CENTRO UNIVERSITÁRIO DA FEI

Werner Fukuma

O POTENCIAL DE REDES COMPLEXAS PARA ANÁLISE DO MERCADO DE AÇÕES

São Bernardo do Campo
2013

Werner Fukuma

O POTENCIAL DE REDES COMPLEXAS PARA ANÁLISE DO MERCADO DE AÇÕES

Dissertação de Mestrado apresentada ao Centro Universitário da FEI para obtenção do título de Mestre em Engenharia Elétrica, orientada pelo(a) Prof. Dr. Paulo Sérgio Rodrigues.

São Bernardo do Campo
2013

Fukuma, Werner
O potencial de redes complexas para análise do mercado de ações / Werner Fukuma. São Bernardo do Campo, 2013.
130 f. : il.

Dissertação - Centro Universitário da FEI.
Orientador: Dr. Paulo Sérgio Rodrigues.

1. Redes Complexas. 2. Medidas Físicas. 3. Mercado de Ações. 4. Otimização de Carteira. 5. Teoria Clássica de Markowitz. I. Rodrigues, Paulo Sérgio, orient. II. Título.

CDU 531:336.76

Centro Universitário da **FEI**

APRESENTAÇÃO DE DISSERTAÇÃO
ATA DA BANCA JULGADORA

PGE-10

Programa de Mestrado de Engenharia Elétrica

Aluno: Werner Fukuma **Matrícula:** 110125-2

Título do Trabalho: O potencial de redes complexas para análise do mercado de ações.

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues

Data da realização da defesa: 20/02/2013

ORIGINAL ASSINADA

A Banca Julgadora abaixo-assinada atribuiu ao aluno o seguinte:

APROVADO ☒ REPROVADO ☐

São Bernardo do Campo, 20 de Fevereiro de 2013.

MEMBROS DA BANCA JULGADORA

Prof. Dr. Paulo Sérgio Silva Rodrigues

Ass.: _____

Prof. Dr. Rodrigo Filev Maia

Ass.: _____

Prof. Dr. Francisco de Assis Zampirolli

Ass.: _____

VERSÃO FINAL DA DISSERTAÇÃO

ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS RECOMENDAÇÕES DA BANCA EXAMINADORA

_____

Aprovação do Coordenador do Programa de Pós-graduação

_____

Prof. Dr. Carlos Eduardo Thomaz

RESUMO

Nos últimos anos as Redes Complexas têm se mostrado bastante eficiente modelando diversos problemas de entidades e suas relações de diversas áreas do mundo real: pessoas e seus relacionamentos de amizade, aviões e seus voos entre aeroportos, doença e sua propagação de contágio, entre outras. Por outro lado, uma área que é sempre muito discutida, é a área de economia, mais precisamente a Bolsa de Valores, e com a teoria clássica de Markovitz da gestão de risco de carteiras de investimento possibilitou a modelagem relacionando as ações e suas correlações. Com isso em mente, este trabalho propõem modelar, utilizando Redes Complexas, o mercado de ações do ponto de vista de otimização de carteiras sob a teoria clássica de Markovitz. Para tanto, uma matriz de correlações das variações dos preços de fechamento é computado através de pares de ações da Bovespa (Bolsa de Valores de São Paulo) formando uma rede mercadológica. Além disso, juntamente, será construído uma árvore de espalhamento mínimo com base na distância da matriz de correlações, esta árvore é um subdomínio da rede mercadológica com as correlações mais fortes entre as ações. A partir daí, serão investigados propriedades físicas utilizando algumas medidas como: Densidade, Distribuição de Graus, Matriz de Distâncias de Caminhos, Coeficiente de Clusterização e Modularidade. Com essas medidas, é possível verificar o modelo da rede estudada, analisar os agrupamentos da rede, retirar informações topológicas e semânticas da Bolsa de Ações. Além disso, é utilizado um método para estudar o comportamento das redes com o passar do tempo, e analisar suas mudanças. Os resultados sugerem que a árvore de espalhamento mínimo possui comportamento de rede livre de escalas.

Palavras-chave: Redes Complexas, Medidas Físicas, Mercado de Ações, Otimização de Carteira, Teoria Clássica de Markowitz.

ABSTRACT

Recent years, complex networks have proven quite efficient modeling many real world problems of entities and their relations in different areas: people and their relationships of friendship, airplanes and their flight between airports, disease and its spread of contagion, and so on. In the other hand, an area that is always discussed, it is the area of economy, specifically the Stock Exchange, and the classical Markovitz's theory of risk of investment portfolios allowed to model the relation between Stocks and their correlations among them. With this in mind, this paper proposes modeling, using complex networks, the stock market from the viewpoint of portfolio optimization under the classical Markovitz's theory. Therefore, a correlation matrix of changes in closing prices are computed through pairs of the Bovespa (Stock Exchange in São Paulo) forming a finance network. Furthermore, a Minimum Spanning Tree (MST) will be constructed based on the distance of the correlation matrix, this tree is a subdomain of finance network with the strongest correlations among Stocks. Thereafter, it will be investigated using physical measures such as: Density, Degrees Distribution, Distance Path Matrix, Clustering Coefficient and Modularity. With these measures, it is possible to check the model of the network, analyze groups of network topological information and remove semantic information from Stock Exchange. Furthermore, a method is used to study the behavior of the networks over time, and analyzing their changes. The results suggest that the MST behavior a scale free network.

Keywords: Complex Networks, Physical Measures, Stock Exchange, Portfolio Optimization, Classical Markowitz's Theory

LISTA DE TABELAS

4.1 Exemplo de escolha das ações para a rede do período de Dezembro de 2001 a Novembro de 2002. A escolha das ações para esta rede é a intersecção das ações dos Ibovespas dentro do período de análise da rede. 45

5.1 Matriz de correlações cruzada entre pares de ações de acordo com a Equação (4.2). 55

5.2 Tabela das redes específicas. 59

5.3 Número de caminhos minimizados (N^o Cam.) pela ação exibida, juntamente com a força do vértice s_i desta ação. Das redes da data de 20041001, 20090504 e 20110401. 72

5.4 Ligações da árvore *MST* de distâncias de correlações de ações da data de 20041001 a 20050930. 83

5.5 Ligações da árvore *MST* de distâncias de correlações de ações da data de 20090504 a 20100430. 85

5.6 Ligações da árvore *MST* de distâncias de correlações de ações da data de 20110401 a 20120330. 87

5.7 Matriz de correlações cruzada entre pares de ações de acordo com a Equação 4.2. . 94

5.8 Matriz de correlações cruzada entre pares de ações de acordo com a Equação 4.2. . 94

5.9 Matriz de correlações cruzada entre pares de ações de acordo com a Equação 4.2. . 94

5.10 Medidas físicas de Densidade ponderada e Força do vértice das ações TNLP4, VALE5, PETR4, USIM5 e BBDC4 da rede do período de 20041001 a 20050930. . 95

5.11 Medidas físicas de Densidade ponderada e Força do vértice das ações BBDC4, VALE5, PETR4, USIM5 e TLPP4 da rede do período de 20090504 a 20100430. . . 95

5.12 Medidas físicas de Densidade ponderada e Força do vértice das ações CSNA3, VALE5, PETR4, USIM5 e CCRO3 da rede do período de 20110401 a 20120330. . 96

5.13 Classificação setorial das ações. 98

LISTA DE ILUSTRAÇÕES

2.1 Cidade de Königsberg no século 18. 5

2.2 Cidade de Königsberg representado por um grafo. 6

2.3 Modelo de Watts-Strogatz para uma rede de Mundo Pequeno 10

2.4 Distribuição de Graus do Modelo de Watts-Strogatz. 11

2.5 Distribuição de Graus do Modelo de Livre Escala de Barabási e Albert. 12

2.6 Rede complexa com comunidades de vértices. 13

2.7 Distribuição de Graus da Rede de correlações de retornos de séries temporais filtrado com limiar $v_c = 0.95$ e com as coordenadas verticais em escala logarítmica, a linha vermelha é de ajuste de curva. 17

2.8 Coeficiente de clusterizaçao em função do vértice (eixo horizontal) e do coeficiente de correlação (eixo vertical). Direita: rede aleatória, esquerda: rede de ações. . . . 21

2.9 Árvore *MST* de distâncias de correlações de ações da *New York Stock Exchange* (*NYSE*). 24

4.1 Trecho da Base de Dados utilizada, onde cada campo armazena o código da ação negociado na BM&F Bovespa, a data do pregão em formato AAAAMMDD e o preço de fechamento desta ação nesta data. 43

4.2 Trecho do armazenamento da classificação setorial das ações, onde cada campo armazena o código da ação negociado na BM&F Bovespa e o setor de classificação da empresa. 43

5.1 Exemplo do método de janelamento, onde cada janela tem um período de um ano de negociações diárias das ações, com intervalo de um mês entre cada janela. . . . 53

5.2 Preços de fechamento em reais das ações AMBV4, CRUZ3, PETR3, PETR4 e VALE5 da data de 20110401 a 20120330. 54

5.3 Densidade ponderada média de cada rede complexa com o passar do tempo. 57

5.4 Densidade ponderada média de cada rede complexa pelo tempo. Os pontos em destaque representam as redes específicas escolhidas representando diversos instantes da densidade média ponderada. 58

5.5 Histograma de pesos das redes complexas específicas de 1 a 6. 60

5.6 Histograma de pesos das redes complexas específicas de 7 a 12. 61

5.7 Densidade média ponderada (sinal em vermelho contínuo seguindo a escala a esquerda) e Entropia de distribuição de pesos (sinal em azul tracejado seguindo a escala a direita) pelo tempo. 63

5.8 Distribuição de forças de vértices discretizado das redes específicas de 1 ao 6. . . . 65

5.9 Distribuição de forças de vértices discretizado das redes específicas de 7 ao 12. . . 66

5.10 Coeficiente de clusterização ponderado médio proposto por (LOPEZ-FERNANDEZ; ROBLES; BARAHONA, 2004) pelo Tempo. 67

5.11 Densidade ponderada média e coeficiente de clusterização ponderado médio proposto por Lopez-Fernandez, Robles e Barahona (2004) pelo Tempo. 69

5.12 Coeficiente de clusterização ponderado Onnela et al. (2005), Zhang e Horvath (2005) médio pelo Tempo. 71

5.13 Distribuição de graus de vértices das redes *MST* de 1 a 6. 76

5.14 Distribuição de graus das redes *MST* de 7 a 12. 77

5.15 Distribuição de forças de vértices das redes *MST* de 1 a 6. 78

5.16 Distribuição de forças de vértices das redes *MST* de 7 a 12. 79

5.17 Árvore *MST* de ações da data de 20041001 a 20050930. 80

5.18 Árvore *MST* de ações da data de 20090504 a 20100430. 81

5.19 Árvore *MST* de ações da data de 20110401 a 20120330. 82

5.20 Preço de fechamento diário da ações TNLP4, VALE5, PETR4, USIM5 e BBDC4 na data de 20041001 a 20050930. 91

5.21 Preço de fechamento diário da ações BBDC4, VALE5, PETR4, USIM5 e TLPP4 na data de 20090504 a 20100430. 92

5.22 Preço de fechamento diário da ações CSNA3, VALE5, PETR4, USIM5 e CCRO3 na data de 20110401 a 20120330. 93

5.23 Modularidade das 112 redes *MST* de correlações intrínsecas. 104

5.24 Árvore *MST* de correlações cruzadas intrínsecas de ações da data de 20061002 a 20070928. 105

SUMÁRIO

1 INTRODUÇÃO **1**

1.1 OBJETIVO 3

2 CONCEITOS FUNDAMENTAIS **4**

2.1 REDE COMPLEXAS 4

2.1.1 TIPOS DE GRAFOS 6

2.1.2 MODELOS DE REDES 8

2.1.3 MEDIDAS FÍSICAS DAS REDES COMPLEXAS 12

2.1.4 MODULARIDADE 21

2.1.5 ÁRVORE ESPALHADA MÍNIMA 23

2.1.6 ALGORITMO DE DIJKSTRA 23

2.1.7 COEFICIENTE DE CORRELAÇÃO 25

2.2 ÁREA ECONÔMICA 27

2.3 BM&F BOVESPA 27

2.3.1 ÍNDICE BOVESPA 28

2.4 RENTABILIDADE DAS AÇÕES 30

2.4.1 ANÁLISE FUNDAMENTALISTA E ANÁLISE GRÁFICA 31

2.4.2 RISCO 31

2.4.3 CARTEIRA 32

3 TRABALHOS RELACIONADOS **34**

4 PROPOSTA **39**

4.1 BASE DE DADOS 39

4.1.1 BASE DE DADOS DA BM&F BOVESPA 39

4.1.2 PREPARAÇÃO DA BASE DE DADOS 41

4.1.3 EXEMPLO DA BASE 42

4.2 CONSTRUÇÃO DA REDE 42

4.3 ÁRVORE DE ESPALHAMENTO MÍNIMO 46

4.4 EXTRAÇÃO DE CARACTERÍSTICAS FÍSICAS DA REDE 47

4.4.1 DENSIDADE 48

4.4.2 DISTRIBUIÇÃO DE GRAUS / FORÇA DE VÉRTICES 48

4.4.3 MATRIX DE DISTÂNCIAS DE CAMINHOS 49

4.4.4 COEFICIENTE DE CLUSTERIZAÇÃO 50

4.5 MEDIDA DE QUALIDADE DE CLUSTERIZAÇÃO DA REDE 51

5 RESULTADOS EXPERIMENTAIS **52**

5.1 CORRELAÇÃO CRUZADA 52

5.2 DENSIDADE MÉDIA PONDERADA 56

5.3 DISTRIBUIÇÃO DE PESOS 56

5.3.1 ENTROPIA DA DISTRIBUIÇÃO DE PESOS 62

5.4 DISTRIBUIÇÃO DE GRAUS E FORÇA DOS VÉRTICES 64

5.5 COEFICIENTE DE CLUSTERIZAÇÃO 64

5.6 MATRIZ DE DISTÂNCIAS DE CAMINHO 70

5.7 ÁRVORE ESPALHADA MÍNIMA - MST 74

5.7.1 DISTRIBUIÇÃO DE GRAUS E DE FORÇAS DE VÉRTICES 75

5.7.2 MODULARIDADE 97

6 CONCLUSÕES **107**

REFERÊNCIAS **110**

1 INTRODUÇÃO

Há muitos problemas que envolvem o estudo da relação entre entidades. A cadeia alimentar (ALLESINA; PASCUAL, 2008), por exemplo, é um problema que deve ser abordado considerando as relações entre presa e predador. Outro exemplo comum é uma rede social (GIRVAN; NEWMAN, 2002). Neste modelo, uma pessoa é uma entidade e a "amizade" é o relacionamento.

Por outro lado, uma área que tem ganhado a atenção dos pesquisadores são as chamadas Redes Complexas. A área de Redes Complexas pode ser definido como a junção da teoria dos grafos e da matemática estatística. Além disso, muitas propriedades extraídas desse tipo de rede podem ser usadas para estudar e entender comportamentos físicos de problemas do mundo real, tais como topologia livre de escalas, fenômeno de mundo pequeno, percolação, entre outros (NEWMAN; BARABASI; WATTS, 2006).

Um problema que pode ser modelado como uma rede complexa está relacionada com o mercado de ações. Nesse tipo de modelagem, as ações são modeladas como um nó na rede e as dependências entre elas são modeladas como as arestas. Essas dependências são normalmente calculadas pela correlação entre as ações. Na literatura, podemos encontrar várias obras (MANTEGNA, 1999; BONANNO; VANDEWALLE; MANTEGNA, 2000; BONANNO et al., 2003, 2004; ONNELA et al., 2002, 2003b, 2003a; JUNG et al., 2006, 2008) usando essa abordagem. A vantagem de modelar o mercado de ações com Redes Complexas é que as medidas físicas podem fornecer suporte para analisar e estudar o problema sob o relacionamento global das correlações entre as ações.

Na medida em que mudanças de preço estão acontecendo na Bolsa, parece seguro assumir que elas são refletidas em correlações a outros ativos ao mesmo tempo. As ações interagem umas com as outras, criando um sistema complexo em evolução (ARTHUR; DURLAUF; LANE, 1997; STANLEY et al., 1996).

A correlação cruzada, que é a correlação do comportamento da variação do preço entre pares de ações, é um ponto central na teoria de investimento e gestão de risco, e também serve como entrada para o problema de otimização de carteira na teoria clássica de Markowitz (MARKOWITZ, 1952).

A fim de reduzir o risco do mercado, a qual todas ações estão sujeitas a variações devido a influências externas, além de reduzir o risco específico de cada ação, a teoria de Markowitz afirma que uma boa carteira de investimentos é composta de ações que têm comportamentos opostos. Isso significa que se uma das ações cai, deve haver uma outra ação que sobe em sua carteira. Com isso em mente, um portifólio eficiente é uma carteira de investimento composta por ações que maximizam os ganhos, dado uma faixa de risco (MARKOWITZ, 1952).

Então, o estudo e a classificação do tipo de rede poderá ajudar a verificar se as ações seguem uma hierarquia correlacional, onde podem existir ações mais correlacionadas e outras menos correlacionadas ou anti-correlacionadas, as ações mais anti-correlacionadas são ótimas para formar portifólios eficientes, enquanto as ações mais correlacionadas da rede podem ajudar a verificar a tendência do mercado.

De acordo com Tabak, Serra e Cajueiro (2010), o mercado brasileiro de ações é um caso interessante de estudo por várias razões. Primeiro, a BM&F Bovespa (integração das operações da Bolsa de Valores de São Paulo e da Bolsa de Mercadorias & Futuros) é um dos mais importantes da América Latina. Em segundo lugar, o Brasil faz parte do BRICS (Brasil, Rússia, Índia, China e África do Sul), grupo de países em grande crescimento e expansão. Terceiro, tem sido alvo de investimentos de empresas internacionais. Finalmente, há um número significativo de empresas brasileiras envolvidas em fusões e aquisições em todo o mundo, aproximando-se com os mercados de ações internacionais.

Neste trabalho, serão analisadas as redes geradas a partir do mercado brasileiro de ações. Nessas redes, os nós representam as ações e as arestas representam a correlação entre elas. Serão estudadas também diversas medidas físicas das redes financeiras, tais como: Densidade, Distribuição de Graus, Matrix de Distâncias de Caminhos, Coeficiente de Clusterização e Modularidade. Essas

medidas ajudam a indicar o modelo de rede, os agrupamentos das redes e informações topológicas e semânticas. Além disso, serão estudadas essas medidas físicas do ponto de vista temporal, analisando suas mudanças em diferentes períodos de tempo.

Ainda mais, o subdomínio de mínimas distâncias de correlações dos ativos também será analisado, a árvore de espalhamento mínimo. Com isso, terão as ligações mais fortes das redes que darão auxílio para analisar as ações controladoras de mercados e a hierarquia das ações.

1.1 OBJETIVO

Mostrar como a teoria das Redes Complexas pode ser usada para estudar um mercado de ações, através da análise de seus parâmetros físicos.

2 CONCEITOS FUNDAMENTAIS

2.1 REDE COMPLEXAS

De uma forma simples, podemos dizer que uma rede é um grupo de entidades, pontos, nós ou vértices e um grupo de conexões, ligações ou arestas que conectam as entidades par a par por algum tipo de relacionamento, sejam elas laços de amizades ligando pessoas (RAPOPORT; HORVATH, 1961; BAUM; ROWLEY; SHIPILOV, 2004), linhas de comunicação ligando computadores (FALOUTSOS; FALOUTSOS; FALOUTSOS, 1999; GOVINDAN; TANGMUNARUNKIT, 2000), produtos químicos e suas reações (JEONG et al., 2000; WAGNER; FELL, 2000; FARKAS et al., 2002; SHEN-ORR et al., 2002), artigos científicos e citações (PRICE, 1965; REDNER, 1998) e ações e suas correlações (MANTEGNA, 1999; BONANNO; VANDEWALLE; MANTEGNA, 2000; ONNELA et al., 2003b; TABAK; SERRA; CAJUEIRO, 2010).

Com isso em mente, é possível dizer que as redes podem ser encontradas em todos os lugares, desde a internet até a economia, passando por mais diversas áreas, como rede de terrorismo e rede de transmissão de doenças (BARABáSI, 2002; BUCHANAN, 2002; WATTS, 2003).

O estudo das redes começou com a teoria dos grafos em 1736, com o matemático Leonard Euler, que solucionou o problema das pontes de Königsberg. A cidade de Königsberg foi construída às margens do Rio Pregel, incluindo duas ilhas. Sete pontes ligam as terras desta cidade, como ilusta a Figura 2.1, e o problema era saber se existia um caminho que atravessasse todas as sete pontes apenas uma vez visitando todas as terras. Euler descobriu que não era possível, pois de acordo com a Figura 2.2, todas as terras (vértices) têm três ou cinco pontes ligando elas, ou seja, grau ímpar, mostrando que o problema da ponte não teria solução. Este problema recebeu o nome de Caminho Euleriano e é precisamente um caminho que passa em cada aresta exatamente uma vez (WEST, 1996).

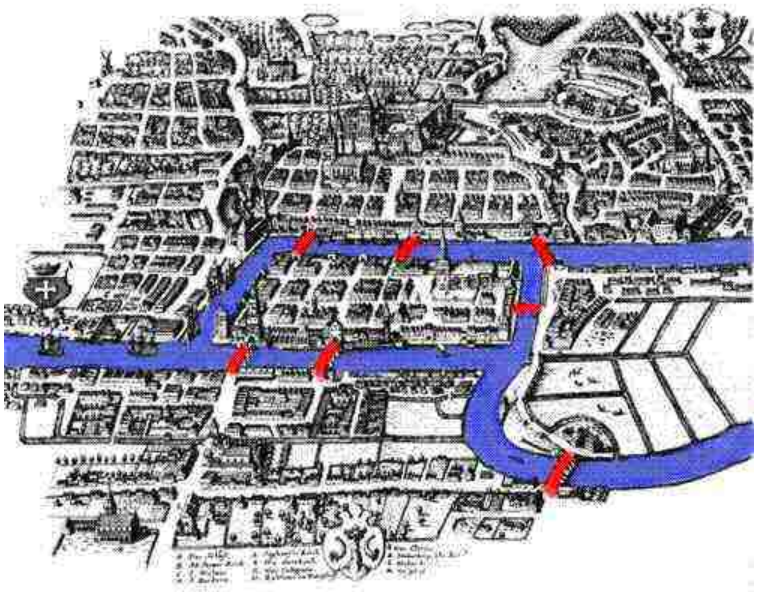

Figura 2.1: Cidade de Königsberg no século 18.
Fonte: (NEWMAN; BARABASI; WATTS, 2006).

No inicio dos anos 50, com o crescimento do estudo da sociologia e da antropologia, a Teoria dos Grafos veio ajudar a compreender os dados de estudos etnográficos, a analisar o comportamento da rede em se tratando de pessoas, centralidade de um indivíduo, comprimento do caminho entre pessoas, grupos de pessoas, conexões entre elas, e assim por diante (WASSERMAN; FAUST, 1994; DEGENNE; FORSé, 1999; SCOTT, 2000). Com isso, matemáticos começaram a analisar o relacionamento de vários modos de influência, como informação e doenças contagiosas (SOLOMONOFF; RAPOPORT, 1951; ERDOS; RENYI, 1960).

A partir daí, pode-se notar que os grafos são considerados objetos estocásticos, ao invés de puramente determinísticos (RAPOPORT, 1963). Com isso, pôde-se analisar o grafo como uma distribuição de probabilidade, uma abordagem muito explorada nos últimos anos.

Com isso, ao abstrair os detalhes de um problema, criando regras de ligações específicas definidas pelo domínio proposto, a teoria dos grafos é capaz de descrever importantes características topológicas de uma rede (NEWMAN; BARABASI; WATTS, 2006).

Nas últimas décadas, a teoria dos grafos têm se expandido para muitos outros domínios, como em aplicações na engenharia (AHUJA; MAGNANTI; ORLIN, 1993), na pesquisa operacional (NAGURNEY, 1993), e em ciências da computação (LYNCH, 1996).

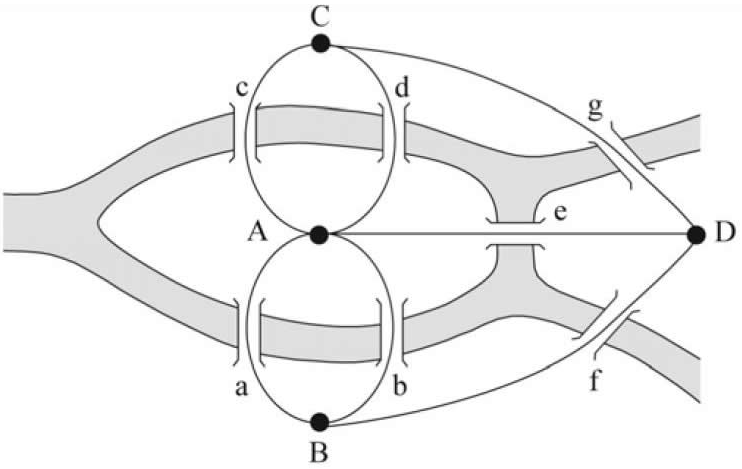

Figura 2.2: Cidade de Königsberg representado por um grafo.
Nota: Os pontos *A* a *D* representam as terras da cidade de Königsberg. De *a* a *g* temos as pontes, as ligações mostram os caminhos possíveis entre as terras da cidade. Fonte: Adaptado de (NEWMAN; BARABASI; WATTS, 2006).

Além disso, nos últimos anos, a ciência das redes tem focado em estudos sobre propriedades das redes naturais ou do mundo real, sobre dinâmismo das redes, e sobre a estrutura a qual sistemas dinâmicos distribuídos são construídos (NEWMAN; BARABASI; WATTS, 2006; WATTS, 1999a).

Ao assumir que a rede é um sistema dinâmico, entende-se que, por exemplo, em uma rede de indivíduos que interagem entre si, as ligações afetam a opinião de outros em decisões coletivas. Em Ball, Mollison e Scalia-Tomba (1997), Kuperman e Abramson (2001) usualmente assumem-se que todos os membros de uma população possuem a mesma probabilidade de interagir com todos os outros. Claramente essa premissa requer modificações, se estivermos considerando a propriedade dos sistemas dinâmicos (MONASSON, 1999; PANDIT; AMRITKAR, 2000).

Assim, é possível dizer que Redes Complexas é a junção da teoria dos grafos com a matemática estatística, tendo em mente que as ligações individuais são de menor importância, mas com ênfase no comportamento físico e topológico da rede e suas propriedades, assim obtendo informações relevantes sobre o domínio estudado.

Nas próximas seções serão mostrados sobre os tipos, os modelos e as medidas físicas de Redes Complexas.

2.1.1 TIPOS DE GRAFOS

Os principais tipos de redes são dígrafos (*directed graphs* ou redes diregidas), onde as arestas vão de um certo vértice para outro, e grafos, que são redes não dirigidas, ou seja, a presença da aresta (i, j), significa que a conexão $(i, j) = (j, i)$ (BANG-JENSEN; GUTIN, 2004; BONDY, 1976).

Ambos, tanto os dígrafos quanto os grafos, podem ser ponderados ou não ponderados. O primeiro, as arestas levam pesos, normalmente entre 0.0 e 1.0, porém dependendo do problema este peso pode variar, podendo ser de -1.0 a 1.0, entre outros. Já os não ponderados, também chamados de binários, as arestas não levam pesos, ou seja, a conexão existe ou não.

Sabendo disso, pode-se dizer que o número total de conexões em um dígrafo pode ser dada por:

$$N(N-1), \tag{2.1}$$

sendo N o número de vértices da rede.

Já em um grafo, este número cai pela metade, pois a aresta de i para j é a mesma de j para i. Então, o número total de arestas para redes não-direcionadas é dada por:

$$\frac{N(N-1)}{2}. \tag{2.2}$$

Uma representação completa de dígrafos ponderados pode ser obtida a partir de uma matriz de pesos W, onde cada campo i, j com certo valor, representa o peso $w_{i,j}$ da aresta entre o vértice i indo para o vértice j.

É possível transformar um dígrafo ponderado em suas contrapartes não-ponderada, para isso pode ser aplicado uma operação de limiarização (*threshoulding*) da Matriz W representado por:

$$A = \delta_T(W). \tag{2.3}$$

Na Equação (2.3) dado um limiar T, se $w_{ij} > T$ então $a_{ij} = 1$, se não, $a_{ij} = 0$.

A operação de transformar um dígrafo em um grafo também é possível, é a operação de

simetria (*symmetry*), dado por:

$$\sigma(W) = W + W^T, \tag{2.4}$$

onde W^T é a transposta de W.

Com isso, pode-se dizer que o tipo de rede dígrafo ponderado pode gerar os outros tipos de redes, a partir dessas operações de simetria e limiarização.

Na Seção 2.1.2 serão apresentados os três modelos de Redes Complexas mais estudados na literatura.

2.1.2 MODELOS DE REDES

Nesta seção é abordado três dos mais conhecidos modelos de redes. Na Subseção 2.1.2.1 apresenta-se as Redes Aleatórias, o mais velho modelo entre elas. Em seguida, na Subseção 2.1.2.2, descreve-se o modelo de Rede de Mundo Pequeno, comportamento encontrado em redes sociais. Por fim, na Subseção 2.1.2.3, as Redes Livre de Escalas, que apresentam um grau de distribuição desigual da rede assim como a *World Wide Web*, são introduzidas.

2.1.2.1 Redes Aleatórias

O modelo Redes Aleatórias ou *Random Graphs* foi introduzido por Solomonoff e Rapoport (1951) e posteriormente estudado por Erdos e Renyi (1960), sendo um dos mais simples e mais antigos modelos de redes encontrados na literatua (BALLOBáS, 2001).

Existem duas representações matemáticas que descrevem as redes aleatórias. A primeira delas, denotada por $G_{n,m}$, representa um conjunto de todos os grafos que consiste em n vértices e m arestas. O grafo é amostrado por arestas uniformemente ao acaso, em pares de vértices escolhidos aleatoriamente a partir de vértices desconexos.

A outra notação, $G_{n,p}$, é o conjunto de todos os grafos que consistem em n vértices, onde cada par de vértices está ligado à probabilidade independente p. Para gerar um grafo do conjunto $G_{n,p}$, toma-se inicialmente n vértices desconexos, a cada par de vértices faz-se uma ligação com probabilidade p (ERDOS; RENYI, 1960).

As redes aleatórias servem como medida de comparação entre os outros modelos de redes. De acordo com Walsh (1999), a Equação (2.5) retorna uma medida de distância, o quão distante sua rede está comparado a uma rede aleatória, utilizando parâmetros de coeficiente de clusterização (ver Seção 2.1.3.6) e diâmetro (ver Seção 2.1.3.5) da rede. Valores próximos a 1, significa que sua rede esta próxima de ser aleatória, valores muito maiores que 1, são indícios que sua rede pode ser de mundo pequeno.

$$\mu = \frac{CC/CC_{rg}}{l/l_{rg}}, \tag{2.5}$$

onde CC e l são o coeficiente de clusterização médio e diâmetro entre todos os nós da rede respectivamente; e CC_{rg} e l_{rg} são coeficiente de clusterização médio e diâmetro estimados para uma rede aleatória (*random graph*) com o mesmo número de nós e arestas da rede estudada, respectivamente.

Além de servir de base na medida de comparação de modelos de redes, as redes aleatórias são úteis para estudar os comportamentos, assim como as características de redes do mundo real, comparando as características físicas de uma rede natural com os valores esperados para uma rede aleatória similar.

Para tanto, existem algumas formas de gerar uma rede aleatória, uma delas é seguindo a Distribuição de Graus, esta importante característica diferenciam as redes aleatórias e as redes modeladas do mundo real (NEWMAN; STROGATZ; WATTS, 2001; NEWMAN, 2005b; CHUNG et al., 2002; AIELLO; CHUNG; LU, 2002). Bender e Canfield (1978) propuseram um modelo para gerar redes aleatórias com Distribuição de Graus pré-definido. Já Molloy et al. (1995), Molloy e Reed (2000) fizeram um novo modelo no qual produziam multigrafos, ou seja, vértices que podem levar conexões de si mesmas, e um mesmo par de vértices pode ter mais de uma conexão, de acordo com a Distribuição de Graus. Nestes métodos, a partir de vértices desconectados, as arestas são conectadas de acordo com a Distribuição de Graus, onde cada vértices é ligado a um número de arestas para formar o grau desejado, até que todas as arestas sejam conectadas.

Outra possibilidade é o método de religação (*rewiring*), a partir de uma rede já montada (geralmente a rede estudada do mundo real) e já com a Distribuição de Graus desejada, iterativamente na rede troca-se as ligações de duas arestas entre elas (MILO et al., 2004). Muitos trabalhos, como

Cohen e Havlin (2003), Chung et al. (2002), Aiello, Chung e Lu (2002), utilizam esta abordagem, devido a sua importância e afabilidade ao tratamento analítico (COSTA et al., 2006).

2.1.2.2 Redes De Mundo Pequeno

As redes de mundo pequeno (*small world*), possuem uma propriedade encontrada em muitas redes do mundo real, chamada propriedade de mundo pequeno (*small world property*). Essa propriedade significa que a maioria dos vértices podem ser alcançados pelos demais vértices através de poucas arestas, ou seja, diâmetro l pequeno (ver Seção 2.1.3.5). Esta característica pode ser encontrada, por exemplo, em redes sociais, onde todas as pessoas do mundo podem ser alcançadas através de poucos níveis de amizade (WATTS, 1999a, 2003).

Em 1967, Milgram realizou um experimento no qual observou que dois cidadãos americanos, escolhidos por acaso, eram conectados por no máximo seis relações de amizade (MILGRAM, 1967). Este experimento parece ter sido a primeira observação de mundo pequeno estudada sobre um problema do mundo real, onde foi encontrado um grau de separação da rede igual a 6, porém atualmente, sabe-se que este número pode ser menor (NEWMAN; BARABASI; WATTS, 2006).

Outra propriedade das redes de mundo pequeno e de muitas redes do mundo real é a presença de muitos *loops* de três vértices (triângulos); isto é, se o vértice i esta conectado aos vértices j e k, então há uma alta probabilidade de j e k estarem conectados entre si; ou seja, coeficiente de clusterização médio alto (ver Seção 2.1.3.6) (COSTA et al., 2006).

Para construir uma rede de mundo pequeno seguindo o modelo de Watts-Strogatz (WATTS; STROGATZ, 1998), conecta-se cada vértice v_i aos k vizinhos mais próximos, totalizando $2k$ conexões, onde $N >> k >> log(N) >> 1$, sendo N o número de vértices. Em seguida, cada aresta é aleatoriamente reconectada com probabilidade p.

Watts e Strogatz (1998) mostraram que seu modelo pode apresentar um diâmetro pequeno e um grande número de *loops* na rede, dependendo da probabilidade p, como mostra a Figura 2.3. Assim como em redes aleatórias, a distribuição de graus para redes de mundo pequeno segue uma distribuição de Poisson, com um pico em $\langle k \rangle = 2k$ como pode-se ver na Figura 2.4. Essa figura

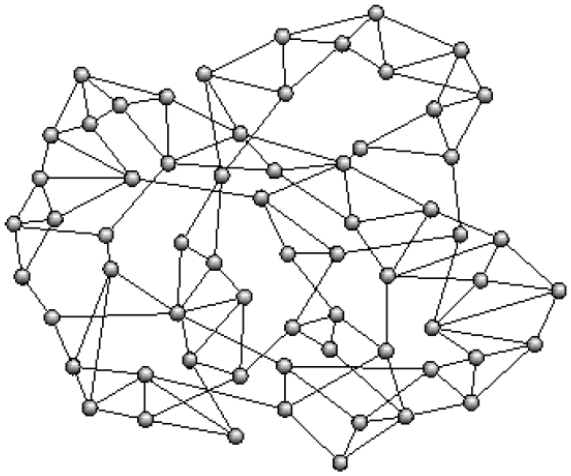

Figura 2.3: Modelo de Watts-Strogatz para uma rede de Mundo Pequeno
Nota: Um exemplo de rede com $N = 64$ vértices, $k = 2$, $p = 0.1$. Fonte: (WATTS; STROGATZ, 1998).

mostra uma distribuição de graus do modelo de Watts-Strogatz (WATTS; STROGATZ, 1998), onde o eixo y mostra a quantidade de nós que têm certo grau na rede, representado pelo eixo x.

2.1.2.3 Redes Livres De Escalas

A principal característica de uma rede livre de escalas é a distribuição de graus dos vértices da rede (ver Seção 2.1.3.3), que segue uma distribuição de lei de potência (*power law*) para os graus k altos, de acordo com a Figura 2.5.

Com isso, pode-se notar que alguns vértices possuem uma significante fração do total de arestas; sendo estes vértices chamados de *hubs*, sendo esta outra característica de redes livres de escalas.

De acordo com Barabasi e Albert (1999), um modelo de rede livre de escala baseia-se em duas regras: crescimento e ligação preferencial. Para gerar este modelo, começa-se com um grupo de vértices m_0. Então, iterativamente, adiciona-se novos vértices; para cada vértice adicionado, são feitas m novas ligações, seguindo a ligação preferencial, de acordo com a Equação (2.6). A probabilidade do vértice i ser conectado a um vértice j é proporcional ao grau k_j do vértice j.

$$P(i \to j) = \frac{k_j}{\Sigma_u k_u}. \tag{2.6}$$

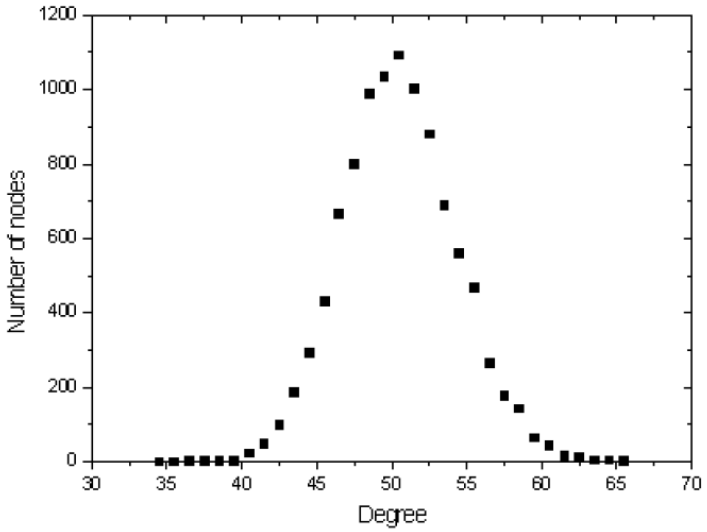

Figura 2.4: Distribuição de Graus do Modelo de Watts-Strogatz.
Nota: Distribuição de Graus médio de 10 redes de Watts-Strogatz com 10000 vértices, $k = 25$ e $p = 0.3$. Fonte: (WATTS; STROGATZ, 1998).

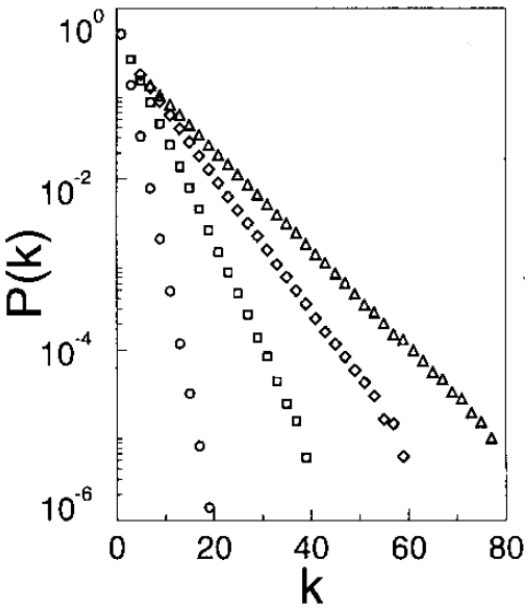

Figura 2.5: Distribuição de Graus do Modelo de Livre Escala de Barabási e Albert.
Nota: Distribuição dos graus de redes de Barabási-Albert com 150000 vértices, usando $m = 1(\bigcirc)$, $m = 3(\square)$, $m = 5(\Diamond)$ e $m = 7(\triangle)$, m representa o número de novas conexões para cada vértice da rede. Podemos notar que as conectividades seguem uma lei de potência. Fonte: (BARABASI; ALBERT, 1999).

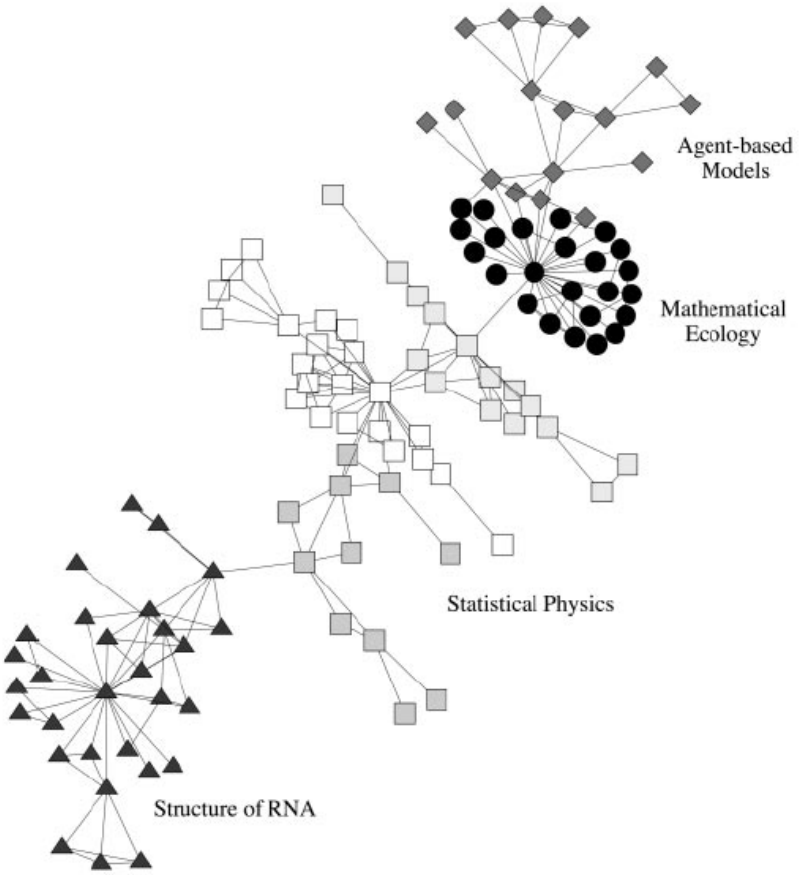

Figura 2.6: Rede complexa com comunidades de vértices.
Nota: Rede de colaborações do Instituto Santa Fé, com as divisões de comunidades indicados por diferentes formas de vértices. Fonte: (GIRVAN; NEWMAN, 2002).

Então, os vértices mais conectados possuem maior probabilidade de receber novas arestas. Este é um paradigma conhecido como "*The rich get richer*".

Além disso, outra característica encontrada em muitas redes naturais ou do mundo real é a presença de comunidades de vértices. Essa característica diz que muitas ligações são encontradas dentro de uma comunidade de vértices, enquanto entre comunidades, as arestas são menos frequêntes (GIRVAN; NEWMAN, 2002). A Figura 2.6 mostra um exemplo de rede com comunidade de vértices específica. Cada comunidade está representada por diferentes formas de vértices.

Note que, na Figura 2.6, as arestas entre os vértices de mesma forma são mais frequêntes, e entre diferentes formas as arestas são mais raras de encontrar.

2.1.3 MEDIDAS FÍSICAS DAS REDES COMPLEXAS

Medidas físicas ajudam a definir o modelo de rede estudada (ver Seção 2.1.2). De acordo com o modelo de rede, as medidas físicas revelam suas propriedades (NEWMAN; BARABASI; WATTS, 2006). Além disso, as medidas físicas ajudam a entender o comportamento e topologia de rede. Existem diversas medidas, as estudadas neste trabalho serão apresentadas a seguir.

2.1.3.1 Densidade Média Ponderada

Densidade é um parâmetro de rede que significa o quão conectado a rede está entre os nós. Matematicamente, a Densidade média ponderada é a soma de todos os pesos das arestas da rede dividido pelo número de possibilidades de conexões entre os nós da rede (KNOKE; KUKLINSKI, 1982). Varia entre 0.0 a 1.0, sendo que próximo a 0 significa que a rede está pouco conectada (esparsa), e próximo a 1 significa que a rede está muito conectada (densa).

A Equação (2.7) apresenta a Densidade média ponderada para uma rede não dirigida (SCOTT, 2000):

$$K_{den} = \frac{\sum_{i,j} w_{i,j}}{N(N-1)/2}, \tag{2.7}$$

onde $w_{i,j}$ é o peso da aresta do nó i para o nó j, somando-se para todos os pares de nós da rede, e N é o número de nós da rede.

Onnela et al. (2003b) utilizou a Densidade média ponderada em seu trabalho de Redes Complexas de 477 ações negociadas na Bolsa de Nova Iorque (*NYSE*) e suas correlações, no período de 02 de Janeiro de 1980 a 31 de Dezembro de 1999. Onnela observou o período de *Black Monday* da Bolsa americana, período em que houve uma crise mundial do mercado de ações, possivelmente pela venda de programas de negociação de ações (FURBUSH, 1993). Neste período de crise econômica mundial, Onnela notou que a correlação (Densidade) média entre as ações sobe consideravelmente, pois todas as ações tendem a cair juntamente em seu preço. Assim como Drozdz et al. (2000) observaram para a Bolsa alemã, em um rede de autovalores da matriz de correlações de ações negociadas na DAX (*Deutsche Aktienindex*).

2.1.3.2 Distribuição De Pesos

A distribuição de pesos mede a frequência com que os pesos das arestas aparecem na rede (BARABASI; BONABEAU, 2003). Para calcular a distribuição de pesos, é computado o histograma discretizado dos pesos da rede. Normalmente, a distribuição de pesos é utilizada em conjunto com outras características físicas para estudar a organização estrutural da rede.

Por exemplo, no trabalho de Barrat et al. (2004) são estudados a relação entre os pesos e a topologia, onde foram utilizados o coeficiente de clusterização ponderado e a distribuição de graus para caracterizar a estrutura de duas redes complexas, uma gerada a partir de voos de transporte aéreos e outra a partir de dados de colaboração científica.

Uma maneira de analisar a estrutura de uma rede associada a sua topologia é calculando a entropia da distribuição de pesos. De acordo com Rodrigues e Giraldi (to appear, 2010), Rodrigues, Giraldi e Araujo (2005), a entropia mede a quantidade de informação em um sistema. Uma baixa entropia significa concentração de informação, ou seja, muitas arestas com pesos semelhantes, levando possivelmente a uma topologia bem definida. Se, por outro lado, a entropia for alta, tem-se uma rede com pouca informação, e um gráfico de distribuição de pesos mais disperso, ou seja, maior variação dos pesos das arestas. Neste caso, não é possível inferir a topologia da rede sem a análise de outras propriedades.

No trabalho de Lopes (2011), foi utilizado a distribuição de pesos para analisar as conexões de Redes de palavras e suas co-ocorrências em uma base médica e outra base da Folha de São Paulo. O autor verifica que ambas as redes são esparsas, pois o número de conexões nulas é cerca de 99% das arestas. Além disso, Lopes verifica que uma base específica, como a base médica, têm um pico em 0.7, pois possuem mais co-ocorrências de palavras. Para uma base genérica, como da Folha de São Paulo, a distribuição de pesos é exponencial crescente, pois há mais ocorrências de novas palavras do que frequência de co-ocorrência.

2.1.3.3 Distribuição De Graus

A distribuição de graus da rede é uma distribuição de probabilidades de ocorrência dos graus da rede. Se a distribuição de graus seguir uma lei de potência (*power-law*), isso é um indício de que essa rede pode ser livre de escalas (BARABASI; BONABEAU, 2003).

O grau k_i de um nó i é o número de arestas diretamente conectadas a este nó i.

Por outro lado, tem-se a lei de potência quando a frequência de algum evento, no caso a ocorrência de um grau k, varia em função exponencial (NEWMAN, 2005a).

Então, para identificar se uma distribuição de graus segue uma distribuição de lei de potências, exibi-se o gráfico de histogramas de ocorrência dos graus da rede em escala logarítmica, tanto no eixo x quanto no eixo y do gráfico. Se o histograma é uma linha reta em escalas log-log, então a distribuição de graus segue uma lei de potência. Esta observação parece ter sido feita primeiramente por Auerbach (AUERBACH, 1913).

Formalmente, seja $p(x)dx$ a fração de vértices com grau entre x e $x+dx$. Se um histograma é uma linha reta em um gráfico de escalas logarítmicas, então $\ln p(x) = -\alpha \ln x + c$, onde α e c são constantes (o sinal negativo é opcional, porém é conveniente pois normalmente, no gráfico, a linha é reta decrescente). Fazendo a exponencial de ambos os lados, isso é equivalente a:

$$p(x) = Cx^{-\alpha}, \tag{2.8}$$

com $C = e^c$.

A distribuição da Equação (2.8) diz-se seguir uma lei de potência (NEWMAN, 2005a).

Um exemplo encontrado na literatura pode ser observado em Zhang et al. (2010). Neste trabalho foi construído uma rede de séries temporais do índice das ações de Shanghai, onde os autores correlacionam par a par séries temporais do período de 5 de março a 17 de março de 2007. A série inteira dos preços do índice é dividida em k partes de segmentos adjacentes, que então é computado o retorno logarítmico dela (ver Seção 2.1.7). Com a matriz de correlações das séries temporais, Zhang limiariza as correlações a $v_c = 0.95$, sendo v_c o termo utilizado por Zhang para

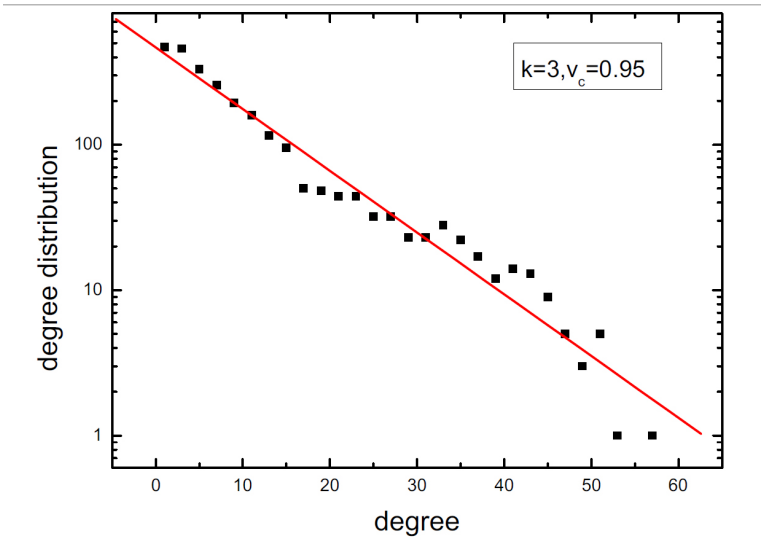

Figura 2.7: Distribuição de Graus da Rede de correlações de retornos de séries temporais filtrado com limiar $v_c = 0.95$ e com as coordenadas verticais em escala logarítmica, a linha vermelha é de ajuste de curva.
Fonte: (ZHANG et al., 2010).

a variável de limiarização, deixando somente as conexões mais fortes.

A Figura 2.7 mostra que a rede proposta por Zhang et al. (2010) de séries temporais correlacionadas é livre de escalas pois sua Distribuição de graus segue uma lei de potência.

2.1.3.4 Força Do Vértice Ou Grau Ponderado

Força do Vértice pode ser obtida através do grau ponderado de um nó i, ou seja, é a soma dos pesos de todas as arestas incidentes ao nó i (BARRAT et al., 2004):

$$s_i = \sum_j w_{i,j} * a_{i,j} \tag{2.9}$$

Na Equação (2.9), $a_{i,j} = 1$ caso haja uma ligação entre i e j; $a_{i,j} = 0$, caso contrário.

A Força do Vértice integra informações de conectividade de um nó, e a importância dela para as ligações da rede. Quanto maior a força de um nó, mais conectado este nó é (BARTHELEMY et al., 2005).

Esta medida física normalmente é utilizada em outras características físicas, como no Coeficiente de Clusterização (ver Seção 2.1.3.6), Distribuição de Graus Ponderado (Seção 2.1.3.3) e

entre outros.

2.1.3.5 Matriz De Distâncias De Caminhos, Excentricidade, Raio, Diâmetro E Comprimento Do Caminho Característico

Caminho (*Path*) é uma sequência ordenada de arestas distintas, ligando um nó i até um nó j. Se $i = j$, ou seja, se o nó fonte e o nó destino forem os mesmos, o caminho é chamado de ciclo. O comprimento r_{ij} de um caminho é definido como o número de arestas distintas ligando o vértices i até o vértice j (BUCKLEY; HARARY, 1990).

A matriz de distâncias $D(G)$ registra as distâncias definido como o comprimento do caminho mais curto entre os pares de nós de toda a rede. Se não existe caminho entre um nó i até um nó j, ou seja $r_{ij} = 0$, então a distância d_{ij} que é a soma dos pesos das arestas dististas que ligam o nó i até o nó j de menor comprimento é $d_{ij} = Inf$. Note que a diagonal principal da matriz $D(G)$ correspondem aos ciclos de menor caminho (HARARY, 1969; BUCKLEY; HARARY, 1990). Computacionalmente, as distâncias da matriz $D(G)$ podem ser obtidas através de algoritmos como Dijkstra (ver Seção 2.1.6) ou Busca em Largura (*breadth-first search*) (CORMEN et al., 2001).

A partir daí, as medidas Excentricidade, Raio, Diâmetro e Comprimento do Caminho Característico podem ser computados através da matriz de distâncias de caminhos $D(G)$.

a) Excentricidade mede o número máximo de passos ou peso que é necessário, a partir de um nó, para alcançar todos os nós da rede. Sendo denotada por:

$$ecc(i) = max(d_i), \tag{2.10}$$

onde d_i são as distâncias contendo o nó i.

b) Raio é a menor Excentricidade de todos os nós da rede, dado por:

$$raio(G) = min(ecc), \tag{2.11}$$

onde ecc são as Excentricidades de todos os nós.

c) O diâmetro é o grau de separação de rede, é o número máximo de passos ou peso que é ne-

cessário, a partir de qualquer nó, para alcançar todos os nós da rede, formalmente:

$$diametro(G) = max(ecc). \tag{2.12}$$

d) Comprimento do Caminho Característico da rede, denotado por $\lambda(G)$, constitui uma medida de tendência central da matriz de distâncias $D(G)$. Formalmente, dado pela média global de $D(G)$ (WATTS; STROGATZ, 1998; WATTS, 1999b).

2.1.3.6 Coeficiente De Clusterização Ponderado

Coeficiente de Clusterização (CC) mede o agrupamento dos nós vizinhos de um nó i (NEWMAN; BARABASI; WATTS, 2006).

Watts e Strogatz (1998) verificaram que as redes esparsas do tipo *small world*, são altamente clusterizados (alto índices de CC) (HOLLAND; LEINHARDT, 1971) e ainda com baixo índice de comprimento de caminho entre pares de vértices.

De acordo com Newman, Barabasi e Watts (2006), para calcular CC em redes não direcionados, verifica-se o número de triângulos formados com o nó i, dividido pelo total de arestas possíveis dos nós vizinhos de i, dado pelo Equação (2.13):

$$CC(i) = \frac{t_i}{k_i(k_i-1)/2}, \tag{2.13}$$

sendo t_i o número de triângulos formados com o nó i e k_i, o grau de i (Seção 2.1.3.3).

Porém, muitas redes do mundo real consideram pesos em arestas. Por isso, existe a necessidade da generalizações da Equação (2.13). Com isso, muitas propostas de se calcular o CC ponderado surgiram (SARAMäKI et al., 2007).

De uma forma geral e direta da Equação (2.13), pode-se ponderar as arestas dos nós vizinhos de i (LOPEZ-FERNANDEZ; ROBLES; BARAHONA, 2004):

$$CC(i)_{LF} = \frac{\sum_{j\neq k\in N(i)} w_{jk}}{k_i(k_i-1)/2}. \tag{2.14}$$

Na Equação (2.14), o termo $\sum_{j\neq k\in N(i)} w_{jk}$ pode ser visto como uma relação de peso entre

os nós vizinhos do nó i.

Barrat et al. (2004) introduziram uma abordagem para o cálculo do *CC* ponderado, que combina informações topológicas com a distribuição de peso da rede:

$$CC(i)_B = \frac{1}{s_i(k_i-1)} \sum_{j,k} \frac{w_{i,j}+w_{i,k}}{2} a_{i,j} a_{i,k} a_{j,k}, \tag{2.15}$$

onde $a_{i,j} = 1$ se existir uma aresta entre i e j e $a_{i,j} = 0$ caso contrário, o mesmo para $a_{i,k}$ e $a_{j,k}$, garantindo assim que serão usados somente triângulos entre o nó i. O fator normalizante $s_i(k_i-1)$ leva em consideração a força do vértice s_i (ver Seção 2.1.3.4), garantindo que $0 \leq C(i) \leq 1$. Esta definição considera somente os pesos das arestas adjacentes ao nó i; porém, não considera os pesos das arestas entre os nós vizinhos de i.

Onnela et al. (2005) propuseram um *CC* ponderado que considera todos os pesos de todas as arestas envolvidas: tanto as arestas adjacentes ao nó i, quanto as arestas entre os vizinhos de i.

$$CC(i)_O = \frac{2\sum_{j,k}(\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk})^{1/3}}{k_i(k_i-1)}, \tag{2.16}$$

onde

$$\hat{w}_{ij} = \frac{w_{ij}}{max(w)}, \tag{2.17}$$

em que os pesos das arestas $\hat{w}_{ij}$ são normalizados pelo maior peso w de todas as arestas da rede $max(w)$ (Equação 2.17).

Na Equação (2.16) o número de triângulos t da Equação (2.13) é substituído pela soma das intensidades dos triângulos de i normalizados.

As Equações (2.14), (2.15) e (2.16) possuem seus limites superiores baseados em uma rede binária (grau k_i). Além disso, o *CC* ponderado falha quando $k_i < 2$. No *CC* ponderado proposto em Zhang e Horvath (2005), usa-se somente pesos da rede:

$$CC(i)_Z = \frac{\sum_{j,k}\hat{w}_{ij}\hat{w}_{jk}\hat{w}_{ik}}{(\sum_k \hat{w}_{ik})^2 - \sum_k \hat{w}_{ik}^2}, \tag{2.18}$$

onde os pesos também são normalizados por $max(w)$, e o denominador ponderado limita superiormente o numerador, garantindo que o $CC(i)_Z \in [0,1]$.

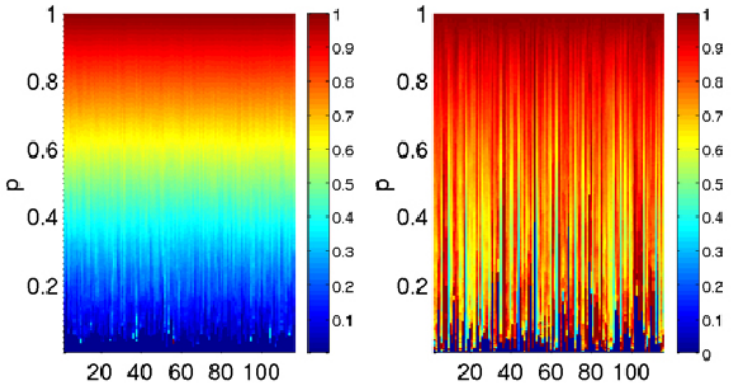

Figura 2.8: Coeficiente de clusterizaçao em função do vértice (eixo horizontal) e do coeficiente de correlação (eixo vertical). Direita: rede aleatória, esquerda: rede de ações.
Fonte: (ONNELA; KASKI; KERTESZ, 2004).

Até agora, foi mencionado como calcular o *CC* para um único nó. Assim, para medir o *CC* médio de toda a rede, soma-se todos os *CC* de cada nó da rede e divide-se pelo total de nós da rede:

$$CC = \frac{1}{N} \sum_{i \in N} CC(i), \tag{2.19}$$

onde N é o número de nós da rede, e $CC(i)$ é o Coeficiente de Clusterização do nó i.

Outra forma de calcular o *CC* médio é proposta em Schank e Wagner (2005). Onde o autor pensa em uma forma ponderada e utiliza uma função peso $w(i)$:

$$CC = \frac{1}{\sum_i w(i)} \sum_i w(i) C(i). \tag{2.20}$$

Na Equação (2.20) $w(i)$ representa uma função peso do nó i. Uma possível escolha para função peso é a força do vértice s_i (Seção 2.1.3.4).

O trabalho de Onnela, Kaski e Kertesz (2004) compara uma rede de correlações de retornos de 116 ações do índice S&P500 (*Standard & Poor's 500* é um índice de Mercado de ações com base em capitalização de 500 empresas líderes de capital aberto das Bolsas dos Estados Unidos (S&P, 2013)) com uma rede aleatória produzida artificialmente com 116 nós.

A Figura 2.8 ilustra a diferença entre o Coeficiente de Clusterização em função do coeficiente de correlação de uma rede aleatória e uma rede de ações.

Onnela analisou os Coeficientes de Clusterização e os Coeficientes de Clusterização Médios,

e descobriu que eles se comportam de forma muito diferente para a rede de ativos e aleatória. Os autores especulam que essa diferença pode ser adequada para estudar qual é a fração de arestas na rede, ou coeficientes de correlação na matriz de correlação relacionada, é informação e qual é ruído. Com base nesta abordagem, eles verificaram que apenas cerca de 10% das arestas parecem carregar informação trivial para verificação dos agrupamentos. Essas correlações são as arestas de peso entre 0 e 0.1.

2.1.4 MODULARIDADE

A Modularidade é uma medida de clusterização que quantifica a qualidade dos agrupamentos dos nós de uma rede (GIRVAN; NEWMAN, 2002; NEWMAN; GIRVAN, 2004). A idéia principal da Modularidade é comparar os agrupamentos da rede proposta com as conexões de uma rede aleatória de mesmas características (mesmo número de nós e mesma densidade de arestas). Quanto mais conexões intra-clusters e menos conexões inter-clusters, melhor a qualidade dos agrupamentos da rede proposta. Formalmente a Modularidade é dada por:

$$Q_s = \frac{1}{W}\sum_{ij} w_{ij} - \frac{s_i s_j}{W} * \delta(c_i, c_j), \tag{2.21}$$

onde w_{ij} é o peso da aresta do nó i para o nó j. Assim, $W = \sum_i \sum_j w_{ij}$, ou seja, a soma de todos os pesos das arestas da rede, sendo chamado de fator de normalização.

Na Equação (2.21) pode-se notar também o uso das forças do vértices s_i e s_j do nó i e j respectivamente (ver Seção 2.1.3.4). Esta parte da equação, $s_i s_j / W$, corresponde à probabilidade de conexão em uma rede aleatória de mesma propriedade.

A função $\delta(c_i, c_j)$ verifica se os dois nós i e j são do mesmo agrupamento. Assim, $\delta(c_i, c_j) = 1$ se o nó i e o nó j pertencerem a mesma comunidade, $\delta(c_i, c_j) = 0$, caso contrário.

Esta medida para quantificar a força de estrutura de comunidades de redes foi proposta por Newman e Girvan (2004). Nesse trabalho de Newman e Girvan (2004) é testado a Modularidade em várias redes: a primeira gerada computacionalmente com 128 nós e 4 comunidades; a segunda rede de relação de amizade entre indivíduos de um clube de karate (ZACHARY, 1977) divididos em duas comunidades; outra rede de colaborações entre pesquisadores sobre redes (PARK; NEWMAN,

2003) com 13 comunidades de vértices; outra rede de interações entre golfinhos (LUSSEAU, 2003) com um total de 5 comunidades; uma rede de interações entre os personagens principais do romance de Victor Hugo compilado por Knuth (KNUTH, 1993) com 11 comunidades; e a última rede de página da internet e seus *links* entre si com 7 comunidades de vértices.

Então, Newman, em seu artigo, utiliza alguns algoritmos de clusterização para agrupar os vértices das redes, são eles: *Shortest-path betweenness*, *Resistor networks* e *Random walks*. Os detalhes sobre esses algoritmos podem ser encontrados em Newman e Girvan (2004). Feito os agrupamentos das redes, os autores utilizaram a Modularidade para medir sua qualidade em várias etapas do algoritmo de clusterização, obtendo os seguintes picos de Modularidade: $Q \approx 0.5$, $Q \approx 0.4$, $Q = 0.72 \pm 0.02$, $Q = 0.52 \pm 0.03$, $Q = 0.54 \pm 0.02$ e $Q = 0.65 \pm 0.02$ respectivamente para as redes acima.

De acordo com Newman e Girvan (2004), uma rede agrupada e não aleatória tem sua Modularidade entre 0.3 a 0.7, valores maiores são raros.

2.1.5 ÁRVORE ESPALHADA MÍNIMA

Árvore Espalhada Mínima ou *Mimimum Spanning Tree* (*MST*) é um conceito da teoria dos grafos, no qual a partir de um grafo encontra-se o menor caminho acíclico que conecta todos os vértices (CORMEN et al., 2001).

Tem-se um grafo conectado não direcionado $G = (V, E)$, onde V é um conjunto de vértices, E é o conjunto de arestas. Cada aresta $(u, v) \in E$, tem um peso $w(u, v)$ ligando o par de vértices u e v.

Então deseja-se encontrar um subconjunto acíclico $T \subseteq E$ que conecte todos os vértices e cuja distância total $w(T)$ seja minimizada (Equação (2.22)).

$$w(T) = \sum_{(u,v) \in T} w(u, v) \tag{2.22}$$

Para computar a árvore de espalhamento mínima existem vários algoritmos que resolvem este problema, alguns deles são: o algoritmo de Kruskal e o algoritmo de Prim (CORMEN et al.,

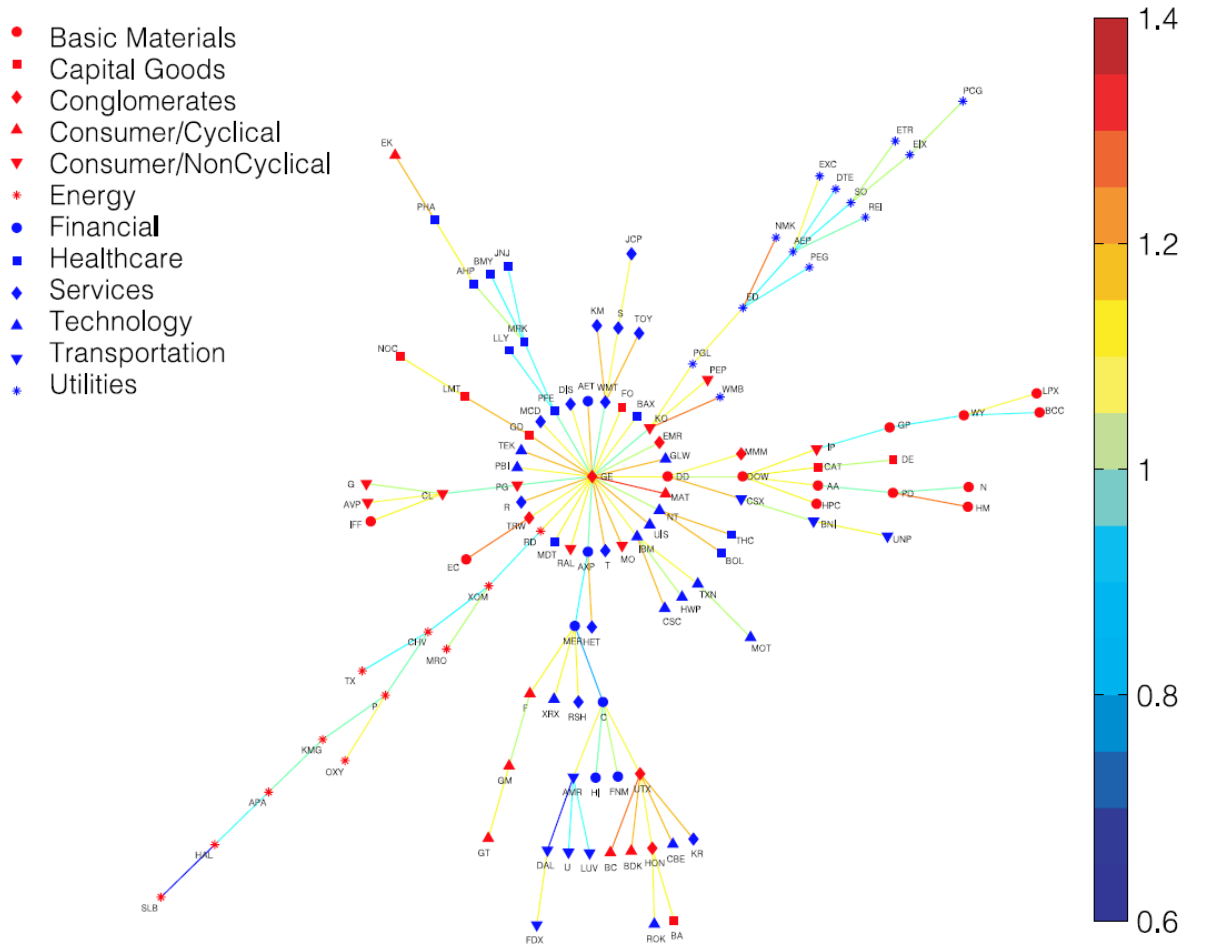

Figura 2.9: Árvore *MST* de distâncias de correlações de ações da *New York Stock Exchange* (*NYSE*).
Fonte: (ONNELA et al., 2002).

2001).

A Figura 2.9 mostra um exemplo de árvore *MST* estudado por Onnela em 2002, a rede é montada a partir de distâncias de correlações de 116 ações negociada na bolsa norte americana *NYSE*. Nesta rede, os nós são representados pelo código da ação, além da forma geométrica que indica o setor a qual a ação é classificada; as arestas são niveladas por cores que representam o quão distante a correlação de pares de ações é.

2.1.6 ALGORITMO DE DIJKSTRA

O algoritmo de Dijkstra encontra o caminho mais curto de uma única origem para um grafo orientado ponderado $G = (V, E)$, sob a restrição que os pesos das arestas sejam positivos $w_{i,j} \geq 0$ (CORMEN et al., 2001).

Este algoritmo utiliza uma estratégia gulosa, a partir de um nó i, ele relaxa todas as arestas incidentes à i. Em seguida, verifica o nó mais próximo para relaxar as arestas ainda não verificadas incidentes a esse nó. Assim, repetindo o processo até que todos os nós sejam verificados e relaxados. Este processo é repetido $|V|$ vezes, $|V|$ é o número de vértices no grafo. Isso garante que se encontre o caminho mais curto sempre (CORMEN et al., 2001).

O relaxamento de uma aresta w_{ij}, significa atribuir a w_{ij} o peso da aresta w_{ik} mais o peso da aresta w_{kj}, se e somente se, $w_{ik} + w_{kj}$ for menor do que w_{ij}. Formalmente $w_{ij} \leftarrow w_{ik} + w_{jk}$, se $w_{ik} + w_{jk} < w_{ij}$.

Analisando a complexidade do algoritmo, tem-se $O(|E| + |V|log|V|)$ no pior caso, pois para cada nó é verificado as arestas ainda não relaxadas, sendo a metade do total de nós. Mais a comparação para o relaxamento, que é o número de arestas da rede.

2.1.7 COEFICIENTE DE CORRELAÇÃO

Também chamado de Correlação Cruzada (*cross correlation*), o Coeficiente de Correlação é utilizado para quantificar o grau de semelhança entre a evolução temporal síncrona de um par de preços das ações (FELLER, 1968).

Na teoria da probabilidade e estatística, o coeficiente de correlação, indica a força e a direção do relacionamento linear entre duas variáveis. Contudo, a correlação não implica causalidade (causa e efeito) (SPIEGEL; SCHILLER; SRINIVASAN, 2000).

Formalmente, existem várias correlações para diferentes situações. Porém, o utilizado neste trabalho é a chamada Coeficiente de Correlação de Pearson. O Coeficiente de Correlação de Pearson é dada pela covariância de duas variáveis divido pelo produto dos desvios padrões delas (SPIEGEL; SCHILLER; SRINIVASAN, 2000).

Por outro lado, pode-se dizer que as ações são séries temporais pois a variação do preço de fechamento é dependente do preço de fechamento vizinho anterior (para verificar se o preço está subindo ou desvalorizando), ou seja, a ordem das observações é de muita importância.

Assim, para calcular a variação de uma série temporal, denotada por r_i^t, da ação i de uma

janela de período t, (por exemplo de 01 de Janeiro de 2012 a 31 de Janeiro de 2012), utiliza-se o chamado retorno logarítmico, dado por:

$$r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1), \tag{2.23}$$

onde $P_i(\tau)$ é o preço de fechamento da ação i no instante τ da janela do período t. Assim, r_i^t é um vetor de retornos de todos os instantes de tempo τ da janela t.

Com isso, o Coeficiente de Correlação de Pearson será utilizado para calcular a correlação de duas séries temporais, formalmente o Coeficiente de Correlação é dado pela equação:

$$p_{ij}^t = \frac{< r_i^t r_j^t > - < r_i^t >< r_j^t >}{\sqrt{(< (r_i^t)^2 > - < r_i^t >^2)(< (r_j^t)^2 > - < r_j^t >^2)}}, \tag{2.24}$$

onde i e j são as ações a serem comparadas; r_i^t é um vetor de retornos caracterizando a evolução temporal síncrona da ação i na janela de tempo t; p_{ij}^t é o Coeficiente de Correlação entre a ação i e j no período de tempo t.

O símbolo $<>$ representa a média empírica estatística. Assim tem-se que a média estatística é uma média temporal realizada em todos os dias negociados dentro do período investigado.

Por definição, p_{ij}^t pode variar entre $[-1,1]$, onde $p_{ij}^t = 1$ significa que as ações i e j são totalmente correlacionadas, ou seja, se a ação i sobe em seu preço, então a ação j sobe igualmente. Por outro lado, se $p_{ij}^t = -1$ as ações i e j são completamente anti-correlacionadas, ou seja, se a ação i varia em seu preço descendo, a ação j compensa, na mesma proporção, subindo em seu preço. Ou então se $p_{ij}^t = 0$ as variações dos preços das ações i e j não são correlacionadas.

2.2 ÁREA ECONÔMICA

Economia é a ciência que estuda a produção, distribuição e consumo de bens e serviços. O termo economia vem do grego, *oikos* que significa casa, e *nomos*, significa lei, costume, gerir ou administrar. Então economia significa "regras da casa" ou "administração da casa".

Todos os agregados econômicos, tais como a inflação, nível de desemprego, nível de reservas cambiais, vendas no atacado/varejo, nível de renda familiar, taxa de juros, produção nacional, taxa de câmbio, déficits e superávits das contas do governo, entre outros, servem para medir o nível de atividade econômica em um país (BODIE; KANE; MARCUS, 2009).

Com a globalização, o mercado acionário brasileiro vem adquirindo uma crescente importância no cenário financeiro econômico internacional. Os países em desenvolvimento e expansão, como os BRICs (Brasil, Rússia, Índia e China), procuram se relacionar com os outros países a fim de receber investimentos externos.

Então, essa captação de recursos é um canal fundamental para o desenvolvimento das empresas, gerando novos empregos e contribuindo para o progresso do país. Além disso o mercado acionário se constitui em uma opção de investimento para pessoas e instituições (BOVESPA, 2008).

Assim, quanto mais desenvolvida é uma economia, mais ativo é o seu mercado de capitais, que é constituído principalmente pela Bolsa de Valores.

2.3 BM&F BOVESPA

A BM&F BOVESPA foi formada em 2008 com a junção das atividades da Bolsa de Mercadorias & Futuro e a Bolsa de Valores de São Paulo. Ela é a principal companhia de capital brasileiro, e a única bolsa de valores, mercadorias e futuros em operação no Brasil (BOVESPA, 2012).

A companhia desenvolve e implanta sistemas para uma negociação segura e clara de valores e mercadorias, tais como: ações, derivativos de ações, títulos de renda fixa, títulos públicos federais, derivativos financeiros, moedas à vista e *commodities* agropecuárias. Ela também atua como depositária central dos ativos negociados em seus sistemas, além de licenciar softwares e

índices.

As plataformas de negociação da BM&F BOVESPA realizam o registro, a compensação e a liquidação de ativos e valores mobiliários transacionais e a listagem de ações e de outros ativos.

A BM&F BOVESPA é regularizada e supervisionada pela Comissão de Valores Mobiliários e pelo Banco Central do Brasil.

As empresas que formam o grupo BM&F BOVESPA são:

a) Banco BM&FBovespa

b) BM&FBOVESPA Supervisão de Mercados

c) Bolsa Brasileira de Mercadorias

d) BM&F USA Inc.

e) Bolsa de Valores do Rio de Janeiro - BVRJ

2.3.1 ÍNDICE BOVESPA

O Índice Bovespa ou Ibovespa foi implementado em 1968, e é composto pelas principais ações da BM&F BOVESPA. Pela sua tradição e composição, o Ibovespa é considerado o indicador médio mais importante do mercado de ações brasileiro (BOVESPA, 2012).

O índice reflete não apenas as variações dos preços das ações, mas também o impacto da distribuição dos proventos, sendo considerado um indicador que avalia o retorno total de suas ações componentes.

A finalidade do Ibovespa é de servir como indicador médio do comportamento do mercado. Para tanto, sua composição procura aproximar-se o mais possível da real configuração do perfil das negociações à vista observadas nos pregões da BM&F BOVESPA.

As ações integrantes da carteira teórica do Ibovespa respondem por mais de 80% do número de negócios e do volume financeiro verificados no mercado à vista da BM&F BOVESPA. Além

disso, as empresas emissoras das ações integrantes da carteira teórica do Índice Bovespa são responsáveis, em média, por aproximadamente 70% do somatório da capitalização bursátil de todas as empresas com ações negociáveis na BM&F BOVESPA.

2.3.1.1 Critérios De Inclusão De Ações No Índice Bovespa

Para uma ação fazer parte da carteira teórica do Ibovespa, ela precisa atender aos seguintes critérios em relação aos doze meses anteriores:

a) estar contido na relação das ações que representam 80% da negociabilidade do mercado;

b) ter volume de participação superior a 0.1% do total;

c) ter sido negociada em mais de 80% do total de pregões do período.

2.3.1.2 Participação Da Ação Na Carteira Teórica

Cada ação na carteira teórica da Ibovespa têm um peso de participação. Este peso tem relação direta com a representatividade deste título no mercado à vista, e é chamado de Índice de Negociabilidade da ação.

O índice de negociabilidade, formalmente, é calculado em termos do número de negócios e volume financeiro da ação em relação ao total negociado na BM&F Bovespa, dado pela seguinte equação:

$$IN = \sqrt{\frac{n_i}{N} * \frac{v_i}{V}} \tag{2.25}$$

onde, IN = índice de negociabidade da ação, n_i = número de negócios com a ação i no mercado à vista (lote-padrão), N = número total de negócios no mercado à vista da BM&F BOVESPA (lote-padrão), v_i = volume financeiro gerado pelos negócios com a ação i no mercado à vista (lote-padrão) e V = volume financeiro total do mercado à vista da BM&F BOVESPA (lote-padrão).

2.3.1.3 Apuração Do Índice Bovespa

O Índice Bovespa é o somatório dos pesos (quantidade teórica da ação multiplicada pelo último preço da mesma) das ações integrantes de sua carteira teórica. Dada pela seguinte equação:

$$Ibovespa_t = \sum_{i}^{n} P_{i,t} * Q_{i,t} \tag{2.26}$$

onde, $Ibovespa_t$ = valor do Índice Bovespa no instante t, n = número total de ações da carteira teórica, $P_{i,t}$ = último preço da ação i no instante t e $Q_{i,t}$ = quantidade teórica da ação i na carteira no instante t.

2.3.1.4 Vigência Da Carteira

A cada quatro meses, a carteira teórica do Índice Bovespa é reavaliada para que sua representabilidade se mantenha. Os períodos de vigor das carteiras são entre janeiro a abril, maio a agosto e setembro a dezembro de cada ano. Além disso, a BM&F BOVESPA lança três prévias da nova composição da carteira, a 30 dias, a 15 dias e a 1 dia do novo ciclo quadrimestral.

2.4 RENTABILIDADE DAS AÇÕES

A rentabilidade é diretamente relacionada ao risco que estamos disposto a correr em função de obter uma maior ou menor lucratividade (BOVESPA, 2008).

Pode-se dizer que as ações seguem a lei da oferta e da procura. Se há muitos compradores, o preço tende a subir, se há muitos vendedores, o preço tende a cair. A partir disso, temos as flutuações do mercado, que são as variações dos preços das ações com a negociação na bolsa de valores. A rentabilidade vem da venda da ação comparado ao valor de sua compra. Se o preço vem subindo, então temos lucro na venda. Alguns fatores fazem com que o investidor compre ou venda, isso se dá a partir da análise fundamentalista e análise gráfica.

Outras formas de rentabilidade também podem ser obtidas a partir dos dividendos, bonificação em ações, direitos de subscrição, venda do direito de subscrição entre outros.

2.4.1 ANÁLISE FUNDAMENTALISTA E ANÁLISE GRÁFICA

A análise fundamentalista tenta medir o valor intrínseco de um ativo. Quanto seria o preço justo a se pagar pelo ativo, de acordo com as atividades da empresa, análise de lucro, faturamento, preço de mercado, perspectiva do setor, sazonalidade, entre outros (BODIE; KANE; MARCUS, 2009).

Para tanto, demanda-se tempo de estudo sobre a empresa a qual se queira investir, além de acompanhar notícias sobre a empresa, saber se o setor em que ela atua é promissor, identificar se é uma empresa sadia, se possui bons líderes.

O processo de estimar o valor de uma empresa é conhecido por *Valuation*, e a partir desta estimativa, decidir quando comprar ou vender um ativo.

Por sua vez, a análise gráfica ou também conhecido por análise técnica, são análises a partir do histórico de um ativo. Nesta análise, o investidor estuda o gráfico do ativo a fim de encontrar padrões que se formam repetidamente, e assim decidir o momento de comprar ou vender um ativo (BODIE; KANE; MARCUS, 2009).

Para tanto, os analistas utilizam indicadores para ajudar a definir padrões, são elas média móvel, bandas, canais, momento, osciladores, volatilidade, volume e entre outros.

Tanto a análise fundamentalista quanto a análise gráfica são amplamente usadas.

Neste trabalho, a análise será baseada na análise gráfica, pois não serão considerados notícias, nem especulações.

2.4.2 RISCO

Existem dois tipos de riscos: risco de mercado e risco específico (SHARPE, 1964).

O risco de mercado, também conhecido como risco sistemático ou não diversificável, é caracterizado pelas influências econômicas, políticas, sociais, ambientais e outras a que todas as ações ou setores de ações estão sujeitas. Podemos citar também fatores macroeconômicos como taxa de juros, câmbio ou qualquer outra variável que afete a economia como um todo, portanto afeta todas as ações do mercado.

O risco específico, não sistemático ou diversificável depende de fatores específicos que afetam uma empresa.

Somando-se o risco sistemático com o não sistemático encontra-se o risco total da carteira, portanto:

$$RiscoTotaldaCarteira = RiscoSistematico + RiscoNaoSistematico. \quad (2.27)$$

2.4.3 CARTEIRA

O grupo de ativos ou ações que um investidor possui é o que chamamos de carteira de investimentos. Para administrar uma carteira é preciso ter em mente a relação risco e retorno. Caso o investidor queira auferir uma taxa de retorno maior, deverá aceitar um nível maior de risco também. No entanto, cabe ao administrador financeiro gerenciar a carteira de investimento de modo que se obtenha o maior retorno para um determinado nível de risco. Uma carteira contendo uma combinação de ativos que tenha o máximo retorno para uma determinada classe de risco, ou o mínimo risco para uma dada classe de retorno, é conhecido por portifólio eficiente. Portanto, o objetivo principal na administração de carteiras é delinear portifólios eficientes (BODIE; KANE; MARCUS, 2009)

Devido às oscilações de mercado em casos que não são previstas ou indevidas, a concentração de investimento em apenas um ativo expõe o investidor a um grande risco, ficando a mercê das oscilações deste ativo. Então, para que este risco seja diluído de forma consistente, o processo de diversificação de ativos é necessário.

De acordo com Aragon (1989), o risco não-sistemático pode ser diminuído com uma simples diversificação aleatória dos ativos no portifólio de investimentos.

Porém, Harry Markowitz Markowitz (1952), ganhador do prêmio Nobel de economia em 1991, desenvolveu um processo científico de alocação de ativos em portifólio de investimentos. O grande diferencial entre a diversificação simples e a diversificação de Markowitz é que nessa, são consideradas as correlações entre os ativos do portifólio. Com esse método, pode-se, além reduzir o risco específico de forma matemática e estatística, pode-se também reduzir o risco de mercado a

que uma setor ou empresa está sujeita.

A diversificação de Markowitz é um procedimento analítico que envolve a combinação de ativos correlacionados, no sentido de formar portifólios eficientes. Assim, não basta diversificar a carteira de investimentos com ativos aleatoriamente para que se consiga uma redução de risco significativa. Uma análise entre correlações de ações é necessária, ativos anti-correlacionados são precisos para que, dado um risco aceitável de um investidor, se consiga o máximo de retorno estimado.

3 TRABALHOS RELACIONADOS

No mercado de ações, um único número (preço da ação) caracteriza o desempenho de uma companhia. Estas companhias interagem umas com as outras, criando um sistema complexo em evolução (ARTHUR; DURLAUF; LANE, 1997; STANLEY et al., 1996). Na medida em que mudanças de preço estão acontecendo na bolsa, parece seguro assumir que elas são refletidas em correlações a outros ativos ao mesmo tempo. Este é um ponto central na teoria de investimento e gestão de risco, e também serve como entrada para o problema de otimização de carteira na teoria clássica de Markowitz (ver Seção 2.4.3).

A teoria de redes fornece uma abordagem para sistemas complexos com muitas unidades interagindo, onde os detalhes das interações são de menor importância, focando-se na sua existência. Recentemente, a utilização das redes provou ser útil em um vasto campo de aplicações que vão desde a Internet à microbiologia, mostrando as capacidades delas também no campo da economia (CALDARELLI et al., 2004; TABAK; SERRA; CAJUEIRO, 2010).

As análises dos índices do mercado de ações podem melhorar a nossa compreensão do comportamento e da estrutura de mercado. Recentemente, muitas pesquisas têm sido feitas aplicando conceitos de física e métodos para o estudo do mercado financeiro e de séries temporais financeiras (STANLEY et al., 1999; PLEROU; GOPIKRISHNAN; STANLEY, 2003; GABAIX et al., 2003). Com isso, um novo foco de atenção tem surgido, tentando interpretar o mercado (KIM et al., 2002a; BONANNO et al., 2004).

Mantegna (1999) foi um dos pioneiros a construir redes baseadas em correlações dos preços das ações. Em seu artigo, o autor modela grafos de coeficientes de correlações de pares de preços das ações (ver Seção 2.1.7). O autor ainda define uma métrica de distância de similaridades de pares de ações em função do coeficiente de correlação.

Assim, Mantegna (1999) constrói uma matriz simétrica de distâncias de todos os pares de ações, e utiliza Árvores Espalhadas Mínimas (*Mimimum Spanning Tree* MST) (ver Seção 2.1.5) para encontrar as conexões de maior correlação entre as ações. O autor conclui que esta topologia é útil na descrição teórica do mercado financeiro e na busca de fatores econômicos comuns que afetam grupos específicos de ações. Além de mostrar que séries temporais dos preços das ações carregam informações econômicas detectáveis.

Seguindo a mesma linha, Bonanno et al. (2004) também montou uma matriz de distâncias utilizando coeficientes de correlações de pares de preços de ações. Porém com outra função métrica (Equação (3.1)), originalmente dado por GOWER (1966):

$$d_{i,j}(\Delta t) = \sqrt{2(1 - p_{ij}(\Delta t))}. \tag{3.1}$$

A Equação (3.1) mede a distância seguindo os axiomas das métrica euclidiana para o coeficiente de correlação p_{ij} das ações i e j em um intervalo de tempo Δt. Por definição, $d_{i,j}(\Delta t)$ vai de $[0,2]$, sendo $d_{i,j}(\Delta t) = 2$ o valor mais distante, ou seja, perfeitamente anti-correlacionado. Caso $d_{i,j}(\Delta t) = 0$, então as ações i e j são totalmente correlacionadas.

A partir daí, no trabalho de Bonanno também é obtida a MST da matriz resultante utilizando o método *nearest neighbor single linkage cluster analysis* conhecido na análise multivariada (MARDIA; KENT; BIBBY, 1979).

Bonanno et al. (2004) conclui que séries temporais de preços em um mercado financeiro refletem informações sobre o setor econômico da atividade da empresa. Esta informação é geralmente ruidosa devido às flutuações estatísticas. Procedimentos de filtragem, como o proposto, é capaz de revelar as informações importantes, como de agrupamento de ativos do mesmo setor.

O autor também verificou que a rede sofre alterações em sua topologia, variando de um grande agrupamento de ações até vários agrupamentos relativamente pequenos, de acordo com a variação do horizonte de tempo investigado. De maneira resumida, o estudo de Bonanno et al. (2004) sobre correlação baseada em redes financeira é um método capaz de filtrar informações econômicas a partir da matriz de coeficientes de correlação de um conjunto de séries temporais

financeiras. A topologia da rede detectada pode ser usada para validar ou falsificar modelos de mercado.

Kim et al. (2002a, 2002b) também introduziram uma rede ponderada, associada à correlação cruzada sobre os preço das ações entre as 500 empresas do S&P500. Porém, antes do cálculo do coeficiente de correlação, foi subtraído do valor do retorno logarítmico (ver Seção 2.1.7), denotado pelos autores por Y_i, pela média dos retornos de todas as 500 ações:

$$G_i(t) = Y_i(t) - \frac{1}{N}\sum_i S_i(t). \tag{3.2}$$

Esta modificação se deu devido a distribuição dos coeficientes de correlações anteriores, em vários períodos de tempo, terem seu desvio padrão quase constantes, enquanto somente a média da distribuição se alterava.

De acordo com Kim et al. (2002a, 2002b), esta modificação exclui o comportamento de variações dos preços causados por fatores econômicos externos (juros bancários, o índice de inflação, a taxa de câmbio, outros). Os autores concluem que a correlação obtida usando esta alteração não dependente do tempo; e os pesos das arestas conseguem extrair propriedades intrínsecas das variações dos preços.

Além disso, Kim verifica que a distribuição das forças dos vértices (ver Seção 2.1.3.4) de sua rede segue uma lei de potência $P_I(\mid q \mid) \sim \mid q \mid^{-} \delta$ com $\delta \approx 1.8$. Kim et al. (2002a, 2002b) concluem que sua rede exibe um comportamento livre de escalas. O fato do expoente ser menor que 2, implica que as mudanças dos preços das ações das empresas estão mais fortemente correlacionadas do que, por exemplo, com a topologia da Internet ou *world-wide web*, refletindo que fenômenos cooperativos e adaptativos aparecem predominantemente nos sistemas econômicos.

Onnela et al. (2003b) estudaram a rede proposta por Mantegna (1999). Neste estudo, os autores introduziram um conceito de camada média de ocupação da árvore de ativos, dado por:

$$l(t, v_c) = \frac{1}{N}\sum_{i=1}^{N} \mathscr{L}(v_i^t), \tag{3.3}$$

onde $\mathscr{L}(v_i)$ é o nível do vértice v_i na árvore MST em relação ao vértice central v_c, cujo nível é

zero. A camada média de ocupação indica a camada onde se concentra a maior parte dos nós da árvore.

De acordo com Onnela et al. (2003b), o vértice central v_c, que é a raiz da árvore, pode ser encontrado das seguintes formas:

a) Nó com o maior grau de vértice.

b) Nó com a maior soma de coeficientes de correlação, que estão associados com a aresta incidente a este nó.

c) Considerando a árvore T^t no instante t, o vértice v_i que produz o menor valor de camada média de ocupação $l(t, v_i)$ é o centro de massa, dado que a todos os nós são atribuídos um peso igual, e camadas consecutivas são equidistântes entre si.

Porém, alterando a forma de encontrar o vértice central, os resultados da camada média de ocupação foram similares.

Com os estudos, pode-se concluir que a camada média de ocupação varia em função do tempo, e em momentos que crise o valor de $l(t, v_i)$ cai. Altos valores de $l(t, v_i)$ podem resultar em um crescimento geral constante nos preços dos ativos durante esse período, como pode ser visto, por exemplo, com as ações do índice S&P 500. Com este estudo, os autores também verificam que as ligações estão concentradas em poucos ativos mostrando a característica livre de escala da rede.

Quanto à análise de portfólio, Onnela et al. (2003b) verificaram que as ações incluídas na carteira de risco mínimo tendem a estar na periferia da árvore de ativos: em média, a camada de portfólios eficientes (ver Seção 2.4.3) pode ser duas camadas ou mais, abaixo do vértice central da árvore de ativos.

No trabalho de Tabak, Serra e Cajueiro (2010), foi construído uma árvore MST a partir de uma rede de correlações de 47 ações do mercado brasileiro. Em seus estudos foram analisados algumas medidas físicas como: disparidade, centralidade, poder de dominação e coeficiente de clusterização. As medidas foram analisadas em diferentes períodos de tempo.

Tabak, Serra e Cajueiro (2010), de acordo com essas características físicas, concluiu que a árvore MST agrupa ações do mesmo setor de atividade. Os setores mais ativos, de acordo com as medidas físicas, são da área financeira, de materiais e de energia. Porém, recentemente, o setor de telecomunicações tem aumentado sua influência na rede. Com isso, os autores concluíram que a árvore MST de ativos pode ser utilizada para analisar o mercado brasileiro, e o estudo sobre o tempo é importante, pois as medidas mudam com o passar do tempo.

Tse, Liu e Lau (2010) utilizaram redes de ações americanas e suas correlações com um limiar de 0.9, em dois períodos distintos de quase dois anos cada. Em seus resultados, os autores obteveram uma rede livre de escalas. E mostraram que o setor financeiro influência muito o mercado dos Estados Unidos.

Huang, Zhuang e Yao (2009) apresentaram um estudo detalhado das propriedades de rede de correlações de ações chinesas. Encontrando componentes, cliques e conjuntos independentes na rede, que gerou uma nova ferramenta para a análise da estrutura do mercado, classificando as ações em diferentes grupos (CORMEN et al., 2001). Em que é de grande ajuda para os investidores para a tomada de decisões sobre suas carteiras. Portanto, esta técnica é útil a partir de pontos de vista teóricos e práticos. Huang, Zhuang e Yao (2009) também fizeram uma análise estatística da distribuição de graus da rede e mostraram que o modelo de lei de potência é válida em redes financeiras do mercado chinês.

Assim, a análise de correlação é importante na otimização do portfólio para investimentos no mercado internacional e nacional. Alguns estudos anteriores mostraram que os benefícios da diversificação podem ser afetados durante os períodos de maior volatilidade nos mercados financeiros. Assim, a análise de correlação tem desempenhado um papel importante no estabelecimento da relação entre os benefícios da diversificação e a volatilidade do mercado. Por exemplo, Longin e Solnik (2001) têm analisado tal correlação, levando à conclusão de que as condições de mercado turbulentas resultariam em uma redução nos benefícios da diversificação da carteira. Estes resultados anteriores estimularam novas pesquisas sobre as características das correlações. Por exemplo, o trabalho de Campbell et al. (2008) levaram a melhor compreensão da "estrutura de correlação" ou distribuição de correlação. Recentemente, também foram estudadas a dependência do tempo

das correlações, resultando em uma rede de influência (KULLMANN; KERTéSZ; KASKI, 2002).

4 PROPOSTA

4.1 BASE DE DADOS

A base de dados utilizada foi montada a partir de cotações históricas das ações da Bovespa (Bolsa de Valores do Estado de São Paulo) no período de 03 de dezembro de 2001 até 30 de março de 2012 (BOVESPA, 2012).

4.1.1 BASE DE DADOS DA BM&F BOVESPA

A base de (BOVESPA, 2012) são arquivos de textos separados por ano com a nomenclatura COTAHIST.AAAA.TXT, onde AAAA corresponde ao ano, exemplo: COTAHIST.2001.TXT, COTAHIST.2002.TXT, e assim por diante. Cada arquivo contém registro de cabeçalho, registros de cotações históricas por papel do mercado e registro de *trailer*. O cabeçalho é representado pela primeira linha do arquivo, e contém informações como: tipo de registro, nome do arquivo, código de origem e data de geração do arquivo. No *trailer* também apresentam estas informações acrescentadas ao total de registros do arquivos, e se localiza na última linha do arquivo. Nos registros de cotações históricas são mostradas informações de papéis negociados diariamente. Cada linha registra informações como:

a) tipo de registro

b) data do pregão

c) código BDI (Benefício e Despesas Indiretas)

d) código de negociação do papel

e) tipo de mercado

f) nome resumido da empresa emissora do papel

g) especificação do papel

h) prazo em dias do mercado a termo

i) moeda de referência

j) preço de abertura do papel-mercado no pregão

k) preço máximo do papel-mercado no pregão

l) preço mínimo do papel-mercado no pregão

m) preço médio do papel-mercado no pregão

n) preço do último negócio do papel-mercado no pregão

o) preço da melhor oferta de compra do papel-mercado

p) preço da melhor oferta de venda do papel-mercado

q) número de negócios efetuados com o papel-mercado no pregão

r) quantidade total de títulos negociados neste papel-mercado

s) volume total de títulos negociados neste papel-mercado

t) preço de exercício para o mercado de opções ou valor do contrato para o mercado de termo secundário

u) indicador de correção de preços de exercícios ou valores de contrato para os mercados de opções ou termo secundário

v) data do vencimento para os mercados de opções ou termo secundário

w) fator de cotação do papel

x) preço de exercício em pontos para opções referenciadas em dólar ou valor de contrato em pontos para termo secundário

y) código do papel no sistema isin ou código interno do papel

z) número de distribuição do papel

Neste trabalho, foi utilizados os arquivos entre os anos 2001 e 2012. Estes arquivos foram filtrados para obter somente informações relevantes ao trabalho como serão vistos na próxima seção.

4.1.2 PREPARAÇÃO DA BASE DE DADOS

Apesar dos arquivos de cotações históricas possuírem diversas informações, somente os seguintes dados serão de interesse:

a) Código da Ação na bolsa de valores

b) Data do Pregão em formato AAAAMMDD

c) Preço de Fechamento em espécie Real brasileiro

A utilização do preço de fechamento é devido à inexistência de *gaps* (diferença entre a cotação de preço de abertura e preço de fechamento do dia anterior). Além disso, o preço de fechamento é a referência utilizada pela própria BM&F Bovespa para verificar se alguma ação ou índice está em alta ou baixa.

Para a seleção das ações foi utilizada a composição do Índice Bovespa (ver Seção 2.3.1). Essas ações representam mais de 80% das negociações da BM&F Bovespa do período anterior a formação do índice.

As datas observadas neste trabalho vão de 03 de dezembro de 2001 até 30 de março de 2012. Totalizando aproximadamente 11 anos e 4 meses de observações.

Além disso, as ações serão classificadas setorialmente, de acordo com a área de atuação da empresa. Essa classificação pode ser obtida em Bovespa (2012).

Cada ação será classificada em um dos dez setores:

a) Bens Industriais

b) Construção e Transporte

c) Consumo Cíclico

d) Consumo não Cíclico

e) Financeiro e Outros

f) Materiais Básicos

g) Petróleo, Gás e Biocombustíveis

h) Telecomunicações

i) Utilidade Pública

j) Tecnologia de Informação

Esta classificação foi utilizada para verificar se as ações podem ser agrupadas de acordo com seu setor, assim como no trabalho de Tabak, Serra e Cajueiro (2010).

4.1.3 EXEMPLO DA BASE

A Figura 4.1 mostra um trecho do armazenamento dos dados coletados da BM&F Bovespa em um *array*. Em que cada elemento do vetor do *array* possui o código da ação negociada na Bovespa, a data do pregão no formato AAAAMMDD e o preço de fechamento desta ação nesta data. O vetor está ordenado sequencialmente pela data do pregão e agrupado pelo código da ação.

Para a classificação setorial é utilizada um *array* onde cada campo armazena o código da ação negociada na Bovespa e o setor a qual essa empresa atua. Um trecho do armazenamento pode ser visto na Figura 4.2.

4.2 CONSTRUÇÃO DA REDE

Para a construção da rede, a base de dados (ver Seção 4.1) tem um total de 2559 observações de preço de fechamento diário para cada ação. Para analisar a evolução da rede em função do

Trecho de armazenamento da Base de Dados

... [‘USIM5’, 20110818.0, 12.2] [‘USIM5’, 20110819.0, 12.26] [‘USIM5’,20110822.0, 12.7] [‘USIM5’, 20110823.0, 13.5] [‘USIM5’,20110824.0, 13.24] [‘USIM5’, 20110825.0, 12.71] [‘USIM5’,20110826.0, 12.25] [‘USIM5’, 20110829.0, 12.02] [‘USIM5’,20110830.0, 11.9] [‘USIM5’, 20110831.0, 12.0] [‘USIM5’,20110901.0, 12.1] [‘USIM5’, 20110902.0, 11.6] [‘USIM5’,20110905.0, 11.23] [‘USIM5’, 20110906.0, 11.72] [‘USIM5’,20110908.0, 11.95] [‘USIM5’, 20110909.0, 11.95] [‘VALE3’,20000103.0, 21.0] [‘VALE3’, 20000104.0, 20.5] [‘VALE3’,20000105.0, 20.5] [‘VALE3’, 20000106.0, 20.5] [‘VALE3’,20000107.0,21.25] ...

Figura 4.1: Trecho da Base de Dados utilizada, onde cada campo armazena o código da ação negociado na BM&F Bovespa, a data do pregão em formato AAAAMMDD e o preço de fechamento desta ação nesta data.

Trecho de armazenamento da classificação setorial das Ações

... [‘ITAU4’, ‘Financeiro e Outros’] [‘ITSA4’, ‘Financeiro e Outros’] [‘KLAB4’, ‘Materiais Básicos’] [‘LIGH3’, ‘Utilidade Pública’] [‘PETR3’, ‘Petróleo, Gás e Biocombustíveis’] [‘PETR4’, ‘Petróleo, Gás e Biocombustíveis’] [‘PLIM4’, ‘Consumo Cíclico’] [‘PTIP4’, ‘Petróleo, Gás e Biocombustíveis’] [‘SBSP3’, ‘Utilidade Pública’] [‘TBLE3’, ‘Utilidade Pública’] [‘TCOC4’, ‘Telecomunicações’] [‘USIM5’, ‘Materiais Básicos’] [‘VALE5’, ‘Materiais Básicos’] [‘VCPA4’, ‘Materiais Básicos’] [‘KLBN4’, ‘Materiais Básicos’] [‘VALE3’, ‘Materiais Básicos’] [‘CMET4’, ‘Materiais Básicos’] [‘GOAU4’, ‘Materiais Básicos’] [‘NETC4’, ‘Consumo Cíclico’] ...

Figura 4.2: Trecho do armazenamento da classificação setorial das ações, onde cada campo armazena o código da ação negociado na BM&F Bovespa e o setor de classificação da empresa.

tempo, esta base é divida em janelas igualmente espaçadas, assim como proposto em Tabak, Serra e Cajueiro (2010).

Os resultados apresentados no próximo capítulo foram calculados em janelas fixas de um ano de negociações espaçados mensalmente. A escolha do tamanho do janelamento é devido as variações das medidas físicas serem de maior intensidade (ruidosas) para períodos de jamelamento curto; e muito suavizadas para grandes períodos de janelamento. Então, a escolha deste tamanho de janelas é uma troca entre dados muito ruidosos e dados muito suavizados.

Tendo isso em mente, a base de dados foi dividida em janelas fixas de $t = 260$ negociações diárias, que corresponde a aproximadamente um ano. E $\delta t = 21$ dias, aproximadamente um mês, que corresponde ao período de deslocamento entre as janelas. Assim, para montar cada rede, foram tomados de 1 a 260 observações de preços de fechamento diário de cada ação, de 22 a 281, de 43 a 302, de 64 a 323, e assim por diante. A utilização de janelas fornece a possibilidade de estudos da evolução da rede em função do tempo.

Para a seleção das ações, foi analisada a intersecção dos ativos que compuseram o índice Bovespa (ver Seção 2.3.1) dentro do período de janelamento da rede. Por exemplo, se a rede observada foi do período de Dezembro de 2001 a Novembro de 2002 (ver Tabela 4.1), então foram observadas todas as ações de mesmo código de negociação que compuseram os índices Bovespa do último quadrimestre de 2001 e dos três quadrimestres de 2002, ou seja, as ações PETR4, USIM5 e VALE5. Se uma determinada ação estiver somente em um ou mais iBovespa e não estiver em outro iBovespa analisado, então esta ação não foi considerada, como as ações AMBV4, EMBR4, CRUZ3 e AMBV4.

Tabela 4.1: Exemplo de escolha das ações para a rede do período de Dezembro de 2001 a Novembro de 2002. A escolha das ações para esta rede é a intersecção das ações dos Ibovespas dentro do período de análise da rede.

3º Quad. 2001	1º Quad 2002	2º Quad 2002	3º Quad 2002	Ações Escolhidas
PETR4	PETR4	USIM5	PETR4	PETR4
USIM5	USIM5	PETR4	USIM5	USIM5
VALE5	VALE5	VALE5	VALE5	VALE5
AMBV4	EMBR4	CRUZ3	AMBV4	
CRUZ3		AMBV4		

Com isso, foram selecionadas as ações de maior negociabilidade na Bovespa para cada rede, havendo uma variação das ações entre as redes, diferentemente do proposto em Tabak, Serra e Cajueiro (2010) que as ações de todas as redes são as mesmas.

Para verificar a variação dos preços de fechamento das ações, foi utilizado uma equação logarítmica, chamada de retorno logarítmico, definido por:

$$r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1) \tag{4.1}$$

(FERGUSSON; PLATEN, 2006).

O retorno logarítmico mede a variação do preço de fechamento P_i da ação i no instante $\tau - 1$ variando para o instante τ. A utilização do log natural na Equação 4.1 é devido os preços poder ter uma distribuição log-normal, com isso é conveniente que a distribuição dos preços seja normal (FERGUSSON; PLATEN, 2006).

Assim, foram computados os coeficientes de correlações de todos as ações em cada janela, par a par, dado por:

$$p_{ij}^t = \frac{< r_i^t r_j^t > - < r_i^t >< r_j^t >}{\sqrt{(< (r_i^t)^2 > - < r_i^t >^2)(< (r_j^t)^2 > - < r_j^t >^2)}}. \tag{4.2}$$

onde r_i^t é um vetor contendo os retornos logarítmicos de todo o período de tempo t do janelamento.

Lembrando que o sinal $<>$ significa uma média temporal (RODGERS; NICEWANDER, 1988).

Então, modelando cada ação como um nó na rede, tem-se uma média de $N = 54$ ações (nós), lembrando que as ações de cada rede variam, assim como o número de ações; as arestas P que ligam pares de nós representam as correlações entre as ações, formando assim uma matriz de correlações entre ações $G = (N, P)$. Essa matriz descreve a estrutura da rede proposta.

Nota-se que a rede proposta é totalmente conectada, uma vez que foram analisadas as correlações de todas as ações entre elas, par a par, obtendo um total de $N(N-1)/2$ arestas na rede, sendo este valor o total de arestas possíveis.

Uma das hipóteses deste trabalho é que haja um agrupamento natural das ações do mesmo setor, uma vez que notícias sobre algum setor, por exemplo mineração ou financeiro, refletem em todas as ações deste setor, como Usiminas, Vale e Braskem, que são ações do setor de mineração, ou ITAU Unibanco, Banco Santander e Banco Bradesco, do setor financeiro. Espera-se que as ações do mesmo setor possuam valor de $w_{i,j}$ alto entre elas (TABAK; SERRA; CAJUEIRO, 2010).

Com o estudo da evolução da rede sobre o tempo, espera-se uma rede dinâmica, que mude as características em certos instantes de tempo, como em momentos de crise do mercado financeiro, onde a maioria das as ações tende a cair, assim como observado em Drozdz et al. (2000), Onnela et al. (2002).

4.3 ÁRVORE DE ESPALHAMENTO MÍNIMO

Antes de gerar a árvore de espalhamento mínimo (*MST*) da rede proposta, uma hipótese adicional sobre a topologia do espaço métrico é necessária. A hipótese é que um espaço útil para conectar as ações é um espaço ultra-métrico, isto é, um espaço no qual todas as distâncias são ultra-métricas e preenche os três axiomas de uma métrica euclidiana. Esta hipótese é motivada a posteriori pela constatação de que a taxonomia associada ao espaço ultra-métrico é significativa do ponto de vista econômico.

Com isso, uma métrica generalizada foi definida como a distância, utilizando uma função

apropriada da correlação coeficiente. A função escolhida foi:

$$d_{ij} = \sqrt{2(1 - p_{ij})}, \tag{4.3}$$

utilizado também em Bonanno et al. (2004), onde p_{ij} é a correlação entre as ações i e j.

Com esta escolha d_{ij} preenche os três axiomas de uma métrica euclidiana, que são:

1. $d_{ij} = 0$ se, e somente se, $i = j$;
2. $d_{ij} = d_{ji}$;
3. $d_{ij} \leq d_{ik} + d_{kj}$;

Onde d_{ij} é a distância da correlação da ação i e j. Equivalentemente para d_{ik} e d_{kj} para as ações i e k, k e j.

O primeiro axioma é válido porque $d_{ij} = 0$ se, e somente se, a correlação (ou o anti-correlação) for perfeita ($p_{ij} = 1$). O segundo axioma também é válido porque a matriz de coeficiente de correlação G e, portanto, a distância matriz D é simétrica, por definição, e o terceiro axioma foi verificada numericamente em todas as redes.

Assim, a matriz de distâncias D foi utilizada para determinar a *MST*. A árvore *MST* é atraente porque seleciona as ligações mais relevantes de cada ação da rede. Além disso, a árvore *MST* dá, de forma direta, a organização hierárquica das ações.

4.4 EXTRAÇÃO DE CARACTERÍSTICAS FÍSICAS DA REDE

As características físicas Densidade, Distribuição de Graus, Distribuição de Força de Vértices, Matriz de Distâncias de Caminhos, Coeficiente de Clusterização e Modularidade foram utilizadas para caracterizar a rede de mercado de ações.

Os experimentos conduzidos visam, entre outras coisas, as verificações das hipóteses consideradas nas próximas seções.

4.4.1 DENSIDADE

Para medir o quanto a rede proposta está conectada ou correlacionada entre os nós, foi calculada a densidade média ponderada (ver Seção 2.1.3.1).

$$K_{den} = \frac{\sum_{i,j} w_{i,j}}{N(N-1)/2}, \tag{4.4}$$

Onde $w_{i,j}$ é a correlação entre a ação i e j, N é o número de ações na rede.

De acordo com a Equação (4.4), quanto maior a densidade, maior a correlação entre as ações. Investir em momentos em que a densidade é alta, pode ser um risco grande, visto que os setores do mercado estão agrupados, e o risco de diversificação não é minimizado para formar portifólios eficientes (ver Seção 2.4.3). Além disso, em Drozdz et al. (2000), Onnela et al. (2002) foi mostrado que a correlação sobe significamente em momentos de crise econômica.

4.4.2 DISTRIBUIÇÃO DE GRAUS / FORÇA DE VÉRTICES

Na rede proposta, para calcular o grau k_i soma-se todos os pesos das arestas que incidem ao nó i (ver Seção 2.1.3.3), ou seja, foi calculada a força de cada nó (ver Seção 2.1.3.4).

Para tanto, é percorrido a matriz de correlações. Para cada ação na matriz, soma-se todos os campos (pesos $w_{i,j}$ das arestas) que estão correlacionadas a esta ação:

$$s_i = \sum_j w_{i,j} \tag{4.5}$$

Feito isso, calcula-se a distribuição de probabilidades dos graus ponderados de todas as arestas (ver Seção 2.1.3.3).

Espera-se, por hipótese, que a distribuição de graus siga uma lei de potência, assim como Kim et al. (2002a, 2002b), que concluíram que as ações do índice S&P500 em uma rede de correlações tem topologia livre de escalas.

Uma hipótese considerada aqui é que os *hubs* (ver Seção 2.1.2.3) sejam formados por ações de primeira linha; ou seja, ações de excelente reputação, e que apresentam grande liquidez e pro-

cura no mercado de ações, como a Petrobrás, VALE, entre outras.

4.4.3 MATRIX DE DISTÂNCIAS DE CAMINHOS

A matriz de distâncias de caminhos (ver Seção 2.1.3.5) $D(G)$, é construída a partir do algoritmo de Dijkstra (ver Seção 2.1.6) da matriz de pesos W (ver Seção 4.2).

Uma vez que uma hipótese é considerar que uma ação minimiza a distância de correlação com outra ação, a partir de outro caminho de comprimento maior. Sporns (2002), em seu trabalho de unidade neurais, verifica que a distância na rede não faz referência à distância física espacial. Isso pode ser um caminho para investigação, motivando o uso da Matriz de Distâncias de Caminhos.

Com a matriz $D(G)$ montada, é possível verificar a excentricidade (ecc) de cada nó. Espera-se que, quando a excentricidade de um nó for baixa, exista uma correlação maior do nó correspondente com outros nós. Para uma boa diversificação da carteira, pode ser melhor escolher nós com excentricidade elevada. Para estimar qual o melhor limiar da excentricidade para uma boa diversificação, pode-se considerar o Comprimento do Caminho Característico $\lambda(G)$, que é a média de $D(G)$.

Por outro lado, o diâmetro da rede proposta pode informar qual o comprimento de caminho máximo que se deve percorrer para alcançar qualquer nó da rede a partir de um nó qualquer. Este é o número que define o grau de separação entre as ações da rede (WATTS, 2003).

Outro ponto discutido, é que baixos valores de Comprimento do Caminho Característico e de Diâmetro pode significar um bom período para investir, com boa liquidez e alto volume de negociação; ou seja, longe de uma crise financeira. Esta hipótese dá-se pelo fato de que as distâncias das correlações entre as ações são maiores quando o Comprimento do Caminho Característico e o Diâmetro são baixos. Para computar o Diâmetro da rede foi verificado a maior excentricidade entre todos os nós, dado pela equação:

$$diametro(G) = max(ecc). \tag{4.6}$$

Já, o Comprimento do Caminho Característico é dado pela média da matriz $D(G)$.

4.4.4 COEFICIENTE DE CLUSTERIZAÇÃO

Para medir o coeficiente de agrupamento de um nó i foi utilizado o Coeficiente de Clusterização Ponderado (ver Seção 2.1.3.6), neste caso foi analisado qual equação melhor descreve os agrupamentos dos nós da rede proposta.

As equações de coeficiente de clusterização que foram utilizadas e analisadas são as propostas por Lopez-Fernandez, Robles e Barahona (2004), Barrat et al. (2004), Onnela et al. (2005) e Zhang e Horvath (2005) (ver Seção 2.1.3.6), dadas pelas Equações (4.7), (4.8), (4.9) e (4.11) respectivamente.

$$CC(i)_{LF} = \frac{\sum_{j \neq k \in N(i)} w_{jk}}{k_i(k_i-1)/2}, \tag{4.7}$$

onde w_{jk} é o peso da aresta entre a ação j e k; e k_i é o grau da nó i (LOPEZ-FERNANDEZ; ROBLES; BARAHONA, 2004).

$$CC(i)_B = \frac{1}{s_i(k_i-1)} \sum_{j,k} \frac{w_{i,j}+w_{i,k}}{2} a_{i,j} a_{i,k} a_{j,k}, \tag{4.8}$$

onde $a_{i,j} = 1$ se o peso w_{ij} for diferente de 0, e $a_{i,j} = 0$ caso contrário; s_i é a força do vértice i (BARRAT et al., 2004).

$$CC(i)_O = \frac{2\sum_{j,k}(\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk})^{1/3}}{k_i(k_i-1)}, \tag{4.9}$$

onde

$$\hat{w}_{ij} = \frac{w_{ij}}{max(w)}, \tag{4.10}$$

em que os pesos das arestas $\hat{w}_{ij}$ são normalizados pelo maior peso w de todas as arestas da rede (ONNELA et al., 2005).

$$CC(i)_Z = \frac{\sum_{j,k} \hat{w}_{ij}\hat{w}_{jk}\hat{w}_{ik}}{(\sum_k \hat{w}_{ik})^2 - \sum_k \hat{w}_{ik}^2}, \tag{4.11}$$

onde os pesos também são normalizados por $max(w)$, e o denominador ponderado limita superiormente o numerador, garantindo que o $CC(i)_Z \in [0,1]$ (ZHANG; HORVATH, 2005).

Com isso, é possível verificar o quanto um nó i está agrupado entre seus vizinhos.

Uma das hipóteses consideradas é quando o coeficiente de clusterização CC_i for alto em relação a média, pode ser indício de que o nó i seja uma ação anti-correlacionada e boa para diversificação na carteira, pois os vizinhos de i estão mais agrupados entre si do que em relação ao nó i.

Além disso, o CC médio é calculado a partir da seguinte equação:

$$CC = \frac{1}{N} \sum_{i \in N} CC(i), \tag{4.12}$$

onde N é o número de nós da rede, e $CC(i)$ é o coeficiente de clusterização do nó i.

4.5 MEDIDA DE QUALIDADE DE CLUSTERIZAÇÃO DA REDE

Para medir a qualidade dos agrupamentos da rede proposta, foi utilizada a Modularidade Ponderada (ver Seção 2.1.4), formalmente dada por:

$$Q_s = \frac{1}{W} \sum_{ij} [w_{ij} - \frac{s_i s_j}{W}] * \delta(c_i, c_j), \tag{4.13}$$

onde $W = \sum_{i,j} w_{ij}$.

Para tanto, foi necessário uma base supervisionada. Na modelagem proposta as ações foram classificadas setorialmente, utilizada para a função $\delta(c_i, c_j)$, onde $\delta = 1$ se as ações i e j pertencerem ao mesmo setor, $\delta = 0$ caso contrário.

A Modularidade foi calculada após computar a Árvore Espalhada Mínima da rede (ver Seção 2.1.5). Com isso foi verificar se a árvore *MST* agrupa as ações de acordo com as maiores correlações por setores.

5 RESULTADOS EXPERIMENTAIS

A base de dados utilizada são as cotações históricas da BOVESPA (Seção 4.1.1) do período de 03 de Dezembro de 2001 a 30 de Março 2012, totalizando 10 anos e 4 meses, com as seguintes informações utilizadas: data do pregão, código de negociação do papel e preço do último negócio do papel-mercado no pregão, em reais. É importante notar que essas informações correspondem ao preço de fechamento diário de cada ação do indíce BOVESPA (iBovespa).

Então, com o método de janelamento as matrizes de coeficientes de correlação são geradas. A Figura 5.1 ilustra esse método, onde as Matrizes G^t com período de um ano de negociações são computados através da correlação cruzada entre todos os pares de nós da iBovespa deste período. $t = 0, 1, 2, 3, 4, ..., 112$ é representado por um número, que o número da rede. Foram geradas um total de 113 redes com intervalo de um mês entre elas. A escolha desse intervalo é para garantir certo suavizamento nas variações das características físicas da rede. Outros tamanhos de janelas e intervalos podem ser analisados em um trabalho futuro.

A Tabela 5.1 mostra um exemplo da Matriz de Correlações G^{112} da data de 20110401 a 20120330. Neste capítulo as datas serão exibidas em formato AAAAMMDD (ano mes dia). Cada tupla exibe o coeficiente de correlação entre os pares de ações. Note que a diagonal principal é composta por valores iguais a 1.0, que é a correlação perfeita, pois se trata de ações iguais.

Assim, considerando vários instantes de tempo, pode-se verificar as mudanças das características físicas temporalmente, como serão mostrados nas próximas seções.

5.1 CORRELAÇÃO CRUZADA

De acordo com a Seção 2.1.7, a correlação cruzada avalia a correlação na variação dos preços de pares de ações ponto a ponto. Com isso, nesta seção serão mostrados os comportamentos da

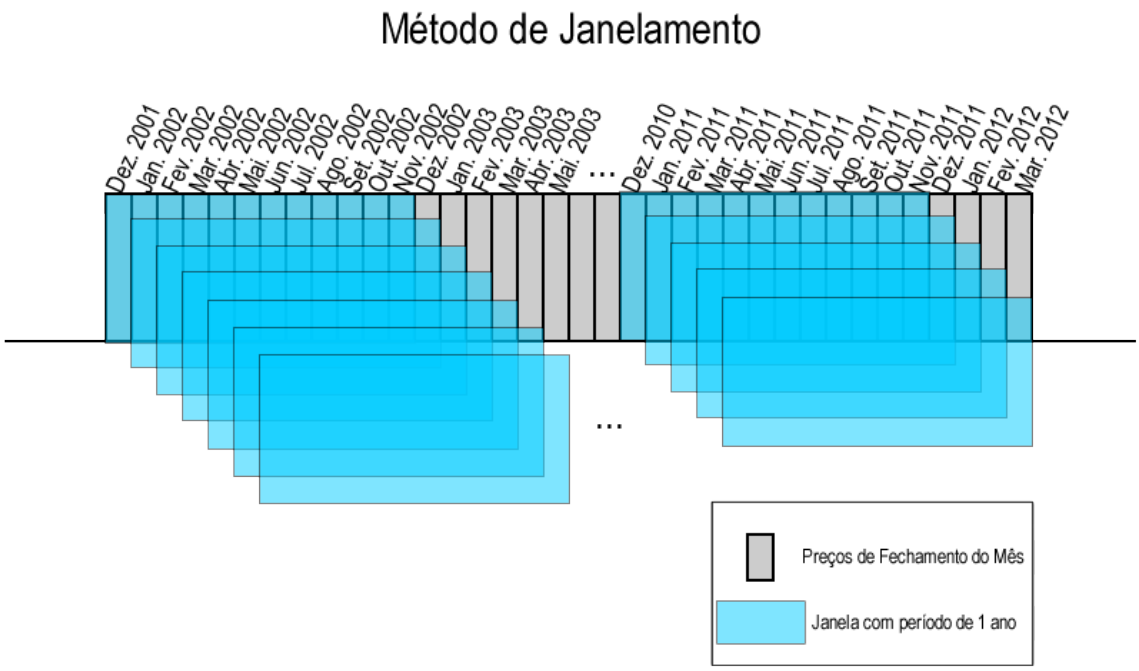

Figura 5.1: Exemplo do método de janelamento, onde cada janela tem um período de um ano de negociações diárias das ações, com intervalo de um mês entre cada janela.

correlação cruzada de diferentes pares de ações.

Neste experimento, será utilizada a rede mais recente analisada neste trabalho, da data de 20110401 a 20120330, porém, o comportamento da correlação cruzada para sequências de preços semelhantes entre pares de ações exibe um resultado semelhante também para qualquer data.

As ações analisadas são AMBV4, CRUZ3, PETR3, PETR4 e VALE5. As variações dos preços de fechamento dessas ações podem ser observados na Figura 5.2.

Com isso, os resultados das correlações cruzadas são mostrados na matriz da Tabela 5.1.

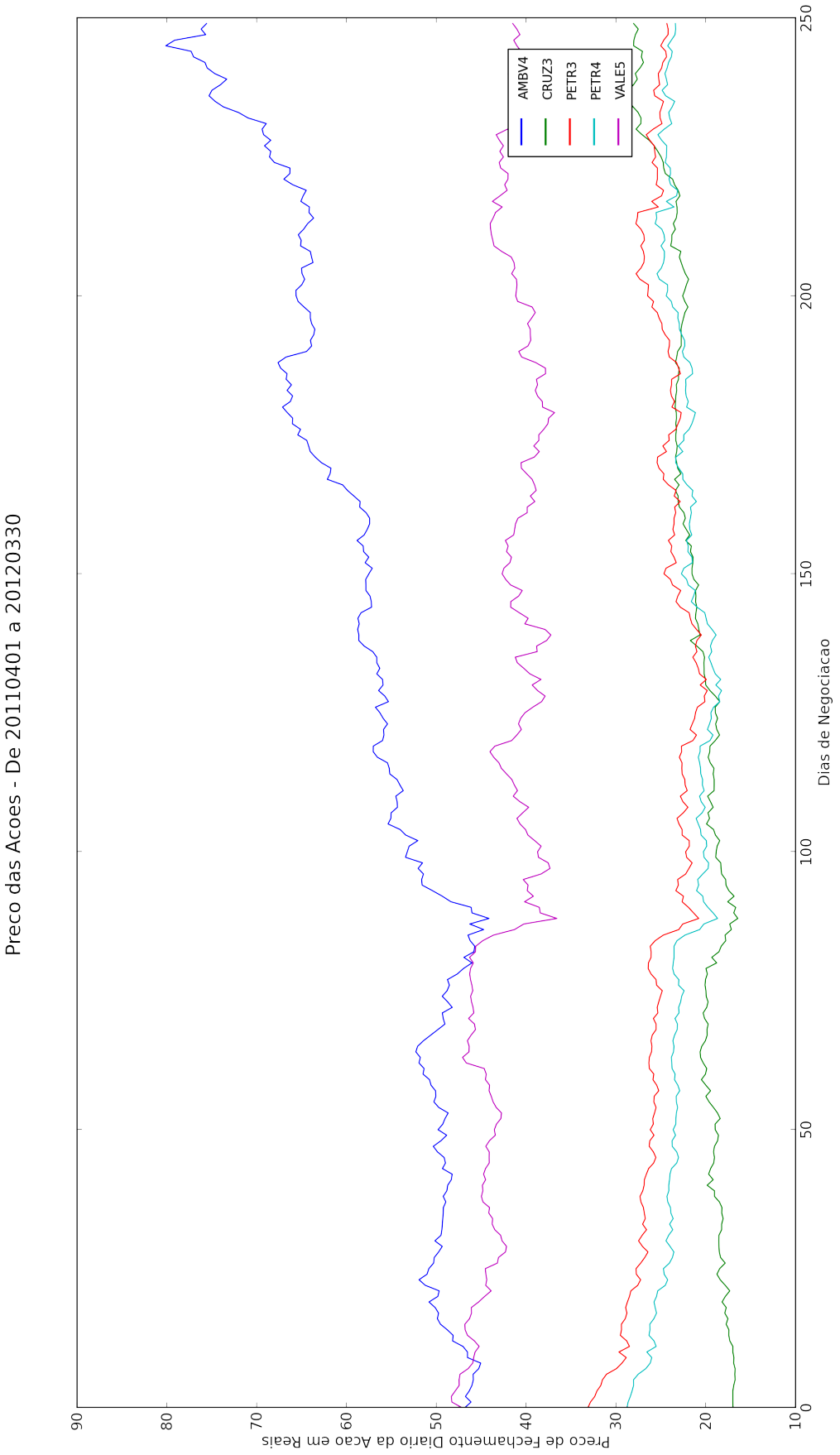

Figura 5.2: Preços de fechamento em reais das ações AMBV4, CRUZ3, PETR3, PETR4 e VALE5 da data de 20110401 a 20120330.

Tabela 5.1: Matriz de correlações cruzada entre pares de ações de acordo com a Equação (4.2).

Correlação Cruzada					
	AMBV4	CRUZ3	PETR3	PETR4	VALE5
AMBV4	1.0	0.5	0.27	0.27	0.29
CRUZ3	0.5	1.0	0.31	0.32	0.34
PETR3	0.27	0.31	1.0	0.98	0.69
PETR4	0.27	0.32	0.98	1.0	0.69
VALE5	0.29	0.34	0.69	0.69	1.0

Observando a Figura 5.2 e a Tabela 5.1, é possível notar que as ações PETR3 e PETR4, que são ações da mesma empresa Petrobrás, possuem alta correlação entre si, igual a 0.98. Como essas duas ações pertencem à mesma empresa, a correlação delas é alta, uma vez que a empresa está bem, as ações sobem. Se a empresa estiver passando por crise, ambas as ações tendem a cair.

A ação VALE5, que é outra ação bastante negociada na BM&F Bovespa, assim como a PETR3 e PETR4, obteve alta correlação com as ações da empresa Petrobrás, isso se dá porque possivelmente essas ações controlavam ou intermediavam as variações de preço do mercado neste período.

Por sua vez, as ações AMBV4 e CRUZ3 obtiveram valores baixos de correlação entre as ações PETR3, PETR4 e VALE5. As melhores ações para otimizar uma carteira, e assim evitar risco específico de uma ação, são ações com correlação baixa (MARKOWITZ, 1952). Se uma ação de sua carteira cair em seu preço por qualquer motivo específico, a outra ação de sua carteira anti-correlacionada poderá evitar maiores prejuízos.

Outra observação importante é que as ações AMBV4 e CRUZ3 são do mesmo setor de classificação: consumo não cíclico, e essas ações obtiveram alta correlação entre si, de acordo com a Tabela 5.1. Uma melhor investigação sobre a correlação entre ações do mesmo setor está na Seção 5.7.2.

Nas próximas seções serão mostrados os resultados das medidas físicas da rede.

5.2 DENSIDADE MÉDIA PONDERADA

Como descrito na Seção 2.1.3.1, esta característica permite analisar a correlação média entre as ações consideradas. Com isso, pode-se verificar a evolução desta característica com o passar do tempo.

A Figura 5.3 mostra o gráfico da densidade média ponderada pelo tempo. Cada ponto (x, y) do gráfico da Figura 5.3 mostra uma rede x com sua primeira data de negociação em formato AAAAMMDD (ano mês dia), e sua correspondente densidade média ponderada y.

Podemos notar que, nas redes entre o período de 20071203 até 20081001, houve um aumento significativo da correlação média. Neste período aconteceu a crise economia mundial, tendo seu início pela falência do banco de investimento estadunidense Lehman Brothers. Em tempos de crise, todas as ações tendem a desvalorizar, por esse motivo, as correlações entre si aumentam.

5.3 DISTRIBUIÇÃO DE PESOS

Segundo a Seção 2.1.3.2, a distribuição de pesos é dada pelo histograma dos pesos das ligações entre os nós de uma rede. Ela permite uma análise abrangente de como os nós estão conectados entre si.

Porém, como são 112 redes complexas na modelagem proposta, serão mostrados somente algumas redes específicas, totalizando 12. Essas redes representam instantes de mínimo local, forte subida, máximo local, forte descida, mínimo local, pouca variação, mínimo local, forte subida, máximo global, forte descida, mínimo local e mínimo global das variações da densidade média ponderada respectivamente de acordo com a Figura 5.4. Essas redes específicas foram numeradas de 1 a 12.

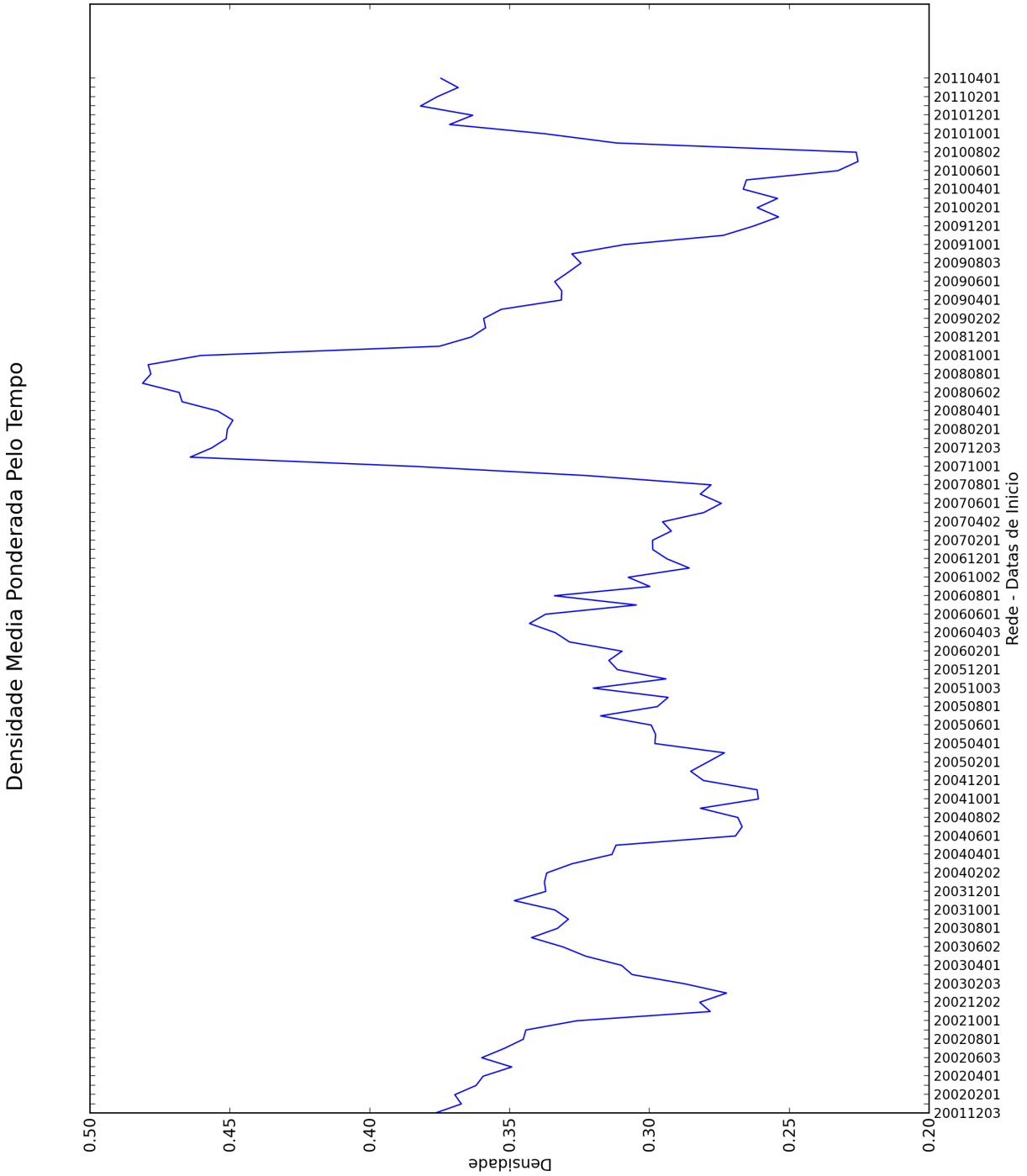

Figura 5.3: Densidade ponderada média de cada rede complexa com o passar do tempo.

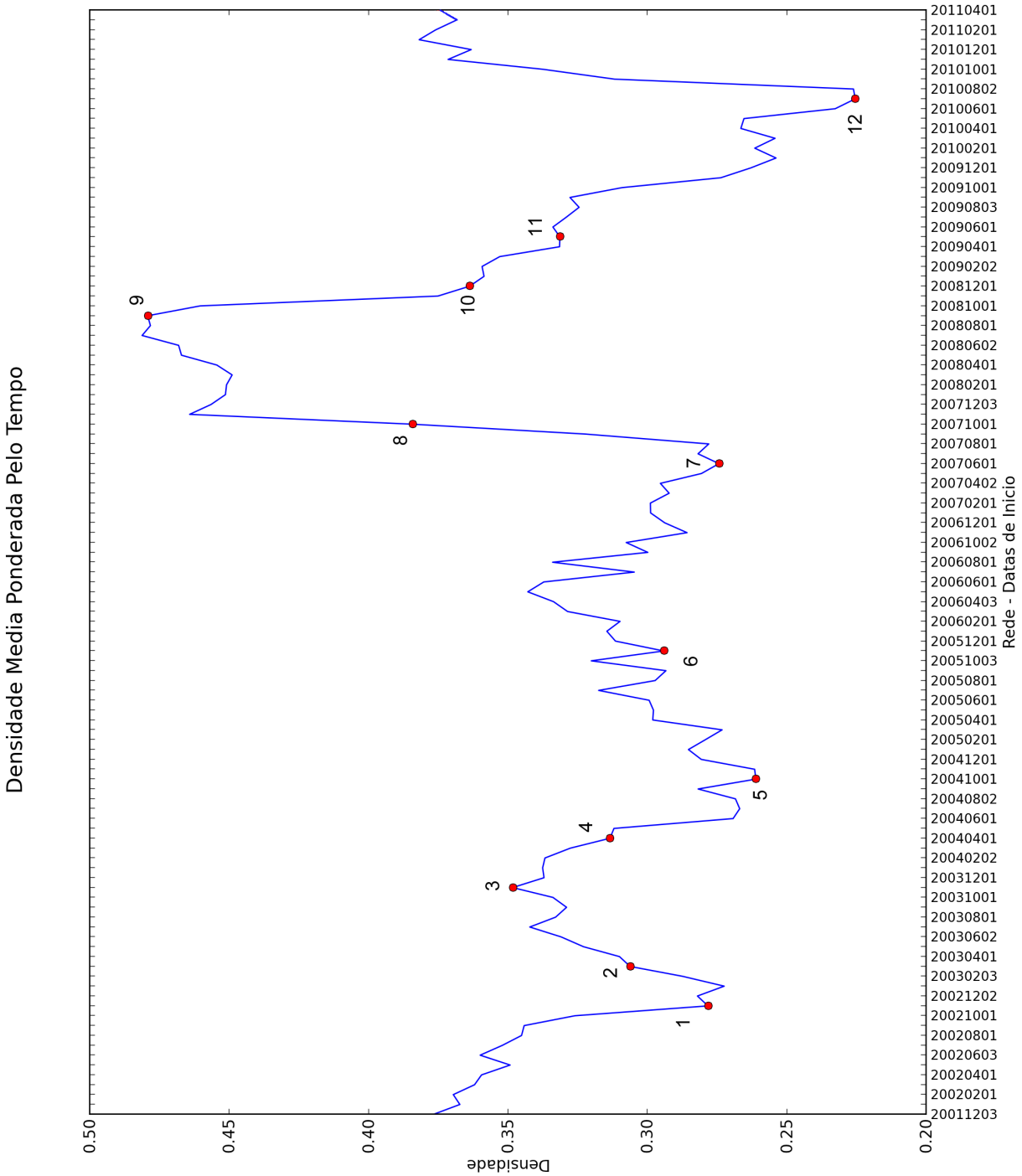

Figura 5.4: Densidade ponderada média de cada rede complexa pelo tempo. Os pontos em destaque representam as redes específicas escolhidas representando diversos instantes da densidade média ponderada.

Tabela 5.2: Tabela das redes específicas.

Rede	Primeiro Pregão	Último Pregão	Densidade Média Ponderada
1	20021101	20031031	0.2734
2	20030305	20040227	0.3053
3	20031103	20041029	0.3472
4	20040401	20050331	0.3121
5	20041001	20050930	0.2542
6	20051101	20061031	0.2932
7	20070601	20080530	0.2716
8	20071001	20080930	0.384
9	20080901	20090831	0.4791
10	20081201	20091130	0.3636
11	20090504	20100430	0.3312
12	20100701	20110630	0.2225

A Tabela 5.2 mostra especificações das redes escolhidas, como a primeira e a última data de negociação e a densidade média ponderada desta rede.

As Figuras 5.5 e 5.6 mostram as distribuições de probabilidades de pesos nas redes específicas de ações. Note que as probabilidades se distribuem nos pesos entre os valores de coeficiente de correlação de 0.0 a 0.7. Ou seja, as ações são mais correlacionadas entre si (valores positivos), do que anti-correlacionadas. Com isso, pode-se dizer que um investimento em ações quando o mercado está a favor à valorização é menos arriscado do que investir em ações anti-correlacionadas contra a tendência de baixa do mercado. Já que as ações são mais correlacionadas e estatisticamente uma perda é mais difícil, se a tendência de alta se concretizar.

Outro ponto notável é que o pico da distribuição de pesos está em torno de 0.25 a 0.5; ou seja, a maioria dos coeficientes de correlações entre pares de nós estão bem correlacionados. Com isso é possível estimar a formação de agrupamentos, que é verificado na Seção 5.7.2.

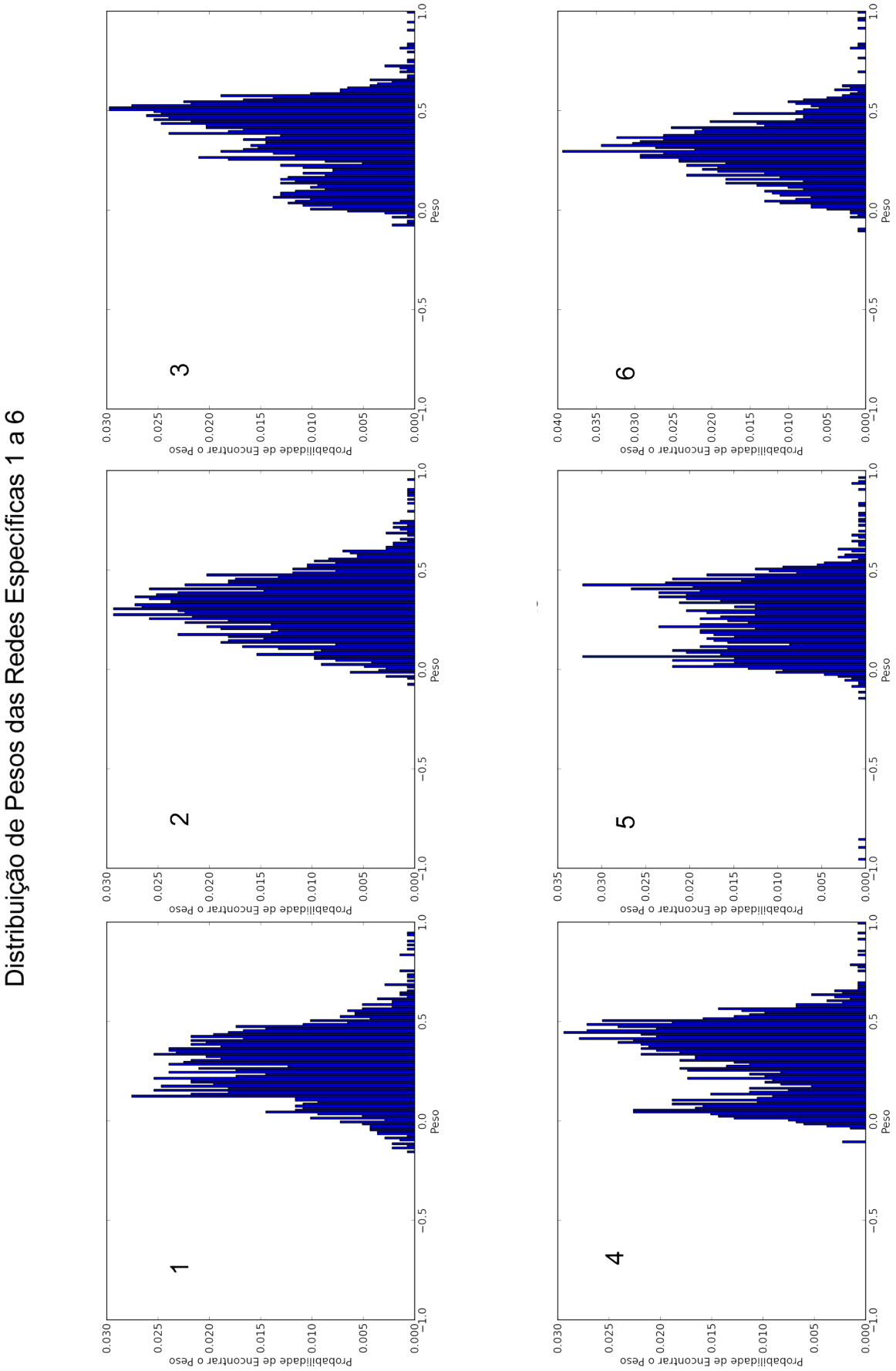

Figura 5.5: Histograma de pesos das redes complexas específicas de 1 a 6.

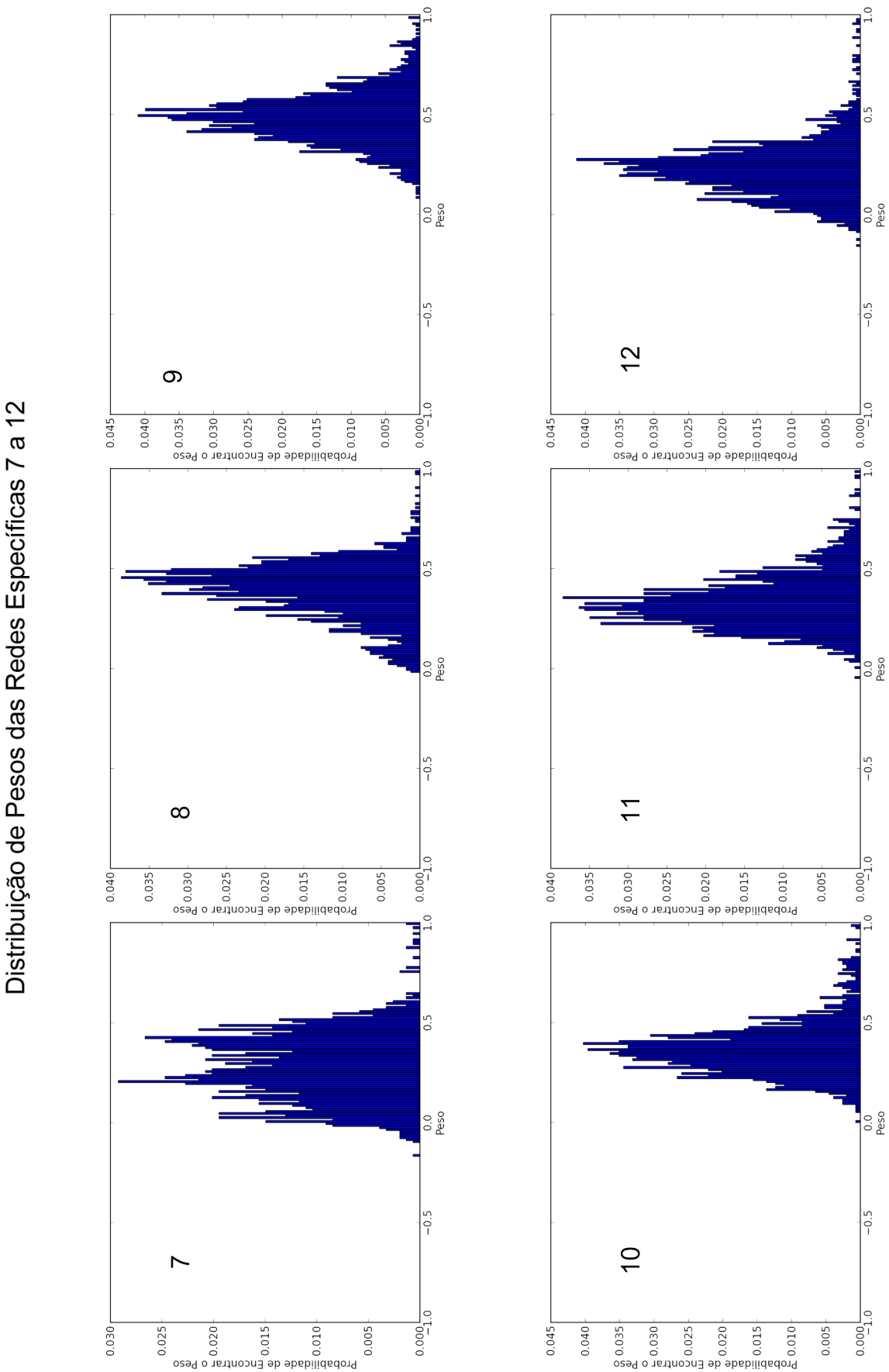

Figura 5.6: Histograma de pesos das redes complexas específicas de 7 a 12.

5.3.1 ENTROPIA DA DISTRIBUIÇÃO DE PESOS

De acordo com a Seção 2.1.3.2, uma maneira de calcular a dispersão de uma distribuição é calcular sua entropia. Uma entropia baixa significa concentração de informação, ou seja, muitas arestas com pesos semelhantes. No caso contrário, se a entropia da distribuição for alta, isso significa uma distribuição de pesos mais heterogênea, significando maior dispersão de probabilidades na distribuição (SHANNON; WEAVER, 1948).

Hipoteticamente, a distribuição de pesos com uma entropia elevada é melhor do ponto de vista econômico. Pois neste caso, as correlações entre as ações são mais diversificadas; ou seja, a rede terá coeficientes de correlações baixo, médio e alto quase igualmente (entre 0.25 a 0.6, de acordo com a distribuição de pesos das Figuras 5.5 e 5.6). Essa diversificação das correlações parece ser bom para os agrupamentos entre as ações e também para a otimização do portifólio. Pois se na rede, a entropia de distribuição de pesos foi baixa, então haverá predominância de correlações médias (0.4, ver Figuras 5.5 e 5.6), que provavelmente sejam ruídos para a rede, pois as ações não serão nem tão correlacionadas e nem tão anti-correlacionadas entre si. Ou ainda, poderá haver a predominância de correlações alta (0.5), que é uma característica de crise econômica, onde as correlações entre as ações são altas entre si.

A Figura 5.7 mostra o gráfico da variação de entropias das distribuições de pesos, juntamente com o gráfico das Densidades médias das 112 redes propostas neste trabalho. Esta figura mostra que no período da crise econômica mundial de 2008 a entropia da distribuição de pesos é baixa.

Quando a Entropia de Distribuição de Pesos é baixa, tem-se que essa distribuição é mais concentrada; ou seja, arestas com pesos semelhantes. Sabendo disso, pode-se dizer que investir em momentos em que a Entropia de Distribuição de Pesos é baixa é mais arriscado, pois se os pesos forem semelhantes e de alta correlação, então o investimento seguirá as variações de mercado, pois não haverá ações anti-correlacionadas para reduzir o risco específico de outra ação de acordo com a Teoria de Markowitz (1952).

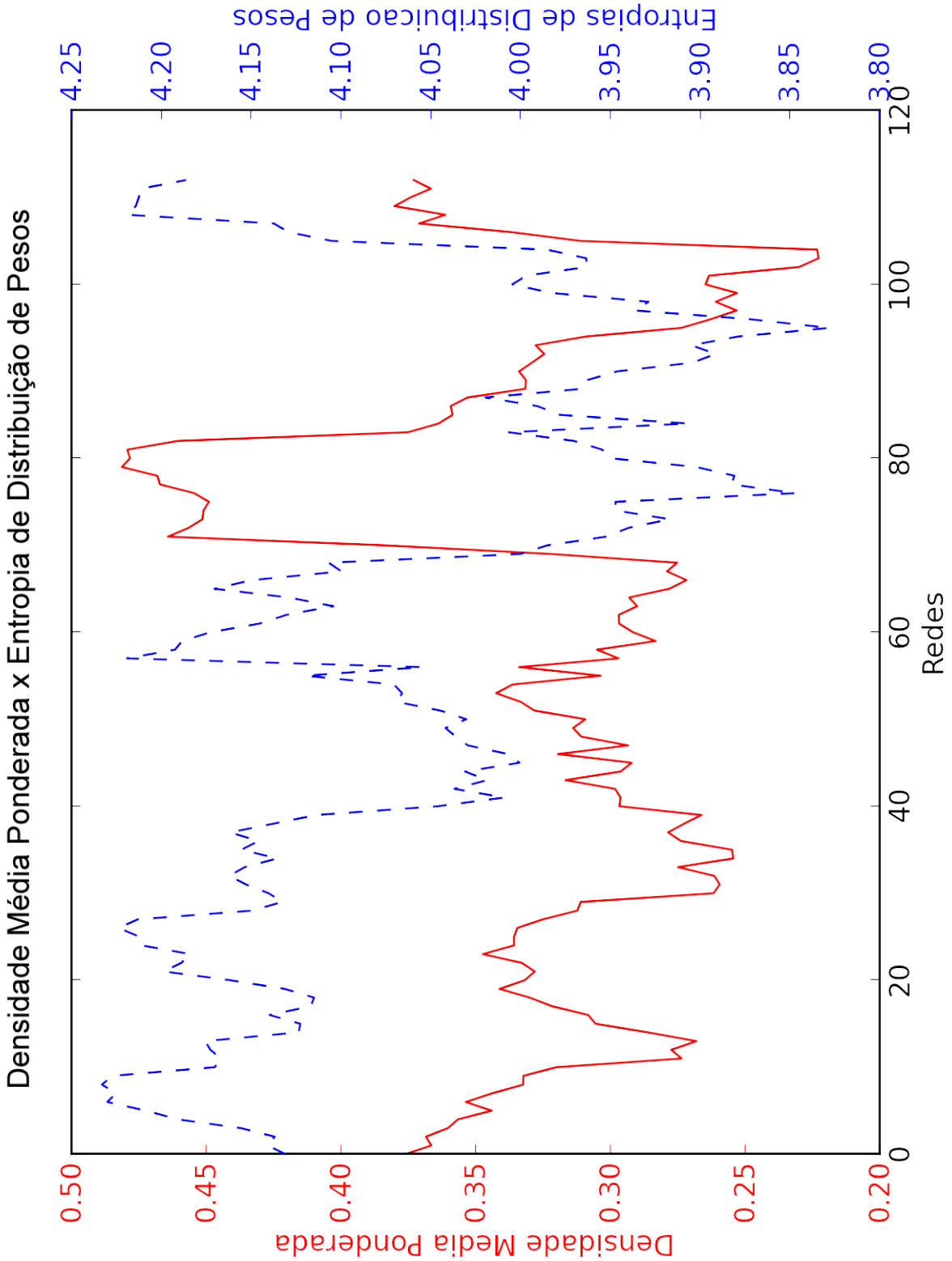

Figura 5.7: Densidade média ponderada (sinal em vermelho contínuo seguindo a escala a esquerda) e Entropia de distribuição de pesos (sinal em azul tracejado seguindo a escala a direita) pelo tempo.

5.4 DISTRIBUIÇÃO DE GRAUS E FORÇA DOS VÉRTICES

De acordo com a Seção 2.1.2.3, a principal característica de uma rede livre de escalas é que a distribuição de graus segue uma distribuição de lei de potências.

Porém, como as redes propostas são totalmente conectadas entre os nós, todos os nós terão grau $N-1$, sendo N o número de ações da rede.

Então, para o modelo proposto, é utilizado a Força do Vértice (Seção 2.1.3.4), que é a soma dos pesos das arestas incidentes ao nó estudado.

Com isso são obtidas a distribuição de forças das redes específicas das Figuras 5.8 e 5.9. Cada ponto (x,y) dos gráficos das figuras mostra a quantidade de nós y com a força do vértice discretizada x.

Nas Figuras 5.8 e 5.9 é possível notar que a distribuição não segue uma lei de potência. Uma explicação pode ser devido as arestas não-correlacionadas (peso 0.3 a 0.4) poderem ser ruído para a rede.

5.5 COEFICIENTE DE CLUSTERIZAÇÃO

Nesta seção, são apresentados os resultados obtidos através dos experimentos de quatro coeficientes de clusterização (*CC*) ponderado (ver Seção 4.4.4).

Neste experimento, é, primeiramente, observado o comportamento dos diferentes *CC* ponderados encontrados na literatura (LOPEZ-FERNANDEZ; ROBLES; BARAHONA, 2004; BARRAT et al., 2004; ONNELA et al., 2005; ZHANG; HORVATH, 2005). Essa comparação poderá servir como referência para outros trabalhos futuros.

Uma importante observação é que os coeficientes de correlações serão utilizados com seu valor absoluto; ou seja, se a correlação for -0.1, então este será considerado como 0.1.

O primeiro *CC* ponderado é uma transformação direta do cálculo de *CC* binário, e foi também utilizado por Lopez-Fernandez, Robles e Barahona (2004).

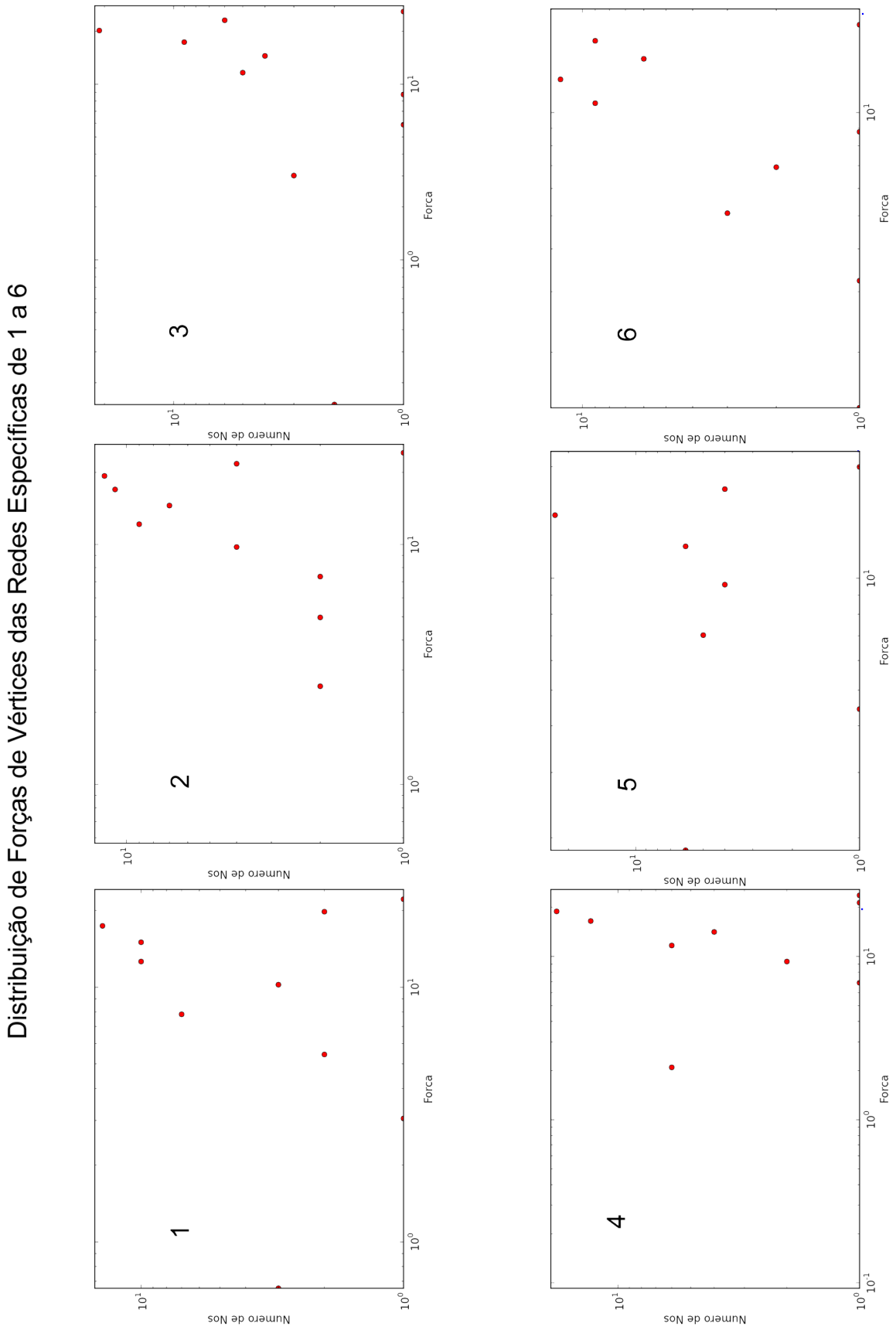

Figura 5.8: Distribuição de forças de vértices discretizado das redes específicas de 1 ao 6.

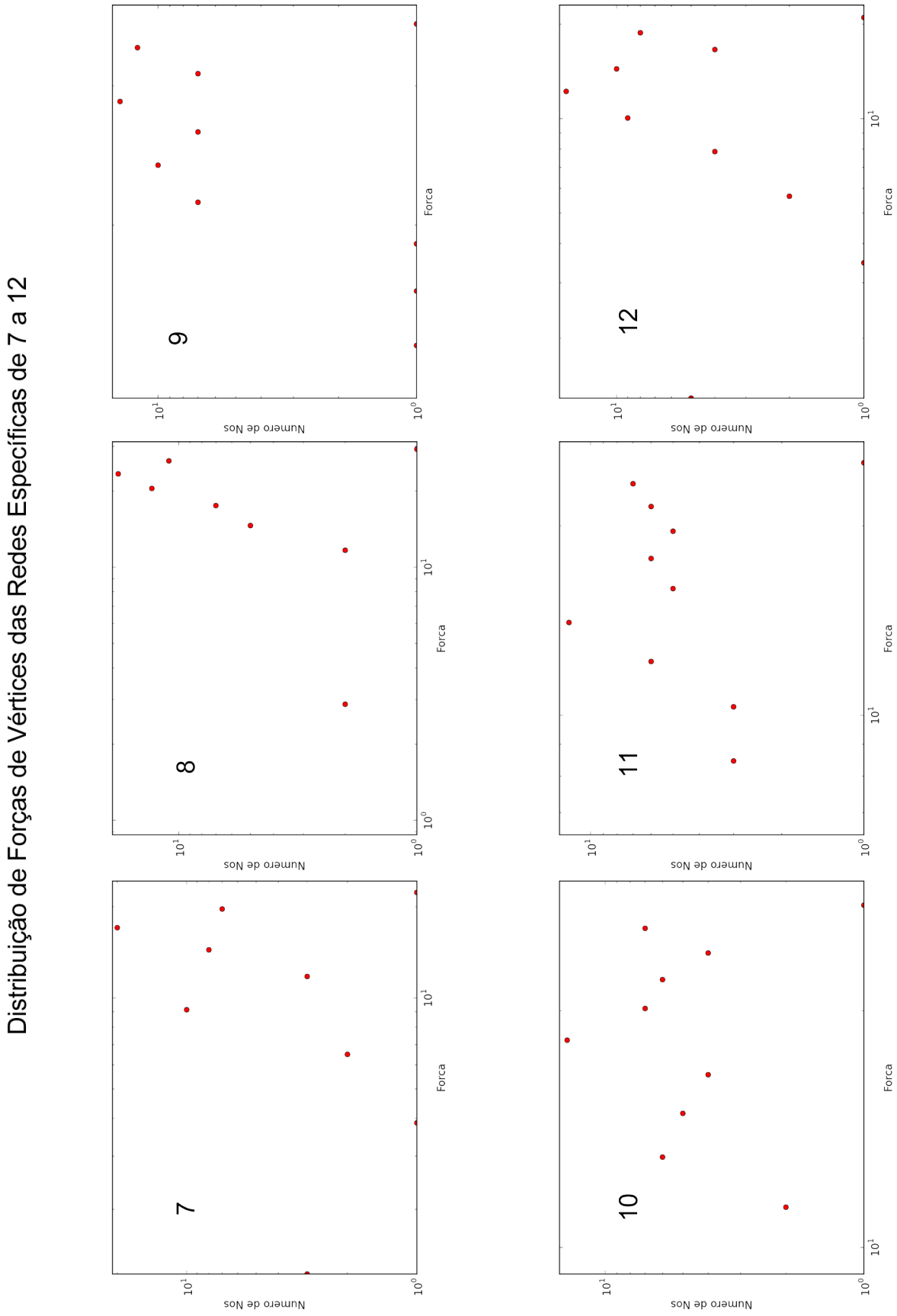

Figura 5.9: Distribuição de forças de vértices discretizado das redes específicas de 7 ao 12.

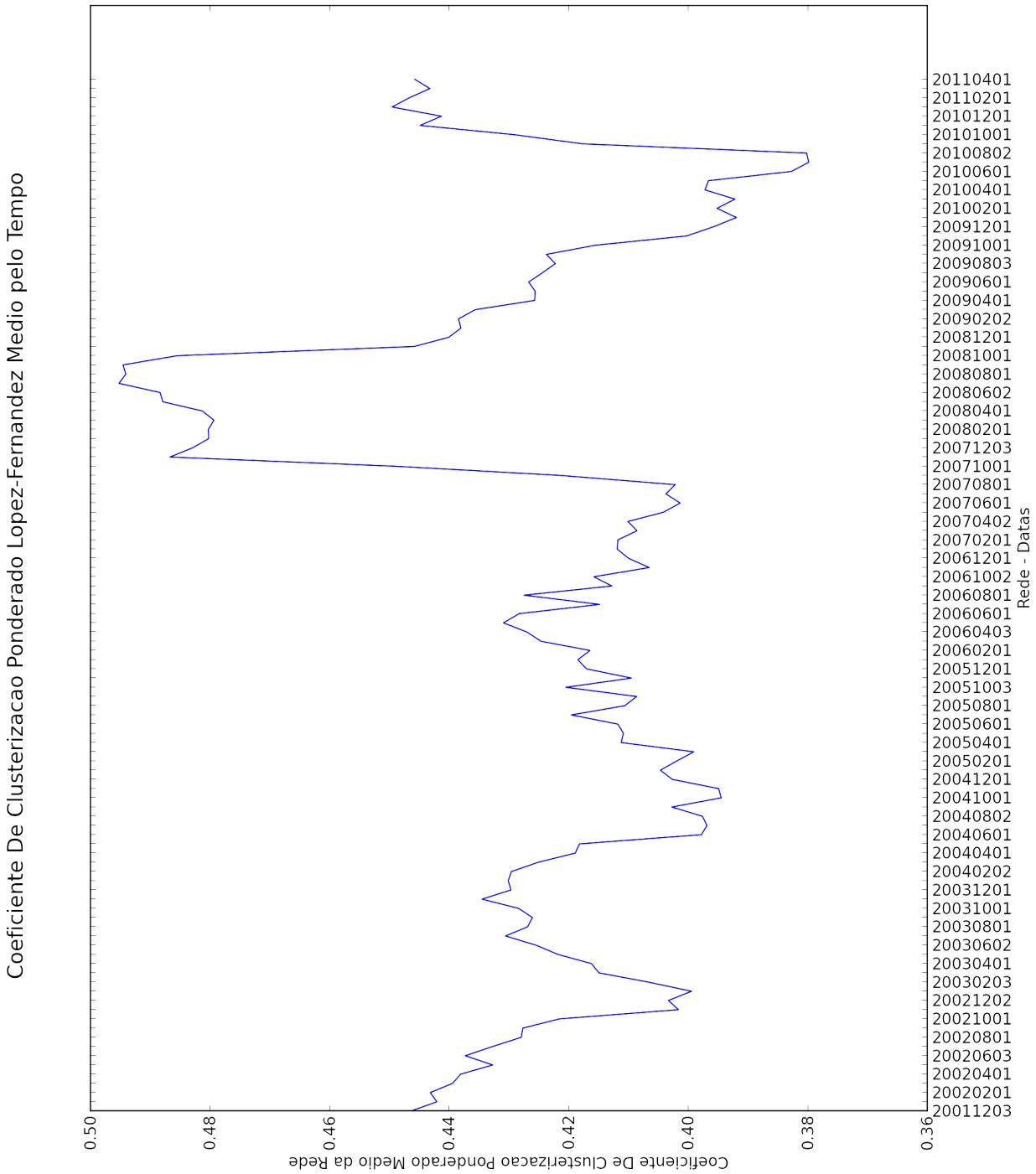

Figura 5.10: Coeficiente de clusterização ponderado médio proposto por (LOPEZ-FERNANDEZ; ROBLES; BARAHONA, 2004) pelo Tempo.

A Figura 5.10 mostra o CC_{LF} médio pelo tempo. É possível notar que este CC_{LF} obteve o mesmo comportamento de variações da densidade ponderada média (Seção 5.2) pelo tempo. Além disso, os valores entre essas duas características físicas ficaram aproximados, como pode ser visto na Figura 5.11.

Uma explicação para este comportamento é demonstrado na Comparação (5.1) a seguir, que se o peso for uniforme ($w_{jk} = w_{ij} = w$) na rede, o CC_{LF} médio é igual a Densidade média ponderada. Além disso, a distribuição de graus uniforme da rede (ver Seção 5.4) facilita que a média do CC_{LF} e a média da densidade sejam aproximados, pois todos os nós têm o mesmo número de conexões.

$$\begin{aligned}
\frac{\sum_{i=0}^{n}}{n}\overbrace{\frac{\sum_{k\neq i}^{n}\sum_{j\neq k,i}^{n} w_{jk}}{k_i(k_i-1)}}^{C(i)_{LF}} &\approx \overbrace{\frac{\sum_{i=0}^{n}\sum_{j\neq i}^{n} w_{ij}}{n(n-1)}}^{K_{den}} \\
\frac{\sum_{i=0}^{n}}{n}\frac{\sum_{k\neq i}^{n}\sum_{j\neq k,i}^{n} w_{jk}}{(n-1)(n-2)} &\approx \frac{\sum_{i=0}^{n}\sum_{j\neq i}^{n} w_{ij}}{n(n-1)} \\
\frac{n}{n}\frac{(n-1)(n-2)w_{jk}}{(n-1)(n-2)} &\approx \frac{n(n-1)w_{ij}}{n(n-1)} \\
w_{jk} &\approx w_{ij}
\end{aligned} \tag{5.1}$$

Por sua vez, o CC_B proposto em Barrat et al. (2004) (Seção 2.1.3.6) dará sempre igual à 1.0, pois como todos os nós estão conectados entre si, $a_{i,j}a_{i,k}a_{j,k} = 1$. O CC_B não leva em consideração as conexões entre a vizinhança de i, por isso para uma rede totalmente conectada o $C(i)_B$ ficará sempre no limite superior, como pode ser visto no Desenvolvimento (5.2).

$$\begin{aligned}
C(i)_B &= \frac{1}{s_i(k_i-1)}\sum_{j,k}\frac{w_{i,j}+w_{i,k}}{2}a_{i,j}a_{i,k}a_{j,k} \\
C(i)_B &= \frac{1}{s_i(k_i-1)}\sum_{j,k}\frac{w_{i,j}+w_{i,k}}{2} * 1 \\
C(i)_B &= \frac{1}{s_i(k_i-1)}(k_i-1)\frac{2*s_i}{2} \\
C(i)_B &= 1.0
\end{aligned} \tag{5.2}$$

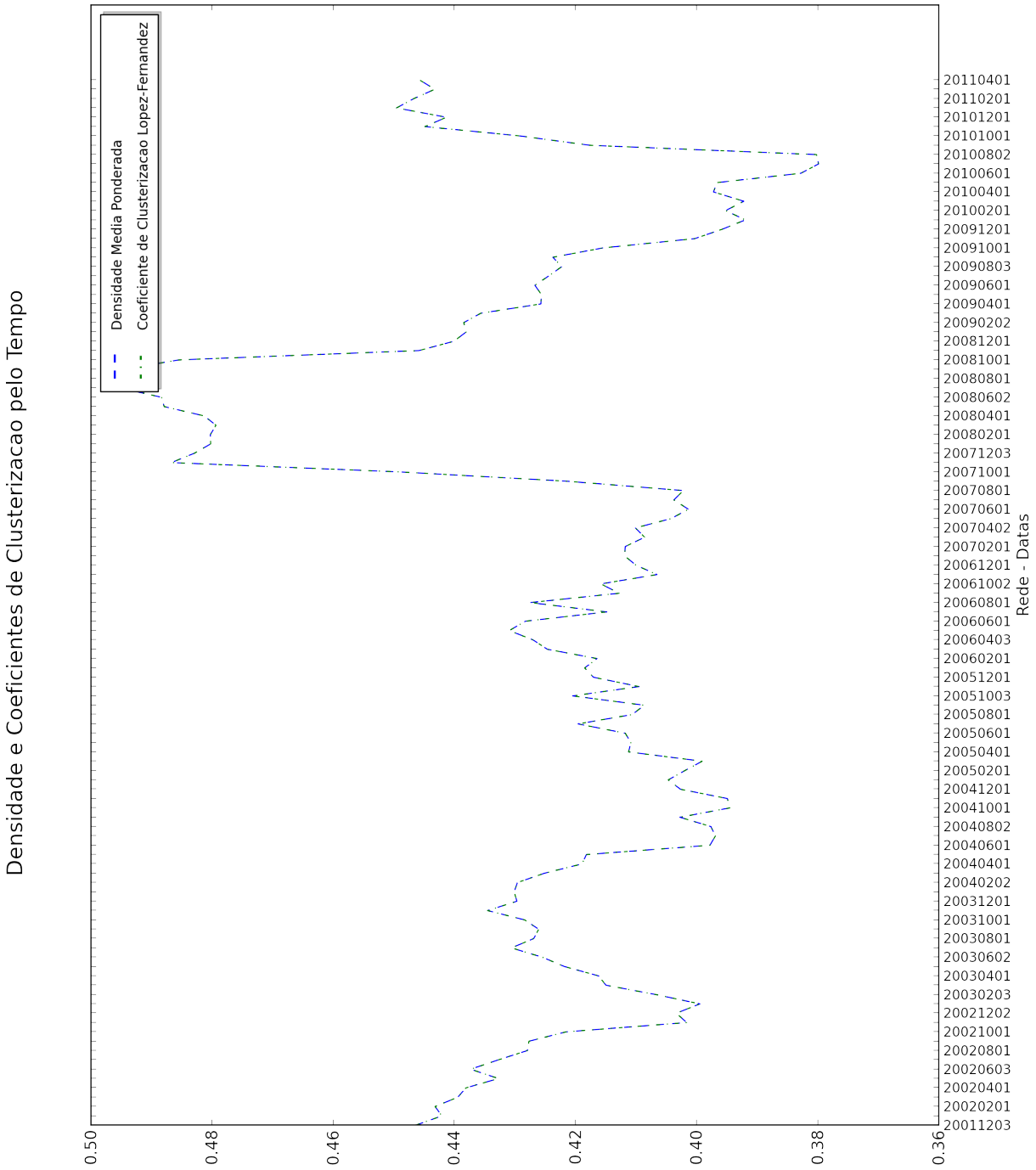

Figura 5.11: Densidade ponderada média e coeficiente de clusterização ponderado médio proposto por Lopez-Fernandez, Robles e Barahona (2004) pelo Tempo.

Agora, os $CC(i)_O$ e $CC(i)_Z$ proposto por Onnela et al. (2005), Zhang e Horvath (2005) respectivamente, consideram as conexões entre as vizinhas de *i* (Seção 2.1.3.6). Além disso, ambos os CC_O e CC_Z normalizam os pesos das conexões pelo peso máximo da rede.

Como pode-se notar na Figura 5.12, o comportamento de variações de CC_O e CC_Z médios são similares, havendo um deslocamento entre eles. Isso se dá devido ao limite superior do $CC(i)_Z$ ser também ponderado, ao contrário do $CC(i)_O$, tem seu limite binário. Em $CC(i)_O$ reflete o quão forte é o peso dos triângulos formados por *i* comparado com o máximo da rede.

Assim, pode-se dizer que o $CC(i)_Z$ é o mais apropriado para a modelagem proposta por ser totalmente ponderado, além de considerar as ligações entre as vizinhas do nó *i* normalizados pela variância das ligações de *i*.

5.6 MATRIZ DE DISTÂNCIAS DE CAMINHO

A matriz de distâncias de caminho é uma matriz contendo os menores caminhos entre todos os pares de nós da rede, como descrito na Seção 2.1.3.5. Para isso, foi utilizado o algoritmo *Dijkstra* (ver Seção 2.1.6) $|V|$ vezes, sendo $|V|$ o número de vértices na rede, para cada rede. Neste experimento são considerados os coeficientes de correlação em seu valor absoluto, pois assim evita-se ciclos negativos. Além disso, os coeficientes negativos são próximos a zero, como pode ser visto nas distribuições de pesos das redes (ver Seção 5.3.1).

Com a Matriz de distâncias de caminho, nota-se que existem algumas ações que diminuem as distâncias entre pares de ações.

Serão exibidos os caminhos de correlações de ações minimizados das redes específicas 5 e 11 (ver Tabela 5.2), que são redes com entropia de distribuição de pesos moderado (ver Seção 5.3.1) e Densidade Média baixa (ver Seção 5.2) Além dos caminhos da rede de data mais recente estudada neste trabalho, de 20110401 a 20120330.

A Tabela 5.3 mostra as ações que minimizaram os caminhos na matriz de distâncias de caminho e a quantidade de caminhos minimizados por essa ação, ordenados pela força do vértice da ação para as rede descritas acima. Com isso, é possível notar que algumas ações que minimizam

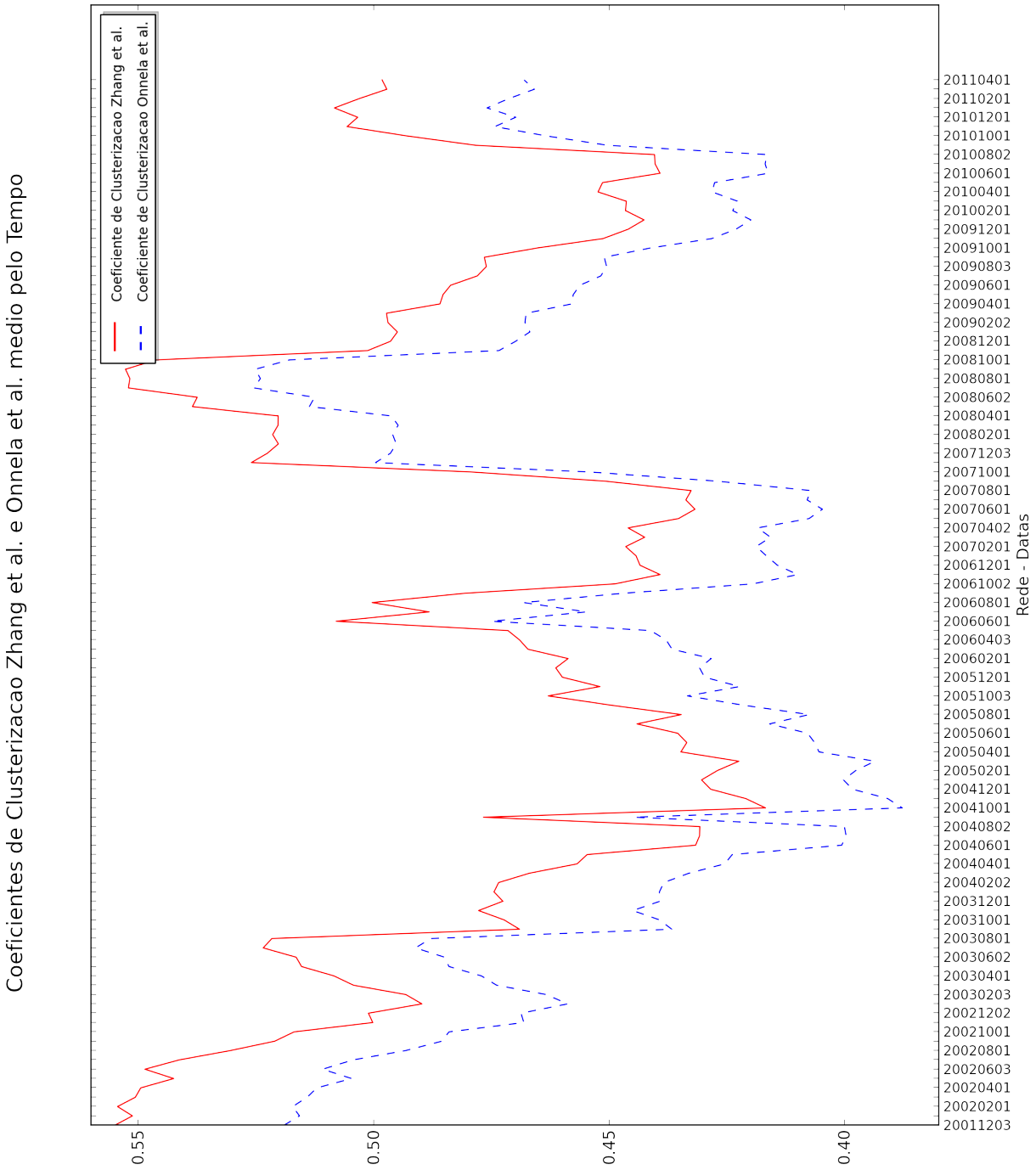

Figura 5.12: Coeficiente de clusterização ponderado Onnela et al. (2005), Zhang e Horvath (2005) médio pelo Tempo.

o peso dos caminhos têm maior frequência de que outras ações. Além disso, as ações com menor força de vértices são as que minimizam mais os caminhos; ou seja, essas ações são mais anti-correlacionadas da rede. Sabendo disso, pode-se dizer que quanto mais uma ação minimiza pesos, melhor ela é para otimizar uma carteira de ações, minimizando o risco específico de outras ações (MARKOWITZ, 1952).

Tabela 5.3: Número de caminhos minimizados (Nº Cam.) pela ação exibida, juntamente com a força do vértice s_i desta ação. Das redes da data de 20041001, 20090504 e 20110401.

Rede 20041001			Rede 20090504			Rede 20110401		
s_i	Ação	Nº Cam.	s_i	Ação	Nº Cam.	s_i	Ação	Nº Cam.
2.42	BBDC4	583	8.46	TLPP4	714	2.19	CCRO3	1334
2.88	VCPA4	65	9.01	ELET3	484	2.59	CIEL3	1267
2.98	ACES4	255	10.03	CSNA3	511	5.47	CPFE3	88
3.28	BRKM5	401	12.09	CRUZ3	0	11.55	ELPL4	0
3.93	PETR3	122	12.1	LIGT3	6	12.34	RDCD3	0
4.14	PETR4	92	12.11	CMIG4	5	13.52	TRPL4	60
4.26	CRTP5	551	12.29	CCRO3	2	16.36	AMBV4	900
7.66	LIGH3	0	13.28	GFSA3	0	16.95	LIGT3	0
8.0	BRAP4	28	13.34	RDCD3	0	17.81	BRTO4	0
8.05	EMBR4	0	13.54	CSAN3	0	17.92	NATU3	0
8.22	AMBV4	178	13.64	TAMM4	0	18.47	USIM3	0
8.57	TLCP4	0	13.97	ELET6	0	18.51	TMAR5	0
11.0	TCOC4	0	14.12	CESP6	0	18.7	CRUZ3	0
11.18	ARCZ6	0	14.13	ELPL6	0	19.58	BRFS3	0
11.46	GGBR4	0	14.37	TRPL4	0	19.72	MRFG3	0
12.14	EMBR3	0	14.48	CPFE3	0	19.97	CMIG4	0
12.32	TSPP4	0	14.8	EMBR3	0	20.07	ELET6	0
12.95	BRTP3	0	14.8	GOLL4	2	20.23	HYPE3	0
13.22	VALE3	0	14.85	TCSL3	0	21.0	TNLP4	0

continua

Tabela 5.3 – *continuação*

Rede 20041001			Rede 20090504			Rede 20110401		
s_i	Ação	Nº Cam.	s_i	Ação	Nº Cam.	s_i	Ação	Nº Cam.
13.63	TMCP4	0	14.86	NATU3	0	21.11	CPLE6	0
13.94	TLPP4	0	15.13	ALLL11	0	21.24	PCAR4	0
14.08	CRUZ3	0	15.37	BRKM5	0	21.44	CESP6	0
14.22	BRTO4	6	15.57	NETC4	0	21.54	SBSP3	0
14.95	PTIP4	0	15.8	UGPA4	0	21.55	JBSS3	0
14.99	TRPL4	0	16.3	SBSP3	0	22.0	KLBN4	0
15.12	ITSA4	0	16.41	CPLE6	0	22.08	BTOW3	0
15.13	BRTP4	0	16.57	TNLP4	0	22.57	EMBR3	0
15.15	VALE5	0	16.72	JBSS3	0	22.81	ELET3	0
15.23	SBSP3	0	16.8	TNLP3	0	22.93	GOLL4	0
15.48	ITAU4	0	18.56	TCSL4	0	23.17	TAMM4	0
15.61	BBAS3	0	18.63	VIVO4	0	23.34	TNLP3	0
15.71	TCSL3	0	18.81	KLBN4	0	23.83	CSAN3	0
15.72	CLSC6	0	18.9	AMBV4	0	23.9	USIM5	0
15.74	KLBN4	0	19.07	LREN3	0	24.56	LAME4	0
15.86	EBTP4	77	19.29	BTOW3	0	25.06	DTEX3	0
15.95	ELPL4	0	19.78	BRTO4	0	25.19	ALLL3	0
16.24	CMIG4	0	20.23	TMAR5	0	25.67	BVMF3	0
16.37	CPLE6	0	20.3	USIM3	0	26.02	LREN3	0
16.51	CGAS5	32	20.85	RSID3	0	26.05	SANB11	0
16.79	TMAR5	0	21.01	USIM5	0	26.29	MRVE3	0
17.05	CSNA3	0	21.6	ITUB4	0	26.44	OGXP3	0
17.12	CMET4	0	21.8	CYRE3	0	26.45	BRKM5	0
17.17	CMIG3	0	22.08	LAME4	0	26.55	FIBR3	0

continua

Tabela 5.3 – *continuação*

Rede 20041001			Rede 20090504			Rede 20110401		
s_i	Ação	Nº Cam.	s_i	Ação	Nº Cam.	s_i	Ação	Nº Cam.
17.19	ELET3	0	22.22	BVMF3	0	26.69	GFSA3	0
17.25	ELET6	0	22.25	BBAS3	0	27.33	BBAS3	0
17.45	CSTB4	0	23.0	PETR3	0	27.66	PETR4	0
17.45	TCSL4	0	23.6	ITSA4	0	27.78	PETR3	0
17.62	CESP4	0	23.9	PETR4	0	28.21	PDGR3	0
18.01	USIM5	0	24.05	BBDC4	0	28.31	CYRE3	0
18.08	TNLP3	0	24.06	BRAP4	0	28.48	LLXL3	0
19.96	TNLP4	0	24.43	VALE3	0	28.55	RSID3	0
			24.63	VALE5	0	28.79	VALE5	0
			24.83	GOAU4	0	28.87	GOAU4	0
			25.21	GGBR4	0	29.02	VALE3	0
						29.14	BRAP4	0
						29.51	BISA3	0
						29.8	GGBR4	0
						30.09	ITUB4	0
						30.15	MMXM3	0
						30.63	BBDC4	0
						31.01	ITSA4	0
						31.35	CSNA3	0

5.7 ÁRVORE ESPALHADA MÍNIMA - MST

De acordo com a Seção 2.1.5, a árvore *MST* seleciona as ligações mais relevantes para cada ação afim de gerar um grafo acíclico de peso mínimo. Com isso a árvore *MST* dá, de forma direta, a organização hierárquica das ações.

Por outro lado, seguiremos a hipótese de que o espaço ultra-métrico será útil para a taxonomia da rede. Para isso, o coeficiente de correlação será transformado em uma distância que respeita os três axiomas da métrica Euclidiana, para isso a Equação 4.3 será utilizada.

Com a transformação dos coeficientes em distância, a Matriz *D* de distâncias será utilizada para gerar a árvore *MST* para todas as redes.

Nas próximas seções serão mostrados os experimentos de medidas físicas executados nas árvores *MST* da matriz de distância de correlações.

5.7.1 DISTRIBUIÇÃO DE GRAUS E DE FORÇAS DE VÉRTICES

Com a construção das árvores *MST*, nesta seção serão mostradas as distribuições de graus e distribuição de forças de vértices, assim como feito na Seção 5.4 porém com a rede *MST*. A distribuição de graus e de força de vértices são interessantes pois se elas seguirem uma lei de potência então a rede verificada é livre de escalas.

Serão mostrados os experimentos no mesmo instante de tempo das 12 redes específicas da Seção 5.3.

As Figuras 5.13 e 5.14 apresentam as distribuições de graus da rede *MST* de 1 a 12.

Enquanto as Figuras 5.15 e 5.16 exibem as distribuições de forças de vértices da rede *MST* no mesmo instante de tempo das Figuras 5.13 e 5.14.

É possível notar que ambas as distribuições, de graus e de força de vértices, parecem seguir uma lei de potência. Isso significa que a rede *MST* proposta é livre de escala, e com isso pode-se dizer que as ações são correlacionadas entre si e seguem uma hierarquia.

As Figuras 5.17, 5.18 e 5.19 mostram as árvores *MST* 5 e 11, além da rede de data mais recente estudada neste trabalho.

Cada figura representa a rede *MST*, onde cada nó representa uma ação, descrita pelo seu código de negociação na BM&F Bovespa. Além disso, a cor de cada nó representa o setor de classificação a qual cada ação se encaixa.

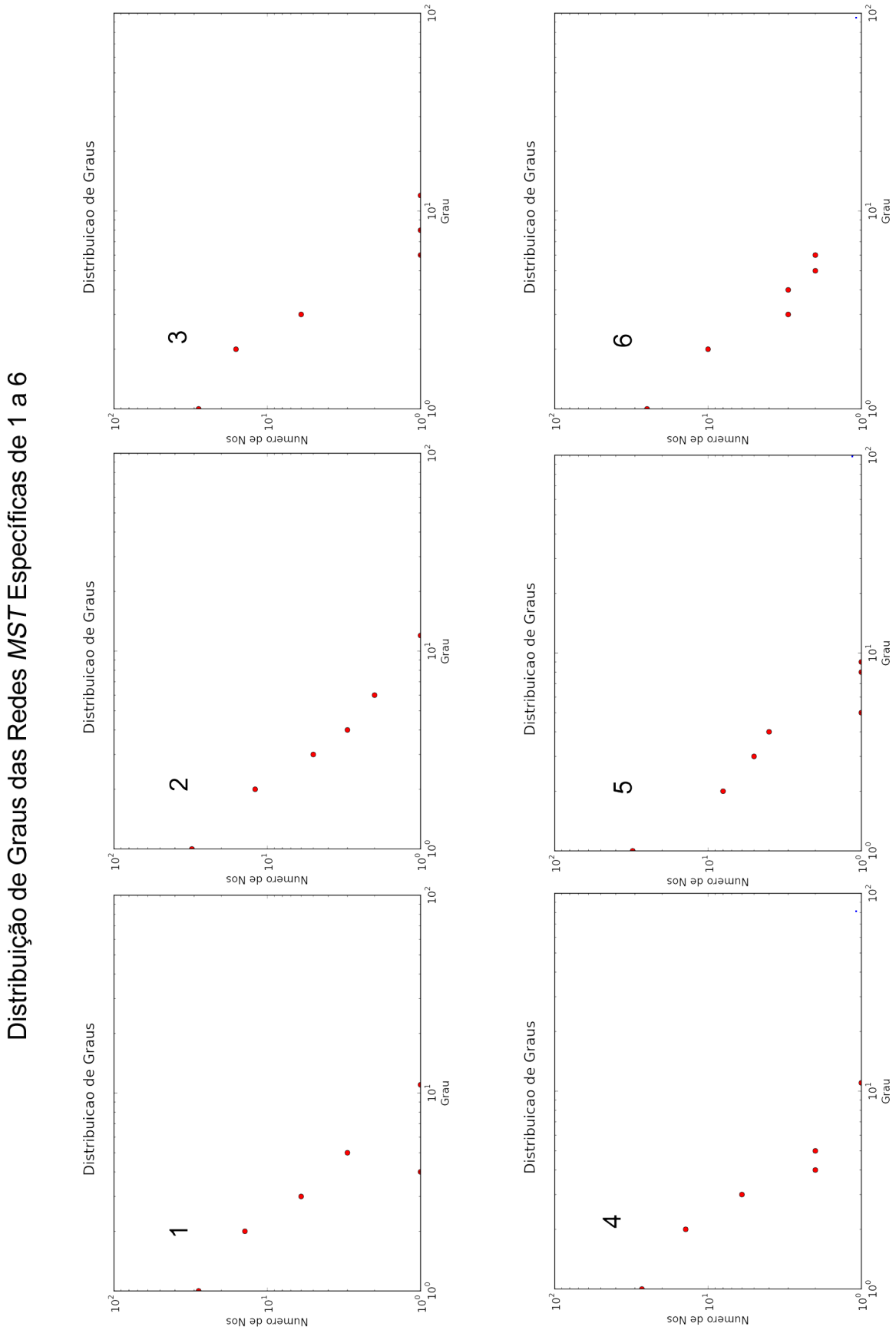

Figura 5.13: Distribuição de graus de vértices das redes *MST* de 1 a 6.

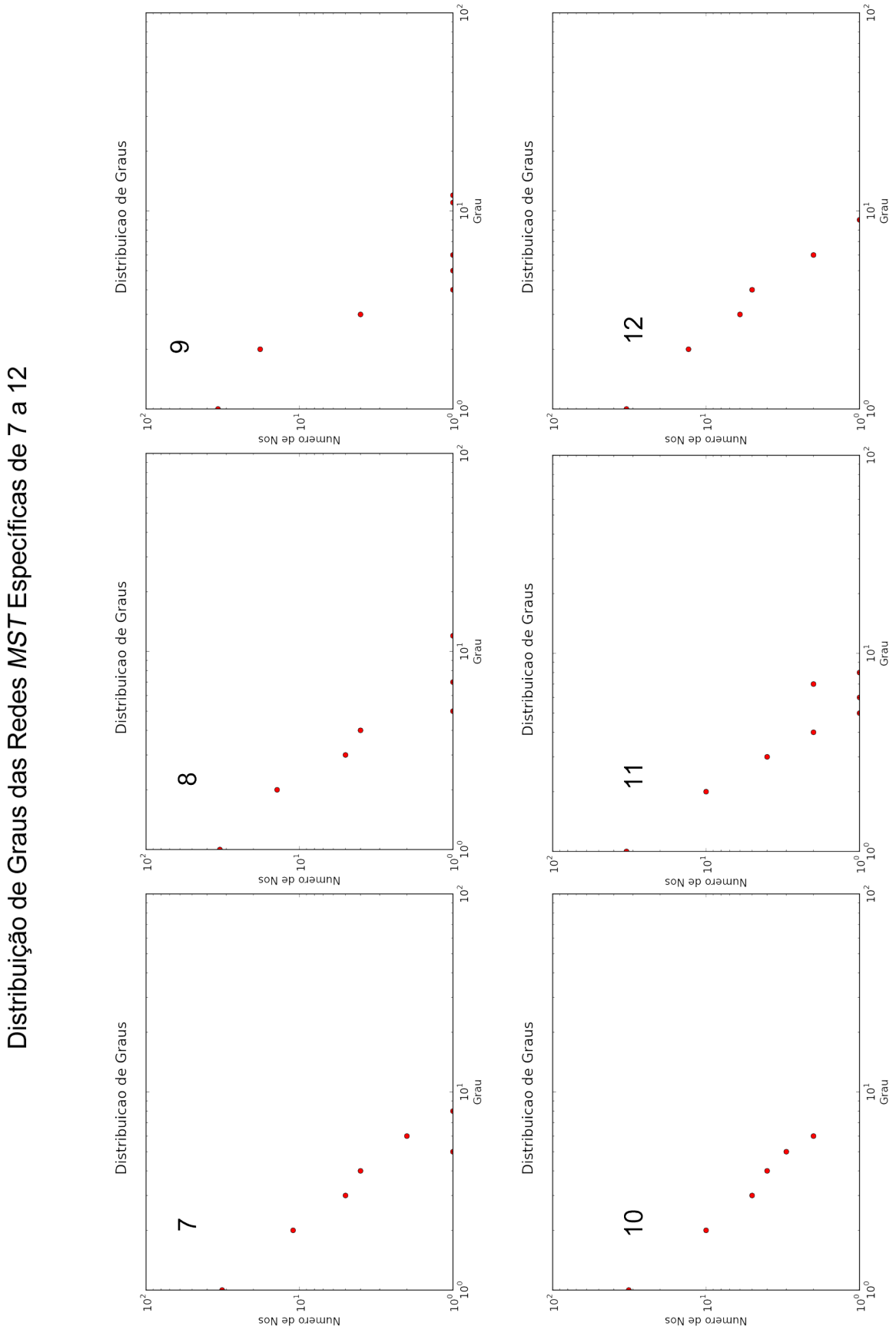

Figura 5.14: Distribuição de graus das redes *MST* de 7 a 12.

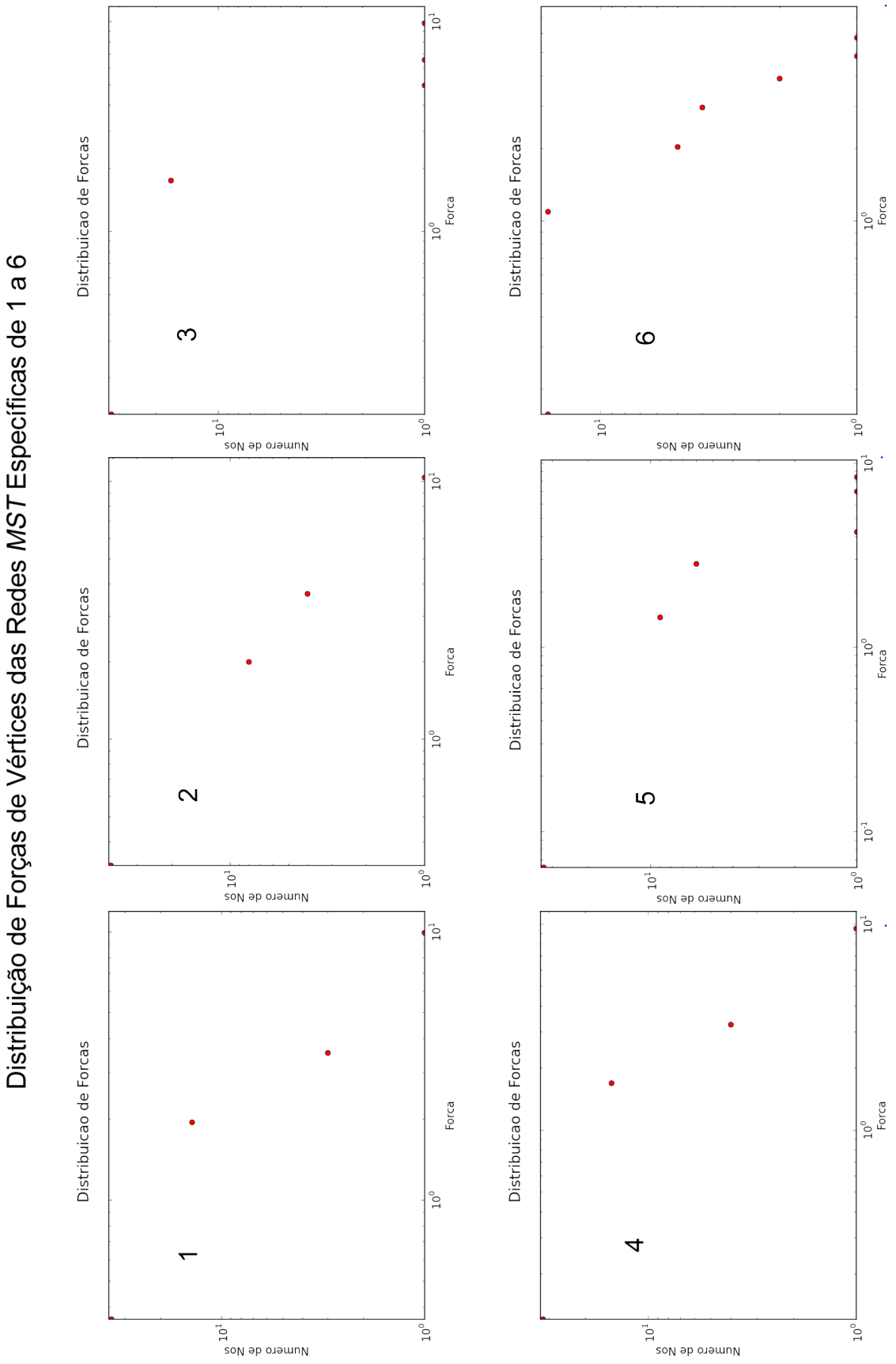

Figura 5.15: Distribuição de forças de vértices das redes *MST* de 1 a 6.

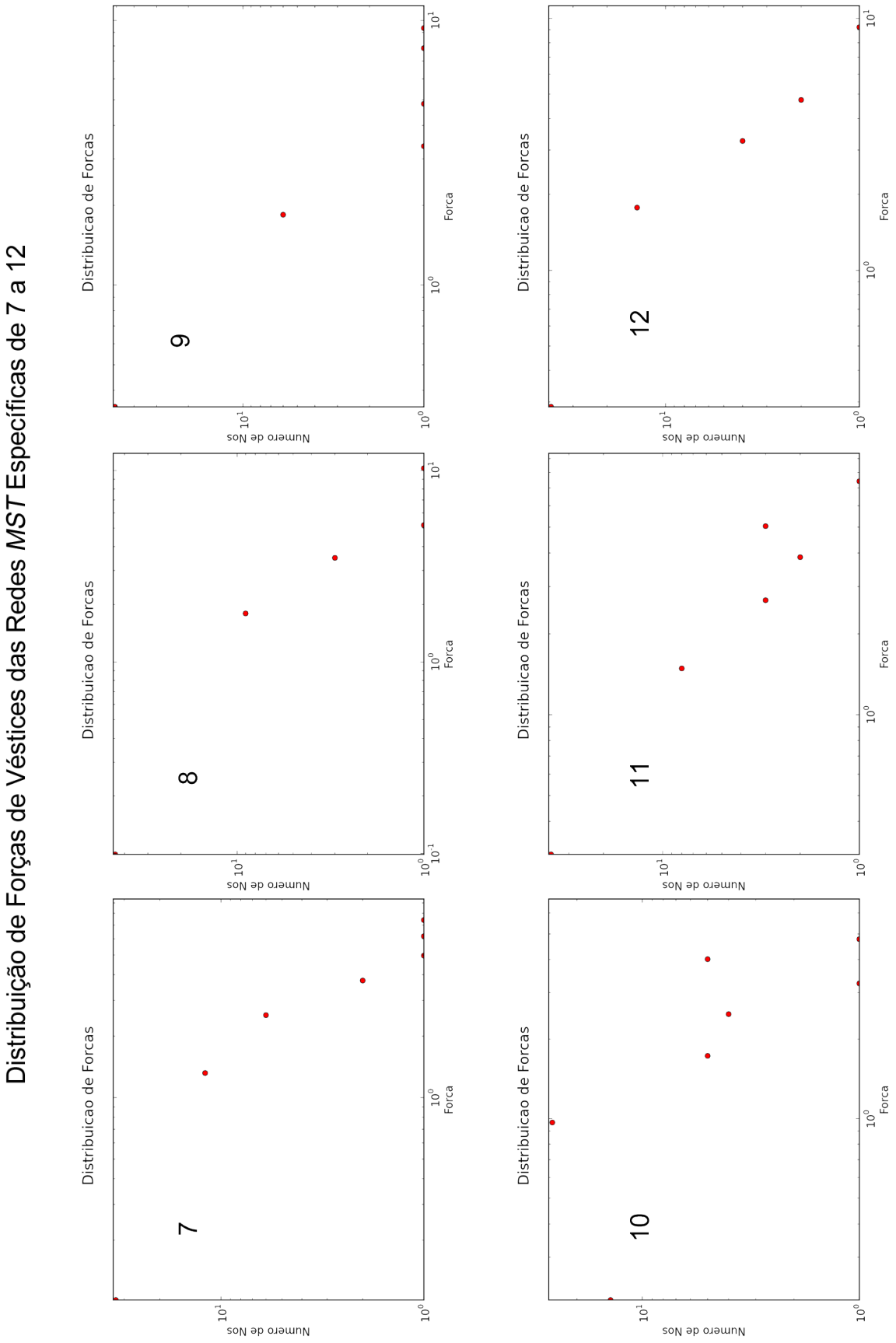

Figura 5.16: Distribuição de forças de vértices das redes *MST* de 7 a 12.

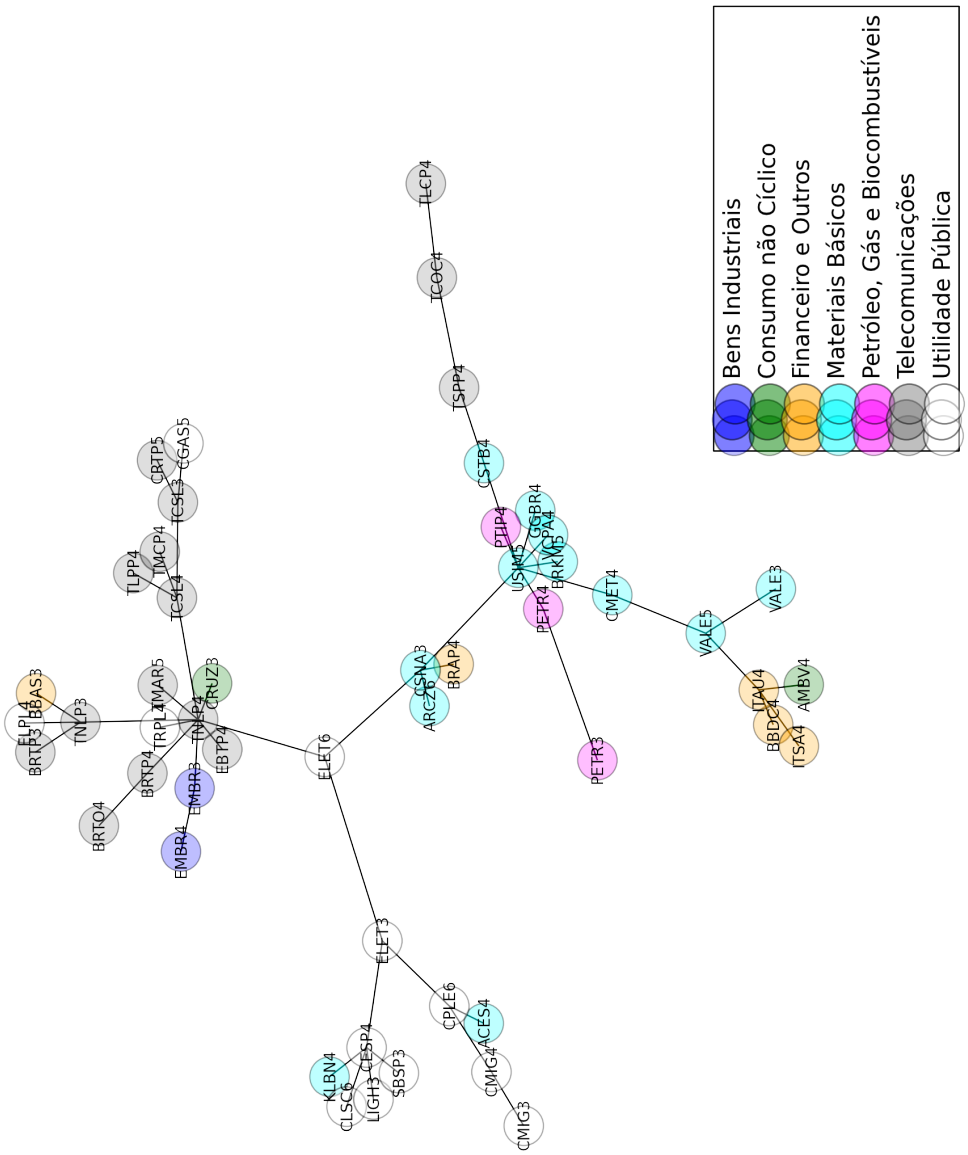

Figura 5.17: Árvore *MST* de ações da data de 20041001 a 20050930.

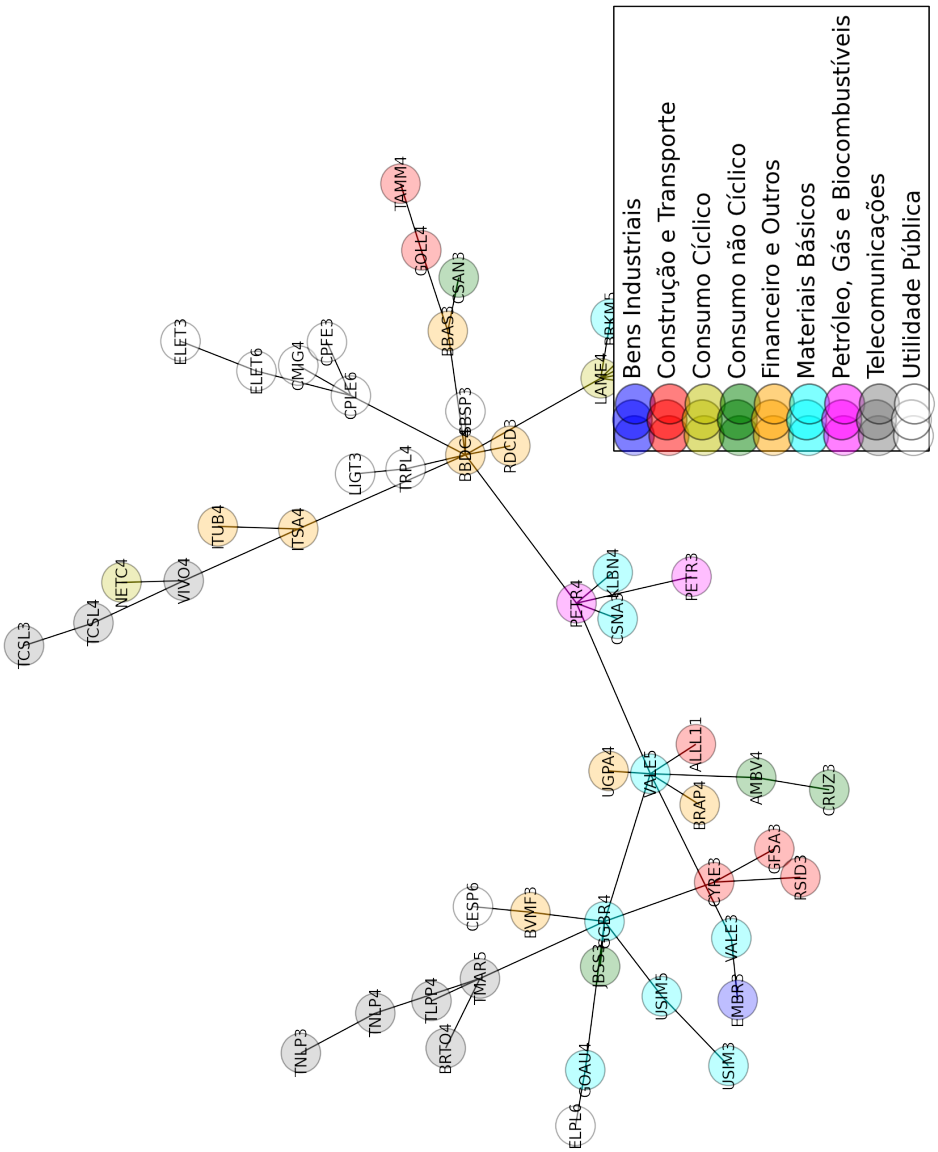

Figura 5.18: Árvore *MST* de ações da data de 20090504 a 20100430.

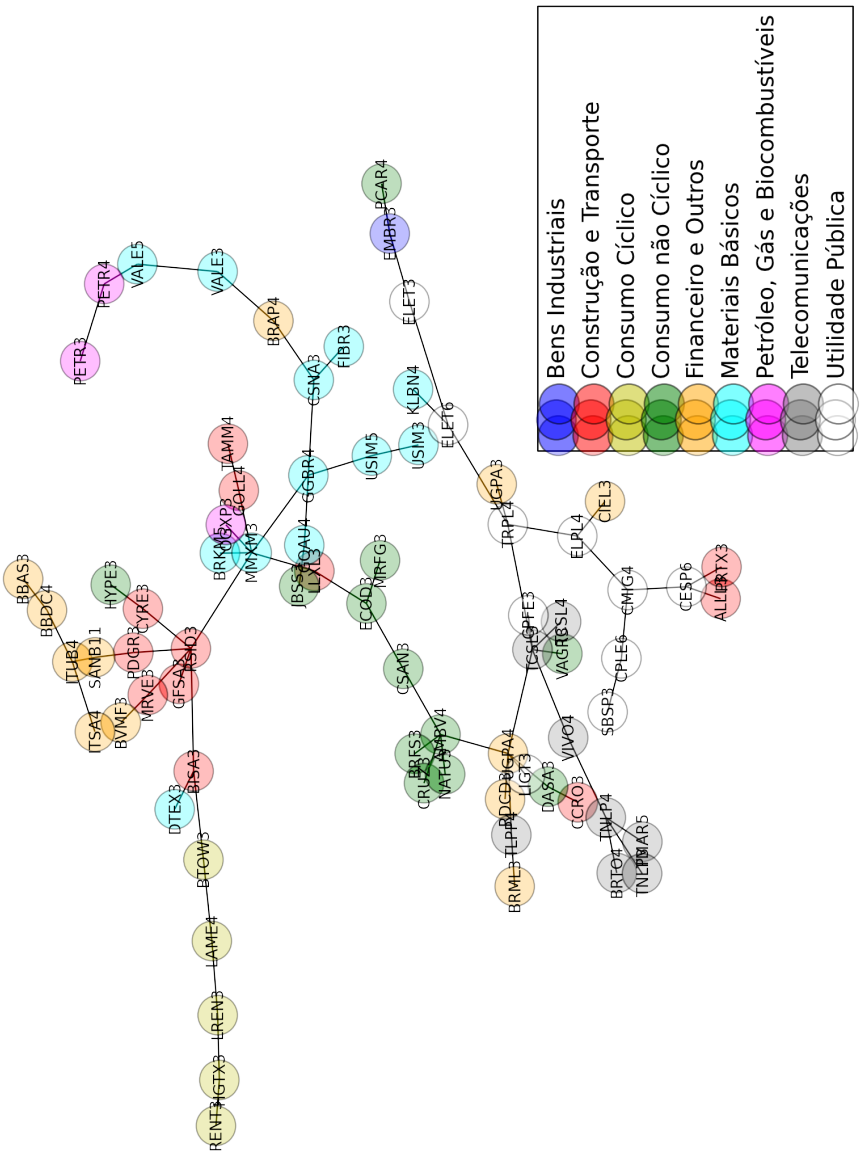

Figura 5.19: Árvore *MST* de ações da data de 20110401 a 20120330.

As Tabelas 5.4, 5.5 e 5.6 mostram o número de ligações das ações e o setor a qual cada empresa das ações se encaixam, descrevendo as Figuras 5.17, 5.18 e 5.19 respectivamente.

Tabela 5.4: Ligações da árvore *MST* de distâncias de correlações de ações da data de 20041001 a 20050930.

Rede - 20041001 a 20050930			
Num. Ligações	Cod. Ação	Nome Ação	Setor
9	TNLP4	TELEMAR PN	Telecomunicações
8	USIM5	USIMINAS PNA	Materiais Básicos
5	CESP4	CESP PN *	Utilidade Pública
4	TNLP3	TELEMAR ON	Telecomunicações
4	TCSL4	TIM PART S/A PN	Telecomunicações
4	ITAU4	ITAUUNIBANCO PN	Financeiro e Outros
4	CSNA3	SID NACIONAL ON	Materiais Básicos
3	VALE5	VALE PNA	Materiais Básicos
3	TCSL3	TIM PART S/A ON	Telecomunicações
3	ELET6	ELETROBRAS PNB	Utilidade Pública
3	ELET3	ELETROBRAS ON	Utilidade Pública
3	CPLE6	COPEL PNB	Utilidade Pública
2	TSPP4	TELESP CL PA PN	Telecomunicações
2	TCOC4	TELE CTR OES PN	Telecomunicações
2	PETR4	PETROBRAS PN	Petróleo, Gás e Biocombustíveis
2	EMBR3	EMBRAER ON	Bens Industriais
2	CSTB4	SID TUBARAO PN *	Materiais Básicos
2	CMIG4	CEMIG PN	Utilidade Pública
2	CMET4	CAEMI PN	Materiais Básicos
2	BRTP4	BRASIL T PAR PN	Telecomunicações
1	VCPA4	V C P PN	Materiais Básicos
1	VALE3	VALE ON	Materiais Básicos

continua

Tabela 5.4 – *continuação*

Rede - 20041001 a 20050930			
Num. Ligações	Cod. Ação	Nome Ação	Setor
1	TRPL4	TRAN PAULIST PN	Utilidade Pública
1	TMCP4	TELEMIG PART PN	Telecomunicações
1	TMAR5	TELEMAR N L PNA	Telecomunicações
1	TLPP4	TELESP PN	Telecomunicações
1	TLCP4	TELE LEST CL PN	Telecomunicações
1	SBSP3	SABESP ON	Utilidade Pública
1	PTIP4	IPIRANGA PET PN	Petróleo, Gás e Biocombustíveis
1	PETR3	PETROBRAS ON	Petróleo, Gás e Biocombustíveis
1	LIGH3	LIGHT ON *	Utilidade Pública
1	KLBN4	KLABIN S/A PN	Materiais Básicos
1	ITSA4	ITAUSA PN	Financeiro e Outros
1	GGBR4	GERDAU PN	Materiais Básicos
1	EMBR4	EMBRAER PN EJ	Bens Industriais
1	ELPL4	ELETROPAULO PN	Utilidade Pública
1	EBTP4	EMBRATEL PAR PN *	Telecomunicações
1	CRUZ3	SOUZA CRUZ ON	Consumo não Cíclico
1	CRTP5	CRT CELULAR PNA	Telecomunicações
1	CMIG3	CEMIG ON	Utilidade Pública
1	CLSC6	CELESC PNB	Utilidade Pública
1	CGAS5	COMGAS PNA	Utilidade Pública
1	BRTP3	BRASIL T PAR ON	Telecomunicações
1	BRTO4	BRASIL TELEC PN	Telecomunicações
1	BRKM5	BRASKEM PNA	Materiais Básicos
1	BRAP4	BRADESPAR PN	Financeiro e Outros

continua

Tabela 5.4 – *continuação*

Rede - 20041001 a 20050930			
Num. Ligações	Cod. Ação	Nome Ação	Setor
1	BBDC4	BRADESCO PN	Financeiro e Outros
1	BBAS3	BRASIL ON	Financeiro e Outros
1	ARCZ6	ARACRUZ PNB	Materiais Básicos
1	AMBV4	AMBEV PN	Consumo não Cíclico
1	ACES4	ACESITA PN	Materiais Básicos

Tabela 5.5: Ligações da árvore *MST* de distâncias de correlações de ações da data de 20090504 a 20100430.

Rede - 20090504 a 20100430			
Num. Ligações	Cod. Ação	Nome Ação	Setor
8	BBDC4	BRADESCO PN	Financeiro e Outros
7	VALE5	VALE PNA	Materiais Básicos
7	GGBR4	GERDAU PN	Materiais Básicos
6	LAME4	LOJAS AMERIC PN	Consumo Cíclico
5	PETR4	PETROBRAS PN	Petróleo, Gás e Biocombustíveis
4	TMAR5	TELEMAR N L PNA	Telecomunicações
4	CPLE6	COPEL PNB	Utilidade Pública
3	VIVO4	VIVO PN	Telecomunicações
3	ITSA4	ITAUSA PN	Financeiro e Outros
3	CYRE3	CYRELA REALT ON	Construção e Transporte
3	BBAS3	BRASIL ON	Financeiro e Outros
2	VALE3	VALE ON	Materiais Básicos
2	USIM5	USIMINAS PNA	Materiais Básicos
2	TRPL4	TRAN PAULIST PN	Utilidade Pública
2	TNLP4	TELEMAR PN	Telecomunicações

continua

Tabela 5.5 – *continuação*

Rede - 20090504 a 20100430			
Num. Ligações	Cod. Ação	Nome Ação	Setor
2	TCSL4	TIM PART S/A PN	Telecomunicações
2	GOLL4	GOL PN	Construção e Transporte
2	GOAU4	GERDAU MET PN	Materiais Básicos
2	ELET6	ELETROBRAS PNB	Utilidade Pública
2	BVMF3	BMFBOVESPA ON	Financeiro e Outros
2	AMBV4	AMBEV PN	Consumo não Cíclico
1	USIM3	USIMINAS ON	Materiais Básicos
1	UGPA4	ULTRAPAR PN	Financeiro e Outros
1	TNLP3	TELEMAR ON	Telecomunicações
1	TLPP4	TELESP PN	Telecomunicações
1	TCSL3	TIM PART S/A ON	Telecomunicações
1	TAMM4	TAM S/A PN	Construção e Transporte
1	SBSP3	SABESP ON	Utilidade Pública
1	RSID3	ROSSI RESID ON	Construção e Transporte
1	RDCD3	REDECARD ON	Financeiro e Outros
1	PETR3	PETROBRAS ON	Petróleo, Gás e Biocombustíveis
1	NETC4	NET PN	Consumo Cíclico
1	NATU3	NATURA ON	Consumo não Cíclico
1	LREN3	LOJAS RENNER ON	Consumo Cíclico
1	LIGT3	LIGHT S/A ON	Utilidade Pública
1	KLBN4	KLABIN S/A PN	Materiais Básicos
1	JBSS3	JBS ON	Consumo não Cíclico
1	ITUB4	ITAUUNIBANCO PN	Financeiro e Outros
1	GFSA3	GAFISA ON	Construção e Transporte

continua

Tabela 5.5 – *continuação*

Rede - 20090504 a 20100430			
Num. Ligações	Cod. Ação	Nome Ação	Setor
1	EMBR3	EMBRAER ON	Bens Industriais
1	ELPL6	ELETROPAULO PNB	Utilidade Pública
1	ELET3	ELETROBRAS ON	Utilidade Pública
1	CSNA3	SID NACIONAL ON	Materiais Básicos
1	CSAN3	COSAN ON	Consumo não Cíclico
1	CRUZ3	SOUZA CRUZ ON	Consumo não Cíclico
1	CPFE3	CPFL ENERGIA ON	Utilidade Pública
1	CMIG4	CEMIG PN	Utilidade Pública
1	CESP6	CESP PNB	Utilidade Pública
1	CCRO3	CCR SA ON	Construção e Transporte
1	BTOW3	B2W VAREJO ON	Consumo Cíclico
1	BRTO4	BRASIL TELEC PN	Telecomunicações
1	BRKM5	BRASKEM PNA	Materiais Básicos
1	BRAP4	BRADESPAR PN	Financeiro e Outros
1	ALLL11	ALL AMER LAT UNT	Construção e Transporte

Tabela 5.6: Ligações da árvore *MST* de distâncias de correlações de ações da data de 20110401 a 20120330.

Rede - 20110401 a 20120330			
Num. Ligações	Cod. Ação	Nome Ação	Setor
11	CSNA3	SID NACIONAL ON	Materiais Básicos
7	ITSA4	ITAUSA PN	Financeiro e Outros
6	RSID3	ROSSI RESID ON	Construção e Transporte
5	BISA3	BROOKFIELD ON	Construção e Transporte
4	GGBR4	GERDAU PN	Materiais Básicos

continua

Tabela 5.6 – *continuação*

Rede - 20110401 a 20120330			
Num. Ligações	Cod. Ação	Nome Ação	Setor
4	ELET3	ELETROBRAS ON	Utilidade Pública
3	VALE3	VALE ON	Materiais Básicos
3	TNLP3	TELEMAR ON	Telecomunicações
3	MMXM3	MMX MINER ON	Materiais Básicos
3	LAME4	LOJAS AMERIC PN	Consumo Cíclico
3	ITUB4	ITAUUNIBANCO PN	Financeiro e Outros
3	CRUZ3	SOUZA CRUZ ON	Consumo não Cíclico
3	CMIG4	CEMIG PN	Utilidade Pública
3	BRAP4	BRADESPAR PN	Financeiro e Outros
3	BBDC4	BRADESCO PN	Financeiro e Outros
2	USIM5	USIMINAS PNA	Materiais Básicos
2	USIM3	USIMINAS ON	Materiais Básicos
2	TNLP4	TELEMAR PN	Telecomunicações
2	TAMM4	TAM S/A PN	Construção e Transporte
2	PETR4	PETROBRAS PN	Petróleo, Gás e Biocombustíveis
2	PETR3	PETROBRAS ON	Petróleo, Gás e Biocombustíveis
2	PCAR4	P.ACUCAR-CBD PN	Consumo não Cíclico
2	GOAU4	GERDAU MET PN	Materiais Básicos
2	ELPL4	ELETROPAULO PN	Utilidade Pública
2	DTEX3	DURATEX ON	Materiais Básicos
2	CYRE3	CYRELA REALT ON	Construção e Transporte
1	VALE5	VALE PNA	Materiais Básicos
1	TRPL4	TRAN PAULIST PN	Utilidade Pública
1	TMAR5	TELEMAR N L PNA	Telecomunicações

continua

Tabela 5.6 – *continuação*

Rede - 20110401 a 20120330			
Num. Ligações	Cod. Ação	Nome Ação	Setor
1	SBSP3	SABESP ON	Utilidade Pública
1	SANB11	SANTANDER BR UNT	Financeiro e Outros
1	RDCD3	REDECARD ON	Financeiro e Outros
1	PDGR3	PDG REALT ON	Construção e Transporte
1	OGXP3	OGX PETROLEO ON	Petróleo, Gás e Biocombustíveis
1	NATU3	NATURA ON	Consumo não Cíclico
1	MRVE3	MRV ON	Construção e Transporte
1	MRFG3	MARFRIG ON	Consumo não Cíclico
1	LREN3	LOJAS RENNER ON	Consumo Cíclico
1	LLXL3	LLX LOG ON	Construção e Transporte
1	LIGT3	LIGHT S/A ON	Utilidade Pública
1	KLBN4	KLABIN S/A PN	Materiais Básicos
1	JBSS3	JBS ON	Consumo não Cíclico
1	HYPE3	HYPERMARCAS ON	Consumo não Cíclico
1	GOLL4	GOL PN	Construção e Transporte
1	GFSA3	GAFISA ON	Construção e Transporte
1	FIBR3	FIBRIA ON	Materiais Básicos
1	EMBR3	EMBRAER ON	Bens Industriais
1	ELET6	ELETROBRAS PNB	Utilidade Pública
1	CSAN3	COSAN ON	Consumo não Cíclico
1	CPLE6	COPEL PNB	Utilidade Pública
1	CPFE3	CPFL ENERGIA ON	Utilidade Pública
1	CIEL3	CIELO ON	Financeiro e Outros
1	CESP6	CESP PNB	Utilidade Pública

continua

Tabela 5.6 – *continuação*

Rede - 20110401 a 20120330			
Num. Ligações	Cod. Ação	Nome Ação	Setor
1	CCRO3	CCR SA ON	Construção e Transporte
1	BVMF3	BMFBOVESPA ON	Financeiro e Outros
1	BTOW3	B2W VAREJO ON	Consumo Cíclico
1	BRTO4	BRASIL TELEC PN	Telecomunicações
1	BRKM5	BRASKEM PNA	Materiais Básicos
1	BRFS3	BRF FOODS ON	Consumo não Cíclico
1	BBAS3	BRASIL ON	Financeiro e Outros
1	AMBV4	AMBEV PN	Consumo não Cíclico
1	ALLL3	ALL AMER LAT ON	Construção e Transporte

Como tem-se uma rede correlacionada de ações livre de escala, devido à lei de potência, isso sugere que, hierarquicamente, as ações que são mais conectadas na rede *MST* controlam ou intermedem as flutuações de preços diários de suas ações vizinhas. As ações que ligam as ações mais conectadas com as ações das extremidades da *MST* intermedem as variações de preços. Enquanto as ações da extremidade da árvore são ações menos influentes, e que são menos influenciadas pelas variações do mercado, essas ações não necessariamente seguem os índices econômicos.

Com isso em mente, as Figuras 5.20, 5.21 e 5.22 mostram as flutuações de preço de fechamento diário da ação com mais ligações da rede *MST*, juntamente com as ações VALE5, PETR4 e USIM5 que são ações de alta liquidez no mercado. Além disso, também será mostrado a ação intermediária mais frequentes da matriz de distâncias de caminho (ver Seção 5.6) nas datas das Figuras 5.17, 5.18 e 5.19, respectivamente.

Nas Figuras 5.20, 5.21 e 5.22 é importante ressaltar que os picos extremos de baixa significam que a ação sofreu *split*; ou seja, a ação é dividida em outras ações e com isso seu preço também é dividido. O *split* acontece quando uma ação está muito valorizada no mercado. Porém, no gráfico das Figuras 5.20, 5.21 e 5.22 o preço, quando a ação sofre *split*, é reajustado manualmente para

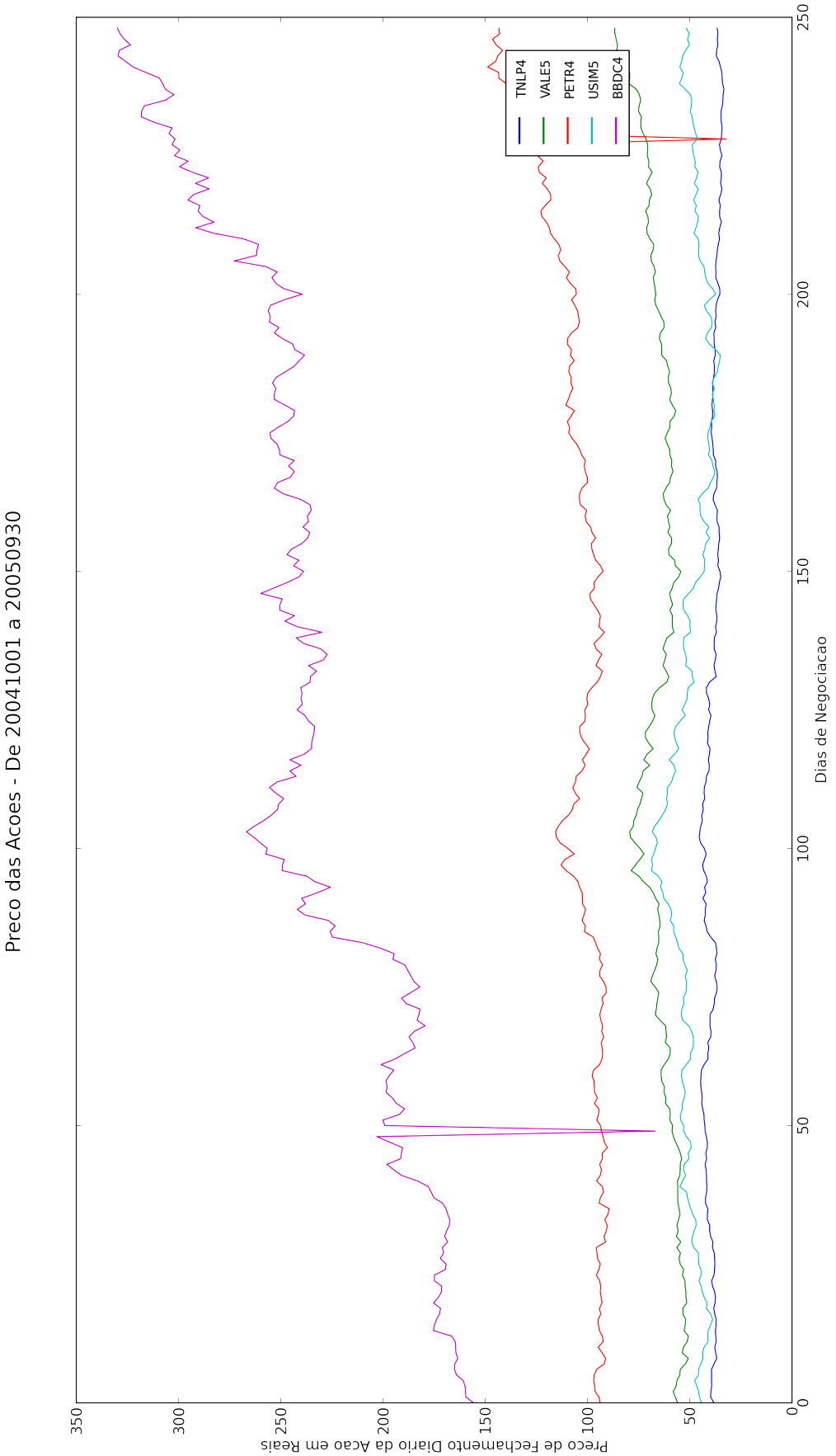

Figura 5.20: Preço de fechamento diário da ações TNLP4, VALE5, PETR4, USIM5 e BBDC4 na data de 20041001 a 20050930.

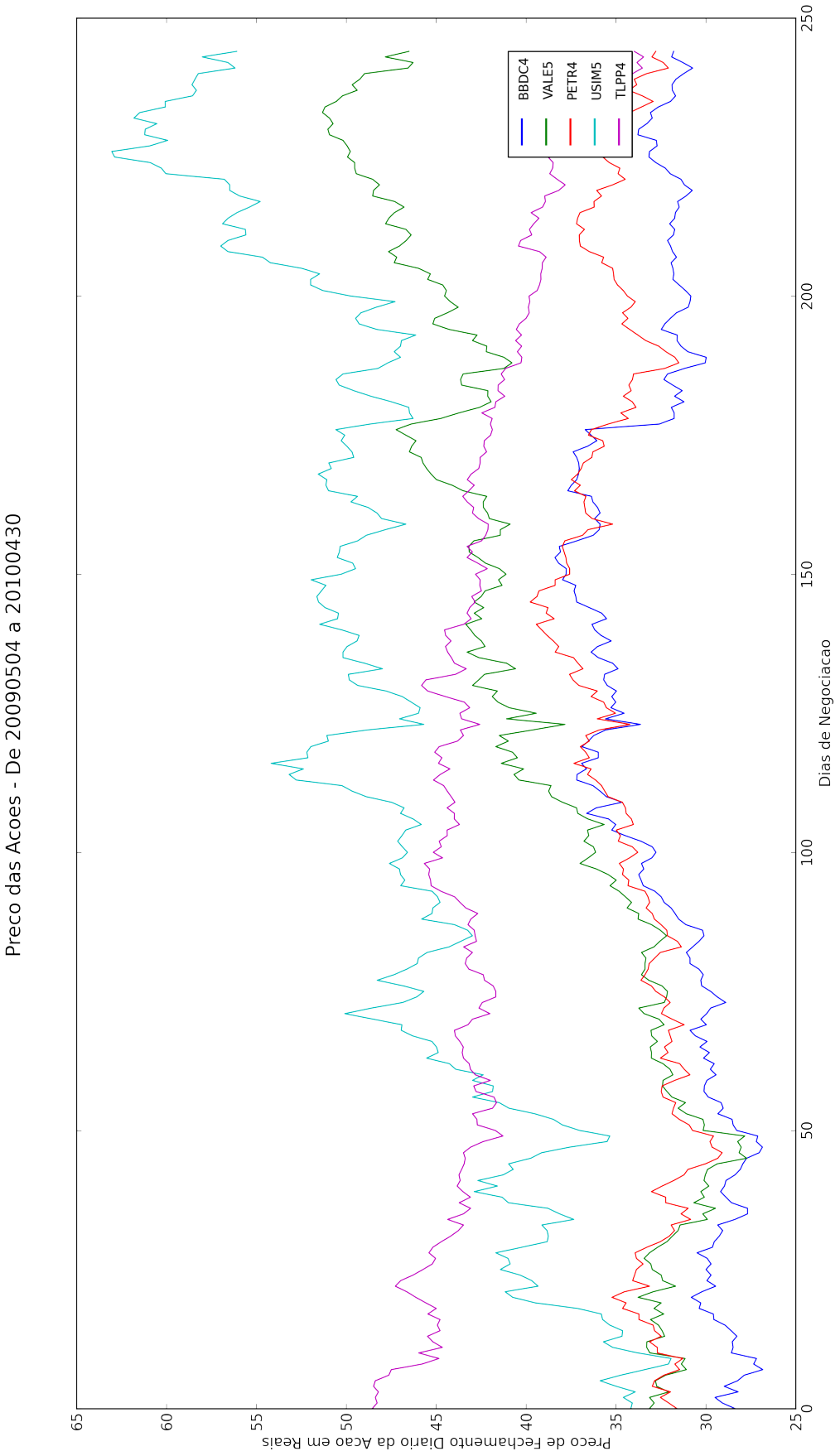

Figura 5.21: Preço de fechamento diário da ações BBDC4, VALE5, PETR4, USIM5 e TLPP4 na data de 20090504 a 20100430.

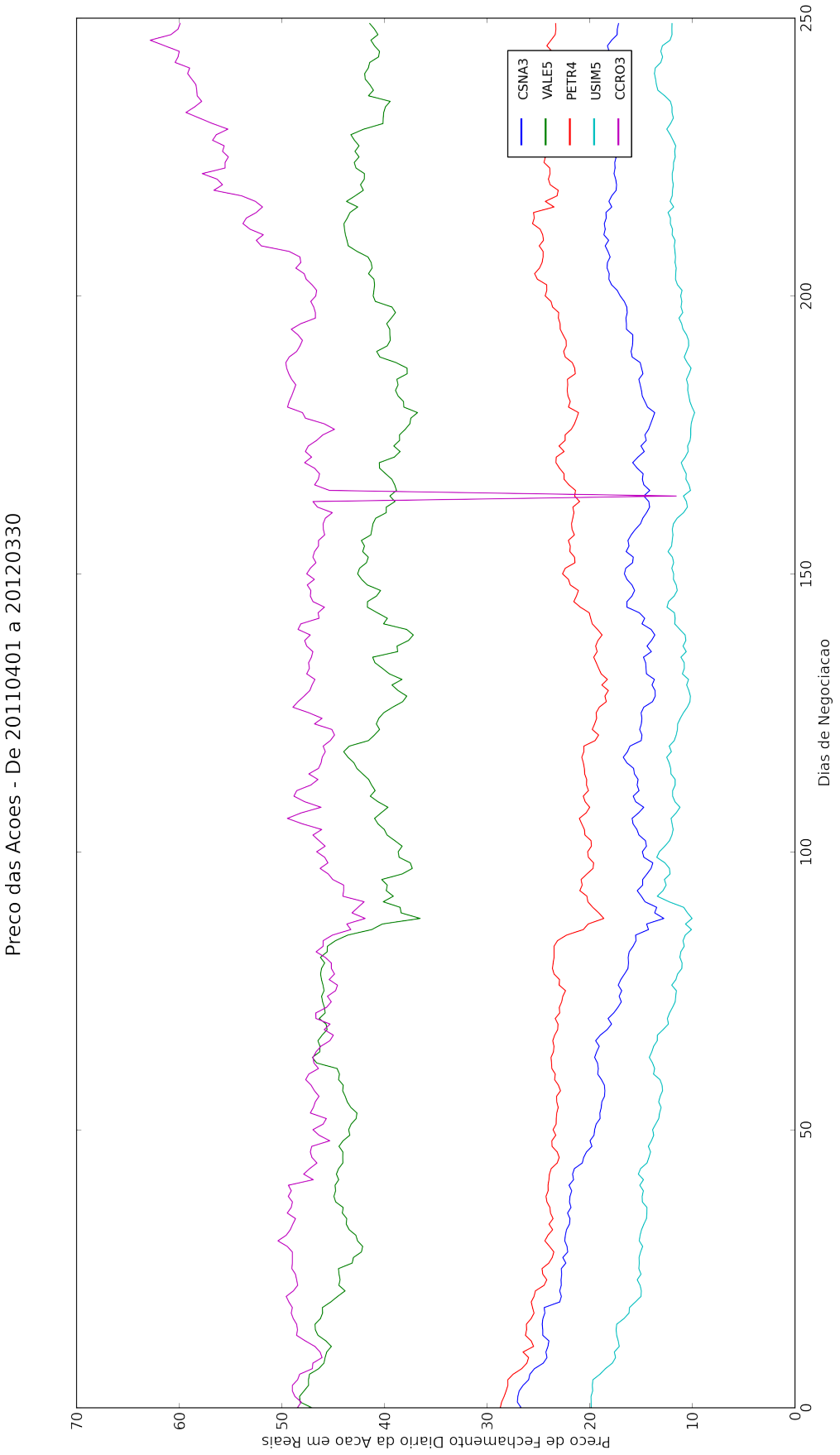

Figura 5.22: Preço de fechamento diário da ações CSNA3, VALE5, PETR4, USIM5 e CCRO3 na data de 20110401 a 20120330.

melhor visualização das correlações entre as ações.

Tabela 5.7: Matriz de correlações cruzada entre pares de ações de acordo com a Equação 4.2.

Correlação Cruzada					
	TNLP4	VALE5	PETR4	USIM5	BBDC4
TNLP4	1.0	0.43	0.18	0.47	0.04
VALE5	0.43	1.0	0.08	0.48	0.06
PETR4	0.18	0.08	1.0	0.2	0.01
USIM5	0.47	0.48	0.2	1.0	0.03
BBDC4	0.04	0.06	0.01	0.03	1.0

Tabela 5.8: Matriz de correlações cruzada entre pares de ações de acordo com a Equação 4.2.

Correlação Cruzada					
	BBDC4	VALE5	PETR4	USIM5	TLPP4
BBDC4	1.0	0.62	0.66	0.54	0.18
VALE5	0.62	1.0	0.73	0.64	0.16
PETR4	0.66	0.73	1.0	0.61	0.24
USIM5	0.54	0.64	0.61	1.0	0.17
TLPP4	0.18	0.16	0.24	0.17	1.0

Tabela 5.9: Matriz de correlações cruzada entre pares de ações de acordo com a Equação 4.2.

Correlação Cruzada					
	CSNA3	VALE5	PETR4	USIM5	CCRO3
CSNA3	1.0	0.73	0.69	0.72	-0.04
VALE5	0.73	1.0	0.69	0.52	-0.0
PETR4	0.69	0.69	1.0	0.51	-0.03
USIM5	0.72	0.52	0.51	1.0	-0.03
CCRO3	-0.04	-0.0	-0.03	-0.03	1.0

As Tabelas 5.7, 5.8 e 5.9 mostram as correlações das ações par a par das Figuras 5.20, 5.21 e 5.22 respectivamente.

Por sua vez, as Tabelas 5.10, 5.11 e 5.12 mostram as propriedades físicas de Densidade do Vértice e Força do Vértice das redes de coeficientes de correlações das datas analisadas, além disso a média dessas propriedades da rede também serão exibidas.

Tabela 5.10: Medidas físicas de Densidade ponderada e Força do vértice das ações TNLP4, VALE5, PETR4, USIM5 e BBDC4 da rede do período de 20041001 a 20050930.

	Densidade	Força do Vértice
TNLP4	0.3992	19.9592
VALE5	0.3007	15.0336
PETR4	0.0785	3.9272
USIM5	0.3601	18.0067
BBDC4	0.0443	2.2161
Média da Rede	0.2542	12.7091

Tabela 5.11: Medidas físicas de Densidade ponderada e Força do vértice das ações BBDC4, VALE5, PETR4, USIM5 e TLPP4 da rede do período de 20090504 a 20100430.

	Densidade	Força do Vértice
BBDC4	0.4539	24.0542
VALE5	0.4648	24.6338
PETR4	0.4509	23.897
USIM5	0.3963	21.0054
TLPP4	0.1595	8.456
Média da Rede	0.3312	17.5528

Tabela 5.12: Medidas físicas de Densidade ponderada e Força do vértice das ações CSNA3, VALE5, PETR4, USIM5 e CCRO3 da rede do período de 20110401 a 20120330.

	Densidade	Força do Vértice
CSNA3	0.5128	31.2822
VALE5	0.4718	28.7772
PETR4	0.4526	27.6092
USIM5	0.3888	23.7168
CCRO3	-0.0103	-0.629
Média da Rede	0.3728	22.7423

De acordo com a Figura 5.20 e a Tabela 5.7 pode-se notar que a ação com mais ligação na rede *MST* mostra-se bem correlacionadas com as ações VALE5, PETR4 e USIM5. Verificando a força do vértice e a densidade ponderada na Tabela 5.10 nota-se que essa ação, TNLP4, é muito bem correlacionada na rede.

Por outro lado, observando a ação BBDC4 na Figura 5.20 e na Tabela 5.7 é notável que essa ação é pouco correlacionada com as outras ações, e de acordo com a Tabela 5.10 essa ação tem a menor Densidade e Força de Vértice da rede. Além disso, a Tabela 5.4 mostra que essa ação está na extremidade da árvore *MST*.

Por sua vez, a Figura 5.21 ilustra que a ação que minimiza os pesos na matriz de distâncias de caminho (ver Seção 5.6) é visivelmente anti-correlacionadas com as outras ações, assim como pode ser visto na Tabela 5.8 de correlações.

As ações BBDC4, PETR4, VALE5 e USIM5 mostraram-se bem correlacionados entre si, como pode ser visto na Figura 5.21 e na Tabela 5.8. Entretanto, neste período, a ação USIM5 obteve alta volatilidade, o que pode ser ruim para o investidor, uma vez que o que se procura na otimização de carteiras é a segurança no investimento (MARKOWITZ, 1952).

A Tabela 5.11 mostra que as ações BBDC4, VALE5, PETR4 e USIM5 são bem correlacionadas na rede, enquanto a ação TLPP4 possui a mais baixa Densidade e Força do Vértice da

rede.

Na Figura 5.22, tanto a ação com mais ligações na árvore *MST* quanto a ação que mais minimiza os pesos na matriz de distâncias de caminho, são possíveis de verificar que é correlacionada e pouco correlacionada respectivamente com as outras ações. Isso também é comprovado também pelas Tabelas 5.9 de correlações e Tabela 5.12 de propriedades físicas.

Assim, é possível dizer que as ações que possuem mais conexões na rede *MST*, ou seja, as ações mais correlacionadas da rede, são as ações controladoras de mercado dentro do período analisado. Então, com elas, pode-se tomar como base para verificar se o mercado tende a alta ou a baixa nos preços, pois suas variações de preço fazem com que a maioria das ações sigam para a mesma direção.

Outra constatação é que as ações que minimizam as matrizes de distâncias de caminho (ver Seção 5.6) são poucos correlacionadas com as outras ações da rede, obtendo valores de correlação próximos a zero, porém nenhuma outra ação possuiu valor de correlação negativo expressivo, nem próximo a -0.5. Por isso, pode-se chamar essas ações com baixa correlação de anti-correlacionadas, uma vez que elas são as mais anti-correlacionadas da rede.

As ações anti-correlacionadas podem minimizar os prejuízos de uma variação causada a uma outra ação da carteira, ou seja, elas podem diminuir o risco específico (MARKOWITZ, 1952). Porém, é recomendado que não se negocie quando as ações controladoras de mercado tendem a baixa de preço, pois essas ações anti-correlacionadas são mais não-correlacionadas (correlação próximo a zero) do que uma correlação expressamente negativa (ver Seção 5.3.1).

Além disso, nas Figuras 5.17, 5.18 e 5.19 é possível notar que as cores, que representam os setores de cada ação, estão agrupadas, ou seja, estão próximas entre si e no mesmo ramo da árvore. Porém, para medir este agrupamento será utilizada a Modularidade, vista na próxima seção.

5.7.2 MODULARIDADE

De acordo com a Seção 2.1.4, a Modularidade quantifica a qualidade de um agrupamento. Ela compara as ligações entre os nós da mesma comunidade, menos as ligações esperadas de uma

rede aleatória com o mesmo número de arestas. De acordo com Newman e Girvan (2004), na prática, estruturas de comunidades boas têm sua modularidade entre 0.3 a 0.7. Além disso, as redes com Modularidade abaixo de 0.3 não apresentam boa estruturas de comunidades de acordo com a classificação supervisionada. Os valores de Modularidade acima de 0.7 são raramente encontrados (NEWMAN; GIRVAN, 2004).

Com isso, será medida a modularidade das 112 redes *MST* propostas neste trabalho. Portanto, as ações são classificadas de acordo com Bovespa (2012) no setor a qual cada ação pertence.

Tabela 5.13: Classificação setorial das ações.

Código da Ação	Nome da Ação	Setor
ACES4	ACESITA PN	Materiais Básicos
AELP3	AES ELPA ON	Utilidade Pública
AMBV4	AMBEV PN	Consumo não Cíclico
ARCZ6	ARACRUZ PNB	Materiais Básicos
BBAS3	BRASIL ON	Financeiro e Outros
BBAS4	BRASIL PN	Financeiro e Outros
BBDC4	BRADESCO PN	Financeiro e Outros
BRAP4	BRADESPAR PN	Financeiro e Outros
BRDT4	PETROBRAS BR PN	Petróleo, Gás e Biocombustíveis
BRKM5	BRASKEM PNA	Materiais Básicos
BRTO4	BRASIL TELEC PN	Telecomunicações
BRTP3	BRASIL T PAR ON	Telecomunicações
BRTP4	BRASIL T PAR PN	Telecomunicações
CESP4	CESP PN *	Utilidade Pública
CGAS5	COMGAS PNA	Utilidade Pública
CLSC6	CELESC PNB	Utilidade Pública
CMIG3	CEMIG ON	Utilidade Pública
CMIG4	CEMIG PN	Utilidade Pública

continua

Tabela 5.13 – *continuação*

Código da Ação	Nome da Ação	Setor
CPLE6	COPEL PNB	Utilidade Pública
CPNE5	COPENE PNA	Petróleo, Gás e Biocombustíveis
CRTP5	CRT CELULAR PNA	Telecomunicações
CRUZ3	SOUZA CRUZ ON	Consumo não Cíclico
CSNA3	SID NACIONAL ON	Materiais Básicos
CSTB4	SID TUBARAO PN *	Materiais Básicos
EBTP3	EMBRATEL PAR ON *	Telecomunicações
EBTP4	EMBRATEL PAR PN *	Telecomunicações
ELET3	ELETROBRAS ON	Utilidade Pública
ELET6	ELETROBRAS PNB	Utilidade Pública
ELPL4	ELETROPAULO PN	Utilidade Pública
EMBR3	EMBRAER ON	Bens Industriais
EMBR4	EMBRAER PN EJ	Bens Industriais
GGBR4	GERDAU PN	Materiais Básicos
GRSU3	GERASUL ON	Utilidade Pública
INEP4	INEPAR PN	Bens Industriais
ITAU4	ITAUUNIBANCO PN	Financeiro e Outros
ITSA4	ITAUSA PN	Financeiro e Outros
KLAB4	KLABIN. PN	Materiais Básicos
LIGH3	LIGHT ON *	Utilidade Pública
PETR3	PETROBRAS ON	Petróleo, Gás e Biocombustíveis
PETR4	PETROBRAS PN	Petróleo, Gás e Biocombustíveis
PLIM4	NET PN N2	Consumo Cíclico
PTIP4	IPIRANGA PET PN	Petróleo, Gás e Biocombustíveis
SBSP3	SABESP ON	Utilidade Pública
TBLE3	TRACTEBEL ON	Utilidade Pública

continua

Tabela 5.13 – *continuação*

Código da Ação	Nome da Ação	Setor
TCOC4	TELE CTR OES PN	Telecomunicações
TCSL3	TIM PART S/A ON	Telecomunicações
TCSL4	TIM PART S/A PN	Telecomunicações
TCSP3	BRASIL TELECOM	Telecomunicações
TCSP4	BRASIL TELECOM	Telecomunicações
TDBH4	TEF DATA BRA PN *	Telecomunicações
TEPR4	BRASIL TELEC PN *	Telecomunicações
TLCP4	TELE LEST CL PN	Telecomunicações
TLPP4	TELESP PN	Telecomunicações
TMAR5	TELEMAR N L PNA	Telecomunicações
TMCP4	TELEMIG PART PN	Telecomunicações
TNEP4	TELE NORD CL PN *	Telecomunicações
TNLP3	TELEMAR ON	Telecomunicações
TNLP4	TELEMAR PN	Telecomunicações
TRPL4	TRAN PAULIST PN	Utilidade Pública
TSPP4	TELESP CL PA PN	Telecomunicações
USIM5	USIMINAS PNA	Materiais Básicos
VALE5	VALE PNA	Materiais Básicos
VCPA4	V C P PN	Materiais Básicos
KLBN4	KLABIN S/A PN	Materiais Básicos
VALE3	VALE ON	Materiais Básicos
CMET4	CAEMI PN	Materiais Básicos
GOAU4	GERDAU MET PN	Materiais Básicos
NETC4	NET PN	Consumo Cíclico
UBBR11	UNIBANCO UNT	Financeiro e Outros
CTAX3	CONTAX ON	Bens Industriais

continua

Tabela 5.13 – *continuação*

Código da Ação	Nome da Ação	Setor
CTAX4	CONTAX PN	Bens Industriais
SDIA4	SADIA S/A PN	Consumo não Cíclico
BELG3	BELGO MINEIR ON *	Materiais Básicos
ARCE3	ARCELOR BR ON	Materiais Básicos
VIVO4	VIVO PN	Telecomunicações
CCRO3	CCR SA ON	Construção e Transporte
PRGA3	BRF FOODS ON	Consumo não Cíclico
ALLL11	ALL AMER LAT UNT	Construção e Transporte
CESP6	CESP PNB	Utilidade Pública
ELPL6	ELETROPAULO PNB	Utilidade Pública
NATU3	NATURA ON	Consumo não Cíclico
PCAR4	P.ACUCAR-CBD PN	Consumo não Cíclico
TAMM4	TAM S/A PN	Construção e Transporte
CSAN3	COSAN ON	Consumo não Cíclico
CYRE3	CYRELA REALT ON	Construção e Transporte
GOLL4	GOL PN	Construção e Transporte
LIGT3	LIGHT S/A ON	Utilidade Pública
SUBA3	SUBMARINO ON	Consumo não Cíclico
CPFE3	CPFL ENERGIA ON	Utilidade Pública
LREN3	LOJAS RENNER ON	Consumo Cíclico
CESP5	CESP PNA	Utilidade Pública
ELPL5	ELETROPAULO PNA	Utilidade Pública
BTOW3	B2W VAREJO ON	Consumo Cíclico
CCPR3	CYRE COM-CCP ON	Financeiro e Outros
DURA4	DURATEX PN	Materiais Básicos
GFSA3	GAFISA ON	Construção e Transporte

continua

Tabela 5.13 – *continuação*

Código da Ação	Nome da Ação	Setor
LAME4	LOJAS AMERIC PN	Consumo Cíclico
BNCA3	NOSSA CAIXA ON	Financeiro e Outros
UGPA4	ULTRAPAR PN	Financeiro e Outros
JBSS3	JBS ON	Consumo não Cíclico
RSID3	ROSSI RESID ON	Construção e Transporte
USIM3	USIMINAS ON	Materiais Básicos
BVMF3	BMFBOVESPA ON	Financeiro e Outros
RDCD3	REDECARD ON	Financeiro e Outros
ITUB4	ITAUUNIBANCO PN	Financeiro e Outros
VCPA3	V C P ON	Materiais Básicos
MMXM3	MMX MINER ON	Materiais Básicos
DTEX3	DURATEX ON	Materiais Básicos
BRTO3	BRASIL TELEC ON	Telecomunicações
FIBR3	FIBRIA ON	Materiais Básicos
BRFS3	BRF FOODS ON	Consumo não Cíclico
LLXL3	LLX LOG ON	Construção e Transporte
MRVE3	MRV ON	Construção e Transporte
OGXP3	OGX PETROLEO ON	Petróleo, Gás e Biocombustíveis
PDGR3	PDG REALT ON	Construção e Transporte
AGEI3	AGRE EMP IMO ON	Construção e Transporte
CIEL3	CIELO ON	Financeiro e Outros
ECOD3	ECODIESEL ON	Consumo não Cíclico
PCAR5	P.ACUCAR-CBD PNA	Consumo não Cíclico
BISA3	BROOKFIELD ON	Construção e Transporte
MRFG3	MARFRIG ON	Consumo não Cíclico
SANB11	SANTANDER BR UNT	Financeiro e Outros

continua

Tabela 5.13 – *continuação*

Código da Ação	Nome da Ação	Setor
ALLL3	ALL AMER LAT ON	Construção e Transporte
PRTX3	PORTX ON	Construção e Transporte
HYPE3	HYPERMARCAS ON	Consumo não Cíclico
UGPA3	ULTRAPAR ON	Financeiro e Outros
BRML3	BR MALLS PAR ON	Financeiro e Outros
HGTX3	CIA HERING ON	Consumo Cíclico
VAGR3	V-AGRO ON	Consumo não Cíclico
DASA3	DASA ON	Consumo não Cíclico
RENT3	LOCALIZA ON	Consumo Cíclico

A Tabela 5.13 mostra a classificação setorial das ações estudadas neste trabalho, de acordo com a Bovespa (2012). Cada ação se encaixa dentro de um dos nove setores: Bens Industriais; Construção e Transporte; Consumo Cíclico; Consumo não Cíclico; Financeiro e Outros; Materiais Básicos; Petróleo, Gás e Biocombustíveis; Telecomunicações; Utilidade Pública.

A Figura 5.23 mostra a medida física Modularidade em relação ao tempo das 112 redes *MST*. Como pode-se notar, grande parte das redes obtiveram valores de modularidade maior que 0.3. Assim como, as redes das Figuras 5.17, 5.18 e 5.19 possuem sua modularidade igual a 0.52, 0.45 e 0.44, respectivamente. É possível perceber que nessas redes, os agrupamentos de nós possuem homogeneidade dentro de cada setor.

A Figura 5.24 mostra a rede *MST* com a menor modularidade das redes propostas, igual a 0.31. Pode-se notar que, apesar de alguns ramos da árvore *MST* não possuírem predominância de ações do mesmo setor, muitos grupos de ações do mesmo setor ainda existem. Por outro lado, do ponto de vista econômico, entre final de 2006 a final de 2007, a Bolsa brasileira batia recordes consecutivos de valorização, além disso, com a redução da taxa de juros, muitas pessoas começaram a investir em ações (BOVESPA, 2012).

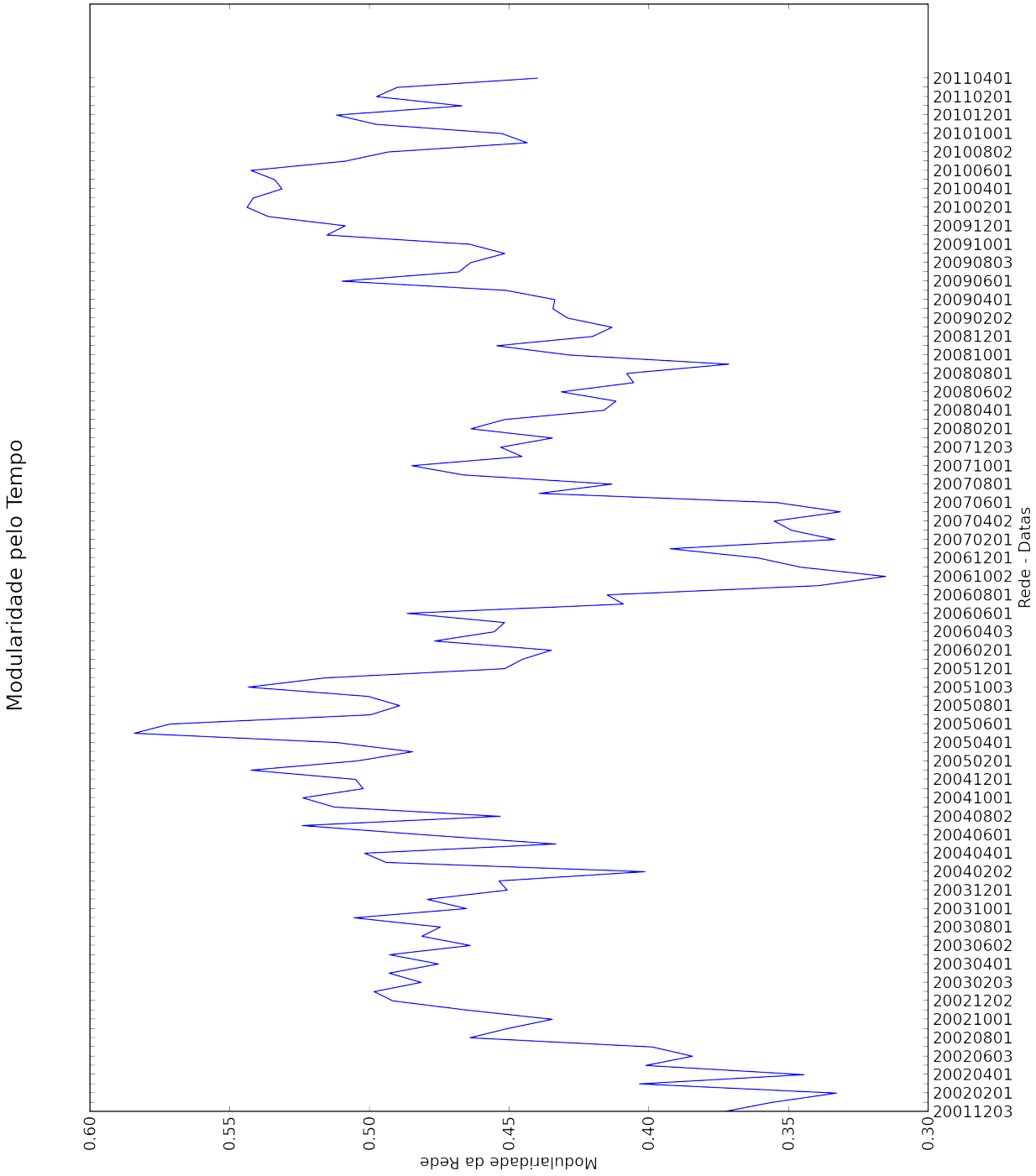

Figura 5.23: Modularidade das 112 redes *MST* de correlações intrínsecas.

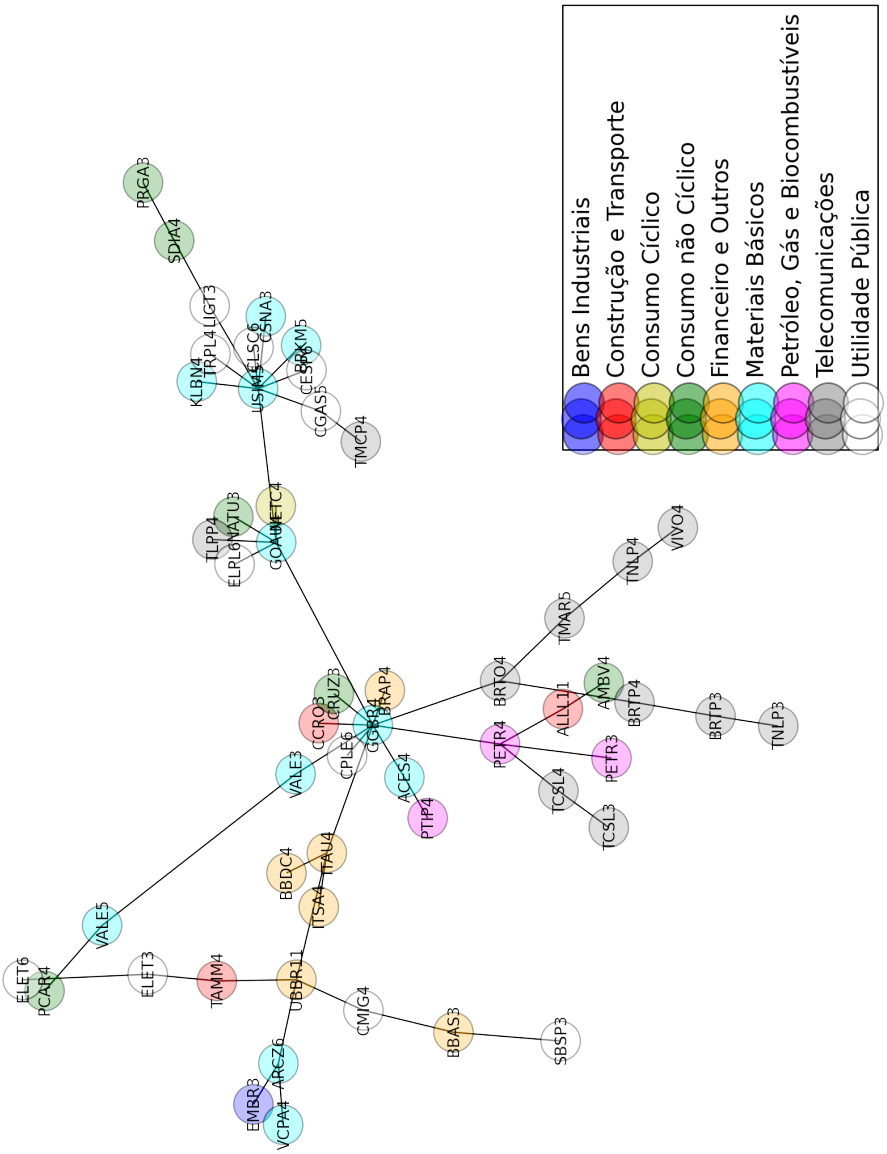

Figura 5.24: Árvore *MST* de correlações cruzadas intrínsecas de ações da data de 20061002 a 20070928.

Então, como a Modularidade das árvores *MST* são maiores que 0.30, as ações são agrupadas por sua classificação setorial. Sendo assim, pode-se dizer que o portfólio eficiente (ver Seção 2.4.3) é formado por ações de diferentes setores, uma vez que as melhores ações para otimizar uma carteira são as anti-correlacionadas entre si.

6 CONCLUSÕES

Neste trabalho foi apresentado um estudo sobre as ações da BM&F Bovespa utilizando medidas físicas de redes complexas. Através de um procedimento que considera correlações entre preços de fechamento dos ativos relacionados a iBovespa, 112 redes de mercado foram geradas entre o período de Dezembro de 2001 a Março de 2012, com uma base de dados de cotações históricas. Com a formação de diversas redes, o método proposto é capaz de indicar as evoluções das características físicas ao longo de um determinado período de tempo.

Os resultados sugerem que a medida física Densidade Média Ponderada consegue descrever momentos em que o Mercado de ações em geral está desvalorizado, de acordo com a Seção 5.2, quando esta medida esteve muito alta. Um fato que exemplifica essa constatação é que, no ano de 2008, ocorreu uma forte crise mundial. Esse período se destaca nos experimentos observados na Seção 5.2.

Além disso, a Densidade, juntamente com a Entropia de Distribuição de Pesos, revela momentos em que o Mercado de ações brasileiro estava em crescimento, assim como também em desvalorização. Quando a Entropia esteve baixa e a Densidade média esteve alta, nos períodos 2005 e 2008, ocorreram a crise política do Governo Lula e a já comentada crise econômica mundial, respectivamente. Por sua vez, nos período de 2003 e 2007, a Entropia apresentava-se alta, período em que o Mercado de ações se recuperava e batia recordes de valorização.

Sobre os coeficientes de clusterização estudados neste trabalho (ver Seção 5.5) a equação proposta por Zhang e Horvath (2005) foi a que obteve melhores resultados na modelagem proposta, por ser puramente ponderada e levar em consideração as ligações entre os nós vizinhos (Seção 5.5).

Por sua vez, na matriz de distâncias de caminhos (ver Seção 5.6) foi constatado que as ações que minimizam as distâncias de caminho; ou seja, as ações intermediárias, são exatamente

as ações mais anti-correlacionadas da rede e de menor força de vértice. Por esse motivo essas ações são as melhores para diminuir o risco específico de outra ação na carteira, otimizando o portifólio.

Outro resultado importante sobre as ações anti-correlacionadas encontradas nas redes é que elas possuem correlação próximo a zero com as outras ações, e não expressamente negativa, conforme visto na Distribuição de Pesos da Seção 5.3. Isso sugere que sejam evitadas as negociações contra a tendência de baixa do Mercado com essas ações anti-correlacionadas à tendência de Mercado, pois o risco pode ser maior neste investimento.

Quanto ao cálculo da *MST* proposto neste trabalho, pode-se observar que a utilização do subdomínio ultramétrico é válida, uma vez que possibilita a caracterização da árvore gerada (ver Seção 5.7). Essas árvores *MST* exibiram um comportamento de distribuição de graus e distribuição de forças de vértices que pôde ser aproximada a uma distribuição de lei de potência, sugerindo que as árvores *MST* sejam livres de escalas (ver Seção 5.7.1).

A topologia livre de escala sugere que, hierarquicamente, as ações que possuem mais ligações na árvore *MST* controlam ou intermedem as flutuações de preços diários de suas ações vizinhas. As ações que ligam as ações mais conectadas com as ações das extremidades da *MST* intermedem as variações de preços. Enquanto as ações da extremidade da árvore são ações menos influentes, as mais anti-correlacionadas da rede, e que são menos influenciadas pelas variações do mercado, são ações que não necessariamente seguem os índices econômicos.

Além disso, foi constatado que as ações com mais ligações na rede *MST* são exatamente aquelas controladoras de mercado no período de análise da rede, conforme visto na Seção 5.7.1, uma vez que possuem alta correlação com a maioria das outras ações da rede; ou seja, uma variação na ação controladora de mercado, variam as outras ações, de acordo com a correlação entre elas. Assim, pode-se verificar tendências de alta ou de baixa para o mercado de ações com a ação controladora de mercado.

Por sua vez, pode-se verificar que as ações brasileiras formam agrupamentos (*clusters*) com base no setor de classificação das empresas. De acordo com a medida física Modularidade, pode-se verificar a qualidade desses agrupamentos. Como foram obtido valores de Modularidade maiores

que 0.3 (ver Seção 5.7.2), pode-se dizer que a qualidade dos agrupamentos das ações pelo setor de classificação é boa. Isso sugere que as melhores ações, para a otimização do portifólio, são de diferentes setores, uma vez que dentro dos agrupamentos as correlações entre as ações são altas, porém, entre os agrupamentos as correlações são menores.

Finalmente, para trabalhos futuros pode-se propor a análise de redes com janelas de períodos menores e de fechamento de preços a cada uma hora, ou janelas maiores de 4 anos com preços de fechamento semanal. Assim, pode-se verificar qual a melhor janela a ser utilizada para as redes de ações e suas correlações, ou ainda fazer análises para rápidas negociações.

Outra possibilidade é estudar as redes divididas por setores, agrupando as ações do mesmo setor, uma vez que constatou-se que as ações realmente agrupam-se pelo setor. Isso possibilita a diminuição de ruídos na rede, além de poder verificar qual setor é mais dominante no mercado de ações para um possível investimento ou verificação de tendências.

Uma análise interessante é mostrada na trabalho de Kim et al. (2002a), onde são desconsiderados fatores externos que influenciam a economia americana, propondo uma rede de correlações de ações com base no valor intrínseco, obtendo puramente as correlações entre as ações. Este também pode ser um caminho de estudo futuro para o mercado brasileiro.

Além disso, os outros dados não utilizados na base de dados da BM&F Bovespa, como as informações de tipo de mercado, os preços de abertura, mínimo, máximo e médio, número de negócios efetuados, entre outros, poderão ser utilizados em um trabalho futuro para especificar ou atribuir informações *a priori* para a modelagem de rede.

REFERÊNCIAS

AHUJA, R. K.; MAGNANTI, T. L.; ORLIN, J. *Network Flows: Theory, Algorithms, and Applications*. [S.l.]: Prentice Hall, 1993.

AIELLO, W.; CHUNG, F.; LU, L. A random graph model for massive graphs. In: *32nd annual ACM synposium od theory of computing*. [S.l.: s.n.], 2002. p. 171–180.

ALLESINA, S.; PASCUAL, M. Network structure, predator-prey modules, and stability in large food webs. *Theoretical Ecology*, v. 1, n. 1, p. 55–64, mar. 2008.

ARAGON, G. A. *Financial Management*. [S.l.]: Allyn & Bacon, 1989.

ARTHUR, W. B.; DURLAUF, S. N.; LANE, D. A. *The economy as an evolving complex system II*. Reading, Massachusetts: Addison-Wesley, 1997. (Santa Fé institute studies in the sciences of complexity).

AUERBACH, F. Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, LIX, p. 73–76, 1913.

BALL, F.; MOLLISON, D.; SCALIA-TOMBA, G. Epidemics with two levels of mixing. In: STATISTICS, I. of M. (Ed.). *The Annals of Applied Probability*. [S.l.: s.n.], 1997. v. 7, p. 46–89.

BALLOBáS, B. *Random Graphs*. 2nd. ed. New York: Academic Press, 2001.

BANG-JENSEN, J.; GUTIN, G. *Digraphs: Theory, Algorithms and Applications (Second Edition)*. [S.l.]: Springer-Verlag, Berlin, 2004.

BARABASI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *SCIENCE*, v. 286, p. 509–512, 1999.

BARABASI, A.-L.; BONABEAU, E. Scale-free networks. *Scientific American*, p. 50–59, May 2003.

BARABáSI, A.-L. *Linked: The New Science of Networks*. Cambridge, CA: Perseus, 2002.

BARRAT, A.; BARTHéLEMY, M.; PASTOR-SATORRAS, R.; VESPIGNANI, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Sciences, v. 101, n. 11, p. 3747–3752, mar. 2004. ISSN 1091-6490.

BARTHELEMY, M.; BARRAT, A.; PASTORSATORRAS, R.; VESPIGNANI, A. Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and its Applications*, v. 346, n. 1-2, p. 34–43, fev. 2005. ISSN 03784371.

BAUM, J. A.; ROWLEY, T. J.; SHIPILOV, A. V. The small world of canadian capital markets: Statistical mechanics of investment bank syndicate networks, 1952 1989. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, Blackwell Publishing, v. 21, n. 4, p. 307–325, 2004. ISSN 1936-4490.

BENDER, E. A.; CANFIELD, E. R. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, v. 24, n. 3, p. 296 – 307, 1978. ISSN 0097-3165.

BODIE, Z.; KANE, A.; MARCUS, A. J. *Investments*. 8. ed. [S.l.]: McGraw-Hill/Irwin, 2009.

BONANNO, G.; CALDARELLI, G.; LILLO, F.; MANTEGNA, R. N. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, APS, v. 68, n. 4, p. 046130+, 2003.

BONANNO, G.; CALDARELLI, G.; LILLO, F.; MICCICHé, S.; VANDEWALLE, N.; MANTEGNA, R. N. Networks of equities in financial markets. *European Physical Journal B*, v. 38, p. 363–371, 2004.

BONANNO, G.; VANDEWALLE, N.; MANTEGNA, R. N. Taxonomy of stock market indices. v. 62, n. 6, p. R7615–R7618, December 2000.

BONDY, J. A. *Graph Theory With Applications*. [S.l.]: Elsevier Science, 1976. ISBN 0444194517.

BOVESPA. *Mercado de Capitais*. 2008.

BOVESPA. *BM&F Bovespa*. 2012. Disponível em: <http://www.bmfbovespa.com.br/>. Acesso em: 14 Jan. 2013.

BUCHANAN, M. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. [S.l.]: W. W. Norton & Company, 2002. ISBN 0393041530.

BUCKLEY, F.; HARARY, F. *Distance in Graphs*. [S.l.]: Addison-Wesley, 1990.

CALDARELLI, G.; BATTISTON, S.; GARLASCHELLI, D.; CATANZARO, M. *Emergence of Complexity in Financial Networks*. [S.l.: s.n.], 2004. 399–423 p.

CAMPBELL, R. A.; FORBES, C. S.; KOEDIJK, K. G.; KOFMAN, P. Increasing correlations or just fat tails? *Journal of Empirical Finance*, v. 15, n. 2, p. 287–309, March 2008.

CHUNG, F.; CHUNG, F.; CHUNG, F.; LU, L.; LU, L. The average distances in random graphs with given expected degrees. *Internet Mathematics*, v. 1, p. 15879–15882, 2002.

COHEN, R.; HAVLIN, S. Scale-free networks are ultrasmall. *PHYS.REV.LETT*, v. 90, p. 058701, 2003.

CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. *Introduction to Algorithms*. 2nd revised edition. ed. [S.l.]: The MIT Press, 2001. Taschenbuch. ISBN 0262531968.

COSTA, L. F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: A survey of measurements. *Advances in Physics*, Taylor & Francis, v. 56, n. 1, p. 167–242, ago. 2006.

DEGENNE, A.; FORSé, M. *Introducing Social Networks*. London: Sage, 1999.

DROZDZ, S.; GRÜMMER, F.; GÓRSKI, A. Z.; RUF, F.; SPETH, J. Dynamics of competition between collectivity and noise in the stock market. *PHYSICA A*, v. 287, p. 440–449, 2000.

ERDOS, P.; RENYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, v. 5, p. 17–61, 1960.

FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. In: *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*. New York, NY, USA: ACM, 1999. (SIGCOMM '99), p. 251–262. ISBN 1-58113-135-6.

FARKAS, I. J.; JEONG, H.; VICSEK, T.; BARABáSI, A. L.; OLTVAI, Z. N. The topology of the transcription regulatory network in the yeast, s. cerevisiae. *Physica A*, v. 318, n. cond-mat/0205181, p. 3–4. 18 p, May 2002.

FELLER, W. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition*. 3. ed. [S.l.]: Wiley, 1968. Hardcover. ISBN 0471257087.

FERGUSSON, K.; PLATEN, E. On the distributional characterization of daily logâreturns of a world stock index. *Applied Mathematical Finance*, v. 13, n. 1, p. 19–38, 2006.

FURBUSH, D. *"Program Trading." The Concise Encyclopedia of Economics.* [S.l.]: Library of Economics and Liberty, 1993.

GABAIX, X.; GOPIKRISHNAN, P.; PLEROU, V.; STANLEY, H. E. A theory of power-law distributions in financial market fluctuations. *Nature*, Nature Publishing Group, v. 423, n. 6937, p. 267–270, maio 2003. ISSN 0028-0836.

GIRVAN, M.; NEWMAN, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA., v. 99, n. 12, p. 7821–7826, jun. 2002. ISSN 0027-8424.

Heuristics for Internet map discovery, v. 3. 1371–1380 vol.3 p.

GOWER, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, v. 53, n. 3-4, p. 325–338, 1966.

HARARY, F. *Graph Theory*. [S.l.]: Addison-Wesley, 1969.

HOLLAND, P. W.; LEINHARDT, S. Transitivity in structural models of small groups. *Comparative Group Studies 2*, p. 107124, 1971.

HUANG, W.-Q.; ZHUANG, X.-T.; YAO, S. A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications*, v. 388, n. 14, p. 2956 – 2964, 2009. ISSN 0378-4371.

JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z. N.; BARABáSI, A.-L. *The Large-Scale Organization of Metabolic Networks*. 2000.

JUNG, W.-S.; CHAE, S.; YANG, J.-S.; MOON, H.-T. Characteristics of the korean stock market correlations. *Physica A: Statistical Mechanics and its Applications*, v. 361, n. 1, p. 263 – 271, 2006. ISSN 0378-4371.

JUNG, W.-S.; KWON, O.; WANG, F.; KAIZOJI, T.; MOON, H.-T.; STANLEY, H. E. Group dynamics of the japanese market. *Physica A: Statistical Mechanics and its Applications*, v. 387, n. 2-3, p. 537–542, jan. 2008.

KIM, H.-J.; KIM, I.-M.; LEE, Y.; B., K. Scale-free network in stock markets. *Journal of the Korean Physical Society*, v. 4, n. 6, p. 1105–1108, June 2002.

KIM, H.-J.; LEE, Y.; KAHNG, B.; KIM, I. mook. Weighted scale-free network in financial correlations. *Journal of the Physical Society of Japan*, v. 71, p. 2133, 2002.

KNOKE, D.; KUKLINSKI, J. H. *Networks Analysis*. Beverly Hills, CA: Sage Publications, 1982.

KNUTH, D. E. *The Stanford GraphBase - a platform for combinatorial computing.* [S.l.]: ACM, 1993. I-VII, 1-576 p. ISBN 978-0-201-54275-2.

KULLMANN, L.; KERTéSZ, J.; KASKI, K. Time-dependent cross-correlations between different stock returns: a directed network of influence. *Phys Rev E Stat Nonlin Soft Matter Phys*, v. 66, n. 2 Pt 2, p. 026125, 2002. ISSN 1539-3755.

KUPERMAN, M.; ABRAMSON, G. Small world effect in an epidemiological model. *PHYS.REV.LETT.*, v. 86, p. 2909, 2001.

LONGIN, F.; SOLNIK, B. Extreme Correlation of International Equity Markets. *The Journal of Finance*, Blackwell Publishers, v. 56, n. 2, p. 649–676, 2001.

LOPES, G. A. W. *Um Modelo de Redes Complexas para Análise de Informações Textuais*. Dissertação (Mestrado) — Centro Universitário da FEI, 2011.

LOPEZ-FERNANDEZ, L.; ROBLES, G.; BARAHONA, G. Applying Social Network Analysis to the Information in CVS Repositories. In: *Proceedings of the Mining Software Repositories Workshop. 26th International Conference on Software Engineering*. [S.l.: s.n.], 2004.

LUSSEAU, D. The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, Department of Zoology, University of Otago, PO Box 56, Dunedin, New Zealand. lusda563@student.otago.ac.nz, v. 270, n. Suppl 2, p. S186–S188, nov. 2003.

LYNCH, N. *Distributed Algorithms*. San Francisco, CA: Morgan Kaufmann, 1996.

MANTEGNA, R. N. Hierarchical structure in financial markets. *European Physical Journal B*, v. 11, p. 193–197, 1999.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. *Multivariate analysis*. [S.l.]: Academic Press; New York, 1979.

MARKOWITZ, H. Portfolio Selection. *The Journal of Finance*, Blackwell Publishing for the American Finance Association, v. 7, n. 1, p. 77–91, 1952. ISSN 00221082.

MILGRAM, S. The small-world problem. *Psychology Today*, v. 1, n. 1, p. 61–67, 1967.

MILO, R.; KASHTAN, N.; ITZKOVITZ, S.; NEWMAN, M. E. J.; ALON, U. On the uniform generation of random graphs with prescribed degree sequences. maio 2004.

MOLLOY, M.; A, P. P. U. S.; REED, B.; COMBINATOIRE, E. A critical point for random graphs with a given degree sequence. In: . [S.l.: s.n.], 1995. p. 161–179.

MOLLOY, M.; REED, B. The size of the giant component of a random graph with a given degree sequence. *COMBIN. PROBAB. COMPUT*, v. 7, p. 295–305, 2000.

MONASSON, R. *Diffusion, localization and dispersion relations on "small-world" lattices*. 1999.

NAGURNEY, A. *Network Economics: A Variational Inequality Approach*. Dordrecht: Kluwer Academic Publishers, 1993.

NEWMAN, M.; BARABASI, A.-L.; WATTS, D. J. *The structure and dynamics of networks*. [S.l.]: Princeton University Press, 2006. ISBN 978-0-691-11356-2.

NEWMAN, M. E.; STROGATZ, S. H.; WATTS, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E Stat Nonlin Soft Matter Phys*, Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA., v. 64, n. 2 Pt 2, ago. 2001. ISSN 1539-3755.

NEWMAN, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, Taylor & Francis, v. 46, n. 5, p. 323–351, maio 2005. ISSN 0010-7514.

NEWMAN, M. E. J. Random graphs as models of networks. In: ____. *Handbook of Graphs and Networks*. [S.l.]: Wiley-VCH Verlag GmbH & Co. KGaA, 2005. p. 35–68. ISBN 9783527602759.

NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. *Physical Review E*, American Physical Society, v. 69, n. 2, p. 026113+, Feb 2004.

ONNELA, J. P.; CHAKRABORTI, A.; KASKI, K.; KERTéSZ, J. Dynamic asset trees and portfolio analysis. *The European Physical Journal B*, v. 30, p. 285–288, 2002.

ONNELA, J.-P.; CHAKRABORTI, A.; KASKI, K.; KERTESZ, J. Dynamic asset trees and black monday. v. 324, p. 247–252, 2003.

ONNELA, J. P.; CHAKRABORTI, A.; KASKI, K.; KERTÉSZ, J.; KANTO, A. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, APS, v. 68, n. 5, p. 56110, 2003.

ONNELA, J. P.; KASKI, K.; KERTESZ, J. Clustering and information in correlation based financial networks. *European Physical Journal B*, v. 38, p. 353–362, 2004.

ONNELA, J.-P.; SARAMäKI, J.; KERTéSZ, J.; KASKI, K. Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, American Physical Society, v. 71, n. 6, p. 065103+, jun. 2005.

PANDIT, S. A.; AMRITKAR, R. E. Random spread on the family of small-world networks. *arXivcondmat0004163*, APS, v. 63, n. 4, p. 41104, 2000.

PARK, J.; NEWMAN, M. E. J. The origin of degree correlations in the internet and other networks. v. 68, n. 026112, 2003.

PLEROU, V.; GOPIKRISHNAN, P.; STANLEY, H. E. Econophysics: Two-phase behaviour of financial markets. *Nature*, v. 421, p. 130+, jan. 2003.

PRICE, D. d. S. Network of scientific papers. *Science*, n. 149, p. 510–515, 1965.

RAPOPORT, A. Mathematical models of social interaction. In: PSYCHOLOGY, H. of M. (Ed.). *R. D. Luce, R. R. Bush, E. Galanter*. [S.l.: s.n.], 1963.

RAPOPORT, A.; HORVATH, W. J. A study of a large sociogram. *Behavioral Science*, John Wiley e Sons, v. 6, n. 4, p. 279–291, 1961. ISSN 1099-1743.

REDNER, S. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, v. 4, n. 2, p. 131–134, ago. 1998. ISSN 1434-6028.

RODGERS, J. L.; NICEWANDER, W. A. Thirteen ways to look at the correlation coefficient. *The American Statistician*, v. 42, n. 1, p. 59–66, 1988.

RODRIGUES, P. S.; GIRALDI, G. A. Improving the non-extensive medical image segmentation based on tsallis entropy. *Pattern Analysis and Applications*, to appear, 2010.

RODRIGUES, P. S.; GIRALDI, G. A.; ARAUJO, A. A. Using tsallis entropy into a bayesian network for cbir. In: IEEE (Ed.). *Proceedings of International International Conference on Image Processing (ICIP'05)*. Genova, Italy: [s.n.], 2005. v. 3, p. 1028–1031.

SARAMäKI, J.; KIVELä, M.; ONNELA, J. P.; KASKI, K.; KERTéSZ, J. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, American Physical Society, v. 75, n. 2, p. 027105+, fev. 2007.

SCHANK, T.; WAGNER, D. Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications*, v. 9, p. 2005, 2005.

SCOTT, J. P. *Social Network Analysis: A Handbook*. [S.l.]: SAGE Publications, 2000. Paperback. ISBN 0761963391.

SHANNON, C.; WEAVER, W. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1948.

SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, JSTOR, v. 19, n. 3, p. 425–442, 1964.

SHEN-ORR, S. S.; MILO, R.; MANGAN, S.; ALON, U. Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, Nature Publishing Group, v. 31, n. 1, p. 64–68, May 2002.

SOLOMONOFF, R.; RAPOPORT, A. Connectivity of random nets. *Bulletin of Mathematical Biophysics*, n. 13, p. 107–117, 1951.

S&P. *S&P 500*. 2013. Disponível em: <http://www.standardandpoors.com/indices/sp-500/en/eu/?indexId=spusa-500-usduf–p-us-l–>. Acesso em: 14 Jan. 2013.

SPIEGEL, M. R.; SCHILLER, J. J.; SRINIVASAN, R. A. *Schaum's Outline of Probability and Statistics*. 2. ed. [S.l.]: McGraw-Hill, 2000. ISBN 0071350047.

SPORNS, O. Graph theory methods for the analysis of neural connectivity patterns. *Complex.*, v. 8, n. 1, p. 56–60, 2002. ISSN 1076-2787.

STANLEY, H.; AMARAL, L.; CANNING, D.; GOPIKRISHNAN, P.; LEE, Y.; LIU, Y. *Econophysics: Can physicists contribute to the science of economics?* 1999.

STANLEY, H. E.; AFANASYEV, V.; AMARAL, L. A. N.; BULDYREV, S. V.; GOLDBERGER, A. L.; HAVLIN, S.; LESCHHORN, H.; MAASS, P.; MANTEGNA, R. N.; PENG, C. K.; PRINCE, P. A.; SALINGER, M. A.; STANLEY, M. H. R.; VISWANATHAN, G. M. Anomalous fluctuations in the dynamics of complex systems: from dna and physiology to econophysics. *Physica A: Statistical and Theoretical Physics*, v. 224, n. 1-2, p. 302–321, fev. 1996.

TABAK, B. M.; SERRA, T. R.; CAJUEIRO, D. O. Topological properties of stock market networks: The case of brazil. *Physica A: Statistical Mechanics and its Applications*, v. 389, n. 16, p. 3240 – 3249, 2010. ISSN 0378-4371.

TSE, C. K.; LIU, J.; LAU, F. C. A network perspective of the stock market. *Journal of Empirical Finance*, v. 17, n. 4, p. 659 – 667, 2010. ISSN 0927-5398.

WAGNER, A.; FELL, D. *The Small World Inside Large Metabolic Networks*. 2000.

WALSH, T. Search in a small world. In: *IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1999. p. 1172–1177. ISBN 1-55860-613-0.

WASSERMAN, S.; FAUST, K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994.

WATTS, D. J. Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, v. 105, p. 493–527, 1999.

WATTS, D. J. *Small worlds: The dynamics of networks between order and randomness*. Princeton, NJ: Princeton University Press, 1999. xv, 262 p.

WATTS, D. J. *Six Degrees: The Science of a Conneted Age*. New York: Norton, 2003.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of small-world networks. *Nature*, v. 393, n. 6684, p. 440–442, June 1998.

WEST, D. B. *Introduction to Graph Theory*. Upper Saddle River, NJ: Prentice Hall, 1996.

ZACHARY, W. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, v. 33, p. 452–473, 1977.

ZHANG, B.; HORVATH, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, v. 4, n. 1, 2005. Article17.

ZHANG, J.; ZHOU, H.; JIANG, L.; WANG, Y. Network topologies of Shanghai stock index. *Physics Procedia*, v. 3, n. 5, p. 1733–1740, ago. 2010.