3	PETZ3→IGTI11
0.06915	0.07147	0.07229	0.1973	-0.06654	0.06893	0.07818
PETZ3→NTCO3	PETZ3→VAMO3	PETZ3→LWSA3	MGLU3→PETZ3	MGLU3→AZZA3	MGLU3→LWSA3	ELET6→ITUB4
0.0999	0.11854	0.12236	0.06097	0.07624	0.10166	-0.06109
ELET6→ELET3	BRAV3→RECV3	BRAV3→ALOS3	CRFB3→HAPV3	CRFB3→ASAI3	AURE3→RDOR3	AURE3→ASAI3
0.73264	0.06056	0.1029	0.08136	0.25641	0.0651	0.08965
AURE3→ALOS3	AURE3→IGTI11	BBSE3→CCRO3	BBSE3→COGN3	BBSE3→SANB11	BBSE3→BBAS3	WEGE3→TOTS3
0.09391	0.12843	0.08038	0.08333	0.10428	0.1143	0.06354
WEGE3→ABEV3	RDOR3→IGTI11	RDOR3→LWSA3	RDOR3→VAMO3	RDOR3→PETZ3	SMTO3→RAIZ4	SMTO3→VIVA3
0.0638	0.06928	0.0723	0.12152	0.1638	0.0649	0.07046
SMTO3→CSAN3	COGN3→YDUQ3	RAIL3→CMIG4	RAIL3→MGLU3	RAIL3→CCRO3	RAIL3→BBSE3	ITSA4→ITUB4
0.17494	0.35416	0.0599	0.06574	0.06811	0.09306	0.42287
VIVT3→CPLE6	VIVT3→ABEV3	PRIO3→VIVA3	PRIO3→AZZA3	PRIO3→PETR3	PRIO3→RECV3	UGPA3→ITSA4
0.06082	0.07462	0.06773	0.06922	0.07032	0.1133	-0.09496
GGBR4→BBDC3	GGBR4→CSNA3	GGBR4→BBDC4	GGBR4→GOAU4	VAMO3→IGTI11	CMIN3→RDOR3	CMIN3→VAMO3
0.06302	0.08049	-0.14399	0.54935	0.16728	0.07714	0.0906
CMIN3→CSNA3	CMIN3→RAIZ4	MRVE3→LREN3	MRVE3→CYRE3	EMBR3→EZTC3	EMBR3→CYRE3	EMBR3→PETR4
0.11931	0.15172	0.06864	0.23114	-0.05981	0.06585	-0.07059
EMBR3→AZUL4	VBBR3→VIVA3	VBBR3→AZUL4	VBBR3→BPAC11	ALOS3→AZUL4	ALOS3→VAMO3	ALOS3→ASAI3
0.13598	0.06683	0.07603	0.14808	0.05943	0.07437	0.14212
ALOS3→IGTI11	HYPE3→RENT3	HYPE3→SANB11	HYPE3→LREN3	HYPE3→B3SA3	SUZB3→KLBN11	CSNA3→GOAU4
0.18274	0.06403	0.06427	0.0795	0.12719	0.11822	0.0966
CSNA3→VALE3	BPAC11→VIVA3	BPAC11→AZUL4	CCRO3→RENT3	STBP3→BPAC11	STBP3→AZUL4	B3SA3→CYRE3
0.17145	0.05895	0.06413	0.09852	0.07362	0.07805	0.06704
NTCO3→IGTI11	NTCO3→VBBR3	NTCO3→BPAC11	NTCO3→LWSA3	NTCO3→VIVA3	IRBR3→HAPV3	IRBR3→BPAC11
0.07476	0.07598	0.08775	0.1207	0.18333	0.07497	0.08295
IRBR3→CVCB3	VIVA3→AZUL4	LWSA3→VAMO3	LWSA3→VIVA3	LWSA3→BPAC11	RECV3→IGTI11	RECV3→VAMO3
0.12778	0.15185	0.07695	0.08743	0.13053	0.08476	0.10656
RECV3→CMIN3	RECV3→RAIZ4	BRKM5→BRFS3	BRKM5→CSAN3	MULT3→BBDC3	MULT3→EZTC3	MULT3→IGTI11
0.12389	0.12605	0.06683	0.07829	0.05968	0.07042	0.1071
MULT3→LREN3	CMIG4→CPLE6	SLCE3→MRFG3	SLCE3→SMTO3	BRAP4→PETR3	BRAP4→PETR4	BRAP4→VALE3
0.11447	0.24416	0.05884	0.13004	-0.06337	0.07567	0.69555
BEEF3→SMTO3	BEEF3→BRFS3	BEEF3→JBSS3	BEEF3→MRFG3	BBAS3→BBDC4	BBAS3→PETR4	BBAS3→BBDC3
0.06066	0.09633	0.10808	0.21712	0.06933	0.08255	0.10441
BBAS3→ITUB4	YDUQ3→RENT3	GOAU4→BRAP4	USIM5→BRAP4	USIM5→BBAS3	USIM5→GOAU4	USIM5→CSNA3
0.19365	0.08324	0.08886	0.05999	0.07195	0.21904	0.35977
ITUB4→PETR3	SBSP3→CMIG4	SBSP3→CPLE6	BRFS3→LREN3	BRFS3→ABEV3	BRFS3→MRFG3	BRFS3→JBSS3
0.06335	0.07071	0.12543	0.07384	0.0887	0.11321	0.11602
RENT3→MRVE3	RENT3→LREN3	RENT3→MULT3	VALE3→PETR4	VALE3→PETR3	ENGI11→BPAC11	BBDC3→ITUB4
0.05979	0.06528	0.07475	-0.11262	0.16852	0.06626	0.24541
AZZA3→VIVA3	AZZA3→LREN3	BBDC4→ELET6	BBDC4→ELET3	BBDC4→GOAU4	BBDC4→ITUB4	BBDC4→BBDC3
0.07263	0.09446	0.06607	-0.07427	0.08963	0.15256	0.46081