

Dinâmica de Correlações no Mercado Financeiro Bovespa&BMF



Daniel Penalva

Dinâmica de Correlações no Mercado Financeiro Bovespa&BMF



Esta dissertação contou com o apoio financeiro da CAPES e CNPQ (Brazil).

Dinâmica de Correlações no Mercado Financeiro Bovespa&BMF

DISSERTAÇÃO

apresentada ao Instituto de Física Teórica
Universidade Estadual Paulista, Brasil

em cumprimento aos requisitos para a obtenção do título de

Mestre em Física

na modalidade de mestrado acadêmico

maio de 2011

Daniel Penalva

nascido em Campinas, Brazil

Comissão Examinadora

Prof. Dr. Antonio Fernando Crepaldi
FEB -UNESP

Prof. Dr. Fernando Fagundes Ferreira
EACH -USP

Prof. Dr. Gerson Francisco (orientador)
IFT-UNESP

Instituto de Física Teórica – UNESP
R. Dr. Bento Teobaldo Ferraz 271, bloco II
01140-070 São Paulo, Brasil

Às memórias de Rosângela Lazarini Penalva, Artur Penalva e Benedito Penalva.
Também a Lucas Penalva e Mateus Penalva, irmãos, e a Marília Sorgi, avó,
fontes de minha persistência.

Sumário

Agradecimentos	viii
Resumo	ix
Resumo em Inglês	x
1 Sistemas Complexos: noções básicas intuitivas	1
1.1 Complexidade: classes de regularidades emergentes em sistemas únicos	1
2 Mercado Financeiro, uma estrutura coletiva	3
2.1 Absorção e circulação de informações economicas pelo mercado: o mercado é eficiente ?	4
2.2 O funcionamento do mercado	4
2.2.1 Escala de preços	5
2.2.2 Escala do tempo	5
3 Métodos para tratamento de correlações em séries temporais	7
3.1 Elementos de Probabilidade e Estatística	7
3.1.1 Axiomas da Teoria de Probabilidades	8
3.2 Erros Experimentais	10
3.3 Correlação Cruzada Linear	10
3.4 Espaços Ultramétricos	11
3.4.1 Espaços Ultramétricos Subdominantes	13
3.5 Árvore de mínimo alcance, taxonomia e o algoritmo de Kruskal	14
3.5.1 Teoria de Grafos	15
4 Análise do mercado Bovespa&BMF	20
4.1 Dados e metodologias gerais	20
4.2 Escalas de correlação cruzada: efeito Epps	21

4.2.1	Influência de lead-lag na relação funcional entre retorno de ações	22
4.2.2	A escala do efeito no tempo inter-transações	24
4.2.3	O que o efeito Epps diz sobre a organização do mercado ? O que é o efeito Epps ?	33
4.3	Árvores de mínimo alcance, hierarquias e dinâmica de setores . . .	35
4.3.1	Tratamento de dados da Bovespa&BMF	35
4.3.2	Carteiras examinadas	36
4.3.3	Ocupação de árvore, vértices raízes e camada de ocupação média	36
4.3.4	Equilíbrio polar entre 2 setores	37
4.3.5	Árvores de mínimo alcance nas escalas temporais de 2, 6 e 10 minutos	38
4.3.6	Informação fornecida por árvores de mínimo alcance e taxonomias	53
4.4	Conclusões e Perspectivas	61

Agradecimentos

Reconheço todos amigos do Alojamento USP São Carlos que convivem comigo e velhos amigos de São Carlos, amigos de Franca e São Paulo.

Reconheço também familiares, pessoas especiais, de Ribeirão Preto e Franca pelo apoio sentimental e logístico.

Reconheço e agradeço minha namorada por me estimular a trabalhar.

Reconheço Capes/CNPQ pelo auxílio financeiro.

Agradeço o professor Antônio Crepaldi que me forneceu os dados sem os quais essa dissertação não existiria.

Agradeço ao professor Gerson Francisco por me propiciar esta oportunidade de aprendizagem única.

Resumo

Em sistemas onde muitos agentes interagem, permitindo obter medidas que podem se apresentar intermitentes, muitas vezes podemos extrair padrões que denotam comportamento de grupo destes agentes, este é o caso do mercado financeiro e sua estrutura de correlações emergentes. Este trabalho visa reproduzir e sintetizar o que é entendido como estrutura de correlações no mercado financeiro. A análise da estrutura consistirá de 2 partes, uma dinâmica, acessando dependências temporais, e outra topológica e econômica, acessando a importância das conexões entre ações. Na análise dinâmica são investigadas a correlação instantânea, quanto o comportamento ao longo das escalas de tempo, e a não instantânea, quanto ao decaimento temporal em relação ao máximo de correlação. A topologia é analisada obtendo-se um grafo a partir da matriz de correlação instantânea e analisando a conectividade dos vértices, partindo do mais conectado (chama-se raiz) analisa-se os diversos clusters de ações obtidos, comparando com a classificação econômica conhecida. A análise topológica é feita em várias escalas de correlação instantânea visando a comparação entre elas. Introduzo noções gerais de sistemas complexos no capítulo 1. No capítulo 2 dou uma breve descrição do mercado através das variáveis mais importantes e seu comportamento, i.e. as escalas de preços e de tempo. O capítulo 3 descreve os métodos utilizados para análise da estrutura de correlação do mercado, é apresentado o estimador de Pearson para correlação linear bem como o método de Kruskal, utilizado para obter o grafo árvore que contém todas as ações e minimizar a soma das arestas (ponderadas pela distância definida a partir da correlação). No capítulo 4 apresento os resultados referentes à análise da estrutura de correlações para o mercado Bovespa.

palavras chaves: sistemas complexos, econofísica, mercado, correlações, ultrametrica, efeito epps, árvore de mínimo alcance

Abstract

In systems where many agents interact, allowing for measures that may be erratic, many times we can extract behavior patterns that denote a group of agents, this is the case of financial market and its emerging structure of correlations. This work aims to reproduce and synthesize what is perceived as a correlation structure in financial markets. Analysis of the structure will consist of 2 parts, one dynamic, accessing temporal dependencies, and other topological and economical by accessing the importance of connections between assets. In analysis of dynamics are investigated instantaneous correlation, its behavior across scales of time, and the not instantaneous, its decay from the maximum correlation. The topology is analyzed by a graph from the instantaneous correlation matrix and analyzing the connectivity of vertices, starting from the most connected (called root) analyzes the various clusters of shares obtained by comparing with known economic classification. The topological analysis is performed at several scales of correlation in order to instantly compare them. Introduce general notions of complex systems in Chapter 1. In Chapter 2 give a brief description of market through the most important variables and their behavior, ie the ranges of price and time. Chapter 3 describes the methods used for analysis the correlation structure of the market, the estimator is presented by Pearson's linear correlation and the Kruskal method is used to obtain the graph tree containing all assets and to minimize the sum of edges (weighted by the distance defined from the correlation). In Chapter 4 I present the results of the analysis of correlation structure for the market Bovespa.

keywords: complex systems, econophysics, market, correlation, ultrametrics, epps effect, minimal spanning tree

1

Sistemas Complexos: noções básicas intuitivas

O estudo através da metodologia de sistemas complexos, apesar de já conhecido dos físicos de matéria condensada desde os anos 50 [1], tem se estabelecido em outras áreas, como estudo do mercado financeiro, apenas recentemente - como será visto no capítulo 2, neste caso devido aos recentes fracassos das teorias econômicas clássicas [4] em considerar aspectos emergentes do agregado no fenômeno econômico - com a utilização de ferramental formal já empregado em sistemas físicos, como é o caso da Teoria de Espaços Ultramétricos aqui utilizada para um novo entendimento metodológico do mercado, sob a luz da intuição física de um sistema complexo.

Neste capítulo, 1.1 passa sucintamente pelas noções centrais de sistemas complexos. Segue-se, no capítulo 3, exibindo formalismo necessário para abordagem quantitativa e qualitativa dos conceitos descritos em 1.1, aqui é dada ênfase à Teoria de Espaços Ultramétricos e Teoria de Grafos, ferramental utilizada para a análise de correlações de mercado nesta dissertação e , ainda antes disso, trago uma breve introdução à teoria de probabilidades e estatística elementares.

1.1 Complexidade: classes de regularidades emergentes em sistemas únicos

A noção de complexo vem à mente de um observador na medida em que diferencia padrões, em sistemas para os quais não existem regularidades óbvias como uma sequência de mesmo número, ou são apenas ruídos assim como ouvido na sintonização do rádio. Sendo o primeiro caso ordem total, o segundo total ale-

atoridade, o observador entende as nuances do contraponto musical como sendo mais complexo que os dois casos, a não ser que se prove conter uma quantidade grande de padrões dentro do aleatório.

Uma das teorias elementares das quais oferecem conceitos de informação, a Teoria de Complexidade Algorítmica, fornece uma base intuitiva sobre o qual podemos adquirir noção das dificuldades envolvidas em medir/formalizar tais padrões dentro de um sistema qualquer. Suponha que uma das regularidades do sistema é traduzida por alguma linguagem teórica e existe menor programa de computador que lê, imprime esta propriedade em uma fila de caracteres e interrompe o computador, tem tamanho de bits N , então N é dita complexidade desta regularidade no sistema traduzido pela determinada linguagem (este computador é também chamado de máquina de Turing).¹ As diversas linguagens teóricas produzidas são comparadas através das complexidades resultantes, sendo que aquela com menor complexidade adotada, ou ainda adaptada para melhor descrever as classes de regularidades do sistema. Qualquer ciência encontra este viés prático de duas maneiras: na determinação de previsões teóricas de uma linguagem adotada, na adoção de uma linguagem adequada partindo de conjuntos de dados.

¹Alan Mathison Turing (23/06/1912 - 07/06/1954), foi um matemático, lógico e cientista da computação foi muito influente nas ciências da computação formalizando conceitos como de "algoritmo" e computação.



Mercado Financeiro, uma estrutura coletiva

Até 2008 existia dentro da economia uma crença, apoiada pelas teorias macro-econômicas modernas, que a demanda do agregado é resultado da soma das partes individuais. Estes indivíduos por sua vez agem racionalmente otimizando seu ganho e forma-se então o mecanismo de preço e assim o mercado deve ser levado ao equilíbrio por meio deste mecanismo. Entretanto a crise de crédito, impulsionada pela contaminação interna do mercado - por investimentos sem fundo - nos mais diversificados níveis de risco, levantou dúvidas sobre a capacidade do mercado de absorver riscos. A crise teve origens endógenas, i.e. mercado de crédito interbancário foi congelado devido à incerteza de um banco com relação à posição financeira dos outros, não de um chacoalhão externo. No mercado real, onde há interação direta entre investidores com interesses e estratégias diversas, não há razão para acreditar que agentes com informações limitadas - locais específicas e globais apenas estatísticas - possam obter uma rede de negociações racionalmente justa, de forma racional, em torno do equilíbrio, e mesmo se considerando a emergente coordenação possível, causada pela correlação das negociações entre os agentes, claramente ela não caminha para uma estabilidade ou equilíbrio. Mais do que isso, os eventos mostram que devemos investigar a teoria econômica - cuja descrição de coordenação - capaz de contabilizar os fatos empíricos observados e explique porque essa coordenação, livre de controle, pode levar o mercado a uma crise.

2.1 Absorção e circulação de informações econômicas pelo mercado: o mercado é eficiente ?

A eficiência do mercado, no contexto clássico após os trabalhos de L. Bachelier¹, é obtida quando agentes absorvem informação externa ao mercado e respondem corrigindo suas ações instantaneamente, de forma racional, de modo que todos os agentes tem o mesmo objetivo em posições diversas e interpretam igualmente a mesma informação. No início dos anos 90 as premissas teóricas da eficiência clássica foram questionadas pela posição contrastante à realidade exibida pela estrutura e fatos de mercado conhecidos. Em primeiro lugar destacou-se que o perfil comportamental dos diversos agentes financeiros não é homogêneo, sendo variados horizontes de investimento (por exemplo, muito breve para um especulador e muito longo para bancos centrais), pelo objetivo de diferentes margens de lucro, cada investidor toma o risco referente às estratégias adequadas para este período. Ainda, longe de ser imediatamente interpretada e ocasionar mudança racional, qualquer mudança de informação será interpretada diferenciadamente por agentes diversos que reagirão ao seu tempo, com ações racionais ou não, resultando em diferentes direções e velocidades de propagação de informação entre ações em setores inomogêneos de mercado.

O mercado, apresentando-se assim, deve ser analisado cuidadosamente para que, se houver algum tipo de eficiência, entenda-se como é realizada sua dinâmica. A seguir dou uma breve descrição do mercado físico, suas regras de funcionamento que levam às hipóteses elementares para descrição matemática do fenômeno.

2.2 O funcionamento do mercado

Apesar de cada mercado acionário possuir suas regras particulares, tais regras delimitam apenas o limite de funcionamento (como por exemplo quantia mínima de alteração do preço em uma compra ou venda, horário de funcionamento do mercado) ou ferramentas contratuais disponíveis (e.g. contrato de compra/venda obrigatória de ações, em data futura, sob um determinado valor, contrato de compra/venda opcional, exercível ou não, de ações sob um valor para o futuro).

O mecanismo de formação dá-se pelo processo de proposta-pedido: o vendedor da ação registra um pedido de com um valor, compradores registram propostas, aquela com valor igual ou maior que um pedido de venda finaliza uma negociação, o conjunto de todos esses registros de propostas-pedidos é chamado *livro de ordens*.

Vale notar o papel de um tipo particular de investidor cuja a estratégia reside em apostas à curto e médio prazo, estes são chamados de *especuladores*. Por

¹Louis Bachelier (1887 - 1946), orientado por Henri Poincaré, defendeu sua tese chamada *Theorie de la Spéculation* em 29 de março de 1900. É tido como o primeiro a modelar o processo estocástico chamado de movimento browniano, aplicando ao mercado de opções.

investir nesta frequência, este tipo de participante, fornece continuidade (liquidez) aos preços, isso diminui a chance de haver oportunidade para ganhos livres de riscos em transações, as chamadas oportunidades de *arbitragem*.

A ausência de arbitragem é um dos fundamentos da mencionada teoria do mercado eficiente, como será visto na seção 4.2.1, os resultados, para os últimos anos de transação, tem corroborado a hipótese do mercado eficiente para ações com correlações significantes.

2.2.1 Escala de preços

Talvez a primeira característica notável, até mesmo para quem não trabalha no mercado, é o dinheiro como unidade elementar do preço das ações transacionadas. São três os principais fatores que influenciam no valor da moeda (capítulo 5 da referência [6]): inflação de preços, crescimento econômico e flutuações no mercado de moedas global.

Esses três fatores, por sua vez, podem ser tomados como aleatórios, já que possuem relação complexa entre si e com os diversos fatores econômicos que os causam. A escolha de uma função do preço que represente fielmente as condições de câmbio depende, portanto, da escala de tempo em que a ordem da mudança de preços é comparável com as alterações inflacionárias.

Nas transações de altas frequências é possível tomar o ajuste inflacionário como nulo, já que em um país com situação econômica estável a inflação cresce décimos de porcentagem ao mês. Apesar da análise ter sido feita em dados de alta frequência, a extensão da média para todo um ano foi tomada na hipótese de inflação nula.

A escolha da equação 4.1, como função do preço estudada, traz facilidades como a possibilidade de determinação do retorno em uma escala através da soma dos retornos na escala anterior. Entretanto é importante notar que a amostragem de preços não é feita à unidade de tempo constante, ela depende da concordância entre o preço pedido e o ofertado no livro de ordens, portanto devo entender os possíveis tratamentos de escalas temporais do mercado.

2.2.2 Escala do tempo

No mercado acionário todas transações acontecem no tempo do leilão (tempo de funcionamento do mercado), na Bovespa delimitado por um tempo total de 7 horas.

Como foi dito acima, o tempo de realização das negociações dependem da saída de propostas negociáveis no livro de ofertas, atribuindo caráter aleatório ao tempo de espera entre negociações. A escolha do retorno deve carregar uma hipótese, qual será referente ao caráter aleatório do tempo de espera.

A análise feita, sem levar em conta a dinâmica do livro de ofertas, considera o preço como constante, e determinado pela última transação, quando interior ao

intervalo de espera. Outras abordagens, como considerar o tempo de realização das transações constante (apenas levando em conta o número de transações), desprezam a natureza do mecanismo gerador da série e só devem ser utilizados em escalas maiores que um dia com o devido cuidado ².

²A escolha do preço de abertura, para escalas diárias ou maiores, acarreta na contabilização de valores exteriores ao leilão, pois realiza-se no chamado período after market, também nos 15 minutos antecedentes à abertura do leilão, negociações entre agentes às portas fechadas. Se o preço tomado para análise for o de fechamento esta alteração estará diluída no fenômeno diário estudado.

3

Métodos para tratamento de correlações em séries temporais

Tendo visto em 1.1 que a própria tentativa de construção para um conceito de complexidade leva-me a considerar estatisticamente grandes sistemas, caso contrário não seria possível investigar padrões cuja complexidade cresce com o seu tamanho e, por meio da evolução temporal, possuem inerente incerteza de medida, uma abordagem natural para um sistema deste tipo é através da estatística e da atribuição de variáveis estocásticas, formalizadas aqui neste capítulo.

Introduzo a linguagem elementar para a abordagem dos dados em 3.1, incluindo a metodologia estatística adotada para estimativa de margens de erro. Prossigo introduzindo o formalismo probabilístico básico na seção 3.1.1 para introduzir as estatísticas utilizadas, bem como a interpretação estocástica de algumas delas.

3.1 Elementos de Probabilidade e Estatística

Como foi observado no capítulo 2, a crítica às teorias de mercado, no ponto de vista do equilíbrio eficiente, traz a necessidade do tratamento rigoroso de premissas de modo que, como em toda teoria científica, seja verificável experimentalmente, ou seja, que os dados possam corroborar ou negar, sem maiores pre-assunções, cada premissa teórica. Neste contexto surge naturalmente a necessidade da estatística como instrumento de investigação e validação científica e, em contrapartida, a teoria de probabilidades oferecer *framework* com rigor formal para o tratamento de qualquer ferramenta estatística e acessibilidade lógica

ao desenvolvimento teórico - experimental (experimental - teórico). Tendo em mente esta importância, a teoria de probabilidade será descrita com rigor formal, mas sempre mantendo em foco a prática científica, as noções essenciais de estatística teram tratamento semelhante sendo exemplificadas à luz de sua necessidade. Alguns aspectos específicos da teoria de probabilidades bem como da estatística, concernentes à natureza do fenômeno em questão, serão brevemente abordados, com a intenção de exibir sua importância com relação ao sistema estudado.

A abordagem das teorias elementares deste capítulo é apenas suficiente para uma exposição crítica do sistema mercado financeiro, para que na sequência introduza as teorias quais servirão de base analítica para a investigação pretendida.

3.1.1 Axiomas da Teoria de Probabilidades

Os axiomas de Kolmogorov nos provem da linguagem básica, em teoria de probabilidades, para estudarmos fenômenos aleatórios mais gerais:

1. DEFINIÇÃO. *Seja a tripla $(\Omega, \mathcal{F}, \mathbf{P})$ aonde*

(a) Ω é um conjunto de resultados experimentais possíveis ω ,

(b) \mathcal{F} é uma σ -álgebra de eventos, subconjuntos, de Ω ,

(c) \mathbf{P} é uma medida de probabilidade em \mathcal{F} ,

é dita modelo probabilístico ou espaço de probabilidade, os conjuntos $A \in \mathcal{F}$ são eventos possíveis, cuja probabilidade é dada por $\mathbf{P}(A)$.

Dizer que \mathcal{F} é uma σ -álgebra equivale a dizer que devo poder tirar conclusões sobre eventos, por meio de operações de união e intersecção de conjuntos, relacionando-os logicamente. Formalmente isso se traduz por:

2. DEFINIÇÃO. *Um conjunto \mathcal{F} é chamado σ -álgebra se*

(a) $\Omega \in \mathcal{F}$,

(b) se $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, \dots$ então $\bar{A}_1 \in \mathcal{F}, \bar{A}_2 \in \mathcal{F}, \dots$,

(c) se $A_n \in \mathcal{F}, n = 1, 2, \dots$, então $\bigcup A_n \in \mathcal{F}$,

onde $\bigcup A_n$ é a união enumerável de conjuntos, claramente se $\bigcup A_n \in \mathcal{F}$ então $\bigcap A_n \in \mathcal{F}$, ainda o duplo (Ω, \mathcal{F}) de Ω com uma σ -álgebra de seus subconjuntos é um espaço mensurável.

Devo estender a definição 1 para incluir tempo e dinâmica, já que nosso modelo probabilístico deve descrever um fenômeno que evolui no tempo mudando o conteúdo de informação estatística realizado, ou seja, deve ser um processo estocástico.

3. DEFINIÇÃO. A coleção $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathcal{P})$ é dita *experimento estocástico filtrado* com

(a) $\mathbb{F} = (\mathcal{F}_n)_{n \geq 0}$ é um *fluxo* tal que: $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$,

(b) $\mathcal{P} = \{P\}$ é uma *família de medidas probabilísticas*,

\mathcal{F}_n é o conjunto de eventos, σ -subálgebras de \mathcal{F} ou *filtrações* desta, observáveis até o tempo n .

É visto que a definição formal, da descrição para dinâmica probabilística, contém prescrição para o efeito histórico de quantidade limitada de informação realizada. Como foi visto em 1.1 este é um atributo essencial de sistemas complexos que apresentam intermitência em sua dinâmica, na abordagem de modelagem este é um dos atributos utilizados na descrição estocástica do mercado.

Agora note que nesta descrição ainda temos uma simplificação de hipótese, aqui a série é descrita como uma variável probabilística indexada no tempo discreto. Como mencionado na seção 2.2.2, tal assunção, levará a assumir o tempo de amostragem através do número de elementos da série em tempo constante.

Para que a descrição matemática mais fiel ao sistema - considerando os graus de liberdade efetuados no mercado, desconsiderando o livro de ordens - seja estabelecida, me provendo do formalismo necessário para interpretar os resultados dados no capítulo 4, devo redefinir a filtragem de σ -álgebras em função do índice $t \in [0, 1]$, tal que cada t é definido pelo sorteio do tempo de espera, i.e. os avanços de t são variáveis aleatórias,

4. DEFINIÇÃO. a *dinâmica* é dita *processo estocástico discreto com avanços aleatórios* se, obedece a definição 3 com a modificação,

- $\tau \in (\mathcal{T}, \mathcal{P})$ é *mensurável*, i.e. \mathcal{T} é σ -álgebra e \mathcal{P} uma família de medidas probabilísticas, este espaço define $t_N = \sum_{i=1}^N \tau_i$, tal que 3.(a) fica redefinido $\mathbb{F} = (\mathcal{F}_{t_N})_{t_N \geq 0}$, onde N é o número de sorteios de τ ,

com isso a interpretação das realizações de retornos ficam em função das realizações de tempos de espera entre transações. Como será visto na seção 4.2.2 esta extensão, da definição de processo estocástico, é essencial para que possa entender os resultados obtidos. Os resultados de eficiência do mercado, referentes à sincronia entre transações de duas empresas, deveriam depender da frequência de transações, i.e. da evolução temporal da distribuição do tempo de espera.

3.2 Erros Experimentais

Posso interpretar resultados experimentais por vias de duas categorias básicas consistentes entre si: aquisição empírica de dados, aplicação destes dados ao modelo físico. Para acessar a significância dos dados e previsões é necessário estimar erros estatísticos e teóricos para experimentos criados com objeto de testar uma teoria. A primeira categoria leva erros derivados da limitação na precisão dos instrumentos de medida ou na execução inadequada do experimento. A segunda determina como interpretar os erros experimentais em teorias assim como determinar o erro de previsão de certo sistema descrito teoricamente. A teoria é tão válida quanto mais próximos do resultado observado forem suas previsões, por outro lado o resultado observado é tão confiável quanto menor for seu erro estimado. Assim a estimativa do erro experimental é fundamental no processo de afirmação teórica, pelos testes de hipóteses, pela determinação de parâmetros empíricos e sua confiabilidade.

Nesta dissertação, cujo o estudo visa exibir fenomenologicamente a estrutura de correlações emergente no mercado, foi feita uma análise de erros baseada nos dados disponíveis e na estatística elaborada a partir dos dados. A seção 4.2.2 detalha o tratamento utilizado para atribuição da significância à análise elaborada.

3.3 Correlação Cruzada Linear

Um dos meios mais simples de computar co-movimentação, relação entre duas séries temporais em uma escala de tempo, é a correlação cruzada linear que parte da hipótese de linearidade entre as séries. O estimador de menor desvio quadrático médio é o de Pearson:

$$\hat{C}(r^{(i)}(t), r^{(j)}(t')) = \frac{1}{N-1} \sum_{\substack{0 \leq t \leq N-1 \\ 0 \leq t' \leq N-1}} \eta^{(i)}(t) \eta^{(j)}(t') \quad (3.1)$$

$$\eta^{(i)}(t) = \frac{r^{(i)}(t) - \langle \hat{r}^{(i)} \rangle}{\hat{\sigma}_i} \quad (3.2)$$

Onde uso $\hat{}$ para caracterizar estimativas em amostras finitas da média ¹ e desvio padrão ². Se a série é estacionária, $f(r^{(i)}(t)) = f(r^{(i)}(t + \tau)) \forall t$, então obtenho $C_{ij}(\tau) = \sum_{0 \leq t \leq N-1} \left(\frac{r^{(i)}(t-\tau) - \langle \hat{r}^{(i)} \rangle}{\hat{\sigma}_i} \right) \left(\frac{r^{(j)}(t) - \langle \hat{r}^{(j)} \rangle}{\hat{\sigma}_j} \right)$. No geral tem-se $C_{ij}(\tau) = C_{ji}(-\tau) \neq C_{ji}(\tau)$, expressando a precedência temporal dos coeficientes

¹A média é estimada por $\langle \hat{r}^{(i)} \rangle = \frac{1}{N} \sum_{0 \leq t \leq N-1} r^{(i)}(t)$

²A variância é estimada por $\hat{\sigma}_i^2 = \frac{1}{N-1} \sum_{0 \leq t \leq N-1} (r^{(i)}(t) - \langle \hat{r}^{(i)} \rangle)^2$

de correlação, apenas com $\tau = 0$ obtém-se $C_{ij} = C_{ji}$, onde o coeficiente de correlação dá as componentes das duas séries que, para o mesmo t , estão relacionadas linearmente. No contexto da bolsa Bovespa o coeficiente representa uma medida ao compartilhamento da informação (linear) no instante t .

3.4 Espaços Ultramétricos

A exposição aqui apresentada da teoria de espaços ultramétricos se baseia no amplo *review* por R. Rammal et. al. (1986 [13]), o qual abrange desde axiomatização matemática, passando por problemas arquetípicos em sistemas complexos, até a modelagem baseada em ultrametria, porém seu escopo vai muito além da análise desta dissertação. No seu livro referência (2000 [6]), para econofísica, R. Mantegna e H. E. Stanley exemplificam o uso dos conceitos de ultrametria para identificação de informação econômica, abordagem que será aqui seguida.

A Teoria de Espaços Ultramétricos tem desenvolvimento histórico paralelo ao descobrimento de seu uso em ciências naturais, como por exemplo a metodologia taxonomica para classificação de espécies. Similar ao ocorrido com a Teoria de Fibrados e os desenvolvimentos em Teoria de Campos, em particular a Relatividade Geral de A. Einstein (1916), o encontro da teoria matemática com a aptidão de refinar modelos e classificar sistemas naturais cada vez mais complexos, é resultado da crescente invasão do método geométrico em outras áreas da matemática a partir do século 20 e da naturalidade em explicar alguns sistemas a partir da Ultrametria.

Na matemática a introdução dos números p -ádicos, na transição do século 19 para o 20 (1897) por Kurt Hensel, trouxe um contexto natural em teoria dos números onde conceitos de ultrametria poderiam ser desenvolvidos e naturalmente envolvidos em uma topologia que convenientemente posiciona os elementos medidos segundo uma hierarquia de distâncias.

Em 1906 M. Fréchet introduziu a noção geral de espaços métricos:

5. DEFINIÇÃO. E é espaço métrico se é definida uma função $d : E \times E \mapsto \mathfrak{R}$, para quaisquer $x, y, z \in E$ tal que,

$$(a) \quad d(x, y) = 0 \leftrightarrow x = y,$$

$$(b) \quad d(y, x) = d(x, y),$$

$$(c) \quad d(x, y) \leq d(x, z) + d(z, y),$$

o item 5.c é a desigualdade triangular, relaciona três elementos do conjunto E , dá o significado intuitivo de triângulo. Entendo 5.a e 5.b como fornecendo os elementos geométricos básicos, o ponto e a reta. Qualquer espaço métrico induz uma topologia.

Topologia pode ser entendida como o estudo de propriedades que se conservam em objetos, sob transformações contínuas como esticamento e compressão, mas não corte e colagem do objeto. Não é objetivo desta dissertação o tratamento formal da topologia induzida por uma escolha métrica, entretanto o conceito é importante pois possibilita entender o valor geométrico do estudo sob qualquer escolha métrica, no caso da ultrametria nos induz a um entendimento geométrico particular. A desigualdade triangular pode ser substituída em 5.c por uma desigualdade mais forte, restringindo a geometria do espaço métrico à triângulos equiláteros e isósceles de base pequena, formalmente:

6. DEFINIÇÃO. *E é espaço ultramétrico se é definida uma função $d : E \times E \mapsto \mathfrak{R}$, para quaisquer $x, y, z \in E$ tal que,*

$$(a) \quad d(x, y) = 0 \leftrightarrow x = y,$$

$$(b) \quad d(y, x) = d(x, y),$$

$$(c) \quad d(x, y) \leq \text{Max}\{d(x, z), d(z, y)\},$$

6.c agora apresenta a restrição da geometria do conjunto aos triângulos equilátero ou isósceles de base pequena.

A desigualdade ultramétrica pode ser melhor entendida analisando as suas possibilidades. Se $d(x, z) \neq d(z, y)$ então terei para $d(x, y) < \text{Max}\{d(x, z), d(z, y)\} = d(x, z)$ (supondo $d(x, z) > d(z, y)$), ainda no mesmo triângulo temos, mudando a ordem dos vértices $d(x, z) < \text{Max}\{d(y, z), d(x, y)\}$ chegamos a uma contradição, então $d(x, y) = \text{Max}\{d(x, z), d(z, y)\}$. Chego a conclusão: ou todos os lados são iguais, ou dois lados são iguais e maiores do que a base.

Espaços ultramétricos fornecem a medida natural para descrição de estrutura como árvores taxonômicas, onde a ordem natural estabelece uma hierarquia, agrupando elementos segundo uma indexização. Sob a luz da teoria da evolução espaços ultramétricos adquirem significado intuitivo de separação em demais espécies a partir de um ancestral comum (figura 3.1), no capítulo 4 verificarei que as taxas adquirem significado de proximidade entre as empresas de um setor econômico .

Árvore filogenética da vida

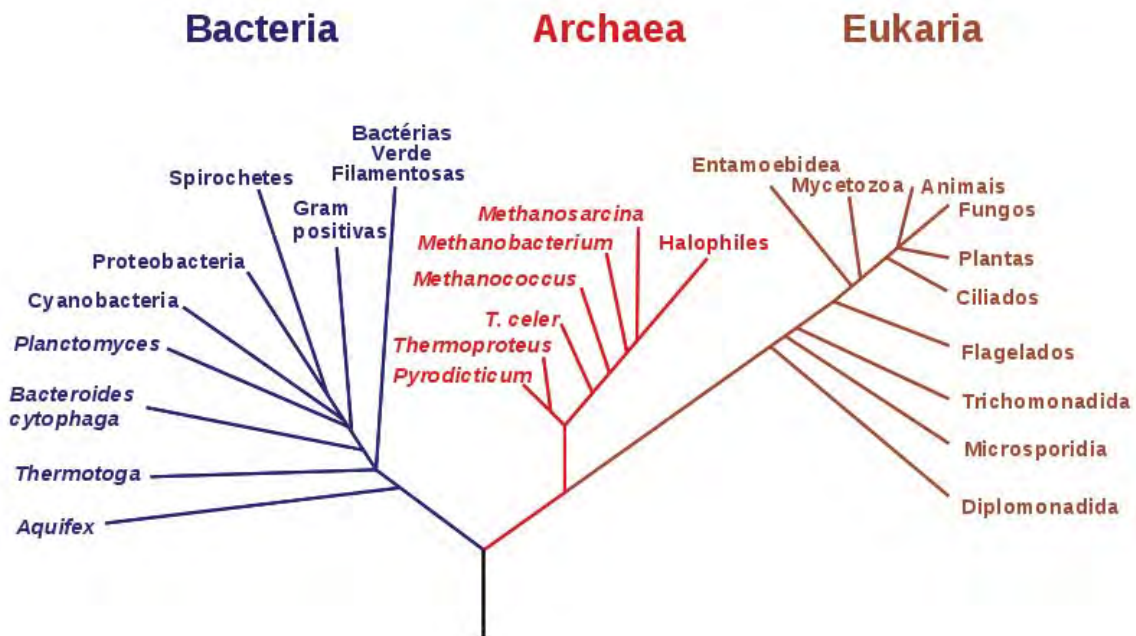


Figura 3.1: Árvore da Vida, exemplo de árvore taxonômica, publicação na internet da Nasa Astrobiologia (2003 [7]). Quaisquer duas espécies no final da árvore tem um ancestral em comum, geometricamente um triângulo isósceles.

3.4.1 Espaços Ultramétricos Subdominantes

A restrição a medidas que obedecem a desigualdade ultramétrica ainda me deixa com várias possibilidades, cada uma gerando um espaço ultramétrico diferente. Utilizarei para a análise do mercado um espaço ultramétrico subdominante, pela simplicidade de obtenção e interpretação através do algoritmo de Kruskal (seção 3.5), mas antes introduzo formalmente o conceito de hierarquia, para então entender como se dá o processo de seleção do espaço do ponto de vista das possíveis hierarquias.

Hierarquias indexadas

7. DEFINIÇÃO. Seja Ω um conjunto finito, uma hierarquia $H = \{h\}$ em Ω é um conjunto das partes de Ω tal que,

- (a) $\Omega \in H$,

(b) $\forall \omega \in \Omega, \{\omega\} \in H,$

(c) para cada par $h, h' \in H,$ se $h \cap h' \neq \emptyset \Rightarrow h \subset h'$ ou $h' \subset h,$

os três itens reference-se a, sucessivamente, todos elementos do conjunto estruturam a ordem da hierarquia, cada elemento individualmente está em um nível final da hierarquia e quaisquer dois níveis hierárquicos contendo mesmos elementos deveram estar no mesmo ramo, determinado pelo h mais abrangente.

Além disso se houver uma função f positiva, definida em $H,$ de forma que $f(h) = 0 \Leftrightarrow h$ está no final de um ramo (é elemento individual de Ω) e se $h \subset h'$ então $f(h) < f(h'),$ então o par (H, f) chama-se *hierarquia indexada* em Ω com f dando o nível de agregação dos elementos.

Construirei uma hierarquia em um conjunto Ω com o auxílio de uma função de agregação δ que leva cada dois elementos $h_1, h_2 \in H$ em um real positivo $\delta(h_1, h_2) \geq 0$ tal que $\delta(h_1, h_2) = \delta(h_2, h_1).$ Haja visto na definição 5 que um espaço métrico permite a escolha de várias possibilidades de índice de agregação por meio de $d(x, y),$ e em particular para selecionar um espaço ultramétrico em $\Omega,$ posso restringir me a uma função δ que tome valores sempre menores ou iguais que uma métrica usual em $\Omega.$ Partindo desta restrição para δ posso obter um espaço ultramétrico subdominante da seguinte forma:

8. DEFINIÇÃO. *Uma função $\delta : H \times H \rightarrow \Re$ é uma ultramétrica subdominante se ela é a maior entre todas funções de agregação, definidas em todas hierarquias possíveis em $\Omega,$ que obedecem a desigualdade $\delta \leq d.$*

Se tornará mais clara a definição 3.4.1 quando utilizar o algoritmo de Kruskal, inicialmente em um grafo totalmente conectado com uma métrica usual em $\Omega,$ para obter uma árvore de mínimo alcance resultando naturalmente, obedecendo a desigualdade supra-citada, em um espaço ultramétrico.

3.5 **Árvore de mínimo alcance, taxonomia e o algoritmo de Kruskal**

Como mostrei na seção anterior a escolha de um espaço ultramétrico implica em uma topologia hierárquica. Para selecionar um espaço subdominante E deverei definir uma regra de agregação (δ na definição 3.4.1), tendo como base a métrica usual no conjunto $\Omega,$ a partir da qual obterei o espaço $E,$ é topologicamente representado por um grafo totalmente conectado, aplicando a regra δ a ultrametria subdominante é topologicamente representada por uma árvore de mínimo alcance quando a regra é o algoritmo de Kruskal. A seguir definirei todo formalismo necessário para descrição do algoritmo e da árvore obtida, equivalente a obtenção de uma taxonomia pela escolha de ultrametria subdominante, natural na teoria de grafos.

3.5.1 Teoria de Grafos

Defino um *ensemble de grafos* como sendo um par de conjuntos $G = (V, E)$, onde o primeiro refere-se aos vértices e o segundo às arestas (G define vértices e arestas, destes componentes posso construir vários grafos). Um elemento do ensemble G é um par (V, E) com uma regra de posicionamento, das arestas em relação aos vértices, definida em (V, E) . O procedimento de associação desta regra depende da natureza do sistema tratado, o importante é que sempre poderei representar um grafo por uma matrix, $N \times N$ em um corpo K , chamada *matriz de adjacência*. A análise trabalhada é sempre com um elemento de G (dito realização de G), utilizarei a notação $G = (V, E)$ para denotar a realização analisada a partir daqui.

O grafo trabalhado nesta dissertação é não direcionado e as arestas tem peso dado pela métrica escolhida. Além disso o gráfico será conectado e sem ciclos, tais gráficos são ditos árvores (figura 3.2).

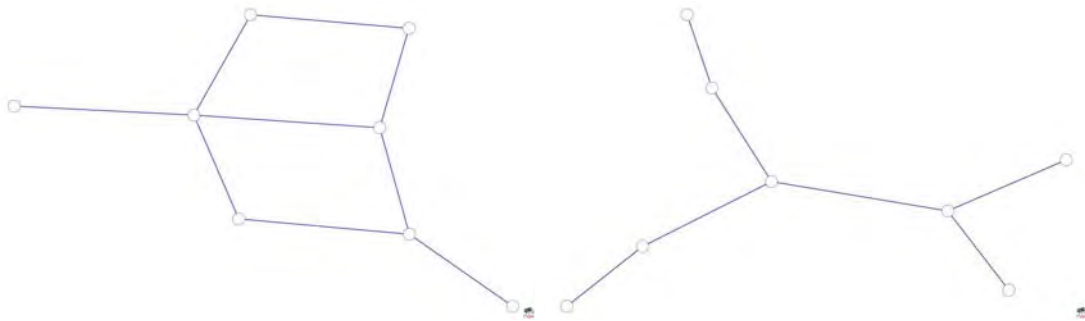


Figura 3.2: A esquerda o grafo é conectado com ciclos e a direita grafo árvore.

Árvore de mínimo alcance através do algoritmo de Kruskal

Defino árvore como sendo um grafo conectado e sem ciclos, é dito que uma árvore $S \subset G$ é *árvore de mínimo alcance* de G se, S contém todos seus vértices e tem a menor soma de arestas. Usarei um grafo G hipotético, com um conjunto $V = \{A, B, C, D, E\}$ de vértices e arestas E (matrix 1), para ilustrar o algoritmo de Kruskal na obtenção da árvore de mínimo alcance.

	A	B	C	D	E
A		1.5	0.9	4.1	3.5
B			2.5	1.1	1.3
C				0.8	3.7
D					0.7
E					

Matrix 1: Matrix de adjacência para o grafo conectado G .

Começo com um sub-grafo S contendo os vértices conectados pela menor aresta, subtrai-se de G estes vértices, prossigo adicionando vértices, sem repetição, ao grafo S em ordem crescente de conexão por arestas, subtraindo de G , até que G fique vazio e S contenha todos os vértices. Nas figuras 3.3, 3.4, 3.5, 3.6, 3.7 e 3.8 reproduzo os passos para obtenção de S , o grafo inicial G é caracterizado por arestas pontilhadas azuis e vértices brancos, a árvore S é caracterizadas por arestas sólidas verdes e vértices verdes.

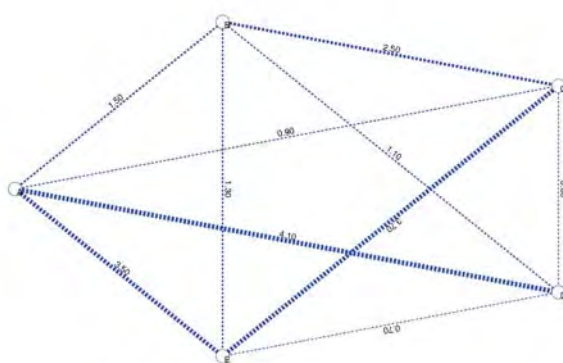


Figura 3.3: Grafo G conectado.

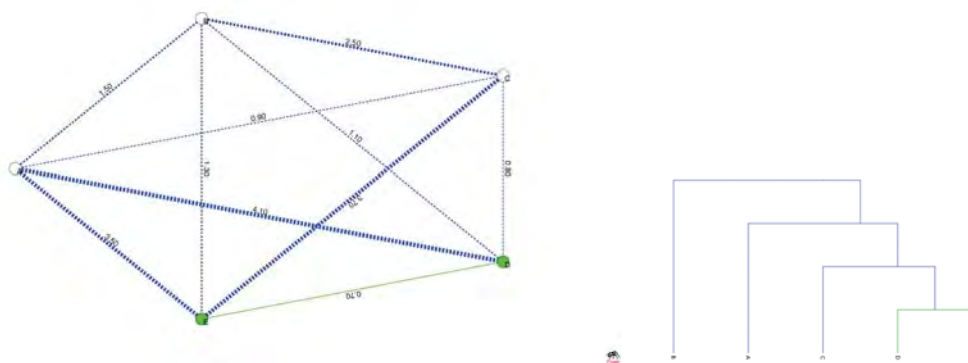


Figura 3.4: Adiciono vértices E e D ao grafo S , aresta de valor 0.7 .

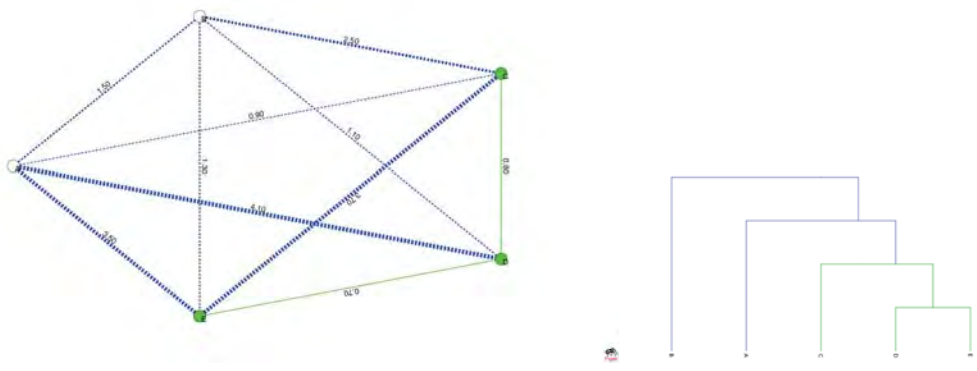


Figura 3.5: Vértice C adicionado ao grafo S , aresta de valor 0.8 .

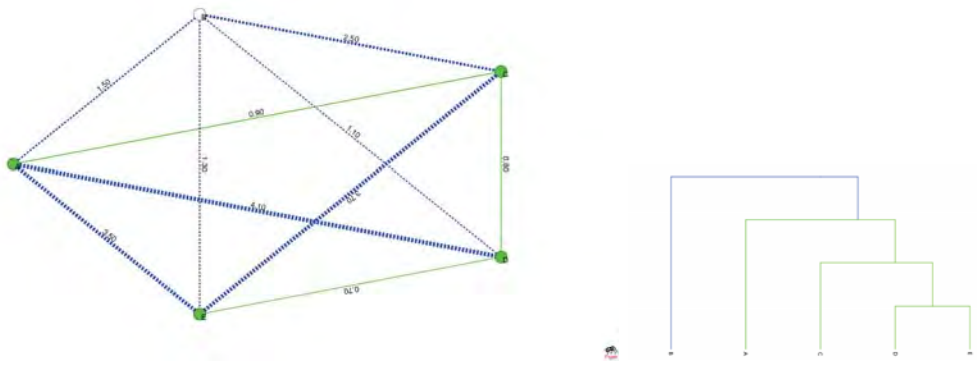


Figura 3.6: Vértice A adicionado ao grafo S , aresta de valor 0.9 .

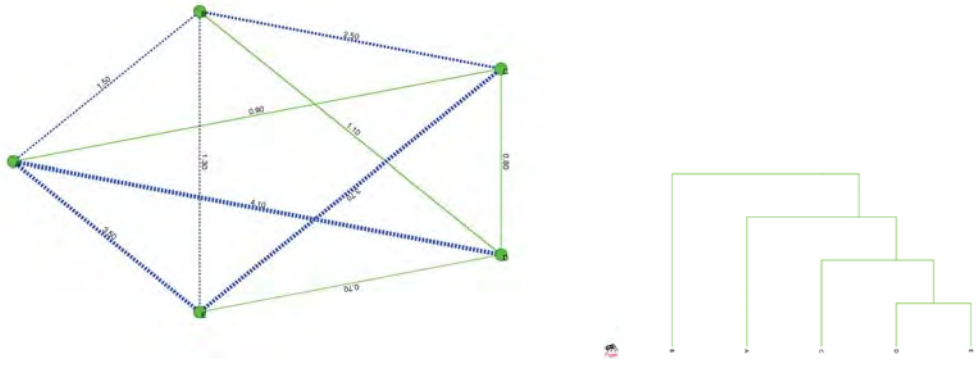


Figura 3.7: Vértice B adicionado ao grafo S , aresta de valor 1.1 .

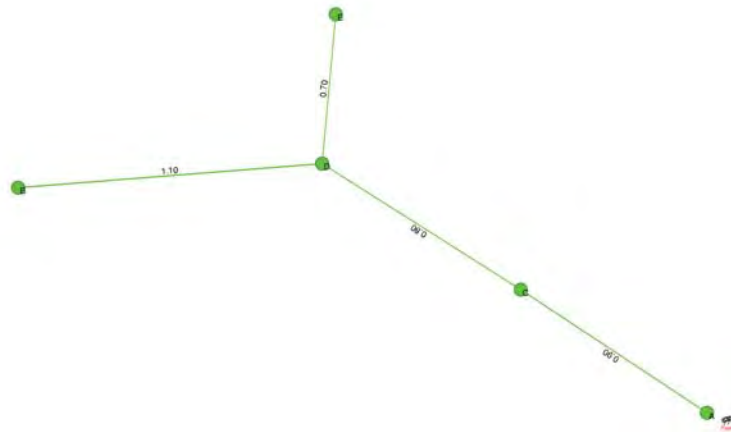


Figura 3.8: Árvore de mínimo alcance S final .

	A	B	C	D	E
A		1.1	0.9	0.9	0.9
B			1.1	1.1	1.1
C				0.8	0.8
D					0.7
E					

Matrix 2 Matrix de distância do espaço ultramétrico subdominante.

Seja $\{h_1\} = (V_1, E_1)$ e $\{h_2\} = (V_2, E_2)$ dois dos grafos árvore obtidos até um passo qualquer do algoritmo de Kruskal. Note que nas figuras acima mostrando o algoritmo, a cada passo, a seleção da menor aresta e de seus respectivos vértices, garante $\delta \leq d$ e ainda, a escolha de $\delta(h_1, h_2) = \text{Max}\{d \in E | E = E_1 \cup E_2\}$, nos garante que a maior função agregação possível, isto permite a relação ultramétrica (triângulo isósceles base pequena ou equilátero) subdominante (aquela que minimiza a soma de arestas) de quaisquer 3 pontos conectados às árvores .

Interpretação das taxas em taxonomia do mercado financeiro

No mercado financeiro defino distância através dos vetores retorno, resultando na proximidade de movimentação de duas empresas, é esperado que duas empresas de mesmo setores se co-movimentem com um motivo financeiro ou tenham o mesmo comportamento para situações similares, então quanto menor a distância entre elas maior a proximidade da mesma característica economica.

Tomando a noção de distância apresentada com relação a uma característica economica que causa a proximidade, taxa de uma taxonomia de mercado pode ser pensada como patamar de proximidade às características economicas que causam proximidade entre duas ações. Neste sentido, duas ações que se ligam no final da ramificação taxonômica possuem maior atividade economica específica de uma classe economica, quando uma ação se liga a uma hierarquia em uma taxa com

maior distância significa que existe menos atributos econômicos, similares aos atributos existentes na hierarquia, na ação em questão.

A taxa próxima à raiz da taxonomia é menos específica, em atributos econômicos, que a taxa próxima ao fim da ramificação. Pode-se utilizar a visualização da árvore taxonomica para obter as ações mais próximas de uma mesma atividade econômica, as ligações mais importantes para a estrutura setorial. Para isso início a busca nas taxas mais específicas, as ações que se ligam diretamente estão no mesmo setor, ações que se ligam à hierarquia seram uma ramificação dessa hierarquia (com grau de proximidade de acordo com a altura da taxa, ações pouco correlatas, ligadas às pontas dos ramos na árvore, apresentaram este padrão), hierarquia se une a hierarquia significando dois setores que tem proximidade econômica dada pela taxa que se ramifica nas duas hierarquias.

4

Análise do mercado Bovespa&BMF

4.1 Dados e metodologias gerais

Para análise do Bovespa&BMF foram utilizados 3 anos de dados (2006, 2007 e 2008), fornecidos pelo próprio Bovespa&BMF, com precisão temporal TANDEM (fração de segundo). As séries de preços são separadas segundo o dia de transação, cada transação efetuada tem horário registrado. Normalizei o tempo de pregão, que no Bovespa&BMF é de 7 horas, para o intervalo $[0, 1]$ de modo identificar um horário de transação com um ponto real pertencente a este intervalo. Assim calculei os retornos das séries utilizando o tempo normalizado, para isso estabeleço a unidade mínima de tempo t_m , sendo que apenas valores $t = t_0 + nt_m$, n natural, são utilizados, com um t_0 arbitrário. Seja $S^{(j)}(t)$ o preço da ação j no instante t , o retorno da ação, na escala Δt , neste instante é dado por:

$$r_{\Delta t}^{(j)}(t) = \ln \frac{S^{(j)}(t)}{S^{(j)}(t - \Delta t)} \quad (4.1)$$

Como não são observadas, necessariamente, transações em t e $t - \Delta t$, é preciso utilizar o esquema de tick anterior. Isso é feito considerando que o preço da ação, na transação imediatamente anterior em relação a t e $t - \Delta t$, permanece constante até a próxima transação.

Como explicado na seção 3.3, o estimador de correlação cruzada linear dá o coeficiente, instantâneo no tempo, da primeira ordem em inter-dependência funcional dos retornos de duas ações, ou ainda a distância informativa entre duas empresas, informação movendo uma empresa é encontrada movendo a outra empresa na mesma direção. Entretanto é necessário trabalhar com uma função

definida em um espaço métrico 5 para as análises de concentração e polarização dentro de uma árvore. Para tanto defino as séries temporais de retornos como sendo vetores transladados pela média e normalizados em um espaço com N dimensões (tamanho da amostra de retornos) com métrica euclidiana, a distância entre dois vetores é dada por:

$$\begin{aligned}
\|\mathbf{r}_{\Delta t}^{(i)} - \mathbf{r}_{\Delta t}^{(j)}\|^2 &= \sum_{t=1}^N \left(\frac{r_{\Delta t}^{(i)}(t) - \bar{r}^{(i)}}{\sqrt{\sum_{t=1}^N (r_{\Delta t}^{(i)}(t) - \bar{r}^{(i)})^2}} - \frac{r_{\Delta t}^{(j)}(t) - \bar{r}^{(j)}}{\sqrt{\sum_{t=1}^N (r_{\Delta t}^{(j)}(t) - \bar{r}^{(j)})^2}} \right)^2 \\
&= \frac{\sum_{t=1}^N (r_{\Delta t}^{(i)}(t) - \bar{r}^{(i)})^2}{\sum_{t=1}^N (r_{\Delta t}^{(i)}(t) - \bar{r}^{(i)})^2} + \frac{\sum_{t=1}^N (r_{\Delta t}^{(j)}(t) - \bar{r}^{(j)})^2}{\sum_{t=1}^N (r_{\Delta t}^{(j)}(t) - \bar{r}^{(j)})^2} \\
&\quad - 2 \frac{(r_{\Delta t}^{(j)}(t) - \bar{r}^{(j)})(r_{\Delta t}^{(i)}(t) - \bar{r}^{(i)})}{\sqrt{\sum_{t=1}^N (r_{\Delta t}^{(i)}(t) - \bar{r}^{(i)})^2 \sum_{l=1}^N (r_{\Delta t}^{(j)}(l) - \bar{r}^{(j)})^2}} \\
&= 1 + 1 - 2\rho_{ij} \\
&= 2(1 - \rho_{ij}) = d_{ij}^2
\end{aligned} \tag{4.2}$$

É visto que a equação de distância 4.2 inclui o termo de correlação como atenuante, quanto mais correlacionadas menor será a distância entre elas. Agora poderei analisar o grafo da bolsa com aresta no espaço métrico definido por 4.2 onde a distância instantânea tem ligação direta com a correlação cruzada instantânea, entre quaisquer duas empresas cujo os retornos foram obtidos.

4.2 Escalas de correlação cruzada: efeito Epps

A dinâmica intra-day das correlações foi primeiramente estudada por Thomas Epps(1979 [3]) que descobriu, aumentando a frequência de amostragem, o decréscimo da correlação de retornos entre ações da New York Stock Exchange. Correlações cruzadas são os principais componentes da teoria clássica de carteiras, então é importante compreender a dinâmica de cada escala temporal, ainda mais quando se considera casos onde ajustes nas carteiras acontecem na escala de minutos.

O efeito descoberto por Epps foi abordado por Renò (2003 [8]), gerando dados através de monte-carlo e análise da série de preços, utilizando um estimador de

fourier para covariância, apropriado para estimativa no espaço das frequências, mostrou que o efeito lag-condutor e de assincronia explicavam parte do efeito Epps. Kwapien et al (2004 [5]) abordaram a mesma questão sob um ponto de vista qualitativo físico, investigando as escalas importantes para o fluxo de informação, entre ações, e a formação da estrutura de correlação no mercado.

O estudo do efeito Epps foi substancialmente revisado e estendido por Tóth e Kertész (2006, 2007 e 2009 respectivamente [10; 11; 12]), aqui seguirei seus três artigos na análise.

São creditados como responsáveis pelo efeito Epps fatores de efeito lag-conduzido, assincronia entre as transações e limite humano de uso da informação de mercado.

Cada um desses fatores serão aqui abordados, com a intenção de elucidar a micro-dinâmica do mercado, para aplicação na estimativa das correlações no mercado BMF&Bovespa.

4.2.1 Influência de lead-lag na relação funcional entre retorno de ações

A existência da relação de atraso entre os retornos de duas ações, em determinada escala, indica a existência de um tempo característico em que a informação econômica, processada no retorno de uma ação, leva até influenciar uma segunda, esta eficiência reduzida faz com que a correlação instantânea seja menor contribuindo para o efeito Epps.

Em seu artigo de 2006 [10] Tóth e Kertész investigaram, ao longo de 10 anos, a existência de atraso na relação funcional do retorno de ações da NYSE através da estimativa da correlação ao longo do tempo. Eles acharam que o máximo da função de correlação se deslocou para o 0 ao longo dos anos, indicando aumento da eficiência do mercado.

Aqui investigo a função correlação dos pares PETR4-VALE5 (fig 4.1) e USIM5-VALE5 (fig 4.2), na escala de 2 minutos, verificamos que no período investigado o mercado se apresenta eficiente com o máximo da correlação no lag 0. Verifica-se ainda a existência de correlação nos lags -1 e 1, mas não são estatisticamente relevantes.

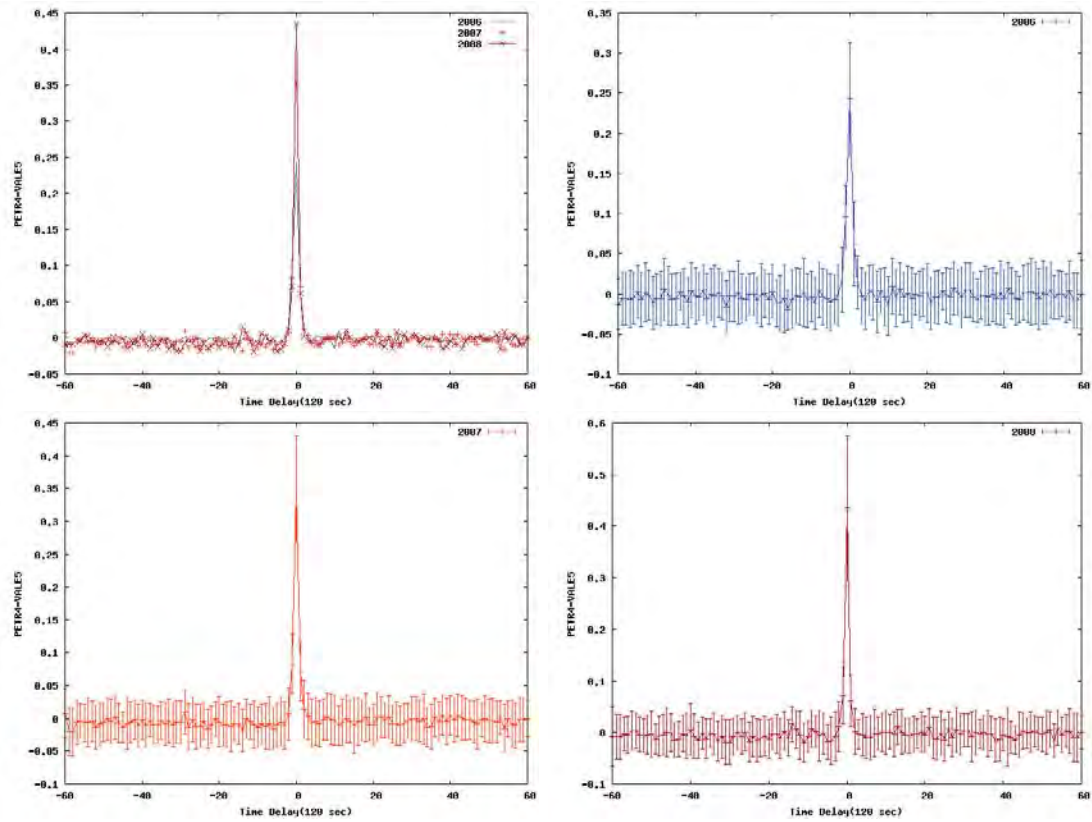


Figura 4.1: Funções de correlação cruzada na escala de 2 minutos do par PETR4-VALE5, as funções sobrepostas mostram que a correlação é predominantemente instantânea no período estudado.

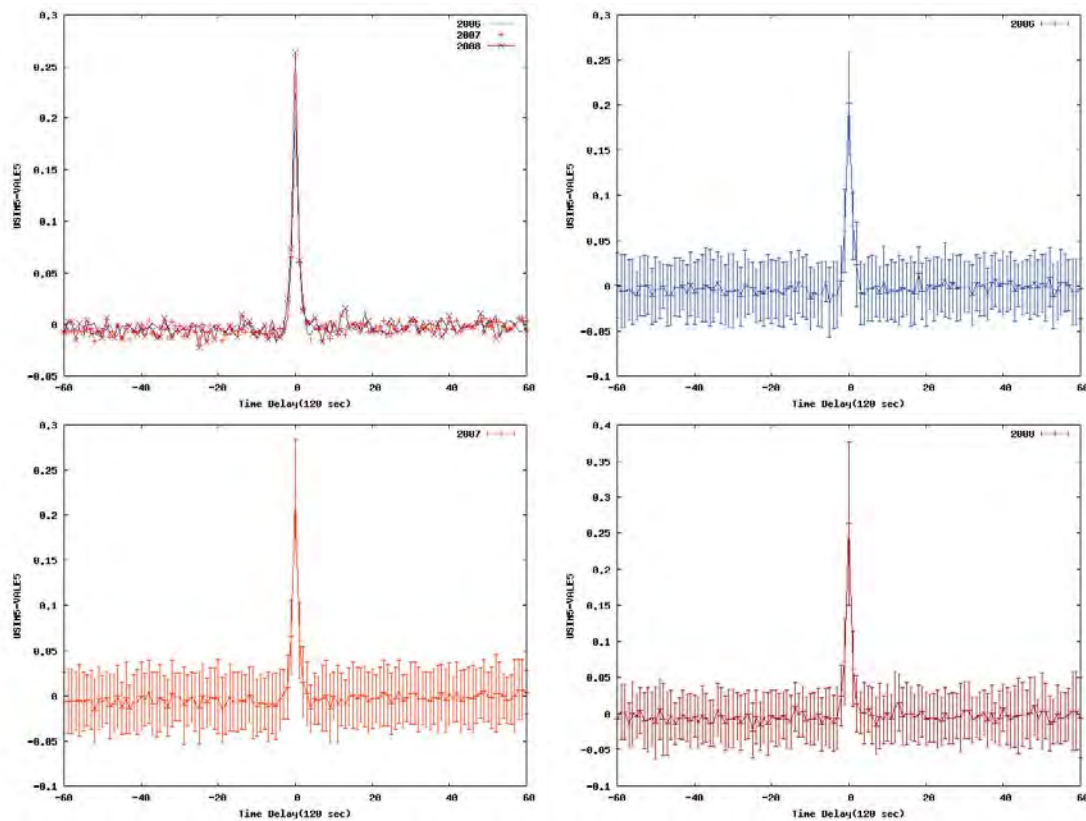


Figura 4.2: Funções de correlação cruzada na escala de 2 minutos do par USIM5-VALE5, aqui também as funções sobrepostas exibem rápido decaimento na correlação.

4.2.2 A escala do efeito no tempo inter-transações

O aumento da frequência de transações, em todas escalas de tempo, é um dos fatores pelos quais podem haver aumento da sincronia de transações, com isso é esperado que a correlação aumente com maior rapidez chegando no regime assintótico em menores escalas. Entretanto o fenômeno não revela diminuição da escala característica, na realidade ele mostra aproximada invariância de escala, o que sugere a atuação de outro fator atenuante para correlação.

As estruturas intra-day de correlações foram estudadas ao longo do calendário utilizado. Determinei a correlação diária das ações, para todos os dias do ano, tomando a média semanal para anular a sazonalidade obtenho um conjunto de dados anuais de 48 semanas, com exceção do ano de 2006 aonde não possuo os dados de janeiro e fevereiro. Ao todo foram contabilizados 207 dias de transação em 2006, 245 em 2007 e 240 em 2008.

Neste estudo exibirei a estrutura de correlação, em várias escalas no tempo,

para os pares de empresa Bradespar e Itaú (BRAP4-ITSA4), Petrobras e Itaú (PETR4-ITSA4), Petrobras e Vale (PETR4-VALE5), Tam e Natura (TAMM4-NATU3), Vale e Usiminas (VALE5-USIM5). As áreas de atuação são: Petrobras atua em petróleo e combustível fóssil, Vale em mineração, Usiminas em siderurgia, Itaú e Bradespar em serviços financeiros, Tam em construção e transporte e Natura em consumo não cíclico. As empresas citadas foram escolhidas por estarem entre as mais negociadas e por estarem em áreas reconhecidas como economicamente relacionadas, desta forma espero que exibam correlação estatisticamente relevantes, exceto no caso Tam e Natura onde setores não relacionados a priori podem ser investigados.

Significância das estimativas para o efeito Epps

O desvio padrão da correlação, para cada escala de tempo, foi calculado considerando as 48 medidas como realizações independentes da correlação em cada escala através do estimador 3.1.

Observo na estimativa do efeito Epps anual que, para quaisquer dos pares de empresas investigadas, o desvio padrão da correlação é menor para escalas temporais de alguns minutos, aumentando com maior rapidez até por volta de 10 minutos, a partir da onde o desvio padrão cresce bem mais lentamente. Variação na medida pode ser em parte entendida como efeito de amostra finita, levando a estatísticas com menos termos em escalas maiores, quando a correlação é calculada por 3.1, por outro lado obtenho maior variação, ao longo do calendário, da correlação medida pela amostra semanal média.

O desvio padrão, comparado com o valor médio da estimativa, me permite saber a nitidez (ou seja a frequência em que a correlação intraday, em uma escala de tempo, tem valor próximo ou igual à μ) do efeito Epps ao longo do ano considerado. É esperado, portanto, que a taxa $\frac{\mu}{\sigma}$ tenha valores altos em boa parte das escalas de tempo para as ações economicamente correlatas, factualmente é observado altos $\frac{\mu}{\sigma}$ para: BRAP4-ITSA4 (fig 4.3), PETR4-ITSA4 (fig 4.4), PETR4-VALE5 (fig 4.5) e VALE5-USIM5 (fig 4.7). O par TAMM4-NATU3 (fig 4.6) apresenta rápida diminuição do $\frac{\mu}{\sigma}$, indicando que o efeito Epps na transição, de não correlacionado para correlacionado, não deve ser observável com a estatística (pearson), ou seja, não tenho correlação alta suficiente, ao longo do ano, para este desvio padrão torne o fenômeno significativo.

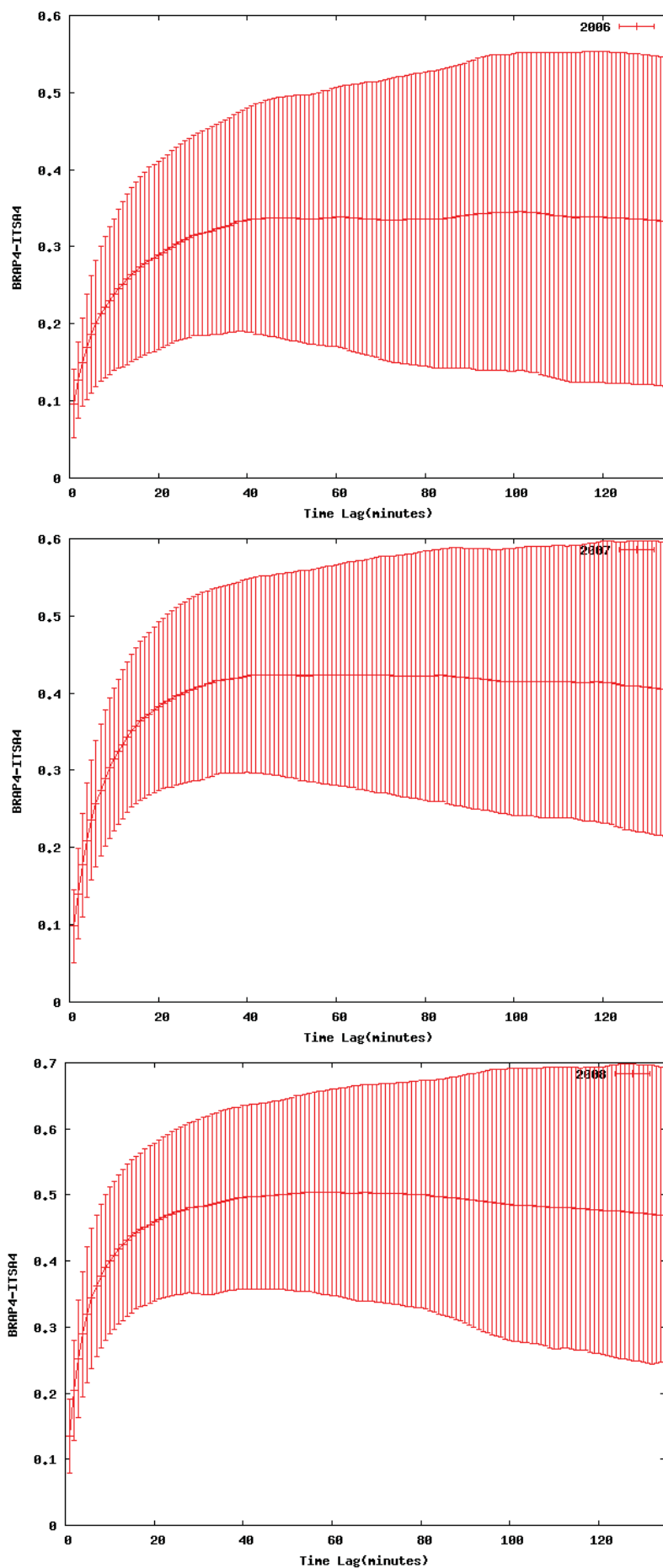


Figura 4.3: Efeito Epps com respectivo desvio padrão para BRAP4-ITSA4

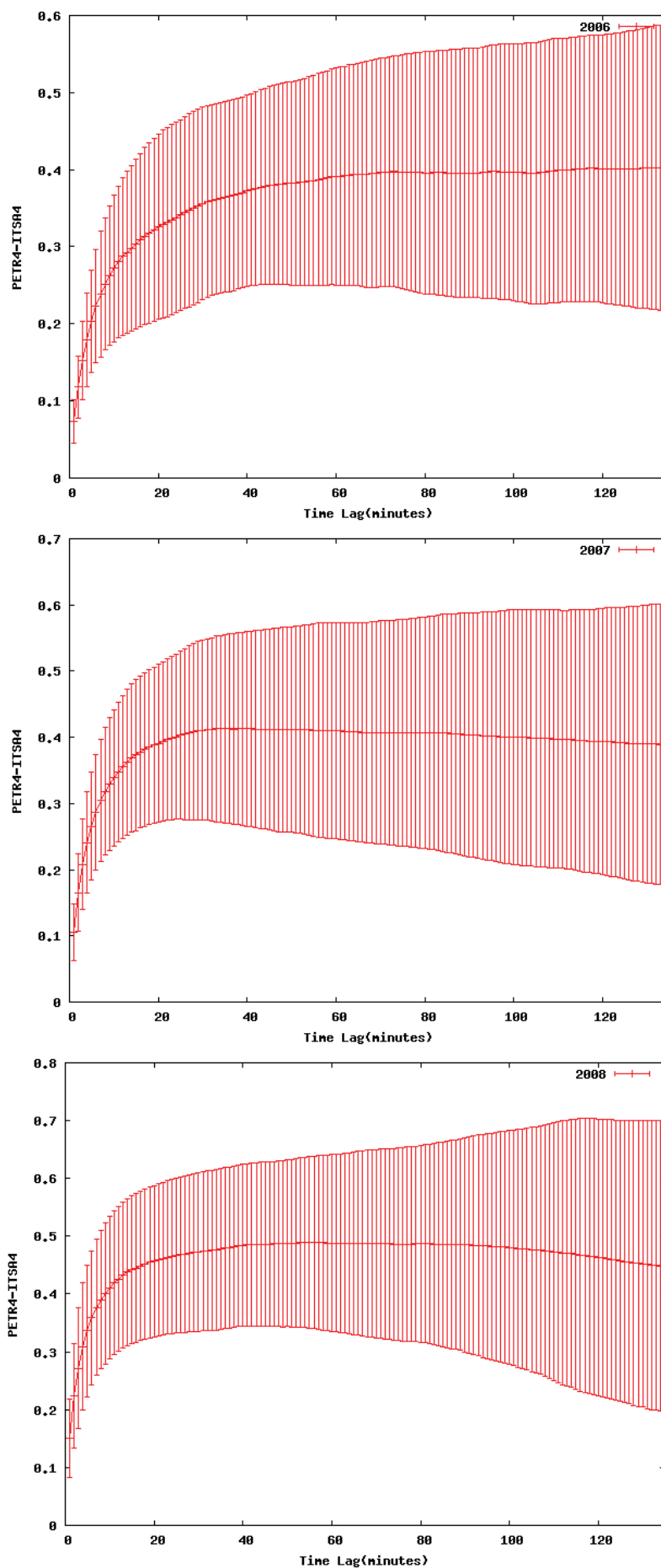


Figura 4.4: Efeito Epps com respectivo desvio padrão para PETR4-ITSA4

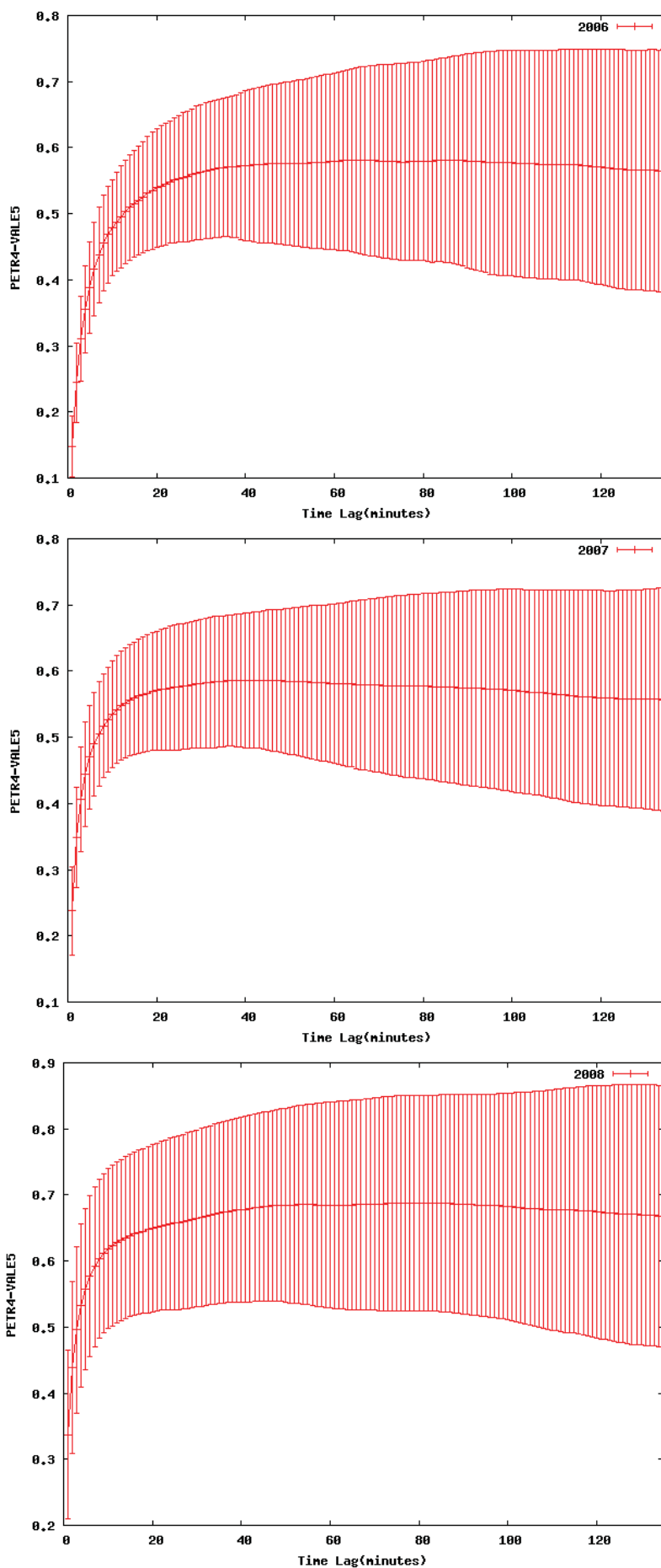


Figura 4.5: Efeito Epps com respectivo desvio padrão para PETR4-VALE5

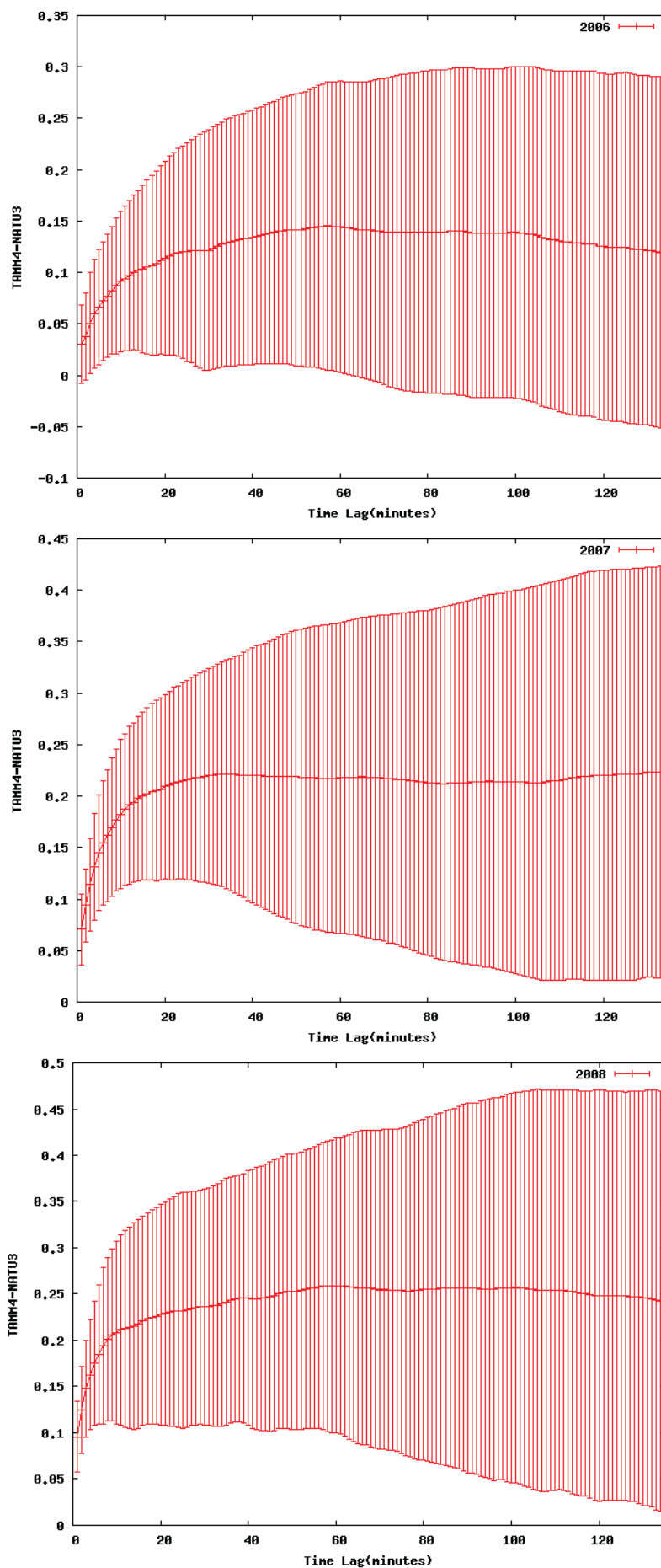


Figura 4.6: Efeito Epps com respectivo desvio padrão para Tamm4-NATU3

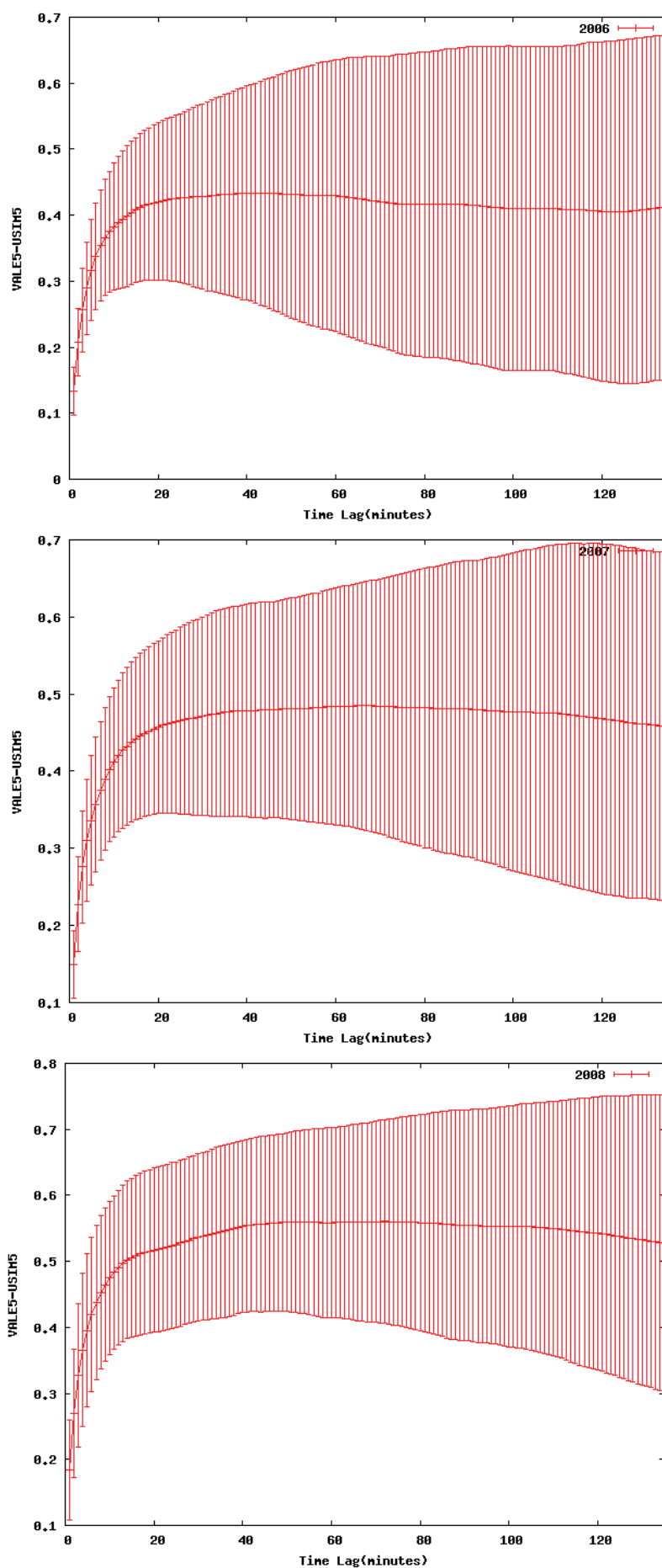


Figura 4.7: Efeito Epps com respectivo desvio padrão para VALE5-USIM5

	Menor valor de $\frac{\mu}{\sigma}$		
ρ	2006	2007	2008
BRAP4-ITSA4	(1.56 , 134)	(2.07 , 1)	(2.08 , 132)
PETR4-ITSA4	(2.16 , 134)	(1.83 , 134)	(1.78 , 134)
PETR4-VALE5	(3.08 , 134)	(3.29 , 134)	(2.64 , 1)
TAMM4-NATU3	(0.7 , 134)	(1.1 , 126)	(1.07 , 134)
VALE5-USIM5	(1.55 , 127)	(2.03 , 134)	(2.36 , 134)
	Maior valor de $\frac{\mu}{\sigma}$		
ρ	2006	2007	2008
BRAP4-ITSA4	(2.63 , 3)	(3.55 , 14)	(3.84 , 21)
PETR4-ITSA4	(3.04 , 5)	(3.31 , 8)	(3.49 , 20)
PETR4-VALE5	(6.67 , 10)	(6.68 , 9)	(5.19 , 16)
TAMM4-NATU3	(1.34 , 13)	(2.64 , 2)	(2.82 , 3)
VALE5-USIM5	(4.20 , 7)	(4.28 , 9)	(4.25 , 40)

Tabela 4.1: Taxa $(\frac{\mu}{\sigma}, \Delta t)$, Os menores $\frac{\mu}{\sigma}$ se posicionam nas maiores escalas estimadas, os maiores se concentram na região abaixo de 40 minutos.

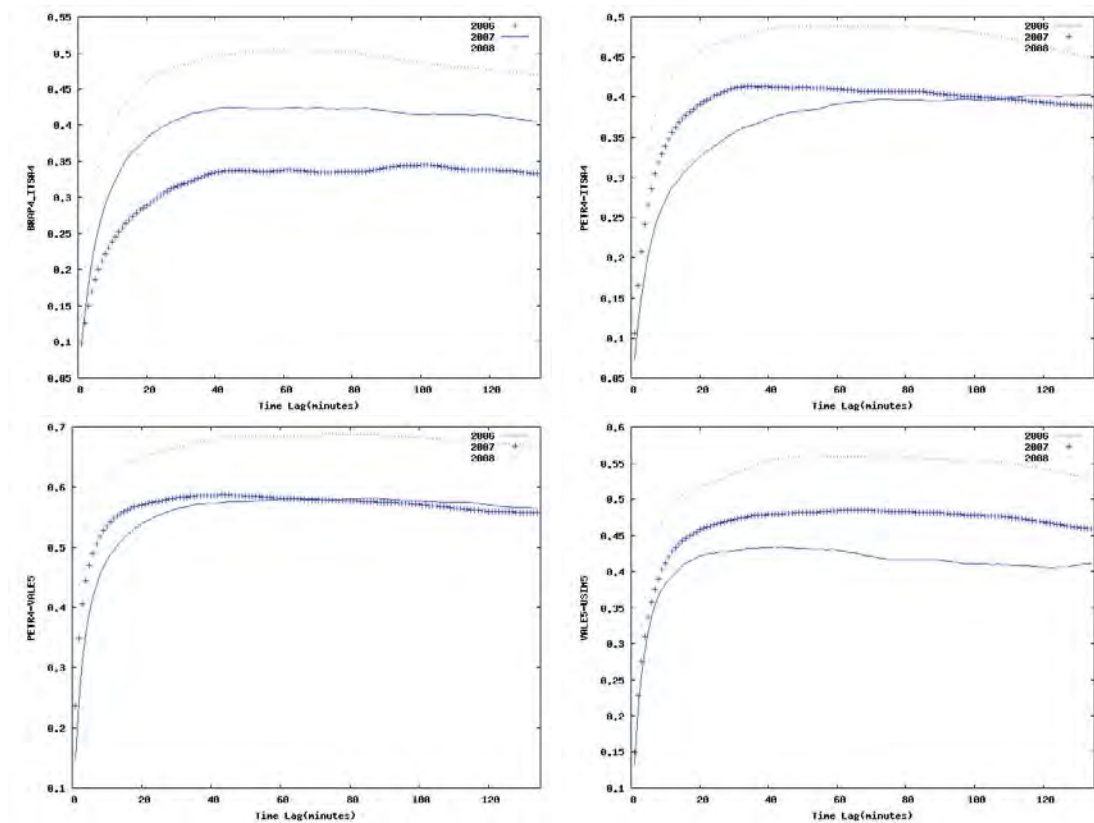


Figura 4.8: Efeitos Epps anuais sobrepostos para os anos de 2006, 2007 e 2008.

É possível ver nos gráficos (fig 4.8) os efeitos Epps anuais sobrepostos para motivos de comparações de sua forma. Todos os pares estudados apresentam o mesmo padrão qualitativo anual, entretanto desejo comparar a escala do efeito segundo o aumento do número de transações, para isso normalizo todas escalas de um efeito anual pela média dos valores considerados assintóticos [11], neste caso a partir da escala de 40 minutos (fig 4.9).

Visto que o tempo entre transações (fig 4.10) tem tendência decrescente, posso testar a hipótese de que, com maior sincronia entre transações, o efeito Epps é atenuado e a correlação entre retorno de ações vai para o patamar assintótico em menores escalas com o passar dos anos.

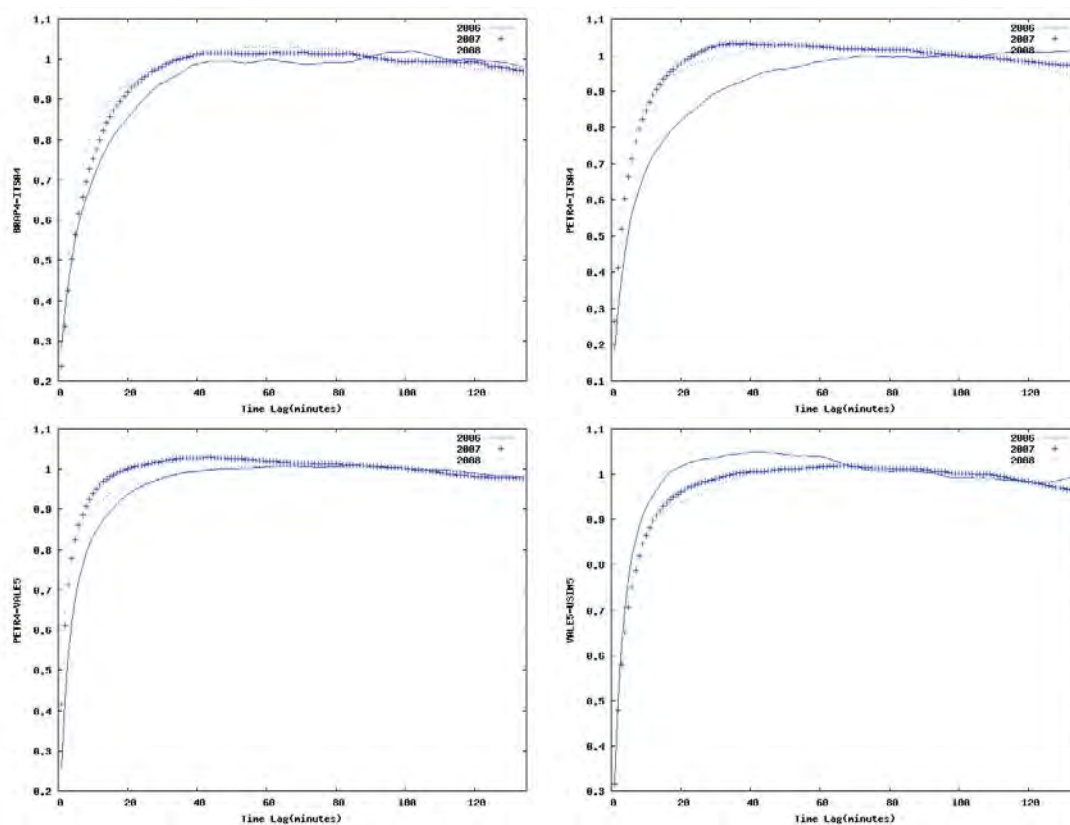


Figura 4.9: Efeitos Epps anuais normalizados e sobrepostos para os anos de 2006, 2007 e 2008.

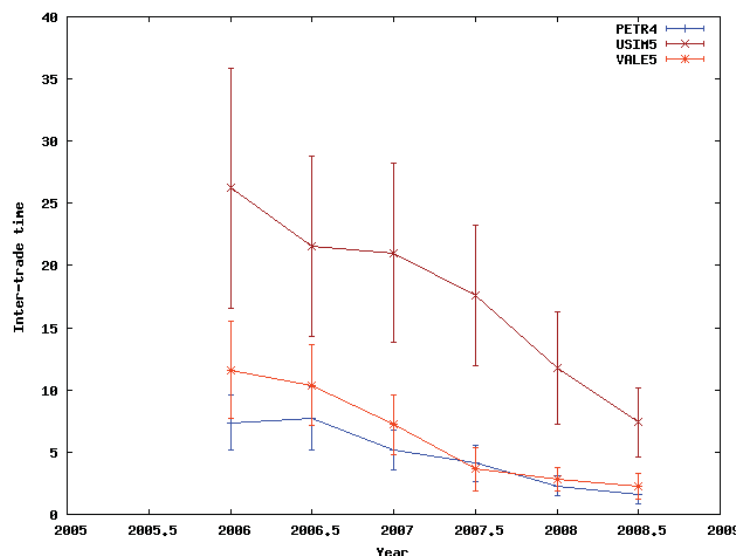


Figura 4.10: Tempo de espera entre transações para Vale, Usiminas e Petrobras, em minutos.

4.2.3 O que o efeito Epps diz sobre a organização do mercado ? O que é o efeito Epps ?

É conveniente começar pela análise e tentativa de resposta da segunda pergunta, de modo a compreender a física do que foi até aqui chamado de efeito Epps, para então colocar o que isso pode dizer a respeito da organização do mercado e as características dinâmicas possibilitadas.

Obtive que, apesar do cálculo anual com estimador de Pearson (3.3) e efeito de amostragem finita limita a significância do fenômeno diário observado, claramente a relação entre ações flutua ao redor de uma função qualitativamente similar para qualquer par de empresas, sendo tão nítida quanto maior a correlação nas escalas (tabela 4.1 e figuras 4.3, 4.4, 4.5, 4.6 e 4.7), portanto o efeito Epps existe como consequência de fatores analisados a seguir.

Na seção 4.2.1 expôs a possibilidade de haver assincronia, a escala investigada era de 2 minutos, levando a uma queda de correlação no período estudado. Entretanto verificou-se que a relação entre Petrobras e Vale e entre Usiminas e Vale já são eficientes, portanto não há influência na correlação e, apesar do estudo restringir-se às empresas, é razoável pensar que o mercado BovespaBMF mantenha a mesma eficiência para as demais correlações onde o efeito Epps foi estudado, devido a similaridade das curvas do efeito.

Prossigui estudando o efeito da assincronia nas transações como determinante do efeito Epps na seção 4.2.2, mas sob a perspectiva da diminuição no tempo entre transações das ações. A queda no tempo entre transações, como explicitado pelo tempo médio de entre transações 4.10, não foi fator determinante para mudança

de escala da curva Epps (fig 4.9), para quase a totalidade dos casos a diferença entre as curvas não é significativa para os níveis de significância estabelecidos em (Tabela 4.1), verifiquei ainda que o patamar assintótico de correlação é livre de escala em todos os casos a não ser no ano de 2006 para as correlações de Petrobras-Vale e Petrobras-Itau, cujas correlações tem regime assintótico em maiores escalas, as duas exceções devem ser entendidas como derivando de acontecimento econômico específico às empresas naquele ano (provavelmente Petrobras) e não dizem a respeito da origem do efeito Epps. O fenômeno é livre de escala no tempo entre transações, que por sua vez indica que o fenômeno não possui dependência exclusiva da sincronia de transações, ainda a medida de correlação de Pearson sendo uma medida de sincronia entre retornos de ações indica portanto que o fator pertinente deve ser limitante na informação econômica processada por investidores negociando papéis das duas empresas. Como já dito aqui e visto na seção 4.2.1, não existe atraso de informação econômica para 2 minutos, escala qual o efeito Epps se manifesta, logo este motivo não pode ser atribuído como único correlato anterior ao efeito. De posse dessas evidências exclusivas Tóth Kertész [12] conjecturam que, se apenas assincronia dos ticks não explica o efeito, então deve haver uma escala temporal humana de reação as mudanças no mercado.

Aceita desta forma as investigações, sobre assincronia de transações, nos dão 3 razões para o efeito Epps: assincronia deve diminuir a correlação, a assincronia deve ser maior se há lag na informação do mercado e, mesmo sem esses dois fatores presentes, o efeito Epps é essencialmente causado de pela assincronia derivada na escala limitante humana de processamento de informação.

Tendo somado as condições, estabelecido um motivo físico para o efeito, podemos agora responder a pergunta sobre a organização do mercado. Consideramos que os agentes tem um tempo mínimo de atualização de suas estratégias, então aqueles que se estabelecem em frequências altas suficientes irão compor um mercado homogêneo e opaco a qualquer possibilidade de estrutura emergente. Por outro lado temos agentes aplicando estratégias, suficientes para surgir ligação entre ações, quais devem remeter a substancial informação econômica, se a ligação é nítida, em 30 minutos e em alguns casos, como de Vale e Petrobras, em escalas ainda menores. Ainda vejo a possibilidade de agentes, utilizando uma teoria de carteiras, agirem em escalas opacas e escalas heterogêneas com a intenção de diminuir o risco total de investimento.

Estas revelações sobre a organização do mercado aponta a favor de um mercado: heterogêneo em perfis de agentes investindo ao longo de diversas escalas, cada qual com uma fase diferente e, em particular, com rápida formação de perfil relacional. Apesar do mercado se mostrar eficiente, ele não se assemelha a visão econômica clássica de agentes racionais maximizando sua utilidade de compra e venda, sendo que o mercado absorve imediatamente essa vontade e atualiza seus preços, é difícil ver como este ponto de vista poderia dar origem a um mercado que transita do opaco para o heterogêneo e estruturado.

4.3 Árvores de mínimo alcance, hierarquias e dinâmica de setores

A Análise topológica, através das árvores de mínimo alcance, do mercado Bovespa aqui conduzida é baseada nas idéias de níveis de ocupações utilizadas em J.-P. Onnela et al (2003 [9]). A seguir defino os conceitos básicos utilizados nesta análise, assim como a metodologia envolvida durante a obtenção dos resultados.

Utilizo o estimador para correlação cruzada instantânea (equação 3.1) considerando uma série temporal de retornos calculada em uma escala de tempo (equação 4.1). Assim como no estudo efeito Epps (seção 4.2) $C_{ij}^{\Delta t}$ dependerá da escala tratada, aqui analiso 3 escalas de tempo intra-day (respectivamente 2, 6 e 10 minutos) ao longo de 2006 , e duas ao longo de 2007 e 2008 (2 e 6 minutos). Apesar da análise contemplar um período de no máximo 10 minutos, substancial diferença na dinâmica estrutural é observada entre as três escalas. Através da análise comparativa das árvores taxonomicas tenho a possibilidade de inferir setores economicos emergentes, níveis de vizinhança em árvores fornecem uma medida de ramificação da estrutura que pode ser interpretada como setor economico.

O processo de obtenção da árvore taxonomica se baseia no algoritmo de kruskal e em sua equivalência com um espaço ultramétrico subdominante (seções 3.5 e 3.4.1), é feita uma inspeção visual nos ramos da árvore resultante para cruzamento com informação setorial economica¹. Acesso os níveis de ocupação das árvores de mínimo alcance determinando uma ou mais vértices raízes e particionando o conjunto de vértices em graus de vizinhança em relação a raíz (a quanto passos um vértice está da raiz).

4.3.1 Tratamento de dados da Bovespa&BMF

Foram extraídas de uma carteira inicial contendo 65 ações, baseada no índice IBrx de 2009 contendo as principais ações negociadas, 9 carteiras sendo 3 em cada ano e uma por cada uma das escalas trabalhadas. Cada carteira foi montada de forma a abranger o maior número de ações que participaram do pregão com maior frequência, todas ações com dias de participação abaixo de 170 dias (aproximadamente 70% dos dias) foram excluídas.

Utilizei o mesmo processo da seção 4.2.2 para tratamento de dados, uma média anual das correlações, desta forma determino a estrutura intraday mais frequente ao longo de cada ano. O desvio padrão corresponde a medida de espalhamento ao redor do fenômeno médio e pode ser retirada do estudo do efeito epps(seção 4.2.2) para as principais empresas. Comparando desvios padrões de correlações com

¹A escolha de uma ultrametria subdominante deve ser cruzada com informação economica representativa, verificando a conveniência da escolha que não pode ser acessada pela teoria de espaços ultramétricos a-priori.

diversas relações μ/σ reconheço o desvio padrão está sempre entre (0.02, 0.23), geralmente indo do menor para o maior valor com aumento da escala de tempo (mas não monotonicamente).

4.3.2 Carteiras examinadas

Seleciono primeiramente uma carteira baseada na composição do índice IBrx de janeiro de 2009(101 ações), a idéia é investigar nos anos anteriores(2006, 2007 e 2008) as ações que são as mais transacionadas no final do período amostrado. Em seguida seleciono uma carteira anual para cada um dos anos estudados, aqui o objetivo foi eliminar aquelas ações com poucos dias de transação, em 2006 seleciono 59 ações das 101, em 2007 seleciono 69 e em 2008 seleciono 85.

4.3.3 Ocupação de árvore, vértices raízes e camada de ocupação média

Aqui apresento os conceitos que nos possibilitará o estudo de proximidade e ocupação dentro da árvore de mínimo alcance. Começo por definir o conceito de vértice raiz (ou vértice central), este vértice deverá ser o que mais influente em mudanças estruturais da árvore. Na literatura encontrei diversas opções para definição de vértice raiz v_r , cada uma dando ênfase em uma propriedade diferente do grafo.

As arestas podem definir os vértices raízes como:

$v_r = \{v \in V | \text{Min}\{\sum_{1 \leq i \leq N_{E_{v_s}}} \frac{1}{N_{E_{v_s}}} e_{v_{s_i}} : e_{v_{s_i}} \in E_{v_s} \subset E, 1 \leq s \leq S\}\}$ (chamado de método do centro de massa na literatura), s define o grau de vizinhança dos vértices (quantos passos até chegar em v) e um subconjunto de E , todas arestas anteriores aos vizinhos s de v e S sendo o maior número de passos entre dois vértices quaisquer, esta definição de v_r irá privilegiar os vértices com menor distância à vizinhança escolhida. Outro critério, que se baseia na quantidade de influências de uma ação, faz a escolha do vértice com maior grau de conectividade k com relação aos primeiros vizinhos. Menciono que é possível escolher um v_r meio termo das duas possibilidades, definindo um parâmetro mínimo através do valor médio das distâncias aos primeiros vizinhos, aplicar parâmetro aos vértices considerando apenas aqueles vizinhos com distância abaixo do parâmetro, por fim escolhe-se o de maior conectividade k . Na minha análise foi utilizada a penúltima definição de v_r , entretanto defino, além do vértice raiz v_{r_0} , os segundos vértices mais importante v_{r_1} (raiz de setores e subsetores do grafo), motivado a posteriori pelos resultados analisados que exibiram aglomeração de ações centradas em mais de um vértice.

Após fixado o vértice raiz v_r posso investigar os níveis de ocupação, a medida da localização da parte média dos vértices em relação ao v_r é definida,

9. DEFINIÇÃO. *é chamado de nível de ocupação médio, v_r o vértice raiz, e dado*

por:

$$l(t, v_r) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(v_i, t|v_r), \quad (4.3)$$

N é o número total de vértices na árvore, $\mathcal{L}(v_i, t|v_r)$ é o nível do vértice v_i com relação a raiz v_r , a dependência temporal pode mudar o valor do nível de cada vértice e/ou pode estar no vértice raiz.

A mesma análise pode ser feita para setores, subconjuntos do conjunto V de vértices da árvore principal. Aqui conduzo esse estudo em paralelo a investigação de ramificação partindo dos vértices raízes com intenção de acessar a importância relativa das raízes a cada setor. Definirei *setores* na árvore de mínimo alcance $T(V, E)$ como $V' \subset V$, obtido através da operação de deleção de um vértice e todas arestas que levam a ele, leva um grafo a outro grafo, $D : T(V, E) \rightarrow F(V', E')$ (isso é equivalente a zerar, linha e coluna de um vértice, na matriz de adjacência). Um *setor* é um subconjunto de V' , se conveniente é possível definir subsetores aplicando aos sub-grafos, definidos nos subconjuntos de V' , a mesma metodologia aplicada para setores.

4.3.4 Equilíbrio polar entre 2 setores

Um dos meios de estudar a árvore de mínimo alcance é definindo uma floresta de duas árvores a partir da deleção de uma aresta dividindo a árvore em duas. Estuda-se as propriedades de ocupação média das duas árvores e a relação existente entre elas nas 3 escalas de tempo estudadas, para isso defino centro de massa e massa reduzida para a ocupação média das duas árvores.

A escolha da aresta define a correlação principal a dividir os dois níveis de ocupação médios estudados, o estudo prossegue definindo um procedimento de contabilização para nível de ocupação médio. Ao invés de contabilizar a totalidade de vértices presentes na árvore, procurei incluir apenas os vértices mais fortemente interagentes com o nível anterior, para isso defino *centro de massa* do nível i , com

relação ao vértice raiz v_r , como $\bar{e}(i, v_r) = \frac{1}{N_i} \sum_{j=1}^{N_i} e_{j, v_r}^{(i)}$, onde $v_r \in V$ e $e_{j, v_r}^{(i)} \in E$,

$F(V, E)$ é a floresta em questão, contabiliza-se o número de ocupação do nível i apenas com os vértices cujas arestas, conectando ao nível anterior, tenham valor menor ou igual a este centro de massa, ainda excluí do cálculo de centro de massa a aresta que une os vértices representando a mesma ação (por exemplo USIM5 e USIM3), na maioria do caso são vértices que possuem mesma origem econômica².

² Não representam uma interação de mercado diversa mas sim uma mesma empresa artificialmente subdividida, de modo a beneficiar dois perfis de investidores, aqueles com ações ditas Ordinárias apenas possuem uma porcentagem da empresa, aqueles com ações ditas Preferências possuem direito a voto em assembléia.

Após calculado o centro de massa das duas árvores contabilizo o *centro de massa da ocupação da árvore* $F(V, E)$ e a *massa reduzida de ocupação da árvore* $F(V, E)$:

$$l_e^{cm}(l(\Delta t, v_r'), l(\Delta t, v_r'')) = l(\Delta t, v_r') + l(\Delta t, v_r'') \quad (4.4)$$

$$l_e^{rm}(l(\Delta t, v_r'), l(\Delta t, v_r'')) = \frac{l(\Delta t, v_r')l(\Delta t, v_r'')}{l(\Delta t, v_r') + l(\Delta t, v_r'')} \quad (4.5)$$

Onde $e \in E$ é a aresta escolhida para deleção e separação das árvores, $v_r' \in V'$ e $v_r'' \in V''$ são os vértices raízes dos subconjuntos complementares de V que estão em cada árvore obtida na operação.

A medida centro de massa 4.4 aponta o quanto cada árvore é mais espalhada ao redor da aresta conectando duas raízes das sub-árvores dentro da árvore de mínimo alcance original. O centro de massa com valor próximo ao de uma das árvores indica que essa árvore tem mais vértices em nível afastado da aresta escolhida.

Massa reduzida 4.5 dá uma medida de afastamento aos níveis de maior ocupação de cada árvore, quanto maior seu valor mais afastados estão um nível de ocupação médio do outro.

4.3.5 Árvores de mínimo alcance nas escalas temporais de 2, 6 e 10 minutos

Em todas árvores estudadas farei primeiramente a análise de vértice raíz, em seguida farei análise setorial para depois relacionar as duas análises procurando contrastar com a informação econômica.

As árvores para o retorno das ações em 2006 foram calculadas e seus vértices foram pintados de acordo com a posição na classe econômica de cada ação³. A correspondência de vértice e classes pode ser encontrada na legenda.

³A relação de empresas e classes econômicas pode ser encontrada diretamente no site da Bovespa, www.bovespa.com.br

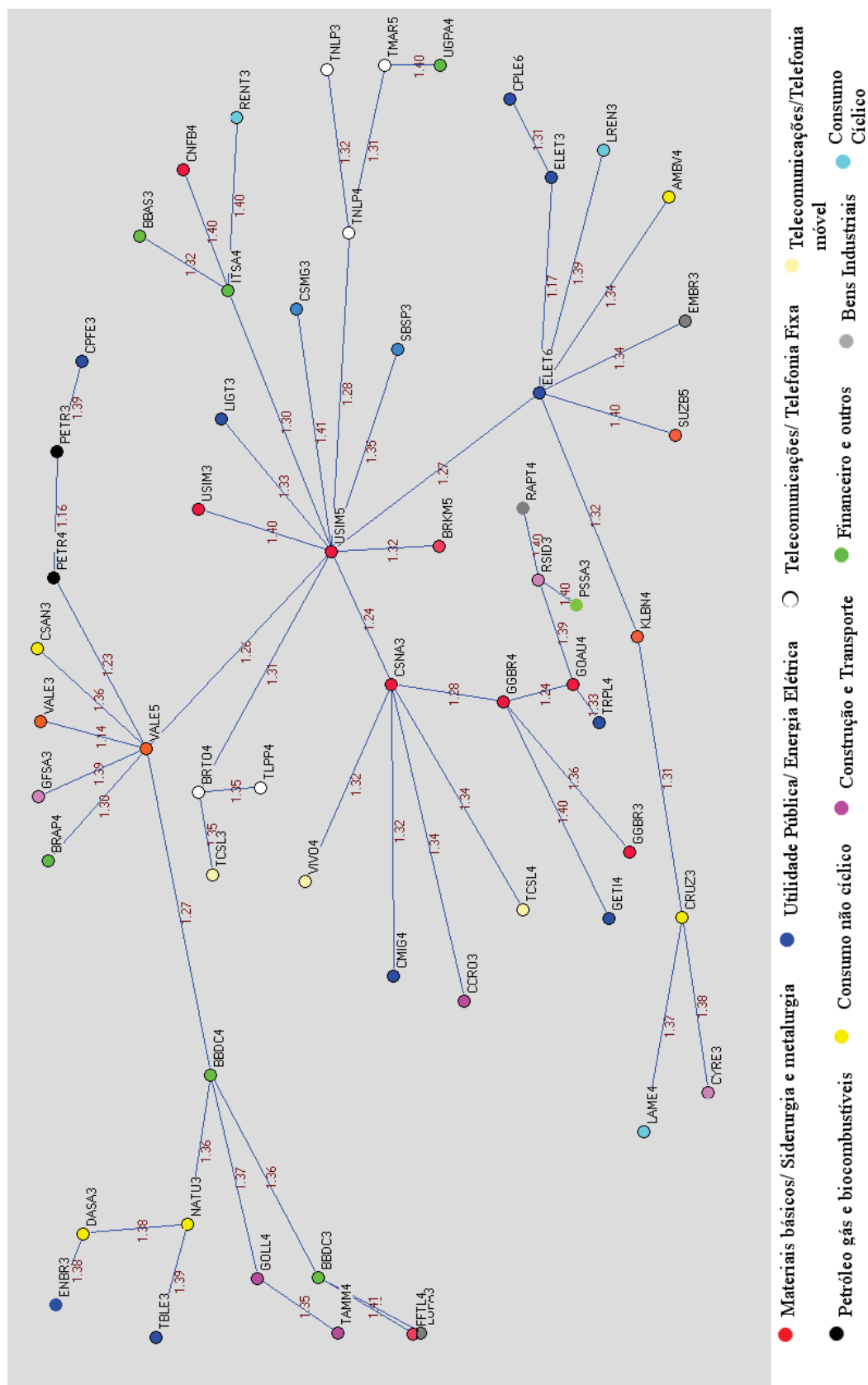


Figura 4.11: Árvore de mínimo alcance para 2006 na escala de retornos 2 minutos.

Em escala de retorno de 2 minutos, o grafo 4.11 apresenta um vértice raiz, USIM5, com a conectividade $k = 11$. O nível de ocupação médio é $l(120s, USIM5) = 2.5$ sugerindo que a maior parte dos vértices estão entre o nível 2 e 3 de vizinhança para USIM5. Reparo também que dos 11 primeiros vizinhos apenas 6 apresentam correlação acima de 0.14 com o vértice raiz e constituem as principais ramificações, contendo todos os segundos vizinhos possíveis.

Deletando USIM5 obtenho os 6 ramos principais, as raízes de cada setor são: CSNA3, VALE5, TNLP4, BRTO4, ELET6 e ITSA4.

O vértice raiz USIM5 pode ser admitido como fazendo parte do setor de CSNA3 lembrando que ambos fazem parte do setor intitulado Materiais Básicos/Siderurgia, este setor apresenta o nível de ocupação médio $l(120s, CSNA3) = 2$ e suas menores distâncias estão entre empresas de classificação Materiais Básicos/Siderurgia, sendo portanto um setor que atribui proximidade econômica a esta classificação, temos ainda vértices como VIVO4, TCSL4, CMIG4 e CCRO3 que estão ligados a menos que 0.13 de correlação à CSNA3, são de classificações heterogêneas entre si, apenas destaco que VIVO4 e TCSL4 são ambas de telefonia móvel.

O setor que tem VALE5 como raiz tem nível de ocupação médio $l(120s, VALE5) = 2.12$ apesar das ligações de segundos vizinhos não serem significativas (correlação aproximada de 0.06), os primeiros vizinhos mais relevantes são de classificações Financeiros e Outros e Petróleo, sendo o raiz posicionado em Materiais Básicos/Mineração posso associar este setor a áreas onde grande exploração de material bruto tem ligação às instituições financeiras.

TNPL4 é raiz de um setor onde $l(120s, TNPL4) = 1.33$ significando que a maioria dos vértices são primeiros vizinhos (2 dos 3), o setor é dominado por Telecomunicações/Telefonia Fixa incluindo o vértice raiz. BRTO4 é central (com $l(120s, BRTO4) = 1.33$) em um grupo de vértices em Telecomunicações.

O setor do vértice raiz ELET6, $l(120s, ELET6) = 1.6$, tem vértices em classificações heterogêneas sendo que, primeiros e segundos vizinhos, as principais são Utilidade Pública/Energia Elétrica (incluindo o raiz), Materiais Básicos/Madeira e Papel e Consumo não Cíclico. Tal setor aproxima neste ano uma correlação de curto prazo entre energia elétrica e materiais básicos.

O setor de ITSA4, $l(120s, ITSA4) = 1$, possui ligação relevante para BBAS3, ambos são da classificação Financeiro e Outros.

A taxonomia 4.18 do grafo mostra as principais ligações do grafo, dando ênfase a proximidade da hierarquia de VALE5 e PETR4 como as que melhor definem suas posições no mercado. Seguindo lendo as taxonomias, verificamos que o setor definido com CSNA3 e seu eixo com a raiz USIM5 define a próxima estrutura econômica à estruturar o mercado, seguido do setor de ELET6 e, ainda no setor de CSNA3, tenho GGBR4-GOAU4 que possuem posição bem definida no setor que ramifica das hierarquias presentes em taxas mais baixas.

Seguindo examinando a árvore de mínimo alcance intraday em 2006, agora a escala analisada é de retornos em 6 minutos, segue o grafo:

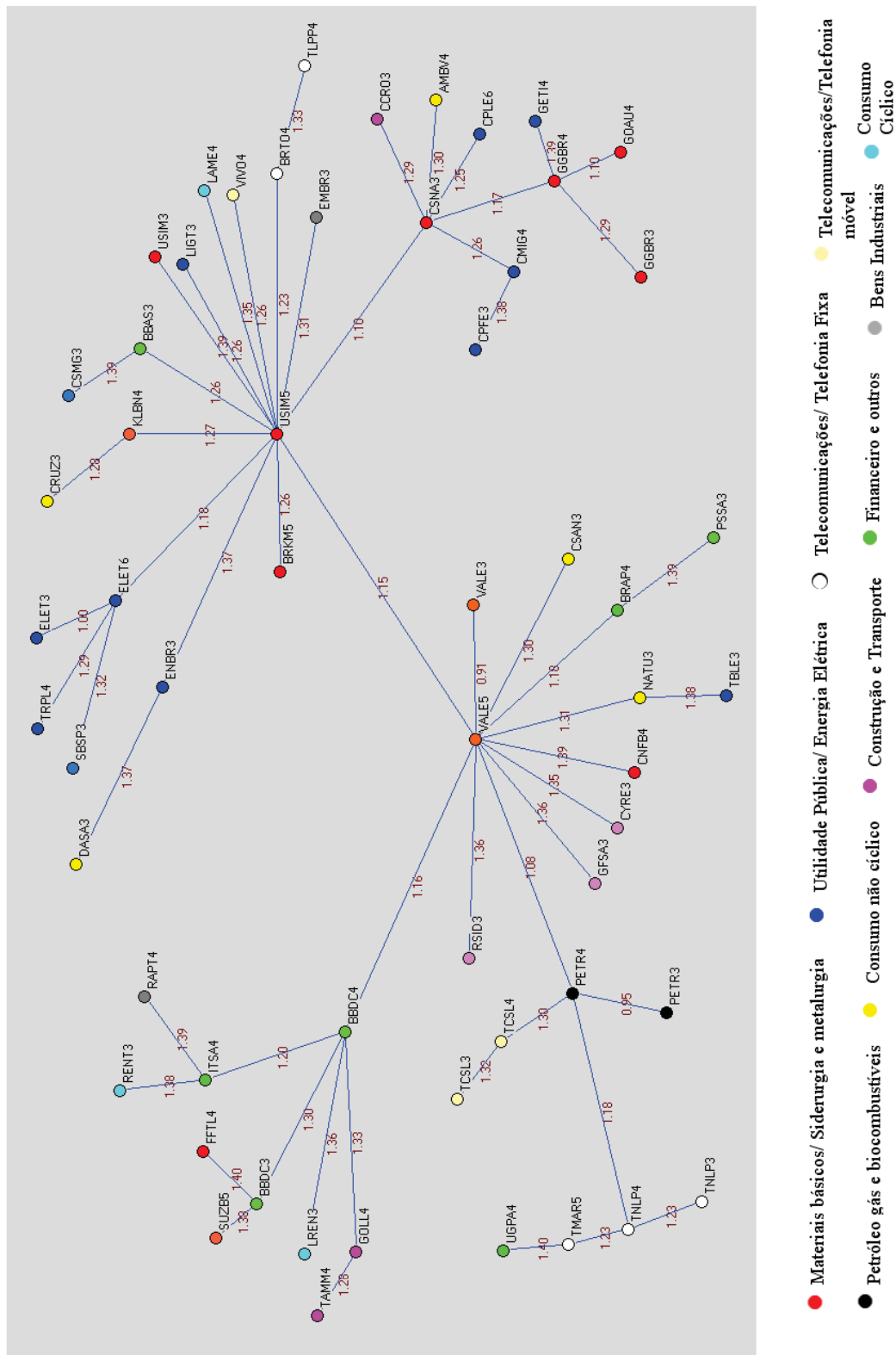


Figura 4.12: Árvore de mínimo alcance para 2006 na escala de retornos de 6 minutos.

Na escala de retorno em 6 minutos o vértice raiz do grafo 4.12 continua sendo USIM5, $k = 13$, mostrando sua estabilidade em influenciar os setores. Este grafo possui um nível de ocupação médio $l(360s, USIM5) = 2.33$, nesta escala os primeiros vizinhos de USIM5 tem em média 0.21 de correlação e maior parte dos segundos vizinhos se conectam por correlação maior que 0.14.

Retirando USIM5 obtenho 10 setores, cujo os vértices raízes são os seguintes: BRKM5, KBLN4, ELET6, BBAS3, BRTO4, LIGT3, VIVO4, EMBR3, CSNA3, VALE5.

BRKM5, VIVO4, LIGT3 e EMBR3 são vértices que são unicamente ligados ao raiz, são de classificações heterogêneas, sendo BRKM5 da mesma classificação do raiz.

KBLN4, BBAS3 e BRTO4 cada um forma um setor de 2 vértices isolados. KBLN4 forma setor com CRUZ3 apesar de estarem em classes economicas diversas, Materiais Básicos e Consumo Não Cíclico respectivamente. BRTO4 forma setor de Telefonia Fixa com TLPP4, apesar da correlação ser baixa para a média de segundos vizinhos em USIM5 (0.21).

ELET6 forma setor com $l(360s, ELET6) = 1.0$ e seus vizinhos estão ligados a Utilidade Pública.

O setor que tem como vértice principal CSNA3, $l(360s, CSNA3) = 1.5$, apresenta a coerência do método utilizado. Achei que este setor é guiado pelas empresas de Materiais Básicos/Siderurgia, com concentração de vértices dessa classificação na região do nível de ocupação médio. O setor apresenta-se heterogêneo, sendo que alguns vértices são da classe de Energia Elétrica e outros dois de Construção e Transporte e Consumo não Cíclico. Lembro agora que o vértice raiz USIM5 pertence a classe Materiais Básicos/Siderurgia, BRMK5 também é da mesma classe, o que sugere que a influência do setor se estende pelos vértices da classe siderurgica.

O setor do vértice VALE5 com $l(360s, VALE5) = 1.89$ exibe um tronco principal, aonde a raiz interagem fortemente com PETR3, da classe Petróleo gás e biocombustíveis, e com vértices que correspondem às classes Financeiras e Outros. A grande quantidade de subsetores possíveis dá o nível de ocupação próximo a 2 para este setor, destaco o agrupamento ligado a PETR4, 3 vértices de classe Telefonia Fixa, ainda as ações de Financeiros e Outros que se ligam a outras classes diversas, por exemplo os dois vértices de Construção e Transporte GOLL4 e TAMM4.

Na taxonomia 4.19 é confirmada a estrutura mais economicamente fixa da árvore, já vista com 2 minutos de retornos. Existe agora o aumento coletivo e simultâneo da proximidade entre ações, acentuando as relações de setores e de arestas com menores distâncias.

A última escala intraday estudada, 10 minutos, deve me mostrar a persistência de vértices com níveis de conectividade diversos, revelando os vértices que compõe a estrutura de influência estável em todas escalas temporais de retornos.

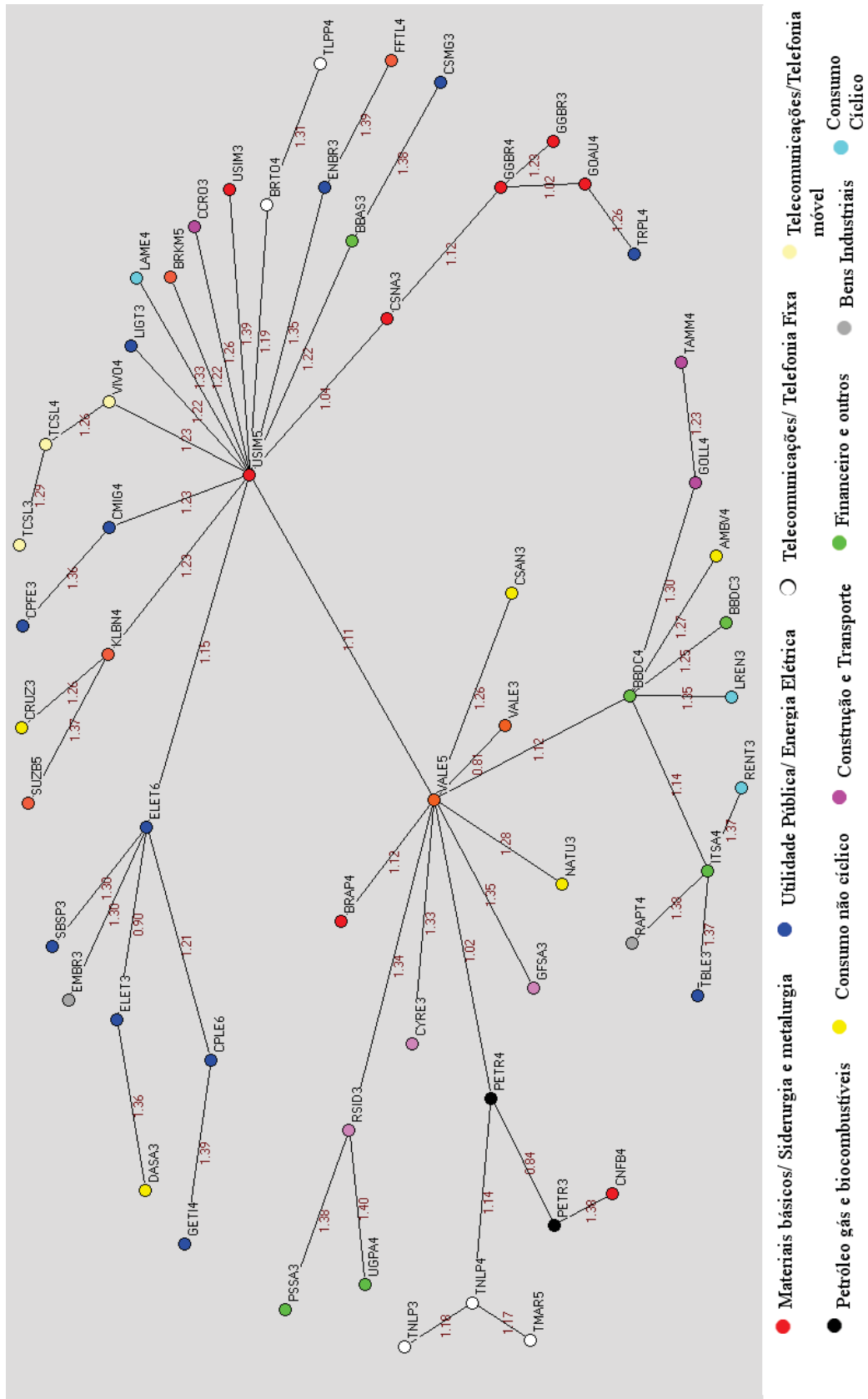


Figura 4.13: Árvore de mínimo alcance para 2006 na escala de retornos de 10 minutos.

O grafo 4.13, escala de tempo em retornos de 10 minutos, mantém o vértice USIM5 como raiz, mostrando sua estabilidade de influência no grafo em longo prazo. O nível de ocupação médio $l(600s, USIM5) = 2.3$ permanece estável com relação à escala de 360 segundos, sendo possível inferir estabilidade de ocupação para maiores escalas.

Deletando o raiz USIM5 obtenho vários setores dos quais posso destacar aqueles com raízes GGBR4, ELET6 ou VALE5. GGBR4 tem como primeiros vizinhos vértices com classificação de Materiais Básicos/ Siderurgia, CSNA3 é vizinho do vértice USIM5 que pertence a mesma classe economica. ELET6 tem primeiros vizinhos mais próximos vértices pertencentes a mesma classe economica Utilidade Pública/Energia Elétrica. VALE5 é raiz de um setor cuja ocupação média $l(600s, VALE5) = 1.92$, os vértices ocupando o segundo nível tem classificação que sugere a subdivisão em subsetores, posso destacar o agrupamento de vértices de Telefonia Fixa e de Financeiros e Outros ao redor da correlação VALE5-PETRA4.

Para esta escala de tempo a taxonomia 4.20 apresenta a mesma estrutura com diminuição nas distâncias, destaque para BRAP4 que faz parte dos eixos de ramificação e, pelo grafo 4.13, vejo que faz parte do setor de outra Financeira e Outros a BBDC4.

Centro de massa e massa reduzida de ocupação nas árvores de 2006

As coordenadas de centro de massa e massa reduzida de ocupação mostram estáveis na escala temporal, notavelmente as medidas são as iguais para 360 segundos e 600 segundos. É possível visualizar nos grafos 4.11, 4.12 e 4.13 o significado das coordenadas de polaridade, na escala de 120 segundos a estrutura visível (no sentido apresentado na seção 4.2.2) envolve poucos vértices que permaneceram exercendo grande influência no sistema, possuindo alta conectividade e ocupação média próxima de 2 (lembramos que isso pode significar que o vértice é ponto de partida para vários subsetores) , o grafo ainda é centralizado em USIM5. Nos grafos de 360 segundos e 600 segundos tenho uma estrutura de árvore polarizada entre USIM5 e VALE5 que não varia com a escala de tempo, massa relativa próxima de 1 indica maior proximidade entre as ocupações médias.

2006 - Centro de massa e massa reduzida de ocupação		
Δt	l_e^{cm}	l_e^{rm}
120s	3.57	0.86
360s	3.71	0.92
600s	3.71	0.92

Tabela 4.2: Tabela com estimativas de polaridade das árvores para 2006

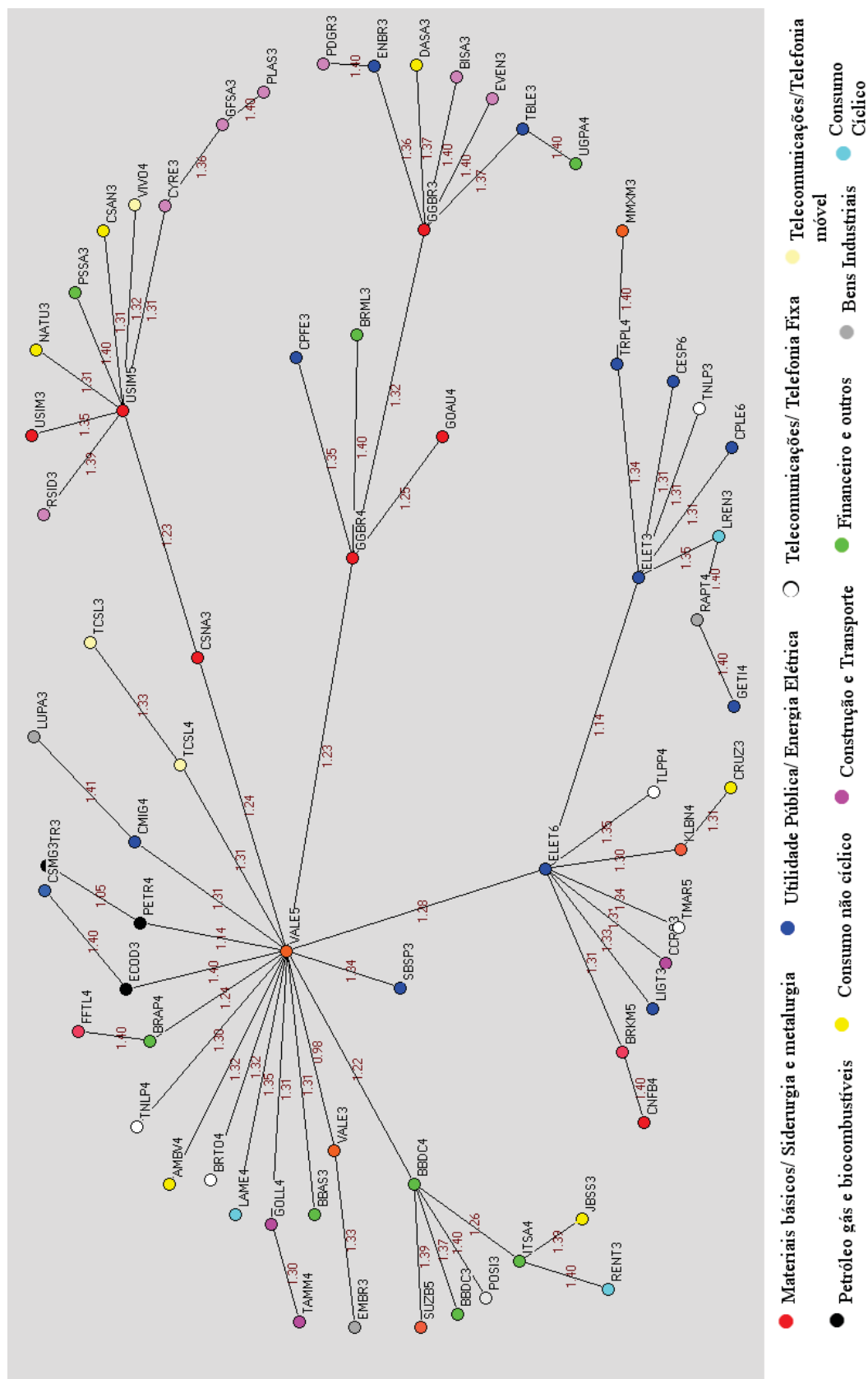


Figura 4.14: Árvore de mínimo alcance para 2007 na escala de retornos 2 minutos.

Em 2007 o vértice raiz v_r passa a ser VALE5, com $k = 17$ e nível de ocupação médio $l(120s, VALE5) = 2.26$ indicando ramificações de setores e sub-setores. Os principais setores tem vértices raízes ligados a Utilidade Pública/Energia Elétrica, Materiais Básicos/Siderurgia e Financeiros e Outros, ainda vejo subsetor de Construção e Transporte ligado ao setor de Siderurgia. O setor ligado à Energia Elétrica, com raiz em ELET6 $l(120s, ELET6) = 1.5$, apresenta uma estrutura ramificada visível já nesta escala.

A taxonomia do período nesta escala(4.21) exhibe a mesma estrutura de setores e ramificações que as de 2006, hierarquias: VALE5-PETR4, USIM5-CSNA3, ELET6-ELET3 e TMM4-GOLL4. Além disso o setor financeiro se apresenta como ramificações das taxas mais baixas.

A análise de polaridade do grafo 4.14 foi feita em 3 arestas centrais diferentes. Esta variação na análise se justifica pelo fato de termos ramificações em sub-setores relevantes para os níveis de ocupação supra-citados, com nível de ocupação médio maior que 2 em v_r , indicando pequenas ramificações de setores em sub-setores(por exemplo o subsetor composto por CYRE3, GFSA3 e PLAS3) e setores mais ramificados como o de Energia Elétrica.

Foi analisada polaridade na escolha das aresta VALE5-CSNA3(1.24), VALE5-GGBR4(1.23) e VALE5-ELET6(1.28). Em VALE5-CSNA3 obtive os valores $l_e^{cm} = 3.05$ e $l_e^{rm} = 0.67$, o primeiro indicando a existência de 3 setores predominantes, 1 com raiz em USIM5 e os outros dois se ramificam de VALE5 (raízes ELET6 e GGBR4), o segundo aponta a distância assimétrica dos níveis de ocupação em relação a aresta escolhida, o vértice VALE5 se posiciona como intermediário de vários setores, enquanto USIM5 está em um setor com nível de ocupação médio 1, tal configuração é característica do grafo mais ocupado nas ramificações setoriais. Deletando VALE5-GGBR4 obtenho uma configuração com $l_e^{cm} = 3.26$ e $l_e^{rm} = 0.61$, similar aquela obtida no vértice VALE5-CSNA3, análise similar pode ser feita. Em torno da aresta VALE5-ELET6 obtenho uma massa reduzida maior, $l_e^{rm} = 0.88$, indicando maior simetria entre ocupações dos dois setores e $l_e^{cm} = 3.57$ aponta para fato de que existem 3 sub-setores dominados por classificação econômica relevante e que eles se ramificam de forma a propiciar outros sub-setores.

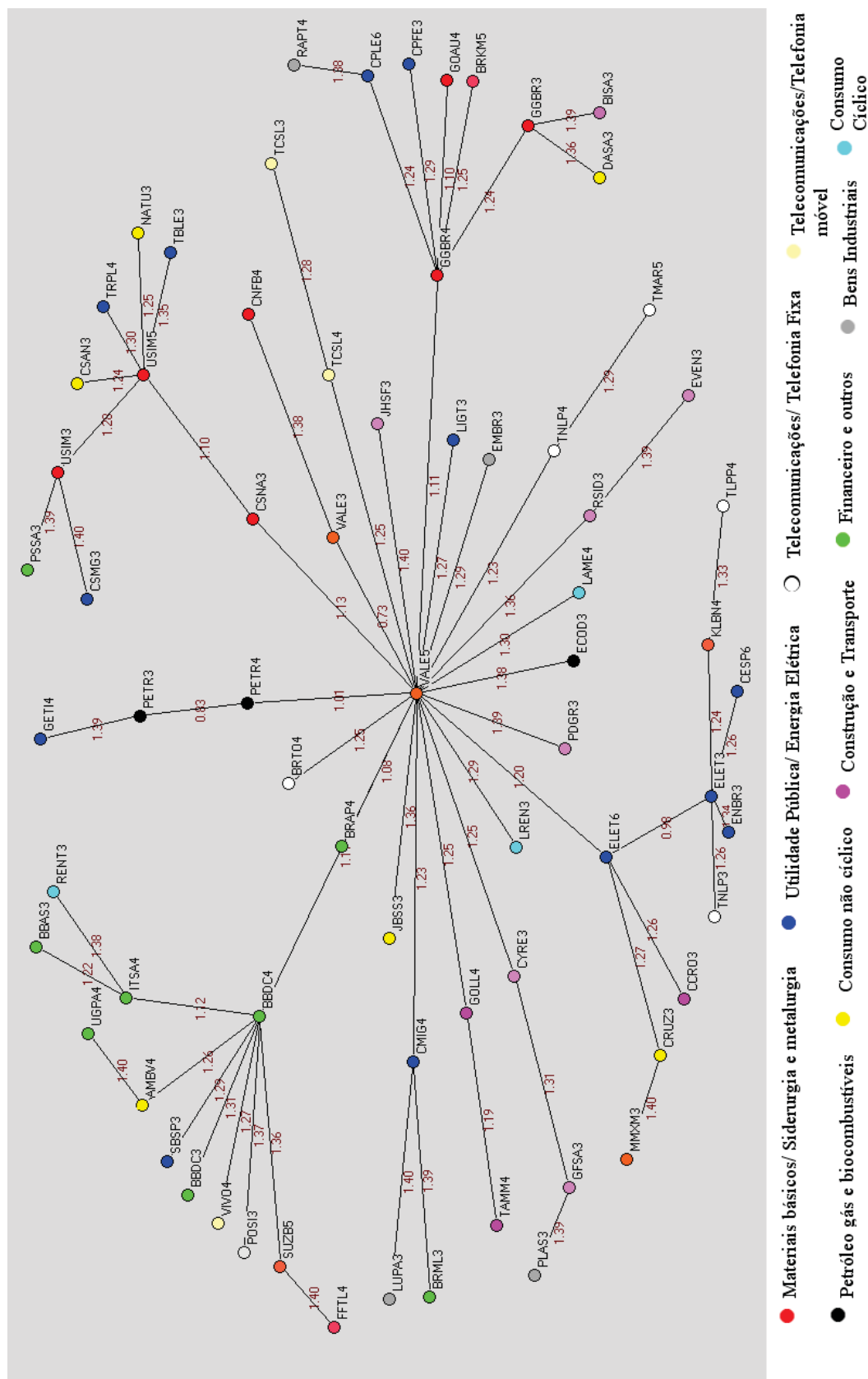


Figura 4.15: Árvore de mínimo alcance para 2007 na escala de retornos de 6 minutos.

Na escala de tempo de 6 minutos, o raiz é VALE5, o grafo tem nível de ocupação médio $l(360, VALE5) = 2.04$ levando à um grafo de setores com poucas ramificações interiores. Dos vários setores, destaca-se o financeiro onde um eixo une BBDC4-ITSA4-BBAS3, três vértices pertencentes à classe Financeiro e Outros. O setor enraizado em ELET6, da classe Utilidade Pública/Energia Elétrica, leva a CESP6 e em outras diversas como TNLP3(Telecomunicações) e KLBN4(Materiais Básicos/Papel), possui $l(360, ELET6) = 1.5$.

Percebi através da taxonomia 4.22 que o eixo do setor financeiro se posiciona como ramificação das principais taxas, exibindo proximidade econômica, sequencialmente nos quatro vértices BRAP4-BBDC4-ITSA4-BBAS3, concordando com o grafo 4.15 que ainda nos exhibe o posicionamento inter-setorial.

Apliquei a análise de polaridade do grafo 4.15 nas arestas dos mesmos pares de vértices da escala anterior(2 minutos) para comparação, ou seja VALE5-CSNA3(1.13), VALE5-GGBR4(1.11) e VALE5-ELET6(1.2).

Obtive, em VALE5-CSNA3, para centro de massa $l_e^{cm} = 2.77$ e $l_e^{rm} = 0.64$ para massa reduzida, indicando a presença de 3 ramos principais, entretanto a ramificação em sub-setores é quase inexistente, o valor da massa reduzida aponta para que a árvore de VALE5 possui ramificação em setores e polariza a relação.

Resultado similar ao de VALE5-CSNA3 foi obtido para VALE5-GGBR4(tabela abaixo), indicando que a relação das árvores enraizadas em VALE5 e GGBR4 é similar a de VALE5 e CSNA3.

Centro de massa e massa reduzida de ocupação nas árvores de 2007

Vejo uma queda no valor do centro de massa enquanto o valor da massa reduzida tem pequena variação para os três casos estudados. Tal comportamento pode ser entendido como rearranjo de algumas empresas ao redor de VALE5 diminuindo a concentração de vértices ocupando níveis maiores em relação a VALE5.

Δt	l_e^{cm}	l_e^{rm}
2007 - Centro de massa e massa reduzida de ocupação VALE5-CSNA3		
120s	3.05	0.67
360s	2.77	0.64
2007 - Centro de massa e massa reduzida de ocupação VALE5-GGBR4		
120s	3.26	0.61
360s	2.85	0.65
2007 - Centro de massa e massa reduzida de ocupação VALE5-ELET6		
120s	3.57	0.88
360s	3.3	0.82

Tabela 4.3: Tabela com estimativas de polaridade das árvores para 2007

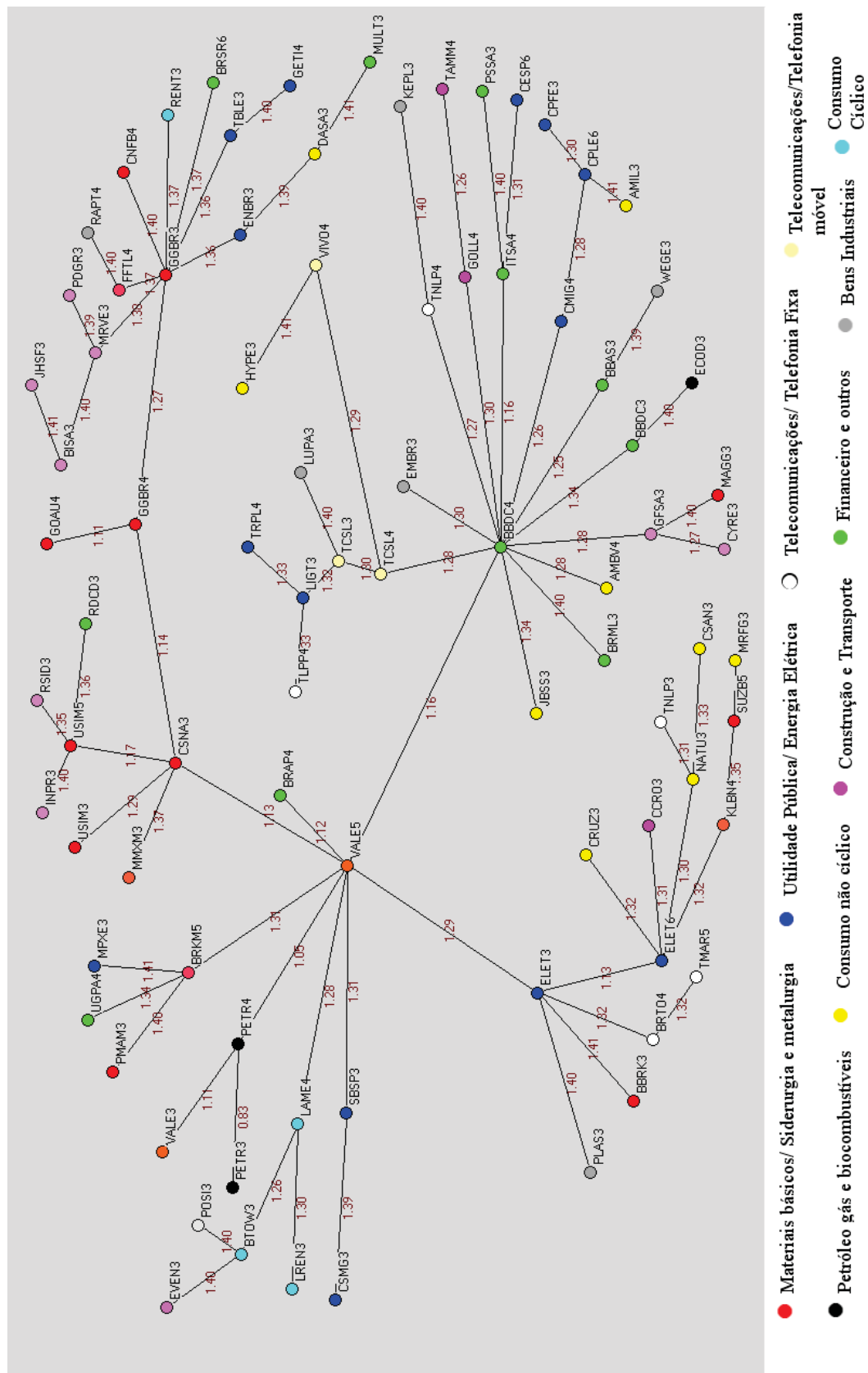


Figura 4.16: Árvore de mínimo alcance para 2008 na escala de retornos 2 minutos.

Em 2008, ano da crise mundial financeira do subprime, no mercado Bovespa as empresas ligadas ao setor Financeiro e Outros adquirem maior importância na estrutura de correlações. O vértice raiz v_r passa a ser BBDC4 em 2008, vértice pertencente a classe Financeiro e Outros, com nível de ocupação médio $l(120s, BBDC4) = 3.19$, indicando um mercado com mais subsectores ramificados de sectores heterogêneos.

O sector que tem como vértice raiz VALE5 possui subsectores, dentro desses destaque o sub-sector com maior ramificação $l(120s, CSNA3) = 1.33$, vértices predominantemente pertencentes a classe de Materiais Básicos/Siderurgia, possui ramificações em sub-setores mas de posição pouco visualizável (por exemplo o sub-sector ligado a Construção e Transporte, estão conectados por distâncias grandes). Pela primeira vez vi um subsector ligado a Consumo Cíclico aparecer com posição visualizável nesta escala de tempo. Tenho ainda sectores de Telecomunicações (TCSL3 e VIVO4) e Energia Elétrica (CMIG4, CPLE6 e CPFE3).

Na taxonomia 4.23 deste grafo é possível verificar novos posicionamentos de taxas em relação à estrutura principal dos anos 2006 e 2007. É notável que a estrutura, formada pelas taxas mais específicas, tenha se modificado para uma ramificação sequencial a partir das gêmeas PETR4 e PETR3, vejo que VALE5, BRAP4, CSNA3 e ELET6 vértices que formam o a posição mais específica de sectores e subsectores, estão antecedendo taxas de classes economicas a quem também pertencem (condizente com a ramificação do grafo 4.16). Ainda posso visualizar as taxas de BTOW4-LAME4, CYRE3-GFSA3, TAMM4-GOLL4 e BBDC4-ITSA4 definindo o eixo de sectores e subsectores diversos.

Fiz a análise de polarização ao redor da aresta BBDC4-VALE5(1.16), obtive o centro de massa $l_e^{cm} = 3.65$ e a massa reduzida $l_e^{rm} = 0.91$. O centro de massa aponta a presença de subsectores dominantes com subsectores menores ramificando a partir destes. A massa reduzida indica que estas duas árvores dividem igualmente o povoamento de seus níveis de ocupação, pelo valor estar próximo de 1.

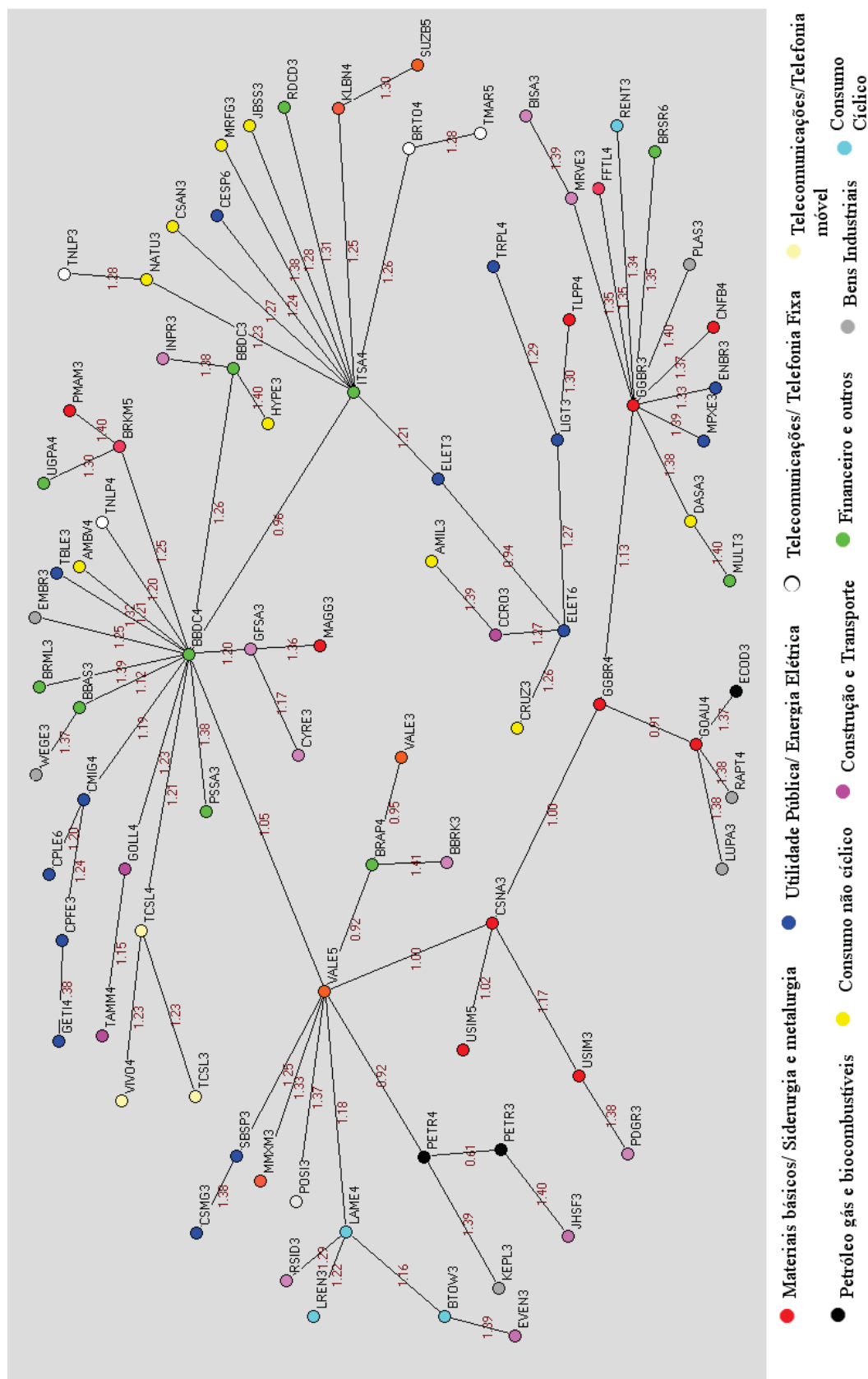


Figura 4.17: Árvore de mínimo alcance para 2008 na escala de retornos de 6 minutos.

Na escala de 6 minutos temos nível de ocupação médio, a partir de BBDC4, de $l(360s, BBDC4) = 2.83$, superior aos níveis medidos em outros anos, contrário do obtido por J.-P. Onnela et al. (2003 [9]) e por Y. Zhang et al. (2011 [14]) para a NYSE, aonde em momentos de crise o nível de ocupação médio (baseado no vértice raiz v_r de maior grau de conectividade) diminui, essa discrepância sugere que o mercado Bovespa&BMF passou de forma diferenciada pelo impacto da crise de 2008. Este nível de ocupação médio sugere, assim como acontece na escala 2 minutos, que o grafo possui ramificação em subsetores importantes, particularmente os vértices da classe Energia Elétrica, Telefonia Móvel e Construção e Transporte ramificados das financeiras BBDC4-ITSA4 e subsetores ramificados de VALE5, como LAME4-BTOW3 para Consumo Cíclico e os vértices da classe Materiais Básicos/Siderurgia, CSNA3 é o vértice mais próximo ao raiz do setor VALE5.

A taxonomia 4.24 do mercado nesta escala revela que a estrutura básica de taxas com maior proximidade econômica permanece a mesma da escala de 2 minutos, reparei que a diferença principal se situa no posicionamento das ações que possuem correlação mais ruidosa, ou de estrutura menos visível na escala utilizada, tal comportamento é também observado nas outras árvores e taxonomias. Apesar disso a estrutura básica é suficiente para determinar setores e, como foi visto, estes setores são coerentes com a classificação econômica empregada no mercado.

Estudei a polarização deste grafo com relação a aresta VALE5-BBDC4(1.05), que apresentou os valores $l_e^{cm} = 4.22$ para o centro de massa e $l_e^{rm} = 1.04$. Ambos os valores são maiores que o da escala antecedente, com o centro de massa sendo bem mais sensível as mudanças nas árvores. O valor do centro de massa indica ramificação de vários subsetores fazendo parte da estrutura estacionária, partindo de VALE5 ou BBDC4 posso chegar até um subsetor em aproximadamente 4 passos. Maior quantidade de sub-setores aumenta o nível de ocupação médio de ambos os lados da aresta, levando a uma massa reduzida maior, tem maior contribuição da árvore na raiz BBDC4 por possuir nível de ocupação médio $l(360s, BBDC4) = 2.3$ maior que a árvore enraizada na VALE5 que possui $l(360s, VALE5) = 1.9$.

Centro de massa e massa reduzida de ocupação no período de 2008

Em 2008 vê-se um pulo, entre as duas escalas, atípico mesmo para o centro de massa, a realocação dos vértices ruidosos na escala 360 segundos pode ser um dos motivos, mesmo a massa reduzida, que se apresenta menos sensível a mudança de escala, varia mais seu valor que nos outros anos (4.4).

2008 - Centro de massa e massa reduzida de ocupação VALE5-BBDC4		
Δt	l_e^{cm}	l_e^{rm}
120s	3.65	0.91
360s	4.22	1.04

Tabela 4.4: Tabela com estimativas de polaridade das árvores para 2008

4.3.6 Informação fornecida por árvores de mínimo alcance e taxonomias

Devo esclarecer aqui a natureza da informação obtida com o método da árvore de mínimo alcance e a visualização das árvores taxonomicas, a pergunta a ser feita é a seguinte: é possível extrair informação na dinâmica das árvores que se ligue diretamente com informações economicas ?

Na seção 3.4.1 defini o espaço ultramétrico subdominante e na seção 3.5, ficou claro que a árvore construída de forma a minimizar a soma das arestas e conter todos os vértices, do grafo original, é um processo equivalente a estruturar uma hierarquia em um espaço ultramétrico subdominante. Esta hierarquia me dá a maior proximidade de ações à uma informação economica em comum, entretanto a hierarquia é construída de forma a priorizar a menor distância de cada vértice à árvore, perdendo a informação da localização dos vértices em relação aos ramos da árvore. A taxonomia é útil para localização dos principais eixos de correlação da bolsa, mas a ocupação da árvore deve ser acessada por meio da definição de um vértice central, vértice raiz v_r , com grande importância para a estrutura da árvore, a partir daí é possível localizar o nível mais povoado do grafo partindo de v_r . Apesar da medida utilizada ter significado de proximidade em movimentação instantânea da ação no mercado, utilizando a força da metodologia em escolher a estrutura mais significativa só pode ser acessada a posteriori através dos resultados.

Os resultados que obtive mostra que a metodologia utilizada é útil para captar os eixos mais importantes de cada setor, e a visualização por meio da árvore mostra que, mesmo algumas ramificações ruidosas ($d_{ij} \geq 1.35$), apresentam-se posicionadas em um setor simétrico a uma classe economica. O nível de ocupação médio forneceu uma medida numérica para ramificação em setores e subsetores, de modo que quanto maior é o nível de ocupação médio maior a ramificação em subsetores a partir do vértice raiz. Para o conjunto de ações estudadas o nível de ocupação médio cai de acordo com o aumento da escala, em um mesmo período.

Uma outra medida escolhida, utiliza nível de ocupação médio para quantificar a relação do nível de ocupação de duas árvores ligadas por uma aresta de tamanho pequeno, desta forma calcula-se uma quantidade chamada centro de massa de ocupação que indica a quantidade de passos aproximada para que se possa visitar qualquer setor, isso se partirmos de qualquer dos dois vértices anterior-

mente ligados pela aresta retirada. A medida do centro de massa cresceu ao longo dos anos, mas em 2007 mostrou comportamento de queda com a escala temporal, indicando um reagrupamento dos vértices ao redor do vértice raiz de cada árvore.

A medida de massa reduzida, ao contrário do centro de massa, mostrou-se estável ao longo das escalas, indicando que o rearranjo do centro de massa é simétrico em toda a árvore, significando que os níveis de ocupação médio, nas duas árvores, ficaram com similar distância um do outro.

O estudo mostra que é possível captar informação econômica através da metodologia utilizada, ainda que no mercado Bovespa&BMF existe uma hierarquia preferencial de proximidade com a informação econômica, e que essa hierarquia dá uma estrutura quase estacionária. Nosso estudo revelou que a estrutura se põe com nitidez variada ao longo das escalas, isso pode ser entendido como o chamado efeito epps agindo na estrutura do mercado.

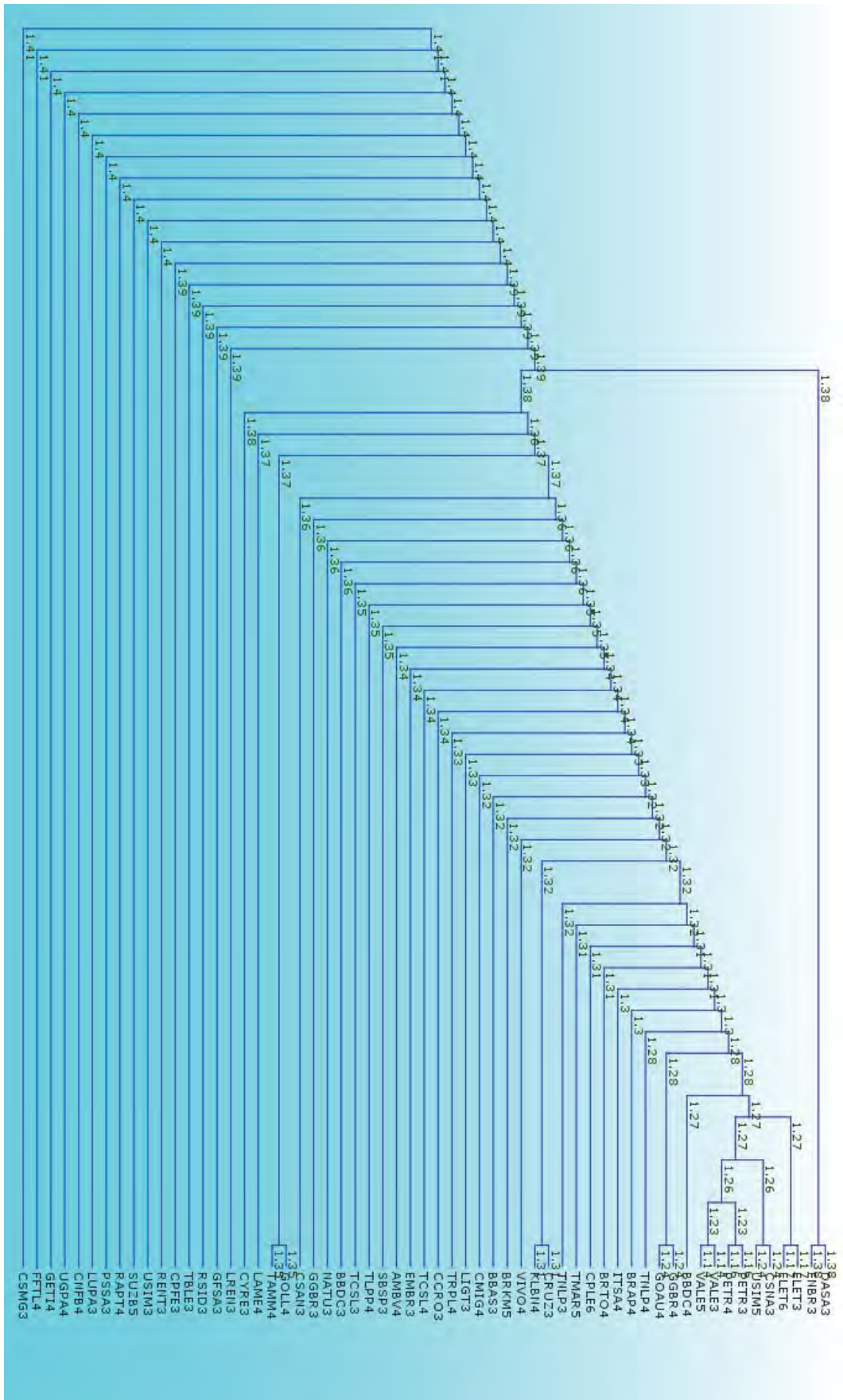


Figura 4.18: Árvore taxonomica para 2006 na escala de retornos 2 minutos.

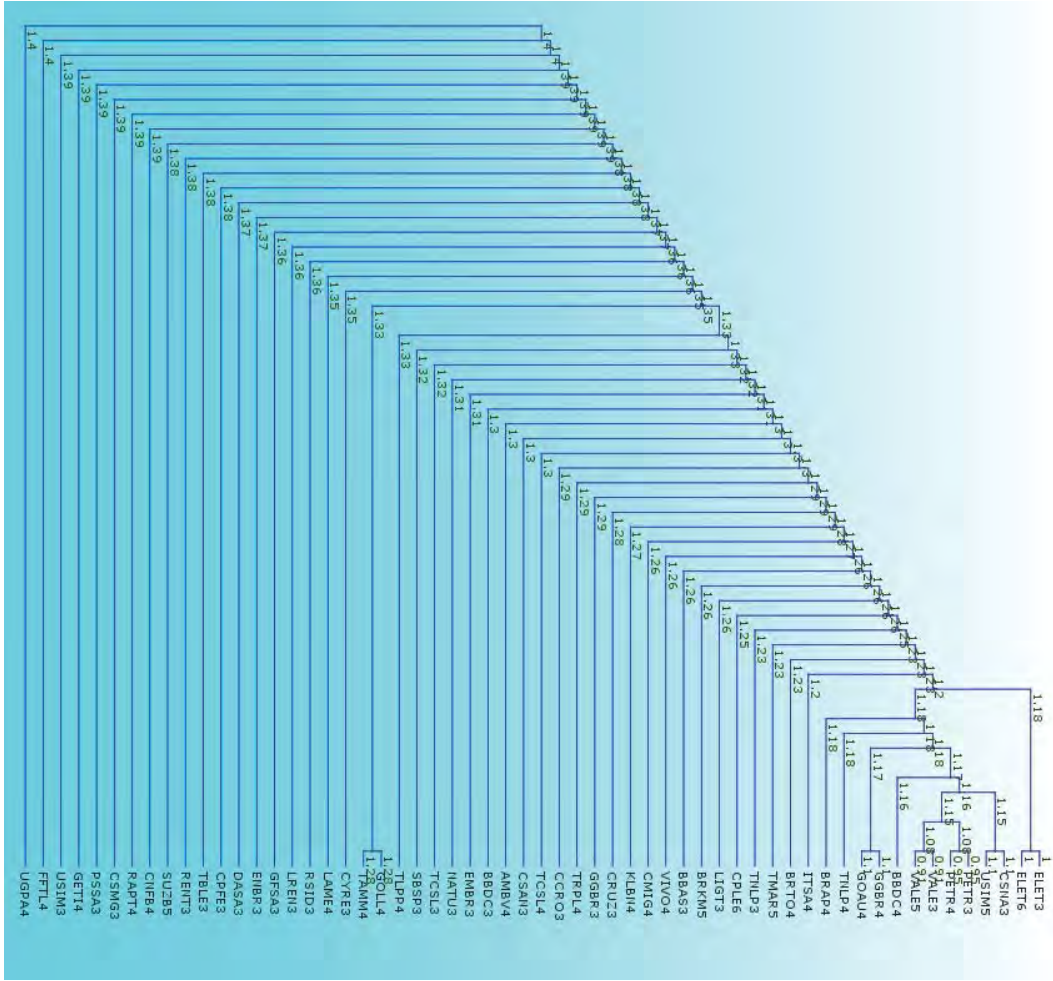


Figura 4.19: Árvore taxonômica para 2006 na escala de retornos de 6 minutos.

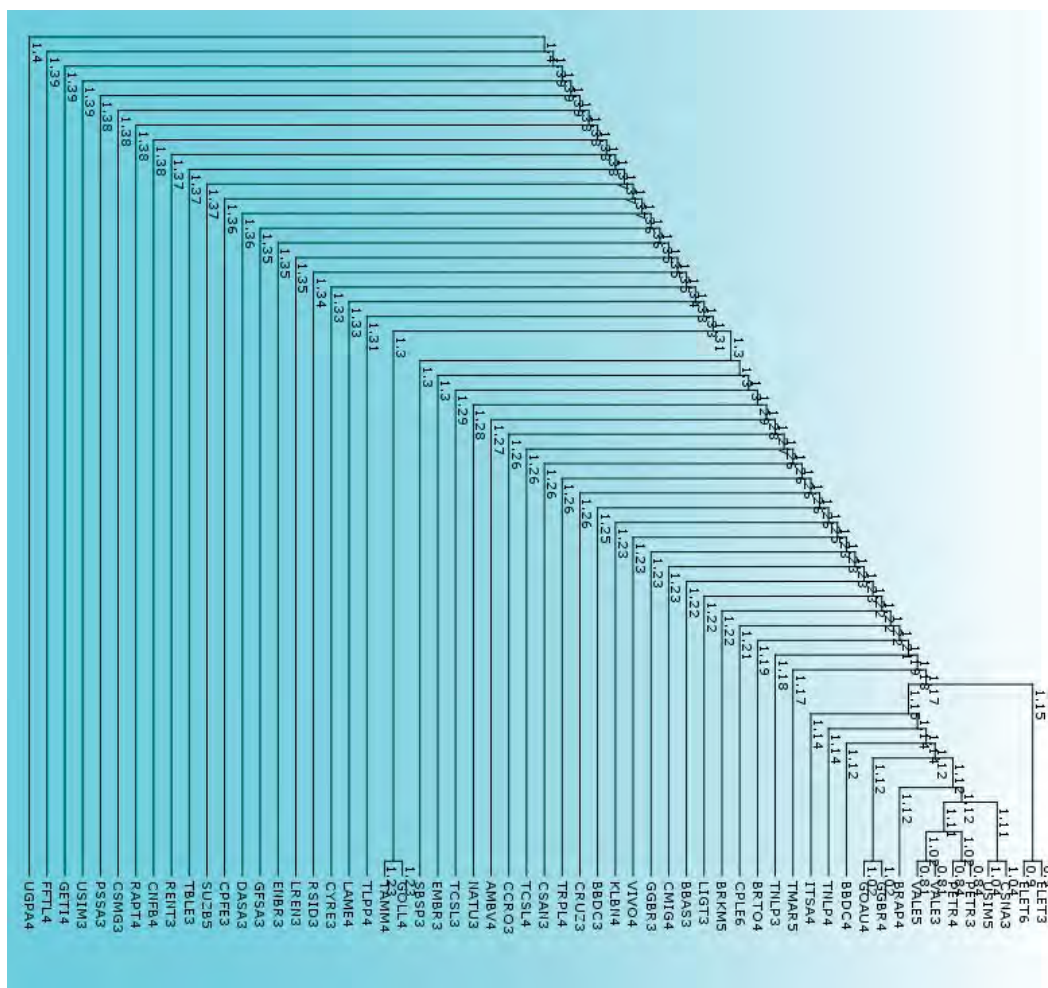


Figura 4.20: Árvore taxonômica para 2006 na escala de retornos de 10 minutos.

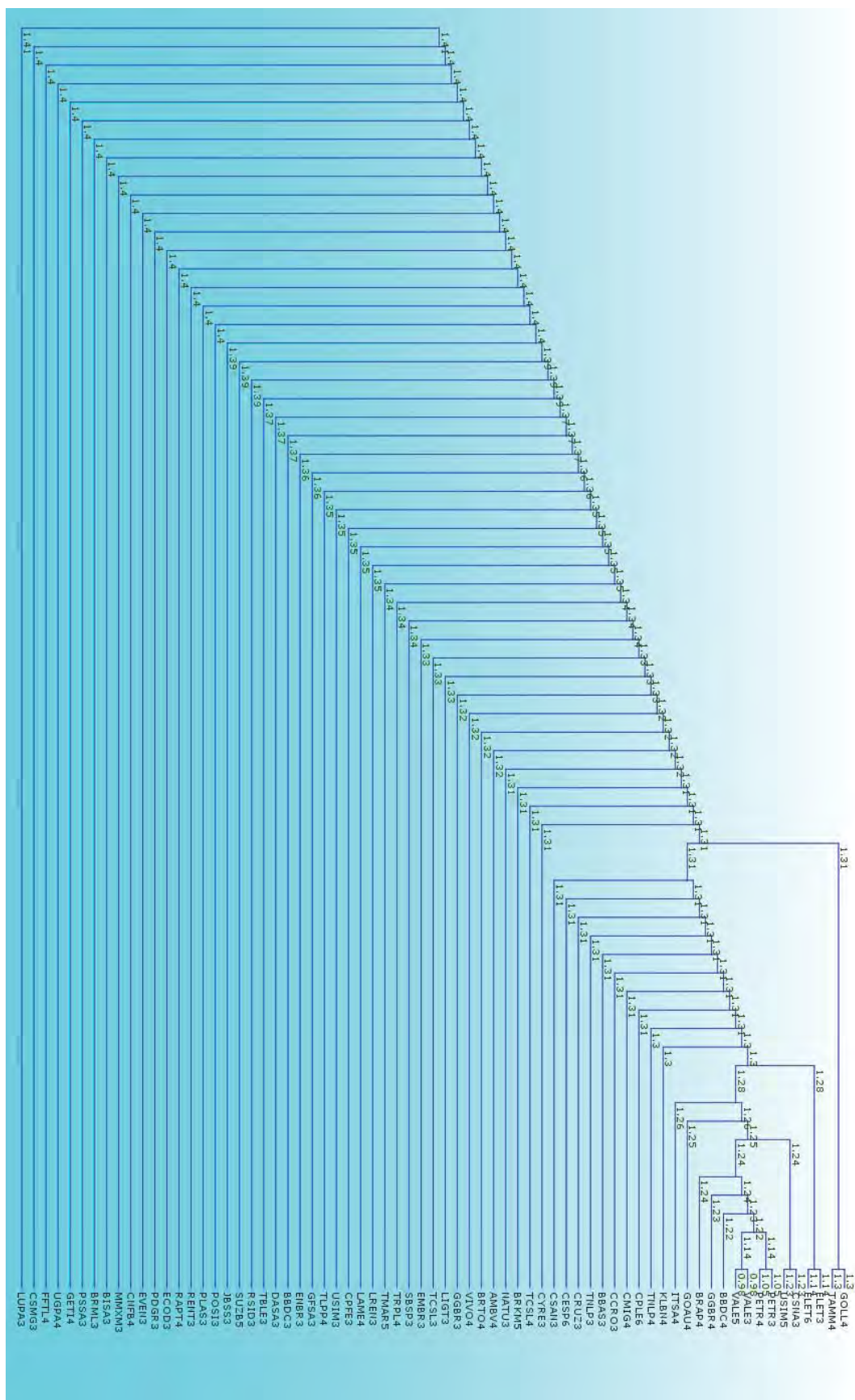


Figura 4.21: Árvore taxonômica para 2007 na escala de retornos 2 minutos.

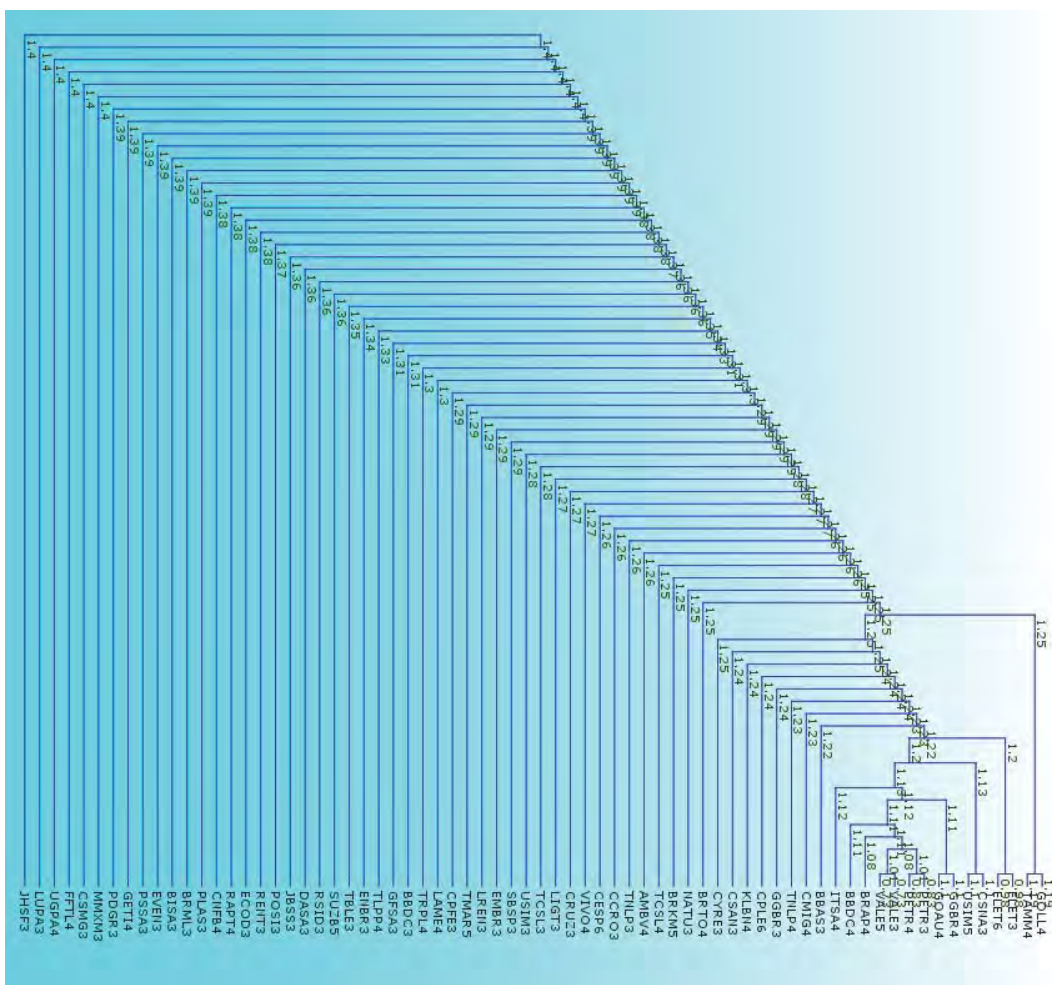


Figura 4.22: Árvore taxonomica para 2007 na escala de retornos de 6 minutos.

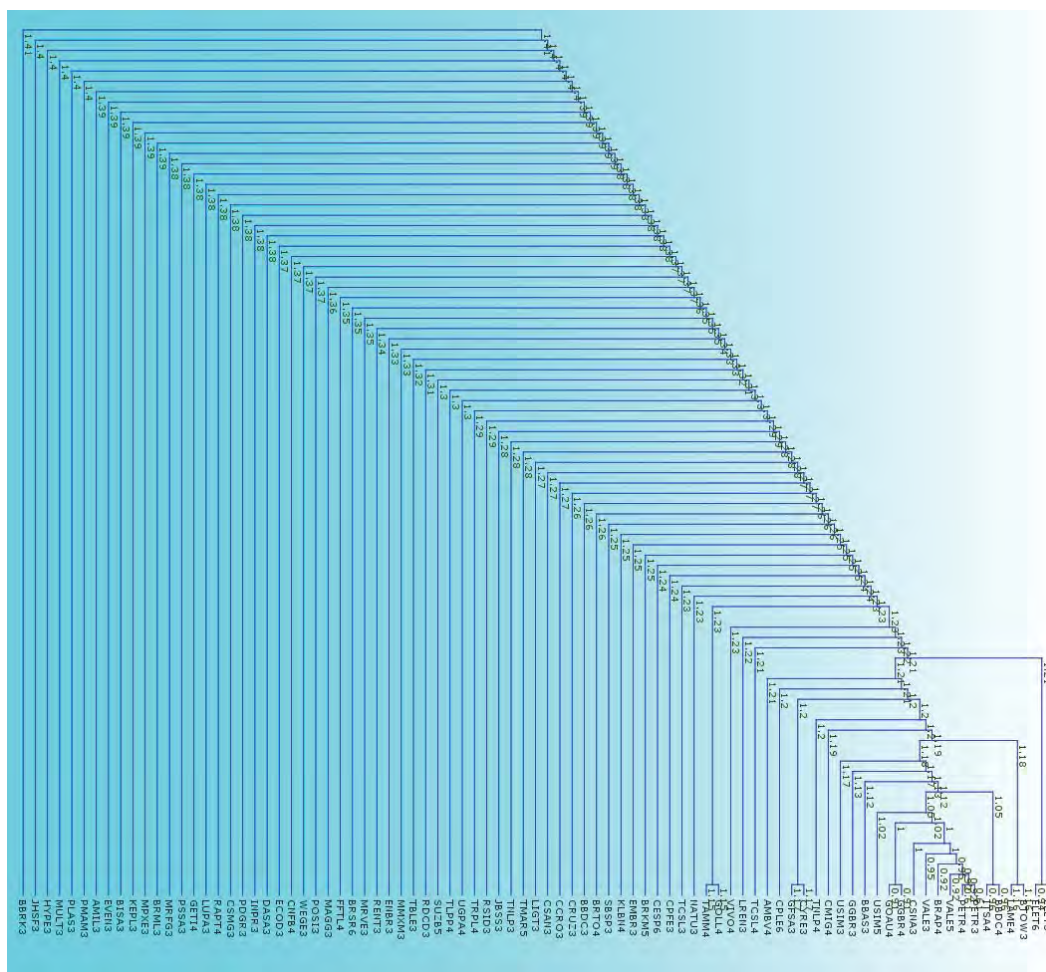


Figura 4.24: Árvore taxonomica para 2008 na escala de retornos de 6 minutos.

4.4 Conclusões e Perspectivas

A construção de um corpo teórico científico, em uma área cuja a tradição teórica é guiada pelo sentido mercadológico do sistema tratado, é um processo que passa pelo exame estatístico do fenômeno na tentativa de extrair padrões descritíveis através do ferramental de modelagem conhecido. O mercado financeiro é um sistema que se beneficia da multi-disciplinaridade no estudo de sistemas complexos e pode fazer uso de ferramentas já conhecidas, como as aqui utilizadas. Com isso mostramos para mais um caso específico, o mercado brasileiro Bovespa&BMF, pode ter informação economicamente significativa extraída, assim como sua estruturação em uma árvore complexa é uma característica típica de sistemas complexos estudados em várias áreas. Ainda é possível extrair fenômeno original e interpretar em termos microscópicos como o Efeito Epps através de um estudo estatístico ao longo do tempo.

A partir daqui as perspectivas que podem ser tomadas abrangem: continuidade do estudo estatístico relacionando as ações por meio de grafos direcionados (utilizando correlação parcial como em Y. Dror et al. [2]), modelagem multivariada que leve em conta o efeito epps e as estruturas de grafo possíveis para análise de proximidade e setores da bolsa.

Referências Bibliográficas

- [1] P. W. Anderson. More is different - broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, August 1972. [1](#)
- [2] Y. Dror, M. Tumminello, M. Asaf, G.-G. Gitit, R. N. Mantegna, and K. B.-J. Eshel. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS ONE*, 5(12):e15032, 12 2010. [62](#)
- [3] Thomas W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298. [21](#)
- [4] A. Kirman. *Complex Economics: Individual and Collective Rationality*. The Graz Schumpeter Lectures. Routledge, first edition, July 2010. [1](#)
- [5] J. Kwapien, S. Drozd, and J. Speth. Time scales involved in emergent market coherence. *Physica A: Statistical and Theoretical Physics*, 337(1-2):231–242, 2004. [22](#)
- [6] R. N. Mantegna and H. E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge, 2000. [5](#), [11](#)
- [7] D. Morrison. Carl woese and new perspectives on evolution - <http://astrobiology.nasa.gov/articles/carl-woese-and-new-perspectives-on-evolution/>. 2003. [13](#)
- [8] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E*, 68(5):056110, Nov 2003. [35](#), [52](#)
- [9] R. Rammal, G. Toulouse, and M. A. Virasoro. Ultrametricity for physicists. *Rev. Mod. Phys.*, 58(3):765–788, Jul 1986. [11](#)

- [10] Roberto Renò. A closer look at the epps-effect. *International Journal of Theoretical and Applied Finance*, 6:87–102, 2003. [21](#)
- [11] B Tóth and J Kerté. On the origin of the Epps effect. *Physica A: Statistical Mechanics and its Applications*, 383(1):54–58, SEP 1 2007. [22](#), [32](#)
- [12] B Tóth and J Kertész. Increasing market efficiency: Evolution of cross-correlations of stock returns. *Physica A: Statistical Mechanics and its Applications*, 360(2):505–515, FEB 1 2006. [22](#)
- [13] B Tóth and J Kertész. The Epps effect revisited. *Quantitative Finance*, 9(7):793–802, FEB 1 2009. [22](#), [34](#)
- [14] Y. Zhang, G. H. T. Lee, J. C. Wong, J. L. Kok, M. Prusty, and S. A. Cheong. Will the us economy recover in 2010? a minimal spanning tree study. *Physica A: Statistical Mechanics and its Applications*, 390(11):2020–2050, 2011. [52](#)