

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE  
RIBEIRÃO PRETO  
DEPARTAMENTO DE ADMINISTRAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO DE ORGANIZAÇÕES

GISELE SIQUEIRA MONTANARI

Árvores geradoras mínimas do mercado financeiro a partir de distâncias entre distribuições de probabilidade do retorno do preço de ações

ORIENTADOR: PROF. DR. GILBERTO  
APARECIDO PRATAVIERA

RIBEIRÃO PRETO  
2019

Prof. Dr. Vahan Agopyan  
Reitor da Universidade de São Paulo

Prof. Dr. André Lucirton Costa  
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Jorge Henrique Caldeira de Oliveira  
Chefe do Departamento de Administração

GISELE SIQUEIRA MONTANARI

Árvores geradoras mínimas do mercado financeiro a partir de distâncias entre distribuições de probabilidade do retorno do preço de ações

Versão corrigida. O original está disponível na FEA-RP

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Organizações da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

ORIENTADOR: PROF. DR. GILBERTO APARECIDO PRATAVIERA

RIBEIRÃO PRETO  
2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Montanari, Gisele Siqueira

Árvores geradoras mínimas do mercado financeiro a partir de distâncias entre distribuições de probabilidade do retorno do preço de ações / Gisele Siqueira Montanari – Ribeirão Preto, 2019.

71f.: il.; 30 cm

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Organizações da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências. – Universidade de São Paulo

ORIENTADOR: PRATAVIERA, GILBERTO APARECIDO

GISELE SIQUEIRA MONTANARI

**Árvores geradoras mínimas do mercado financeiro a partir de distâncias entre distribuições de probabilidade do retorno do preço de ações**

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Organizações da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de Concentração: Administração de Organizações

**Data de Aprovação:**

\_\_\_\_/\_\_\_\_/\_\_\_\_

**Banca Examinadora:**

---

**PROF. DR. GILBERTO  
APARECIDO PRATAVIERA  
ORIENTADOR**

---

**Professor**  
Avaliador 1

---

**Professor**  
Avaliador 2

---

**Professor**  
Avaliador 3

Dedico este trabalho os meus pais Nivaldo e Zélia, meu esposo Alessandro e nossa filha Isadora.

## **AGRADECIMENTOS**

A Deus pelo dom da vida e pelos caminhos que me permitiu trilhar.

A meus pais Nivaldo e Zélia por me ensinarem a não desistir, mas persistir e alcançar.

A meu esposo Alessandro por todo o apoio, compreensão e auxílio durante esta jornada. Por cuidar tão bem de nossa filha Isadora em minhas ausências.

Agradeço IMENSAMENTE a meu orientador professor Dr. Gilberto Aparecido Prativiera por tudo que me ensinou sobre probabilidades, estatísticas e pesquisa de modo geral, agradeço por sua disponibilidade e pelo tanto que me apoiou, se doou e se empenhou para a conclusão deste projeto.

A todos os professores da FEA-USP de Ribeirão Preto e São Paulo que contribuíram para com meu aprendizado e em especial ao professor Dr. Evandro Saidel Ribeiro por todo o conhecimento compartilhado, sobre R, sobre redes e por sua disponibilidade.

A todos os colaboradores da pós-graduação, em especial ao Matheus Carlos da Costa pela compreensão de cada dificuldade que surgiu e proposta de solução para cada caso. Desejo a todos muito sucesso e reconhecimento profissional.

Que Deus os abençoe!

*"Uma longa caminhada começa com o primeiro passo." Lao Tsé*



## RESUMO

MONTANARI, Gisele Siqueira. *Árvores geradoras mínimas do mercado financeiro a partir de distâncias entre distribuições de probabilidade do retorno do preço de ações*. 2019. 71f. Dissertação (Mestrado em Administração de Organizações) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2019.

Neste trabalho propõe-se a obtenção de árvores geradoras mínimas do mercado financeiro a partir de uma distância entre as distribuições de probabilidade do retorno do preço de ações. A distância proposta permite definir uma intensidade de ligação entre os ativos do mercado que é relacionada a similaridade entre suas distribuições de probabilidade. Como aplicação, foram obtidas as árvores geradoras mínimas mensais da carteira de ações do IBOVESPA para os meses de maio, junho, julho e agosto de 2019 a partir das distribuições empíricas de probabilidade dos retornos de preços com transações registradas a cada 15 minutos. As árvores geradoras mínimas obtidas foram caracterizadas por suas distribuições de grau simples e de grau ponderado e comparadas com os resultados para árvores geradoras mínimas baseadas em correlação. A distribuição de grau simples para as árvores mínimas baseadas em distribuição de probabilidade apresentaram menor dispersão e um número de menor de ações na periferia da rede (ações com grau 1) quando comparadas com as redes baseadas em correlação. A distribuição de grau ponderado das redes baseadas em distribuição de probabilidade apresentaram grau médio ponderado praticamente constante ao longo dos períodos e com valores sempre acima do grau médio ponderado das redes baseadas em correlação. A classificação das ações nos graus ponderados das redes baseadas em distribuição de probabilidade tem pouca variabilidade se consideradas ações com mesmo grau simples quando comparadas com as redes baseadas em correlação.

**Palavras-chave:** Redes. Árvore geradora mínima. Mercado de ações. Distribuição de probabilidade. Distância entre distribuições de probabilidade. Correlação.

## ABSTRACT

MONTANARI, Gisele Siqueira. *Financial market minimum spanning trees from distances between probability distributions of stock price returns*. . 2019. 71f. Dissertação (Mestrado em Administração de Organizações) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2019.

This work proposes financial market minimum spanning trees from a distance between probability distributions of stock price returns. The proposed distance allows to define an intensity for the interaction between market assets that is related to the similarity between their probability distributions. As an application, the monthly minimum spanning trees of the IBOVESPA stock portfolio for May, June, July and August 2019 were obtained from the empirical probability distributions of stock price returns with trade recorded every 15 minutes. The minimum spanning trees were characterized by their simple and weighted degree distributions and compared with the minimum spanning trees obtained from correlations between stock prices returns. The simple degree distribution for the minimum spanning trees based on probability distribution shows less dispersion and fewer stocks on the network periphery (stocks with degree 1) when compared to the correlation-based minimum spanning trees. The weighted degree distribution for networks based on probability distributions showed higher and a nearly constant values of average weighted degree over the periods when compared with the average weighted degree of correlation minimum spanning trees. The weighted degree for stocks with the same simple degree has low variation for minimum spanning trees from probability distribution when they are compared to the correlation-based minimum spanning trees.

**Keywords:** Networks. Minimum spanning tree. Stock Markets. Probability distribution. Probability distributions distance. Correlation.

## LISTA DE ILUSTRAÇÕES

Figura 1 – O grafo das pontes de Königsberg . . . . .	27
Figura 2 – Exemplos de redes . . . . .	28
Figura 3 – Exemplo de rede direcionada . . . . .	28
Figura 4 – Exemplo de rede bipartida . . . . .	30
Figura 5 – Exemplo de rede conectada e rede do tipo árvore . . . . .	33
Figura 6 – Algoritmo de Kruskal . . . . .	34
Figura 7 – Exemplo de aplicação do algoritmo de Kruskal . . . . .	35
Figura 8 – Exemplo de função distribuição acumulada empírica . . . . .	39
Figura 9 – Valor em risco a partir da distribuição de probabilidade acumulada. . . . .	40
Figura 10 – Distância entre distribuições de probabilidade . . . . .	44
Figura 11 – Lista de ações da carteira do IBOVESPA . . . . .	47
Figura 12 – Árvore geradora mínima do mês de maio de 2019 . . . . .	49
Figura 13 – Árvore geradora mínima do mês de junho de 2019 . . . . .	50
Figura 14 – Árvore geradora mínima do mês de julho de 2019 . . . . .	50
Figura 15 – Árvore geradora mínima do mês de agosto de 2019 . . . . .	51
Figura 16 – Diagrama de caixa da intensidade das ligações dos elementos das árvores geradoras mínimas . . . . .	52
Figura 17 – Distribuição de grau de simples das árvores geradoras mínimas . . . . .	53
Figura 18 – Desvio padrão das distribuições de grau das árvores geradoras mínimas . . . . .	54
Figura 19 – Robustez das árvores geradoras mínimas . . . . .	55
Figura 20 – Grau das ações para o mês de maio de 2019 . . . . .	56
Figura 21 – Grau das ações para o mês de junho de 2019 . . . . .	57
Figura 22 – Grau das ações para o mês de julho de 2019 . . . . .	58
Figura 23 – Grau das ações para o mês de agosto de 2019 . . . . .	59
Figura 24 – Diagrama de caixa da distribuição de grau ponderado . . . . .	60
Figura 25 – Grau médio e desvio padrão da distribuição de grau ponderado das árvores geradoras mínimas. . . . .	61
Figura 26 – Grau ponderado das ações para o mês de maio de 2019 . . . . .	63
Figura 27 – Grau ponderado das ações para o mês de junho de 2019 . . . . .	64
Figura 28 – Grau ponderado das ações para o mês de julho de 2019 . . . . .	65
Figura 29 – Grau ponderado das ações para o mês de agosto de 2019 . . . . .	66

## LISTA DE QUADROS

Quadro 1 – Estrutura dos dados: preços de ações . . . . .	45
Quadro 2 – Estrutura dos dados: retorno de ações . . . . .	48

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>21</b>
<b>1.1</b>	<b>Problema de Pesquisa</b> . . . . .	<b>24</b>
<b>1.2</b>	<b>Objetivos</b> . . . . .	<b>25</b>
<b>1.2.1</b>	<b>Gerais</b> . . . . .	<b>25</b>
<b>1.2.2</b>	<b>Específicos</b> . . . . .	<b>25</b>
<b>1.3</b>	<b>Justificativa</b> . . . . .	<b>25</b>
<b>1.4</b>	<b>Estrutura do trabalho</b> . . . . .	<b>25</b>
<b>2</b>	<b>UMA BREVE INTRODUÇÃO A TEORIA DE REDES</b> . . . . .	<b>27</b>
<b>2.1</b>	<b>Conceitos básicos</b> . . . . .	<b>27</b>
<b>2.2</b>	<b>Métricas de redes</b> . . . . .	<b>29</b>
<b>2.3</b>	<b>Redes tipo árvore e árvores geradoras mínimas</b> . . . . .	<b>33</b>
<b>2.4</b>	<b>O algoritmo de Kruskal</b> . . . . .	<b>34</b>
<b>3</b>	<b>DISTRIBUIÇÕES DE PROBABILIDADE, RISCO, CORRELAÇÕES E DISTÂNCIAS</b> . . . . .	<b>37</b>
<b>3.1</b>	<b>Distribuição de probabilidades</b> . . . . .	<b>37</b>
<b>3.2</b>	<b>A função distribuição acumulada empírica</b> . . . . .	<b>38</b>
<b>3.3</b>	<b>Função distribuição acumulada e risco financeiro</b> . . . . .	<b>39</b>
<b>3.4</b>	<b>Correlação</b> . . . . .	<b>40</b>
<b>3.5</b>	<b>Distâncias</b> . . . . .	<b>41</b>
<b>4</b>	<b>METODOLOGIA</b> . . . . .	<b>45</b>
<b>4.1</b>	<b>A Base de dados</b> . . . . .	<b>45</b>
<b>4.2</b>	<b>Obtenção das árvores geradoras mínimas</b> . . . . .	<b>47</b>
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>49</b>
<b>5.1</b>	<b>Árvores geradoras mínimas com ligações simples</b> . . . . .	<b>52</b>
<b>5.2</b>	<b>Árvores geradoras mínimas ponderadas</b> . . . . .	<b>60</b>
<b>6</b>	<b>CONCLUSÕES</b> . . . . .	<b>67</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>69</b>

## 1 INTRODUÇÃO

A teoria de redes, que tem origem na área da matemática chamada teoria de grafos (ERDÖS, 1959; KREYSZIG, 1999; BOLLOBÁS, 2013), tem sido amplamente utilizada para o estudo de sistemas compostos por várias partes que interagem mutuamente (CARRINGTON; SCOTT; WASSERMAN, 2005; NEWMAN, 2010; JACKSON, 2010; BARABÁSI, 2016). Rede é um conjunto de itens, chamados de vértices (nós), com conexões entre eles, chamadas por arestas (NEWMAN, 2010; BARABÁSI, 2016). Essa definição simples permite abordar um número extenso de sistemas físicos, biológicos, econômicos e sociais compostos por unidades que interagem entre si. Juntamente com teorias físicas, biológicas e sociais, a teoria de redes é uma das ferramentas que permitem compreender melhor teorias estabelecidas, identificar interações e relações entre as partes de sistemas interagentes e a formação de padrões em sistemas compostos de várias unidades que atuam conjuntamente (BARABÁSI, 2016).

Dentre os vários tipos de redes, as do tipo árvore estão entre as mais importantes. Uma árvore é uma rede conectada sem ciclos e que pode ser usada para expor, organizar ou analisar redes elétricas, relações produtor-consumidor e outras relações comerciais, informações em sistemas de banco de dados, estrutura sintática de programas de computador, etc. (KREYSZIG, 1999). A árvore geradora de uma rede é uma árvore contendo todos os vértices da rede. Em uma rede com  $n$  vértices, a árvore geradora mínima é uma árvore conectando  $n - 1$  vértices em que a soma das distâncias entre os vértices é a menor dentre todas as árvores possíveis (KRUSKAL, 1956; KREYSZIG, 1999).

Um das áreas onde a teoria de redes tem sido intensivamente aplicada é no estudo do comportamento do mercado financeiro (MANTEGNA; STANLEY, 1999; EASLEY; KLEINBERG, 2010; SINHA et al., 2010). Notadamente, a proposta de Mantegna (1999) de obtenção de árvores geradoras mínimas usando uma distância baseada nas correlações entre as séries temporais de ativos do mercado revelou-se uma técnica efetiva para estes estudos. A árvore geradora mínima resultante representa um resumo do mercado nos quais são selecionadas as correlações que representam as interações mais importantes. A árvore geradora mínima obtida pode ser associada a análise hierárquica de grupos em que a hierarquia detectada na estrutura da rede mínima é útil na descrição teórica de mercados financeiros e na busca de fatores econômicos que afetam grupos específicos de ativos (BONANNO; VANDEWALLE; MANTEGNA, 2000). Além disso, a possibilidade de identificação de comunidades pode ajudar na formação das carteiras com diversificação de investimento de modo a reduzir o risco e aumentar o retorno (MANTEGNA, 1999; PERALTA; ZAREEI, 2016).

As redes de mercado financeiro baseadas nas correlações e em particular as árvores geradoras mínimas dessas redes tem sido amplamente usadas em estudos do mercado financeiro. Vandewalle et al. (2001) obtiveram uma árvore geradora mínima para 6358 observações de preços de ações norte-americanas durante o ano de 1999. Estudaram a distribuição no número de vizinhos de cada nó. Verificaram a existência de um conjunto extremamente amplo de configurações locais indicando uma certa ordem na aparente desordem do mercado.

Bonanno et al. (2003) compararam as propriedades estruturais da árvore geradora mínima de dados empíricos do New York Stock Exchange (NYSE) com árvores mínimas obtidas a partir de modelos teóricos simples. Concluíram que a árvore mínima empírica tem características de uma rede complexa que não pode ser reproduzida por um modelo de mercado aleatório. Miccichè et al. (2003) investigaram a estabilidade de árvores geradoras mínimas utilizando um procedimento de agrupamento baseado em correlação de séries temporais de retorno e de volatilidade de ativos financeiros. Verificaram que as árvores geradoras mínimas de retorno são mais estáveis no tempo do que as de volatilidade. Onnela et al. (2003) estudaram a dinâmica de árvores mínimas de ativos aplicando à análise de carteiras e verificaram que o comprimento da árvore diminui ao longo do tempo e permanece baixo durante uma crise. Também mostraram como a quebra da bolsa de 1987, que culminou na chamada Segunda-Feira Negra, poderia ser visto a partir da perspectiva de árvores mínimas dinâmicas.

Matteo et al. (2005) propõem um método geral para estudar a organização hierárquica de dados financeiros, ao incorporar a estrutura de suas correlações em gráficos métricos dentro de espaços multidimensionais. Os gráficos resultantes contêm a árvore geradora mínima como um sub-gráfico que preserva a estrutura hierárquica e permite uma melhor definição de grupos.

Ortega e Matesanz (2006) utilizaram dados de uma amostra de 28 países para classificar as consequências das crises cambiais com base na análise de agrupamentos da série histórica do câmbio. A análise forneceu informações relevantes sobre ligações entre países e um discernimento significativo sobre o fenômeno de contágio de crises financeiras. Jung et al. (2006) apresentam uma estrutura de rede árvore geradora mínima do mercado de ações Coreano. Eles descobriram que o mercado acionário da Coreia, um dos mercados emergentes, não forma aglomerados do setor empresarial ou das categorias da indústria e que estes clusters somente eram formados quando o índice MSCI (Morgan Stanley Capital International Inc.) era explorado, indicando que o mercado Coreano se comportava diferente de mercados amadurecidos (com saturação da taxa de crescimento).

Usando árvores geradoras mínimas Coelho et al. (2007) estudaram o grau e a evolução da interdependência entre os mercados acionários mundiais e descobriram que há uma forte tendência dos mercados se organizarem por localização geográfica e que as correlações médias mostram tendência de aumento ao longo de todo o período. Concluem que diversificação de aplicações em carteiras internacionais tem potencial de redução de benefícios. Garas e Argyrakis (2007) estudaram as propriedades de três diferentes carteiras negociadas na Bolsa de Valores de Atenas (ASE) no período 1987-2004. Os autores utilizaram a árvore geradora mínima para analisar evolução temporal das carteiras e do mercado como um todo e identificaram as mudanças nas carteiras de investimento associadas a uma crise no mercado. O artigo de Naylor, Rose e Moyle (2007) utilizou a árvore geradora mínima e análise de agrupamentos para extrair um mapa de influência das principais moedas para o período de 1995-2001.

Pozzi, Matteo e Aste (2008) utilizaram a árvore geradora mínima gerada a partir de correlações entre os 300 títulos de ações da Bolsa de Valores de Nova Iorque (NYSE) mais capitalizados no período de 2001 a 2003 para estudar a estrutura e a distribuição da rede em

relação aos setores econômicos. Encontraram que os setores financeiros localizam-se no centro, enquanto que os outros setores se distribuem na periferia rede.

Kwapien, Gworek e Drozd (2009) estudaram a estrutura e evolução temporal da moeda estrangeira usando teoria de redes, com base em taxas de câmbio de um conjunto de 63 moedas. Identificaram grupos relacionados as moedas e que estes dependem basicamente de fatores geográficos e econômicos. Sieczka e Hołyst (2009) analisaram as dependências nos mercados de commodities, investigando a correlação dos contratos futuros entre os anos de 1998 e 2007 por meio da árvore geradora mínima. Verificaram o mercado estava tornando-se mais correlacionado durante o período investigado e que as correlações não eram distribuídas uniformemente entre os contratos.

Papadimitriou, Gogas e Tabak (2013) propõe uma metodologia baseada em redes mínimas que pode ser usada adicionalmente aos métodos padronizados de supervisão bancária para identificação de problemas e de contágio, de modo a ser possível revelar o grau de conexão e identificar bancos que são centrais e podem ser usados como indicadores de destes problemas.

Peralta e Zareei (2016) mostraram a relação negativa entre a centralidade dos ativos da árvore geradora mínima de mercados financeiros americano e europeu e suas proporções ótimas segundo a teoria de formação de carteiras de Markowitz (MARKOWITZ, 1952). Mostraram como redes do mercado podem ser úteis para melhorar o processo de seleção de carteiras de investimento.

Muitos trabalhos baseados em redes também têm sido realizados para o mercado brasileiro. Em Cajueiro e Tabak (2008) foi analisada a estrutura de rede interbancária do Brasil. Os resultados sugerem uma fraca evidência da estrutura da comunidade, alta heterogeneidade da rede e que esse mercado é caracterizado por centros monetários com maior exposição de muitos bancos. Posteriormente, em (TABAK; CAJUEIRO; SERRA, 2009) os autores investigaram as propriedades estruturais de redes de bancos com árvores geradoras mínimas baseadas em correlação entre diversas variáveis bancárias. Os resultados obtidos indicaram que os bancos privados e estrangeiros tendem a formar grupos no interior da rede. Posteriormente, Tabak, Serra e Cajueiro (2010) investigaram as propriedades da árvore geradora mínima, utilizando a matriz de correlação para ações de diferentes setores. A dinâmica das medidas da rede obtida mostraram que os setores mais importantes são o financeiro, o energético e o de materiais. Em Tabak, Ludovice e Cajueiro (2011) foram calculadas as probabilidades de inadimplência para 30 diferentes setores econômicos. Mostraram que fatores macroeconômicos domésticos podem explicar essas probabilidades de inadimplência e que variáveis macroeconômicas têm efeitos distintos nestas probabilidades que devem ser considerados na modelagem de risco de crédito. Em Tabak et al. (2012) apresentam-se medidas de mitigação de risco sistêmico adotadas no Sistema de Pagamentos Brasileiro e analisados os pagamentos processados por esse sistema com a finalidade de iniciar um processo de identificação de focos com potencial de riscos sistêmicos. Foram também realizadas análises da rede formada pelos pagamentos utilizando-se conceitos de teoria de redes, tendo sido obtidas informações sobre a interligação de instituições financeiras sistemicamente importantes e a fragilidade das mesmas que podem auxiliar na prevenção



de eventos sistêmicos. Recentemente, Barbi e Prata (2019) usaram árvores geradoras mínimas baseadas em informação mútua para estudar efeitos não lineares no mercado brasileiro de ações. Verificaram que dependências não lineares entre os ativos do mercado, que não são reveladas por redes baseadas em correlação, são relevantes em períodos de alta volatilidade dos retornos de preços.

Um dos interesses no estudo de redes em geral e de redes geradoras mínimas em particular para o mercado financeiro é como ferramenta de apoio em análise de risco. Propostas nesse sentido estudam redes formadas a partir de interações definidas a partir das correlações entre os ativos do mercado (TABAK et al., 2012; PERALTA; ZAREEI, 2016; SORAMAKI; COOK; LAUBSCH, 2016) e árvores geradoras mínimas obtidas a partir de uma distância baseada em correlação. Estudos tradicionais de otimização de risco e retorno, como a otimização de risco-retorno da carteira de investimentos de Markowitz (MARKOWITZ, 1952) usam a variância como medida de risco. Como a variância da soma de variáveis aleatórias depende das correlações entre as variáveis, o risco fica dependente das correlações entre as variáveis. Embora importante e muito usado, o uso de correlações fica limitado aos casos onde existem dependências lineares entre os ativos e portanto nas situações em que as distribuições de probabilidade do retorno dos preços são descritos por uma distribuição normal. Outras propostas de avaliação de risco como o Valor em Risco (Value at Risk)(JORION, 2000), que embora muitas vezes fazem uso da distribuição normal por conveniência, são baseadas diretamente nas distribuições de retorno dos ativos do mercado. O Valor em Risco indica a pior perda em um horizonte de tempo que não será excedida a um dado nível de confiança (JORION, 2000) e pode ser obtido diretamente das distribuições empíricas de probabilidade dos retornos. Nesse sentido, a obtenção de redes do mercado financeiro em que a intensidade das interações são definidas diretamente a partir da distribuição de probabilidade empírica pode ser importante para uma melhor avaliação de risco.

Portanto, nesta dissertação propõe-se a obtenção de árvores geradoras mínimas do mercado financeiro a partir de uma distância entre distribuições de probabilidade do retorno do preço de ações. A distância proposta permite definir uma intensidade de ligação entre os ativos que é relacionada a similaridade entre suas distribuições de probabilidade. Como aplicação, obtivemos as árvores geradoras mínimas da carteira de ações do IBOVESPA para os meses de maio, junho, julho e agosto de 2019 a partir das distribuições empíricas de probabilidade dos retornos de preços com transações registradas em intervalos de 15 minutos. As árvores geradoras mínimas obtidas foram caracterizadas por suas distribuições de grau simples e de grau ponderado e comparadas com as árvores geradoras mínimas baseadas em correlação.

## 1.1 Problema de Pesquisa

O problema de pesquisa deste trabalho é explorar uma distância entre distribuições de probabilidade do retorno do preço de ações para obtenção da árvore geradora mínima da rede de ativos do mercado financeiro. Além disso, a pesquisa pretende comparar as redes baseadas em

distribuições de probabilidade com as redes obtidas a partir de correlações para o caso concreto de uma carteira de ações do mercado brasileiro que pertencem ao índice IBOVESPA.

## **1.2 Objetivos**

### **1.2.1 Gerais**

O objetivo geral desta pesquisa é propor e obter a árvore geradora mínima para o mercado financeiro a partir de distribuições de probabilidade dos retornos de ações e compará-las com árvores geradoras mínimas baseadas em correlações de Pearson.

### **1.2.2 Específicos**

- Propor a obtenção da rede árvore geradora mínima com uma medida de distância entre as distribuição de probabilidades dos retornos de preços de ativos do mercado financeiro.
- Obter as redes árvore geradora mínima para ações da carteira do IBOVESPA.
- Caracterizar as redes obtidas pela distribuição de grau simples e grau ponderado.
- Comparar as redes obtidas a partir das distribuições de probabilidade com as redes obtidas a partir de correlações.

## **1.3 Justificativa**

No estudo de redes financeiras o coeficiente de correlação tem sido usado como medida da intensidade da ligação entre ações das empresas. A correlação é uma medida adequada em situações em que as relações entre variáveis é linear e adequadamente descritas por uma distribuição Normal de probabilidade. Assim, é importante explorar redes financeiras baseadas em interações definidas diretamente pelas distribuições de probabilidade empíricas dos retornos de seus ativos. As distribuições de retorno estão diretamente relacionadas a técnica de avaliação de risco chamada Valor em Risco (JORION, 2000). Assim, a proposta de que a interação entre ativos tenha uma intensidade, nesse caso uma similaridade, definida a partir de uma distância entre distribuições de probabilidade pode ser importante para identificar ativos do mercado com mais semelhança de Valor em Risco.

## **1.4 Estrutura do trabalho**

A dissertação está organizada da seguinte maneira: O Capítulo 2 contém uma introdução sucinta dos conceitos básicos de teoria de redes necessários para a dissertação. No Capítulo 3 expõe-se a teoria básica de probabilidades necessária para entender a proposta de distância

baseada em distribuições de probabilidade para obtenção das redes geradoras mínimas. O Capítulo 4 contém a apresentação dos dados e os métodos de obtenção dos resultados da pesquisa. O Capítulo 5 contém resultados e discussão. Finalmente, o Capítulo 6 apresenta as conclusões da dissertação.

## 2 UMA BREVE INTRODUÇÃO A TEORIA DE REDES

O primeiro trabalho relacionado ao estudo de redes foi publicado em 1736 pelo matemático Leonard Euler ao tratar o Problema das Pontes de Königsberg (EULER, 1953). Na época, a cidade de Königsberg (Prússia) tinha sete pontes e as pessoas perguntavam se era possível ter um caminho que cruzasse as sete pontes sem passar duas vezes por uma mesma ponte. Euler provou que a existência de tal caminho é impossível através de um grafo onde os pontos representam o solo e as linhas representam as pontes, como pode ser visualizado na figura 1. Neste capítulo apresentamos alguns conceitos básicos da teoria de redes.

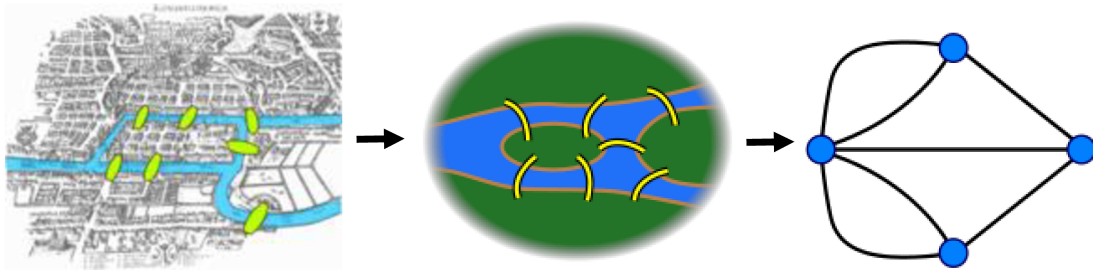


Figura 1 – As pontes de Königsberg e o grafo correspondente.

Fonte: By Bogdan Giușcă - Public domain (PD),based on the image, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=112920>; CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=990540>; CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=851840>

### 2.1 Conceitos básicos

Rede é um conjunto de elementos discretos chamados vértices (ou nós) e um conjunto de arestas (ou ligações) que ligam os vértices (NEWMAN, 2010; BARABÁSI, 2016). Os vértices podem representar os mais variados objetos e as arestas representam a conexão entre eles. Assim, numa rede de amizades, os vértices poderiam representar pessoas e as arestas a existência de amizade entre elas; numa rede de negócios os vértices poderiam representar empresas e as arestas a existência de relações comerciais entre elas.

Redes simples são assim chamadas por ter arestas simples entre dois vértices. Uma rede é completa se todo vértice é ligado por uma aresta. Em alguns casos pode haver mais do que uma aresta entre o mesmo par de vértices, que são chamadas de “arestas múltiplas”. Existem redes que apresentam “auto arestas”, isto é, arestas conectando vértices consigo mesmo. A rede representada na figura 2(a) é chamada rede simples (ou grafo) e a rede representada na figura 2(b) mostra uma rede com arestas múltiplas e auto arestas (ou multigrafo).

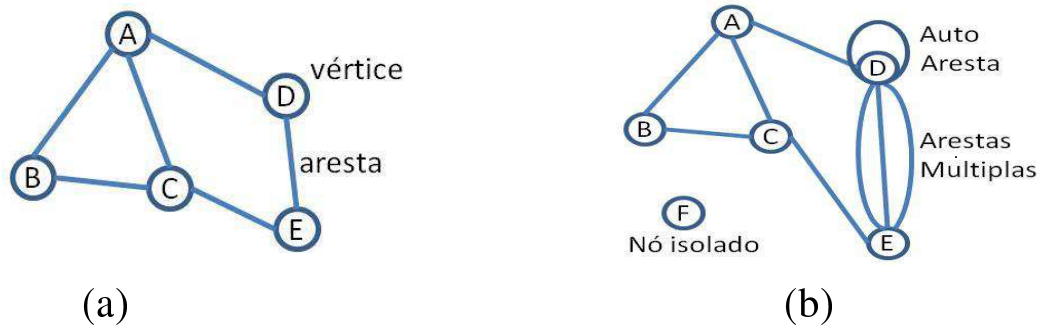


Figura 2 – (a) Exemplo rede simples (grafo). (b) Exemplo de rede com múltiplas e auto arestas (multigrafo)

Fonte: Elaborada pela autora.

Uma rede também pode ser direcionada, isto é, as ligações entre vértices tem uma direção definida. A figura 3 ilustra uma rede direcionada. Nesta dissertação só consideramos redes com arestas simples e não direcionadas.

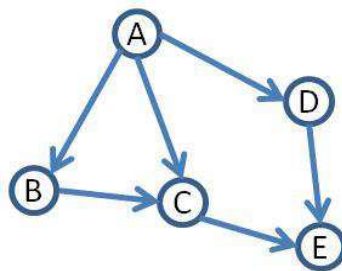


Figura 3 – (a) Exemplo rede direcionada.

Fonte: Elaborada pela autora.

A matriz de adjacência  $A$  é uma das formas de se representar um grafo. A matriz de adjacência tem por elementos  $A_{ij}$  com valores 1 ou 0 indicando a ligação ou a não ligação entre vértices, respectivamente. Para um grafo não direcionado a matriz de adjacência é simétrica com elementos diagonais nulos (ausência de auto arestas). Para a rede não direcionada da figura 2(a) a matriz de adjacência tem a seguinte forma

$$A = \begin{pmatrix} & A & B & C & D & E \\ A & 0 & 1 & 1 & 1 & 0 \\ B & 1 & 0 & 1 & 0 & 0 \\ C & 1 & 1 & 0 & 1 & 0 \\ D & 1 & 0 & 0 & 0 & 1 \\ E & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \quad (2.1)$$

enquanto que para a rede direcionada da figura 3 a matriz de adjacência será

$$A = \begin{pmatrix} & A & B & C & D & E \\ A & 0 & 1 & 1 & 1 & 0 \\ B & 1 & 0 & 1 & 0 & 0 \\ C & 1 & 1 & 0 & 0 & 0 \\ D & 1 & 0 & 0 & 0 & 1 \\ E & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad (2.2)$$

Nas matrizes das equações 2.1 e 2.2 os nomes dos vértices foram inseridos apenas para facilitar a identificação das ligações.

Existe também a possibilidade de que numa rede a ligação entre os vértices tenham uma intensidade. Neste caso temos a chamada rede ponderada. Os elementos da matriz de adjacência de uma rede ponderada indicam a intensidade da ligação entre os vértices.

Em algumas situações pode ocorrer uma rede bipartida. A rede bipartida apresenta dois tipos de vértices, um representa os vértices de origem e o outro, o grupo ao qual os vértices de origem pertencem (NEWMAN, 2010). A figura 4 mostra um grafo representando uma rede bipartida formada pelos vértices  $A, B, C, D$  e os vértices  $1, 2, 3, 4, 5, 6, 7$ .

## 2.2 Métricas de redes

Muitas medidas locais e globais foram desenvolvidas pela teoria de redes. Na sequência serão apresentados os principais indicadores de redes.

- **Grau médio:** Para uma rede simples o grau de um vértice mede o número de arestas nele conectadas, ou seja, o grau de um vértice  $i$ , representado por  $k_i$ , indica o número de conexões de  $i$  com outros vértices da rede (NEWMAN, 2010). O grau de um vértice  $i$  pode ser calculado a partir dos elementos da matriz de adjacência ( $A_{i,j}$ ) pela expressão

$$k_i = \sum_{j=1}^n A_{i,j} \quad (2.3)$$

O grau pode ser usado como uma medida de importância. Interpreta-se que, o vértice com o maior grau é o mais importante da rede, porque tem o maior número de conexões. A distribuição

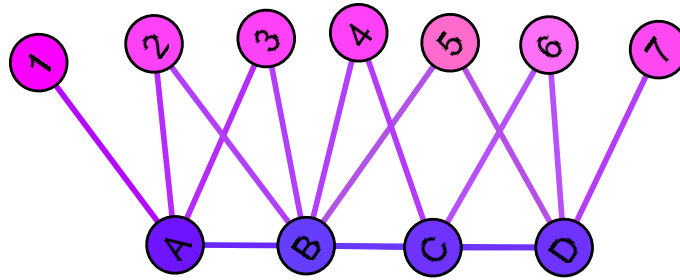


Figura 4 – Exemplo rede bipartida.

Fonte: Elaborada pela autora.

de grau  $P(k_i)$  de uma rede com  $n$  vértices representa as probabilidades (frequências relativas) de ocorrência dos vários graus da rede, ou seja,

$$P(k_i) = \frac{n_{k_i}}{n}, \quad (2.4)$$

onde  $n_{k_i}$  é o número de vértices com grau  $k_i$ .

A partir da distribuição de grau podemos obter o grau médio da rede

$$\bar{k} = \sum_{i=1}^n k_i P(k_i), \quad (2.5)$$

que representa o número médios de ligações que os vértices da rede possuem. Valores médios tais como grau quadrático médio  $\overline{k^2}$  e graus médios de ordem mais altas  $\overline{k^m}$ , se existirem, são dados por

$$\overline{k^m} = \sum_{i=1}^n k_i^m P(k_i) \quad m = 1, 2, 3... \quad (2.6)$$

Em particular, o desvio padrão de grau será dado por  $\sigma_k = \sqrt{\overline{k^2} - \bar{k}^2}$  e representa a variação média de grau em relação ao grau médio.

Para uma rede direcionada precisamos introduzir o conceito de grau de entrada e de grau de saída de um vértice. O grau de entrada de um dado vértice é igual ao número de arestas que chegam no vértice enquanto que o grau de saída é o número de arestas que saem do vértice. De modo análogo define-se a distribuição de grau de entrada e de saída e os graus médios de entrada e de saída da rede, respectivamente.

No caso de redes ponderadas podemos introduzir o conceito de grau ponderado de um dado vértice como sendo a soma das intensidades das ligações incidindo sobre ele. A distribuição de grau ponderado é definida de modo análogo ao grau simples e o grau médio ponderado representa a intensidade média das ligações que os vértices possuem.

- **Densidade:** Mede o quão próximo uma rede está de ser completa. Uma rede completa tem todas as arestas possíveis e a densidade é igual a 1.

- **Número de componentes:** O Número de componentes apresenta a quantidade de grupos com vértices conectados existentes na rede.

- **Diâmetro da rede:** Se dois vértices estão conectados a distância é igual a 1. O diâmetro da rede é o quão separados estão os dois vértices mais distantes.

- **Comprimento de caminho médio:** É a distância média entre os pares de vértices, considerando toda a rede (BARABÁSI, 2016). Este comprimento é uma medida que considera o número total de conexões necessárias para ir de um vértice  $i$  para um vértice  $j$  utilizando a menor rota. Considerando um determinado vértice  $i$ , a média sobre todos os vértices  $j$  com distância  $d_{i,j}$  entre  $i$  e  $j$ , é dada por

$$L_i = \frac{1}{n-1} \sum_{j \neq i} d_{i,j} \quad (2.7)$$

e o comprimento de caminho médio que indica a separação média entre vértices da rede é dado por

$$\bar{L} = \frac{\sum L_i}{n}. \quad (2.8)$$

- **Proximidade:** é a média do inverso das distâncias entre todos os nós, em que a distância a partir de um nó  $i$  para outro  $j$  é definido como o comprimento, medido pelo número de ligações, do caminho geodésico (mais curto) de um para o outro e obtido da expressão

$$C = \frac{\sum_{j \neq i} \left( \frac{1}{d_{i,j}} \right)}{n-1}, \quad (2.9)$$

com a interpretação de que quanto maior a proximidade mais influente é o vértice.

- **Intermediação:** A medida de intermediação (Betweenness, em inglês) é também uma medida de centralidade. Foi inicialmente utilizada no trabalho de Bavelas (1948) e popularizada em Freeman (1977). A intermediação mede o quanto um vértice encontra-se nos caminhos entre outros vértices (NEWMAN, 2010). Para um determinado vértice  $i$ , seja  $n_k^i = 1$



se o vértice  $i$  está no caminho geodésico entre os vértices  $j$  e  $k$ , e  $n_k^i = 0$  se o vértice  $i$  não está em tal caminho geodésico. A intermediação  $x_i$  é definida por

$$x_i = \sum_k n_k^i \quad (2.10)$$

De acordo com Newman (2010), os vértices com alta intermediação podem ter uma influência considerável dentro de uma rede devido ao controle sobre a informação que passa para outros vértices. A intermediação não mede o quão ligado o vértice está, mas mede o quanto o vértice aparece entre os outros. Assim, um vértice com baixo grau pode estar conectado a outros com baixo grau, e até mesmo ter um caminho longo, e ainda pode ter alta intermediação (NEWMAN, 2010).

- **Modularidade:** A modularidade é uma das medidas mais usadas para identificar comunidades na estrutura de redes. Ela é definida como a fração das arestas dentro das comunidades menos o valor esperado da fração de arestas inseridas aleatoriamente. É obtida pela seguinte expressão

$$Q = \frac{1}{2m} \sum_{i,j} (A_{i,j} - \frac{k_i k_j}{2m}) \delta_{c_i, c_j} \quad (2.11)$$

na qual  $A_{i,j}$  é o número de arestas entre os vértices  $i$  e  $j$ , o produto  $k_i k_j$  é o número esperado de arestas entre  $i$  e  $j$ ,  $m$  é o número total de arestas e  $\delta_{c_i, c_j}$  é a delta Kronecker, um símbolo matemático que tem valor 1 se os elementos pertencem a mesma comunidade ( $i = j$ ) e zero se não pertencem a mesma comunidade ( $i \neq j$ ).

- **PageRank (classificação de página):** PageRank é o nome comercial de uma medida de centralidade pertencente a empresa de tecnologia Google. A medida tem o objetivo de gerar listas de páginas da internet baseado no texto pesquisado e em uma classificação de páginas da web. As páginas são classificadas de acordo com a relevância para a pesquisa (NEWMAN, 2010). O objetivo original de PageRank era uma maneira de classificar os resultados da pesquisa se houvesse um grande número de links que direcionavam para determinada página. O PageRank também pode fornecer confiabilidade ao usuário, como por exemplo, pode ser mais confiável uma página que contenha um link a partir de uma página de uma empresa renomada. No estudo de redes, o PageRank pode indicar que um vértice é importante por que recebeu uma aresta de um vértice relevante em um grupo.

- **Robustez:** Diz-se que uma rede tem uma componente gigante se em média cada vértice tem mais de uma ligação. A robustez de uma rede indica quantos vértices devem ser removidos para que ela se fragmente em vários grupos. A existência de uma componente gigante pode ser obtida pelo critério de Molloy-Reed (MOLLOY; REED, 1995; BARABÁSI, 2016).

O critério considera a razão entre o grau quadrático médio  $\overline{k^2}$  e o grau médio  $\overline{k}$  obtidos da distribuição de grau (para redes simples). O critério de Molloy-Reed para ocorrência de uma componente gigante é dado por

$$k_o = \frac{\overline{k^2}}{\overline{k}} > 2. \quad (2.12)$$

A robustez da rede é definida a partir do critério de Molloy-Reed pela expressão (MOLLOY; REED, 1995; BARABÁSI, 2016)

$$f_c = 1 - \frac{1}{k_o - 1}, \quad (2.13)$$

que é a fração crítica de vértices que devem ser removidos para que uma componente gigante seja desfeita (BARABÁSI, 2016). A robustez pode ser interpretada como a rigidez ou a densidade da rede. Altas densidades indicam redes mais rígidas.

### 2.3 Redes tipo árvore e árvores geradoras mínimas

Uma rede (ou grafo) é conectada se para qualquer par de vértices existir um caminho que os conecte. Uma rede conectada sem ciclos (caminhos fechados de pelo menos 3 arestas) é chamada uma árvore (KREYSZIG, 1999). A figura 5(a) mostra um exemplo de grafo conectado e a figura 5(b) mostra o exemplo de uma árvore. As árvores estão entre os tipos mais importantes de redes e ocorrem em várias aplicações. As árvores podem ser usadas para expor, organizar ou analisar vários sistemas como: redes elétricas, relações produtor-consumidor e outras relações comerciais, informações em sistemas de banco de dados, estrutura sintática de programas de computador, etc.

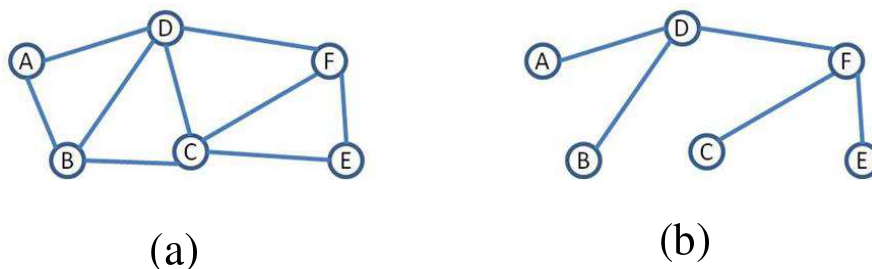


Figura 5 – (a) Rede conectada. (b) Rede tipo árvore.

Fonte: Elaborada pela autora.

A árvore geradora de uma rede é uma árvore contendo todos os vértices da rede. Para uma rede com  $n$  vértices, a árvore geradora mínima é uma árvore conectando  $n - 1$  vértices

em que a soma das distâncias entre os vértices é a menor dentre todas as árvores possíveis. Os pesos das arestas de uma Árvore Geradora Mínima representam a distância entre dois pontos na rede, mas é importante lembrar que os pesos também podem representar tempo, custo ou ainda alguma outra variável relacionada ao estudo.

## 2.4 O algoritmo de Kruskal

Kruskal, na década de 1950 propôs um algoritmo para obtenção da árvore geradora mínima (KRUSKAL, 1956). O algoritmo se baseia no teorema de que há apenas uma árvore de cobertura do grafo cuja soma dos pesos das arestas seja a menor possível. O algoritmo ordena os pares de vértices em ordem crescente de distâncias, escolhe as ligações nesta ordem e rejeita um vértice se ele formar um ciclo com um outro vértice já selecionado, o processo termina quando  $n-1$  vértices ficarem conectados (KRUSKAL, 1956; KREYSZIG, 1999). A figura (6) mostra os passos do algoritmo de kruskal.

ALGORITMO DE KRUSKAL [Grafo com  $n$  vértices e arestas de comprimento  $l_{ij}$  ]

Dado um grafo conectado com  $n$  vértices e arestas  $(i,j)$  de comprimentos  $l_{ij} > 0$ , o algoritmo determina a árvore mínima geradora .

Entrada: as arestas  $(i,j)$  do grafo e seus comprimentos  $l_{ij}$ .

Saída: Árvore mínima do grafo

1. Ordene as arestas do grafo em ordem crescente de comprimento.
2. Escolha eles nesta ordem como arestas da árvore mínima, rejeitando uma aresta somente se ele forma um ciclo com os vértices já escolhidos.

Se  $n-1$  arestas foram escolhidas, então

Saída: o conjunto de arestas escolhidas. Pare.

Fim KRUSKAL

Figura 6 – Algoritmo de Kruskal

Fonte: Tradução adaptada de Kreyszig (1999).

Para ilustrar o algoritmo de Kruskal vamos aplicá-lo em um exemplo simples de rede mostrado na figura 7(a) e que foi adaptado de Kreyszig (1999). As distâncias entre os vértices são indicadas na figura 7(a). A figura 7(b) mostra a tabela com a sequência de ligações seguindo-

se o algoritmo de Kruskal e a figura 7(c) ilustra a sequência de ligações até a formação da árvore mínima.

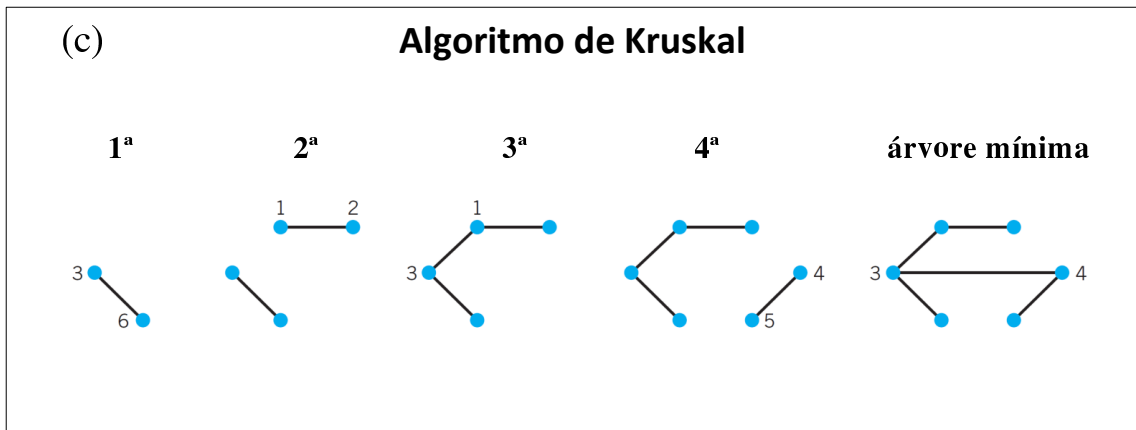
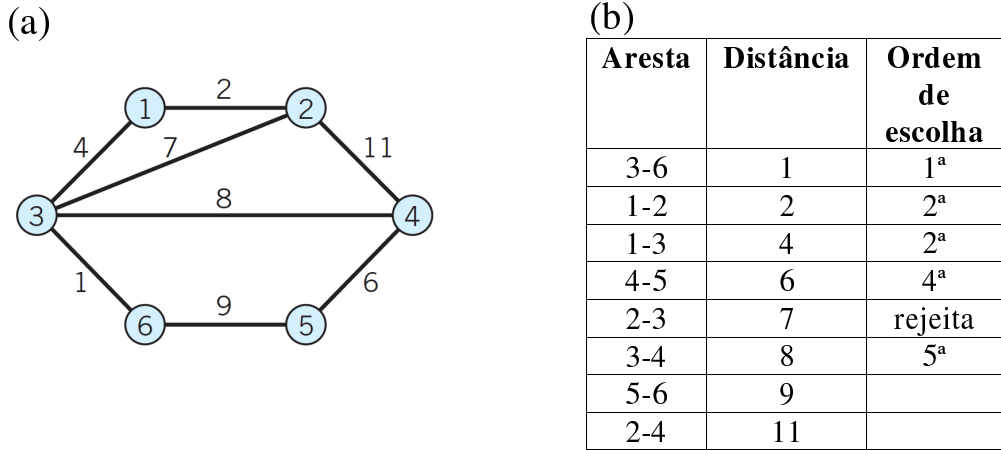


Figura 7 – Exemplo de aplicação do algoritmo de kruskal mostrando a sequência de escolha de ligações para obter a árvore geradora mínima

Fonte: Adaptado de Kreyszig (1999)

### 3 DISTRIBUIÇÕES DE PROBABILIDADE, RISCO, CORRELAÇÕES E DISTÂNCIAS

Em probabilidade, uma função  $X$  que associa a cada evento  $\omega$  do espaço amostral de um experimento aleatório um número real  $X(\omega) = x \in \mathfrak{R}$ , é denominada variável aleatória (V.A.). Desta forma, pode-se escrever os resultados de um experimento aleatório através de números mesmo que os elementos do espaço amostral não sejam números como no caso dos resultados cara e coroa no lançamento de uma moeda. Neste caso podemos definir uma variável aleatória como sendo a contagem do número de coroas. Em outras palavras, a variável aleatória traduz o resultado do experimento em números reais, apesar de não ser possível saber qual o resultado do experimento antes da realização, as probabilidades podem ser previamente estabelecidas por meio de um modelo matemático (por exemplo, eventos igualmente prováveis) ou empiricamente (frequências observadas). A distribuição de probabilidades de uma variável aleatória  $X$  é, portanto, uma descrição das probabilidades associadas com os possíveis valores de  $X$ .

As variáveis aleatórias podem ser classificadas como discretas ou contínuas, ou uma combinação das duas. O número de clientes que chegam numa fila de atendimento num intervalo de tempo é um exemplo de uma variável aleatória discreta enquanto que o intervalo de tempo entre chegadas de clientes é um exemplo de uma variável contínua.

#### 3.1 Distribuição de probabilidades

Uma variável aleatória discreta  $X$  é caracterizada pela distribuição de probabilidades  $P(X = x_i)$  dos valores  $X = x_i$  satisfazendo as seguintes propriedades

$$P(X = x_i) \geq 0 \quad (3.1)$$

$$0 \leq P(X = x_i) \leq 1 \quad (3.2)$$

$$\sum_i^n P(X = x_i) = 1 \quad (3.3)$$

onde a soma na terceira equação pode se estender ao infinito. A função distribuição acumulada (fda)  $F(x_k)$ , isto é, a probabilidade  $P(X \leq x_k)$  é dada por

$$F(x_k) = P(X \leq x_k) = \sum_{i=1}^k P(X = x_i). \quad (3.4)$$

Para uma variável aleatória contínua  $X$  introduzimos a função densidade de probabilidade  $f(x)$ , a probabilidade por unidade de  $x$ , de modo que a integral

$$P(a \leq x \leq b) = \int_a^b f(x) dx \quad (3.5)$$

representa a probabilidade de  $X$  assumir valores no intervalo  $[a, b]$  e satisfazendo as propriedades

$$f(x) \geq 0 \quad (3.6)$$

$$\int_{-\infty}^{+\infty} f(x) dx = 1. \quad (3.7)$$

A equação 3.7 é o equivalente para variável contínua de que a soma das probabilidades de todos os valores de  $X$  deve ser 1. A função distribuição acumulada (fda) é definida como

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x') dx' \quad (3.8)$$

e representa a probabilidade de  $X$  assumir valores no intervalo  $(-\infty, x]$ .

Com as funções distribuição de probabilidade e densidade de probabilidade pode-se calcular o valor esperado de qualquer função  $g(X)$  da V.A.  $X$ . O valor esperado é definido como

$$E(g(x)) = \sum_i g(x_i) P(x_i) \quad (3.9)$$

para uma variável discreta, em que a soma é sobre todos os valores de  $x$ , e

$$E(g(x)) = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (3.10)$$

para uma variável contínua. Em particular, são importantes os momentos de potências de  $x$ ,

$$E(x^n) = \begin{cases} \sum_i x_i^n P(x_i) & \text{variável discreta;} \\ \int_{-\infty}^{+\infty} x^n f(x) dx & \text{variável contínua} \end{cases} \quad (3.11)$$

em que o primeiro momento  $E(x)$  é o valor médio de  $X$ ,  $E(X^2)$  é a média quadrática de  $X$ , e assim sucessivamente. A partir dos momentos de  $X$  pode-se definir a variância como

$$\sigma^2 = E(X^2) - E^2(X) \quad (3.12)$$

e o desvio padrão  $\sigma = \sqrt{\sigma^2}$ .

### 3.2 A função distribuição acumulada empírica

Em geral, as funções de probabilidade são representadas por modelos matemáticos. Mas os modelos propostos podem ser baseados nos aspectos observados em distribuições de valores amostrais de uma variável aleatória. Em nosso trabalho estamos interessados na função distribuição de probabilidade acumulada de retornos de ativos financeiros obtidos empiricamente.

Para um conjunto  $x_1, \dots, x_n$  de valores observados de uma variável aleatória  $x$ , define-se a função distribuição acumulada (fda) empírica  $F_n(x)$  como a proporção de valores observados menores ou iguais a  $x$  (DEGROOT; SCHERVISH, 2012). Assim, se  $k$  valores observados são menores ou igual a  $x$ , então

$$F_n(x) = \frac{k}{n} \quad (3.13)$$

é a fda empírica. A fda empírica  $F_n(x)$  é equivalente a fda de uma variável aleatória discreta com probabilidade  $1/n$  para cada valor  $x_1, x_2, \dots, x_n$ . Portanto  $F_n(x)$  será uma função degrau com um salto de altura  $1/n$  em cada ponto  $x_i, i = 1, \dots, n$  (DEGROOT; SCHERVISH, 2012).

Para ilustrar, a figura 8 mostra a fda para a amostra  $\{-2; -1, 3; 0, 5; 0, 5; 3; 5; 5; 5; 5; 6; 6, 2; 6, 8\}$  com  $n = 12$  valores de uma variável  $x$ . Pode-se observar os degraus em cada valor de  $x$  da amostra.

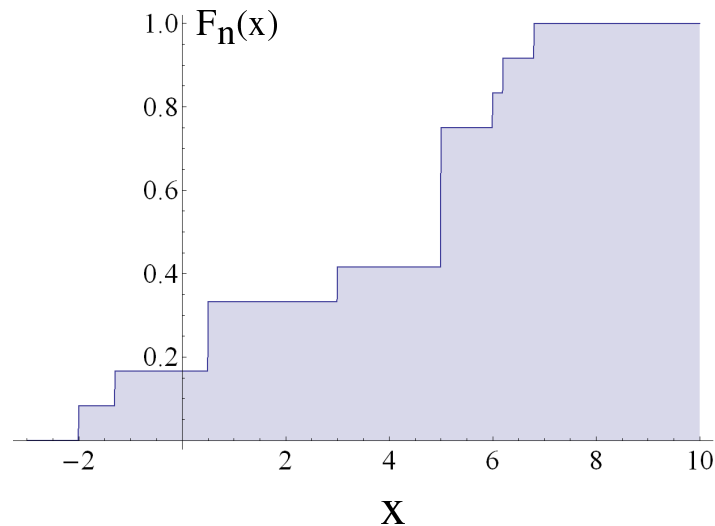


Figura 8 – Exemplo de função distribuição acumulada empírica.

Fonte: Elaborada pela autora.

### 3.3 Função distribuição acumulada e risco financeiro

Em finanças os retornos de ações são tratados com uma variável aleatória e descritos por meio de uma distribuição de probabilidade. Uma das maneiras de medir o risco de uma ação do mercado financeiro é por meio da distribuição de probabilidades dos seus retornos em determinado período. Essa técnica, chamada de "Valor em Risco" (JORION, 2000), abreviada como VAR (do inglês "Value at Risk"), consiste em determinar a partir da fda dos retornos o valor de retorno  $x_V$  que define uma probabilidade menor ou igual a um  $\alpha$  fixo, ou seja, dada uma fda de retornos  $F(x) = P(X \leq x)$ , o VAR ao nível  $\alpha$  define o valor de retorno  $x_V$  tal que

$$F(x_V) = P(X \leq x_V) = \int_{-\infty}^{x_V} f(x) dx = \alpha, \quad (3.14)$$

em que  $f(x)$  é a densidade de probabilidade do retorno. Como perdas correspondem a retornos negativos define-se o VAR pelo valor absoluto de  $x_V$ ,

$$Var(\alpha) = -x_V, \quad (3.15)$$

que representa a maior perda ao nível  $\alpha$ . Resumidamente, O VAR indica a pior perda em um horizonte de tempo que não será excedida a um dado nível de confiança (JORION, 2000). Por

exemplo, se definirmos um VAR diário de 5% ( $\alpha = 0,05$ ) significa que em 5% dos dias teremos uma perda menor ou igual a  $x_V$ . O valor de  $x_V$  vai depender da distribuição de retornos e pode ser obtida da fda empírica ou de algum modelo teórico, como por exemplo uma distribuição normal. A figura 9 ilustra o conceito de VAR.

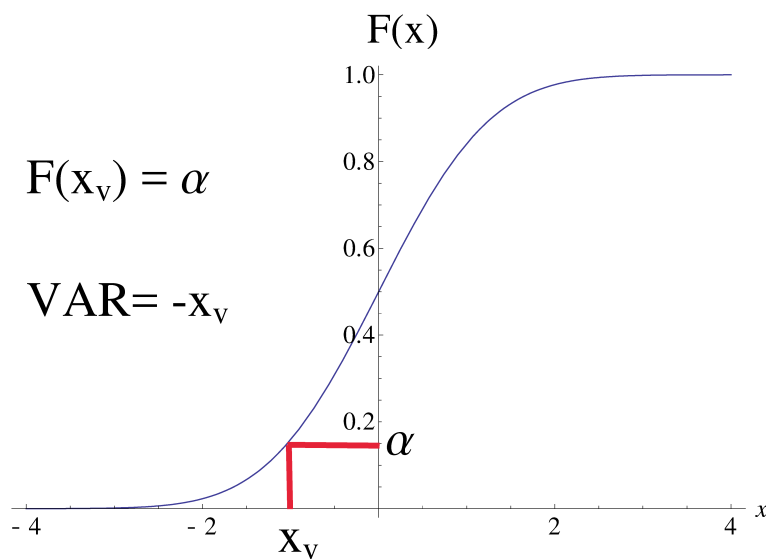


Figura 9 – O conceito de VAR obtido a partir da distribuição de probabilidade acumulada.

Fonte: Elaborada pela autora.

### 3.4 Correlação

O conceito de correlação aparece em problemas multivariados onde é importante obter informação sobre a existência de relações entre as variáveis. Este é um problema difícil e, em geral, é mais simples estudar as relações entre pares de variáveis. Uma das mais usadas medidas da relação entre duas variáveis  $x$  e  $y$  é o coeficiente de correlação de Pearson. Para duas variáveis aleatórias  $X$  e  $Y$  contínuas com distribuição conjunta de probabilidade  $P(X, Y)$  o coeficiente de correlação é definido por

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \quad (3.16)$$

em que

$$\sigma_{xy} = E\{[x - E(X)][y - E(Y)]\} = \sum_i [x_i - E(X)][y_i - E(Y)]P(x_i, y_i) \quad (3.17)$$



é a covariância entre  $x$  e  $y$ . No caso de variáveis contínuas a soma é trocada por integrais ponderadas pela função densidade de probabilidade  $f(x, y)$  sobre todos os valores de  $x$  e  $y$ .

O coeficiente de correlação é simples de calcular a partir de valores empíricos de pares das variáveis  $X$  e  $Y$ . Para um conjunto de medidas  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  das variáveis  $x$  e  $y$  defini-se o coeficiente de correlação pela seguinte expressão (JOHNSON; BHATTACHARYYA, 2009)

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.18)$$

em que  $\bar{x} = \sum_{i=1}^n x_i/n$  e  $\bar{y} = \sum_{i=1}^n y_i/n$  são as médias das variáveis  $x$  e  $y$ , respectivamente.

O coeficiente de correlação de Pearson é uma medida do grau da relação linear entre duas variáveis  $x$  e  $y$ . O coeficiente de correlação assume valores no intervalo de  $-1 \leq \rho \leq +1$  e valor absoluto da correlação indica a intensidade da relação linear entre  $x$  e  $y$ . Valores negativos de  $\rho$  indicam que  $x$  e  $y$  seguem direções opostas, ou seja, se  $x$  cresce,  $y$  decresce e vice-versa, enquanto que valores positivos indicam que  $x$  e  $y$  seguem mesma direção, ou sejam, ambas crescem ou ambas decrescem. Valores próximos de  $-1$ , indica uma correlação negativa forte e, valores próximo de  $+1$ , uma correlação positiva forte. Um valor de  $\rho$  próximo a 0 indica que há correlação fraca, ou ainda, que não há correlação (JOHNSON; BHATTACHARYYA, 2009).

É importante salientar que o coeficiente de correlação mede o grau de relação linear entre duas variáveis e não o grau de dependência entre elas. A independência entre duas variáveis ocorre no caso em que a distribuição de probabilidade conjunta  $P(x, y)$  fatora-se como um produto das distribuições  $P(x)$  e  $P(y)$  de cada variável, ou seja,  $P(x, y) = P(x)P(y)$ . Entretanto, no caso em que as variáveis  $x$  e  $y$  são descritas por uma distribuição conjunta normal a correlação nula implica que as variáveis são independentes (JOHNSON; BHATTACHARYYA, 2009). Portanto, correlação nula não implica em variáveis independentes pois, pode existir uma relação não linear entre elas, exceto no caso em que a distribuição conjunta de probabilidades é uma normal.

### 3.5 Distâncias

Para obtenção da árvore mínima geradora precisamos introduzir uma medida de distância entre os elementos que compõem os vértices da rede. O conceito de distância surge em geometria com a definição de distância entre pontos. Um caso conhecido é a distância entre pontos no plano obtido pelo teorema de Pitágoras e que pode ser generalizado para o caso de um espaço  $n$ -dimensional. Assim, a distância Euclidiana entre dois pontos  $x_i$  e  $y_j$  com coordenadas  $(x_1, x_2, \dots, x_n)$  e  $(y_1, y_2, \dots, y_n)$  é dada por

$$d_{ij} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}. \quad (3.19)$$

Entretanto, a distância Euclidiana pode não ser a mais adequada em todas as situações. Por exemplo, numa caminhada pelas ruas da cidade a distância entre dois pontos fica melhor

caracterizada somando-se as distâncias percorridas pelos lados dos quarteirões, ou seja, a distância entre os pontos  $x$  e  $y$  nesse caso seria dado por

$$d_{ij} = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|. \quad (3.20)$$

As expressões 3.19 e 3.20 tem em comum as seguintes propriedades:

$$d_{ij} = 0, \quad \text{se } i = j, \quad (3.21)$$

$$d_{ij} = d_{ji}, \quad (3.22)$$

$$d_{ij} \leq d_{ik} + d_{kj}, \quad (3.23)$$

onde a primeira propriedade diz que a distância de um ponto a si mesmo é zero, a segunda que a distância de  $i$  até  $j$  e de  $j$  até  $i$  devem ser a mesma e a terceira que a distância para ir de  $i$  até  $j$  é menor ou igual a soma das distâncias indo de  $i$  a  $j$  passando-se pelo ponto  $k$ . Portanto, podemos definir uma distância como uma função que satisfaz as propriedades 3.21, 3.22 e 3.23 que associa um número real a dois elementos  $x$  e  $y$  de um conjunto de pontos.

As expressões 3.19 e 3.20 satisfazem as propriedades 3.21, 3.22 e 3.23 e portanto definem uma distância. Outros exemplos de medidas de distância são:

(a) Para um conjunto arbitrário de pontos  $(x_1, x_2, \dots, x_n)$  a função

$$d = \begin{cases} 1 & \text{se } x_i \neq x_j; \\ 0 & \text{se } x_i = x_j \end{cases} \quad (3.24)$$

define uma distância entre dois pontos  $x_i$  e  $x_j$  quaisquer.

(b) Para os conjuntos de  $n$ -uplas ordenadas de números reais  $x = (x_1, x_2, \dots, x_n)$  e  $y = (y_1, y_2, \dots, y_n)$ ,

$$d = \text{Max}(|x_1 - y_1|, |x_2 - y_2|, \dots, |x_n - y_n|) \quad (3.25)$$

em que  $\text{Max}$  indica que usamos o maior valor dentre as diferenças absolutas das coordenadas como medida de distância entre  $x$  e  $y$ .

(c) Distância entre funções: Sejam  $f(x)$  e  $g(x)$  funções definidas no intervalo  $[a, b]$ . Podemos definir uma distância entre as funções no intervalo  $[a, b]$  como

$$d = \text{Max}|f(x) - g(x)|, \quad (3.26)$$

em que  $\text{Max}$  indica que toma-se o maior valor da diferença absoluta entre as funções no intervalo  $[a, b]$  como medida de distância entre as funções.

As redes que estamos interessados tem conexões definidas a partir do coeficiente de correlação entre os retornos de ações. Pode-se agora definir uma medida de distância a partir do coeficiente de correlação pela expressão (GOWER, 1966; JOHNSON; WICHERN, 2002; MANTEGNA, 1999)

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}, \quad (3.27)$$

onde  $\rho_{ij}$  é o coeficiente de correlação de Pearson. A distância dada pela equação 3.27 tem valores no intervalo  $0 \leq d_{ij} \leq 2$ . A distância máxima  $d_{ij} = 2$  ocorre se  $\rho = -1$  e as variáveis são maximamente anticorrelacionadas, enquanto que a distância mínima  $d_{ij} = 0$  ocorre para  $\rho_p = +1$  e as variáveis são maximamente correlacionadas.

A distância definida pela equação 3.27 tem sido usada para obter árvores mínimas geradoras a partir de correlações em muitos estudos sobre ativos do mercado financeiro. Além disso, é possível redefinir a equação 3.27 para ficar limitada ao intervalo  $[0, 1]$ . Entretanto, na árvore mínima geradora ela atribui menos importância para correlações negativas, mesmo que estas tenham alta intensidade. Alguns trabalhos sugerem que a intensidade pode ser mais importante que o sinal pois, em geral a rede é definida pela intensidade de ligação (SORAMAKI; COOK; LAUBSCH, 2016; ORIGUELA, 2018), enquanto que outros autores indicam que ela não é comparável com medidas de dependência não linear que, em geral, é um número positivo (MALEVERGNE; SORNETTE, 2006; BARBI, 2017; BARBI; PRATAVIERA, 2019). Portanto, neste trabalho vamos considerar uma medida de distância baseada no valor absoluto (intensidade) da correlação e definida pela seguinte expressão

$$d_{ij} = 1 - |\rho_{ij}|. \quad (3.28)$$

A distância definida pela equação 3.28 é limitada ao intervalo  $0 \leq d_{ij} \leq 1$  e, ao contrário da equação 3.27, em que correlações negativas com alta intensidade produzem distâncias grandes, na equação 3.28 correlações negativas altas produzem distâncias pequenas. Além disso, ela é diretamente comparável com a distância baseada na fda que será proposta nesta dissertação.

A proposta de distância alternativa que usaremos neste trabalho é baseada na equação 3.26 de distância entre funções com  $f(x)$  e  $g(x)$  escolhidas agora como as fda das variáveis aleatórias  $x$  e  $y$ . Essa distância representa uma dissimilaridade entre os ativos do mercado relativamente a sua fda. Assim, definimos a distância entre as fda empíricas  $F_n(x)$  e  $G_n(x)$  como (DEGROOT; SCHERVISH, 2012)

$$D_{m,n} = \text{Max}|F_m(x) - G_n(x)|, \quad (3.29)$$

com  $0 \leq D_{m,n} \leq 1$ . A distância entre distribuições 3.29 é a mesma usada no teste de Kolmogorov-Smirnov (DEGROOT; SCHERVISH, 2012). Para ilustrar, a figura 10 mostra  $D_{m,n}$  em um caso hipotético em que  $m = 5$  e  $n = 3$ .

Em nosso estudo estamos interessados na distância entre pares de ativos baseados no seus retornos no mesmo intervalo de tempo. Nesse caso  $m = n$  e podemos definir a distância entre as distribuições  $F_n^{(i)}$  e  $F_n^{(j)}$  do ativo  $i$  e o ativo  $j$  como

$$D_{i,j} = \text{Max}|F_n^{(i)}(x) - F_n^{(j)}(x)|. \quad (3.30)$$

A partir dessa distância podemos definir uma intensidade para ligação entre os elementos da rede como

$$I_{i,j} = 1 - D_{i,j}, \quad (3.31)$$

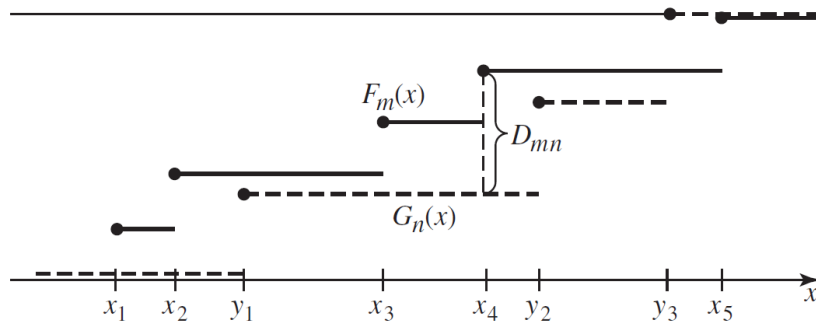


Figura 10 – Ilustração da distância  $D_{m,n}$  entre duas fda empíricas  $F_m(x)$  e  $G_n(x)$  para  $m = 5$  e  $n = 3$

Fonte: DeGroot, Morris H., and Mark J. Schervish. Probability and statistics. Pearson Education, 2012.

que pode ser interpretada como o grau de similaridade entre os elementos da rede baseado nas fda de suas variáveis. Como vimos na seção 3.3 o valor em risco de um ativo financeiro é obtido da distribuição de probabilidade do retorno. A medida de distância proposta é baseada diretamente na distribuição de probabilidade empírica dos retornos. Assim, a rede mínima baseada na distância entre distribuições de probabilidade é uma rede resumo das relações mais importantes entre as ações baseadas na similaridade das distribuições de probabilidades e portanto com mais semelhança de valor em risco.

## 4 METODOLOGIA

Nesta seção são apresentados os dados e os métodos de obtenção dos resultados da pesquisa.

### 4.1 A Base de dados

A carteira do IBOV (IBOVESPA) é a de maior representatividade da bolsa e com ações de alta liquidez. Para o estudo foram coletados preços das ações do IBOVESPA cotados a cada 15 minutos de Maio a Agosto de 2019. Os dados foram obtidos utilizando um código em R para baixar dados em alta frequência das ações da B3 a partir do pacote GetHFData (PERLIN; RAMOS, 2017). Foram coletados 2587 preços de transações de cada ação de uma carteira composta por 65 ações. Obtivemos um total de 169.183 observações e desse total, 2.484 preços não tinham valor de negociação no período. Esses valores faltantes, isto é, o valor do preço em um instante de tempo em que não houve transação, foram substituídos com o valor do preço no instante anterior. Os dados coletados dos preços em função do tempo são organizados de acordo com o Quadro 1, onde  $p_{i,j}$  é o preço da ação  $j$  no instante  $i$ .

Quadro 1 – Estrutura de dados dos preços das ações em que  $p_{i,j}$  corresponde ao retorno da ação  $j$  no instante  $t_i$ .

<b>Tempo</b>	<b>Ação 1</b>	<b>Ação 2</b>	<b>Ação 3</b>	<b>...</b>	<b>Ação m</b>
$t_0$	$p_{0,1}$	$p_{0,2}$	$p_{0,3}$	...	$p_{0,m}$
$t_1$	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	...	$p_{1,m}$
$t_2$	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	...	$p_{2,m}$
...	...	...	...	...	...
$t_n$	$p_{n,1}$	$p_{n,2}$	$p_{n,3}$	...	$p_{n,m}$

Fonte: Elaborada pela autora.

As ações podem ser de diferentes tipos:

- Ordinária Nominativa (ON): dá direito a voto em assembleias;
- Preferencial Nominativa (PN): tem preferência na distribuição de dividendos, pode ser classificados em A, B, C; Segmento de Listagem Novo Mercado (NM); Segmento de Listagem Nível 1 de Governança Corporativa (N1);
- Segmento de Listagem Nível 2 de Governança Corporativa (N2);
- Certificado de Depósito de Ações (UNT): ativos compostos por um conjunto de ações ordinárias e preferenciais

Para identificar quais empresas pertencem ao IBOVESPA no período mencionado, deve-se seguir os seguintes passos: entrar no site BMF Bovespa <[www.b3.com.br](http://www.b3.com.br)> e na aba superior Market Data e Índices > Índices > Ações por índice, e então, baixar o arquivo que fornece os códigos das empresas que pertencem ao IBOVESPA no referido período.

O mercado de ações é um ambiente público e organizado para negociação de títulos mobiliários e imobiliários (ações, opções de ações, fundos imobiliários, etc.). No Brasil, a B3 (Brasil, Bolsa, Balcão) é a única bolsa de valores de mercadorias e futuros em operação atualmente. A Comissão de Valores Mobiliários (CVM) é uma autarquia que controla o mercado acionário, responsável por disciplinar, fiscalizar e promover o mercado de valores mobiliários. Ação corresponde à menor parcela do capital social das companhias de capital aberto, que tem seus papéis negociados no mercado. É um título patrimonial e concede aos seus detentores, os direitos e deveres de um sócio, estes podem receber parte do lucro de uma companhia através do pagamento de dividendos, de juros sobre o capital próprio e de bonificações de acordo com os critérios de cada empresa e também podem ganhar com a valorização das ações. A B3 desenvolve, calcula e administra uma ampla variedade de índices. Os índices servem como referência para investidores nacionais e internacionais. Seguindo critérios rígidos de governança e com base em estudos de mercado, a B3 busca capturar de maneira correta o comportamento dos ativos e dos diversos segmentos da economia brasileira. Os índices são classificados como:

- Índices amplos: Índice Bovespa (IBOVESPA), Índice Brasil 100 (IBRX100), Índice Brasil 50 (IBRX 50), Índice Brasil Amplo BM&FBovespa (IBRA)

- Índices de governança: Índice de Ações com Governança Corporativa Diferenciada (IGCX), Índice de Ações com Tag Along Diferenciado (ITAG), Índice de Governança Corporativa Trade (IGCT), Índice de Governança Corporativa (IGC-NM).

- Índices de segmentos e setoriais: Índice Dividendos BM&FBOVESPA (IDIV), Índice MidLarge Cap (MLCX), Índice Small Cap (SMLL), Índice Valor BM&FBOVESPA (IVBX), Índice BM&FBOVESPA Financeiro (IFNC), Índice de BDRs Não Patrocinados-GLOBAL (BDRX), Índice de Commodities Brasil (ICB), Índice de Consumo (ICON), Índice de Energia Elétrica (IEE), Índice de Fundos de Investimentos Imobiliários (IFIX), Índice de Materiais Básicos BM&FBOVESPA (IMAT), Índice DI, Índice do Setor Industrial (INDX), Índice Geral do Mercado Imobiliário - Comercial (IGMI-C), Índice Imobiliário (IMOB), Índice Utilidade Pública BM&FBOVESPA (UTIL).

- Índices de sustentabilidade: Índice Carbono Eficiente (ICO2), Índice de Sustentabilidade Empresarial (ISE).

- Índices de S&P DowJones: Commodities, Empresas Privadas, Smart Betas, Índice de Futuros e Índice de Renda Fixa.

O IBOVESPA, principal índice da B3, é negociado em diversos países. É um indicador do desempenho médio das cotações dos ativos de maior negociabilidade e representatividade no mercado brasileiro. É chamado de índice de retorno total porque reflete a variação dos preços das ações e também o impacto da distribuição dos proventos. As carteiras teóricas agrupam ações de empresas que possibilitam o investimento teórico representado por um índice. O índice IBOVESPA é o principal indicador médio para representar o desempenho do mercado de ações brasileiro, formado pelas ações das empresas com maior liquidez e maior volume financeiro negociado de todo o volume de negócios da bolsa. O índice é revisado trimestralmente considerando o volume dos últimos 12 meses. As ações que integram o IBOVESPA devem

seguir alguns critérios, dispostos em sua metodologia, como, presença em pregão 2 de 95%, participação maior ou igual a 0,1% no mercado à vista, em termos de volume financeiro; não ter ação com cotação abaixo de R\$ 1,00 (as chamadas Penny Stocks). As companhias que atendam a esses critérios são ordenadas conforme o Índice de Negociabilidade (IN), que considera a quantidade de negociações e o volume financeiro gerado, sendo incluídas, assim, as ações que representem, de forma cumulativa, 85% das negociações ocorridas no período.

A figura 11 mostra a lista de ações que compõem o índice IBOVESPA para o período considerado nesta dissertação.

Código	Ação	Tipo	Qtde. Teórica	Part. (%)	Código	Ação	Tipo	Qtde. Teórica	Part. (%)
ABEV3	AMBEV S/A	ON	4.342.636.342	5,019	IRBR3	IRBRASIL RE	ON NM	218.737.606	1,331
AZUL4	AZUL	PN N2	316.666.637	0,954	ITSA4	ITAUSA	PN ED N1	4.492.854.927	3,339
B3SA3	B3	ON NM	2.044.215.807	5,625	ITUB4	ITAUUNIBANCO	PN ED N1	4.718.960.710	9,76
BBAS3	BRASIL	ON NM	1.331.530.017	3,61	JBSS3	JBS	ON NM	1.603.247.593	2,802
BBDC3	BRADESCO	ON N1	904.516.682	1,592	KLBN11	KLABIN S/A	UNT ED N2	591.009.862	0,529
BBDC4	BRADESCO	PN N1	3.873.597.664	7,681	KROT3	KROTON	ON NM	1.470.274.625	0,959
BBSE3	BBSEGURIDADE	ON EDR NM	671.584.112	1,346	LAME4	LOJAS AMERIC	PN N1	686.775.763	0,737
BRAP4	BRADESPAR	PN N1	222.108.601	0,372	LREN3	LOJAS RENNER	ON NM	778.042.663	2,325
BRDT3	PETROBRAS BR	ON NM	728.125.000	1,233	MGLU3	MAGAZ LUIZA	ON NM	540.780.800	1,237
BRFS3	BRF SA	ON NM	811.139.545	1,877	MRFG3	MARFRIG	ON NM	402.731.780	0,186
BRKM5	BRASKEM	PNA N1	264.632.416	0,464	MRVE3	MRV	ON NM	292.037.854	0,32
BRML3	BR MALLS PAR	ON NM	839.974.219	0,714	MULT3	MULTIPLAN	ON N2	267.482.247	0,437
BTOW3	BZW DIGITAL	ON NM	167.767.522	0,429	NATU3	NATURA	ON NM	172.485.909	0,66
CCRO3	CCR SA	ON NM	1.115.695.556	1,033	PCAR4	P.ACUCAR-CBD	PN N1	155.389.579	0,819
CIEL3	CIELO	ON NM	1.118.386.806	0,491	PETR3	PETROBRAS	ON EJ N2	2.708.734.869	4,38
CMIG4	CEMIG	PN N1	969.719.162	0,865	PETR4	PETROBRAS	PN EJ N2	4.332.323.046	6,296
CSAN3	COSAN	ON NM	155.785.693	0,468	QUAL3	QUALICORP	ON NM	237.835.371	0,389
CSNA3	SID NACIONAL	ON	630.161.054	0,529	RADL3	RAIADROGASIL	ON NM	211.224.469	1,141
CVCB3	CVC BRASIL	ON NM	147.088.800	0,478	RAIL3	RUMO S.A.	ON NM	1.115.172.704	1,554
CYRE3	CYRELA REALT	ON NM	261.043.342	0,374	RENT3	LOCALIZA	ON NM	561.023.808	1,518
ECOR3	ECORODOVIAS	ON NM	195.023.762	0,139	SANB11	SANTANDER BR	UNT	372.112.887	0,953
EGIE3	ENGIE BRASIL	ON NM	254.813.849	0,696	SBSP3	SABESP	ON NM	305.385.123	0,971
ELET3	ELETROBRAS	ON N1	269.667.840	0,663	SMLS3	SMILES	ON NM	58.691.428	0,126
ELET6	ELETROBRAS	PNB N1	225.964.766	0,552	SUZB3	SUZANO S.A.	ON NM	751.848.759	1,451
EMBR3	EMBRAER	ON NM	735.864.045	0,799	TAEE11	TAESA	UNT EDJ N2	218.568.280	0,362
ENBR3	ENERGIAS BR	ON NM	295.342.982	0,353	TIMP3	TIM PART S/A	ON EJ NM	807.711.660	0,585
EQL3	EQUATORIAL	ON NM	198.163.464	1,157	UGPA3	ULTRAPAR	ON NM	1.085.159.631	1,04
FLRY3	FLEURY	ON NM	306.936.057	0,431	USIM5	USIMINAS	PNA N1	513.629.423	0,219
GGBR4	GERDAU	PN ED N1	987.996.350	0,732	VALE3	VALE	ON NM	3.147.736.916	8,309
GOAU4	GERDAU MET	PN ED N1	570.896.065	0,211	VIVT4	TELEF BRASIL	PN	415.131.868	1,316
GOLL4	GOL	PN N2	133.111.745	0,281	VVAR3	VIAVAREJO	ON NM	499.769.567	0,204
HYPE3	HYPERA	ON NM	396.959.288	0,761	WEGE3	WEG	ON NM	740.453.283	1,009
IGTA3	IGUATEMI	ON NM	86.705.835	0,239					

Figura 11 – Lista de ações da carteira do IBOVESPA no período de maio a agosto de 2019.

Fonte: Elaborada pela autora.

## 4.2 Obtenção das árvores geradoras mínimas

Os dados utilizados para elaborar as matrizes de correlação e distâncias entre distribuições foram as variações do logaritmo dos preços diários das 65 ações das empresas selecionadas no índice IBOVESPA nos meses de maio, junho, julho e agosto de 2019. Seja  $P(t)$  e  $P(t - \tau)$

os preços de um ativo no instante  $t$  e  $t - \tau$ . O retorno  $R(t)$  é definido por (MANTEGNA, 1999)

$$R(t) = \ln P(t) - \ln P(t - \tau) = \ln \left( \frac{P(t)}{P(t - \tau)} \right). \quad (4.1)$$

A partir dos preços das ações podemos obter os retornos de cada uma delas usando a equação 4.1. Dessa forma os dados iniciais para obtenção das redes ficam dispostos como mostra o Quadro 2 em que o valor  $R_{i,j}$  corresponde ao retorno da ação  $j$  no instante  $t_i$ . Os retornos são baseados em preços de negociações consecutivas e portanto o primeiro valor nos dados de preços não tem valor de retorno.

Quadro 2 – Estrutura dos dados dos retornos dos preços das ações em que  $R_{i,j}$  corresponde ao retorno da ação  $j$  no instante  $t_i$ .

<b>Tempo</b>	<b>Ação 1</b>	<b>Ação 2</b>	<b>Ação 3</b>	<b>...</b>	<b>Ação m</b>
$t_1$	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	...	$R_{1,m}$
$t_2$	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$	...	$R_{2,m}$
$t_3$	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$	...	$R_{3,m}$
...	...	...	...	...	...
$t_n$	$R_{n,1}$	$R_{n,2}$	$R_{n,3}$	...	$R_{n,m}$

Fonte: Elaborada pela autora.

Com os valores empíricos dos retornos de cada ativo obtém-se a matriz de correlação, que tem por elementos as correlações  $\rho_{i,j}$  entre pares de ações de empresas  $i$  e  $j$ . A matriz de correlação é simétrica e com elementos iguais a 1 na diagonal principal. Para a obtenção da rede foi definida a matriz adjacência ponderada por correlações que tem por elementos o valor absoluto das correlações e atribuindo-se zero aos elementos da diagonal. A matriz de distâncias com elementos  $d_{i,j}$  para a elaboração das árvores geradoras mínimas foi obtida a partir da medida de distância definida pela equação 3.28. As intensidades das ligações dos elementos da árvore geradora mínima foi definida novamente pela intensidade das correlações entre os elementos da árvore mínima.

Para a obtenção das árvores geradoras mínimas a partir da distância entre distribuições de probabilidade obtivemos as distribuições acumuladas do retorno dos preços de cada ação. A partir das distribuições acumuladas obtivemos a distância  $D_{i,j}$  definida pela equação 3.31 entre cada par  $i$  e  $j$  de ações. Na árvore mínima a intensidade de ligação foi definida pela medida  $I_{i,j} = 1 - D_{i,j}$ .

Os cálculos e figuras dessa dissertação foram obtidos com o programa estatístico R-Studio e com o programa para visualização e análise de redes *software* Gephi versão 0.9.2.



## 5 RESULTADOS E DISCUSSÃO

Com dados descritos no capítulo 4 obtivemos as redes geradoras mínimas para os meses de maio, junho, julho e agosto de 2019. Neste capítulo apresentamos e discutimos os resultados obtidos.

As Figuras 12 (a)-(b) mostram as estruturas das redes para o mês de maio obtidas por correlação e por distribuição de probabilidade, respectivamente. As respectivas estruturas de rede para os meses de junho, julho e agosto são apresentados nas figuras 13,14 e 15. O tamanho dos nós em todas as redes são proporcionais ao grau simples e as arestas são proporcionais as intensidades das ligações dadas pelo valor absoluto das correlações no caso de redes de correlação e pela intensidade definida pela equação 3.31 para as redes baseadas em distribuição de probabilidade. Os grupos, representados por nós de mesma cor, foram definidos pela modularidade.

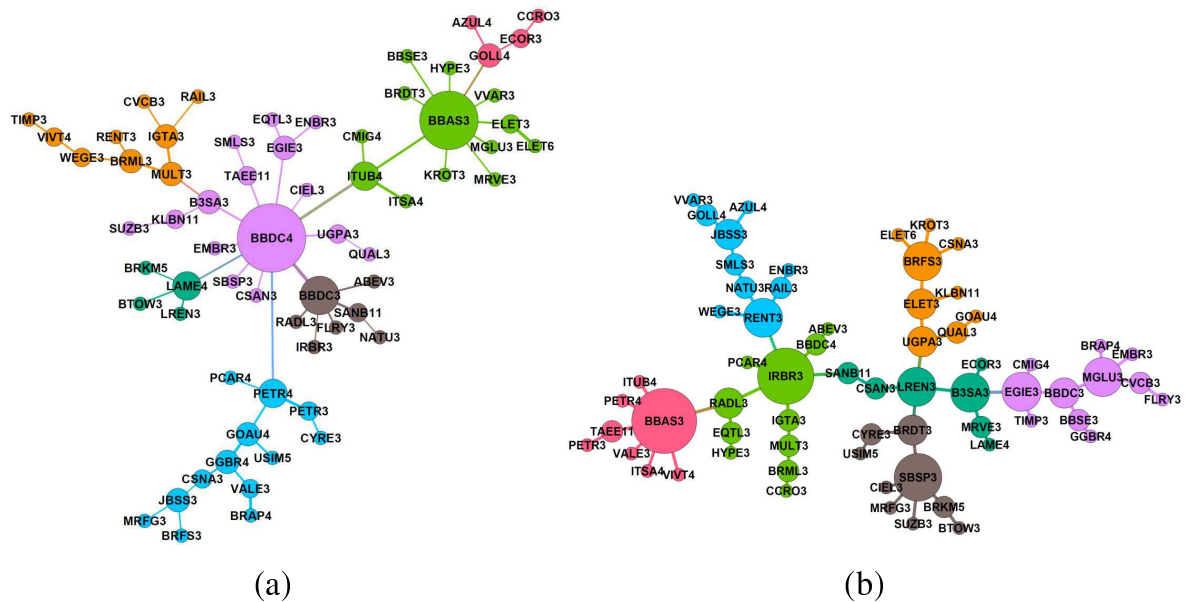


Figura 12 – Árvore geradora mínima para o mês de maio de 2019. (a) Rede baseada em correlação. (b) Rede baseada em distribuição de probabilidade.

Fonte: Elaborada pela autora.

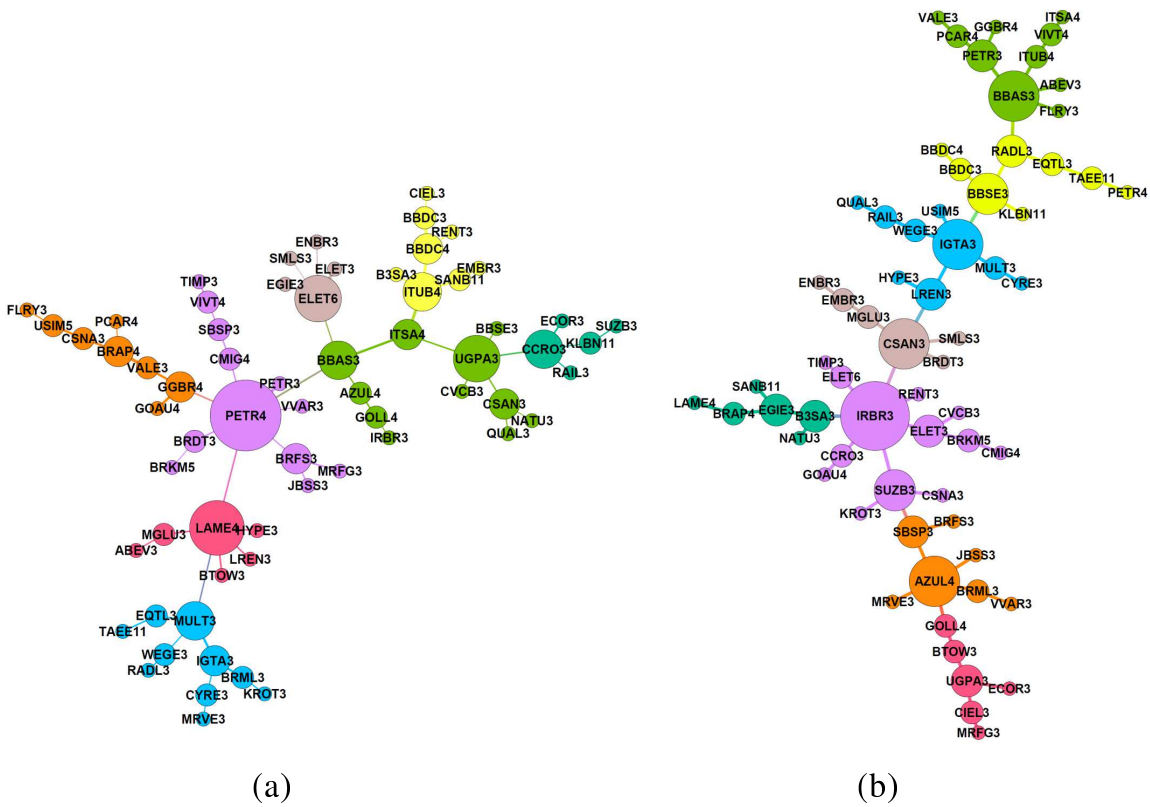


Figura 13 – Árvore geradora mínima para o mês de junho de 2019. (a) Rede baseada em correlação. (b) Rede baseada em distribuição de probabilidade.

Fonte: Elaborada pela autora.

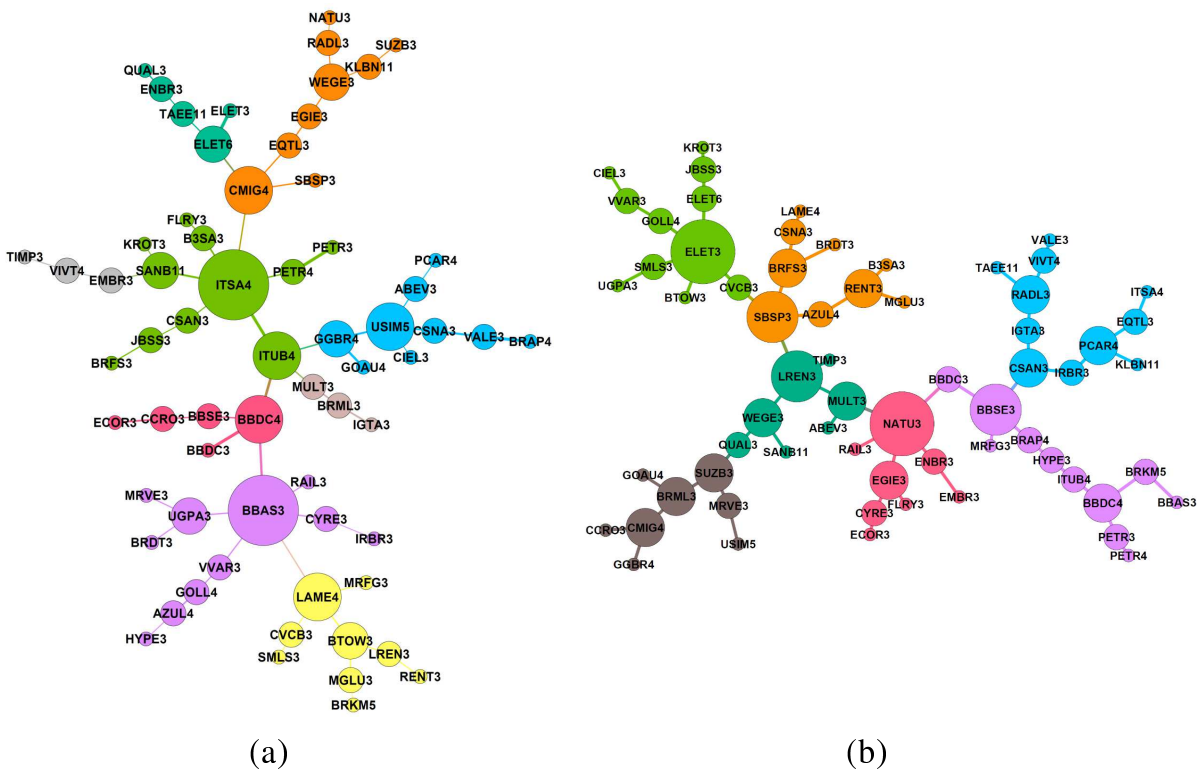


Figura 14 – Árvore geradora mínima para o mês de julho de 2019. (a) Rede baseada em correlação. (b) Rede baseada em distribuição de probabilidade.

Fonte: Elaborada pela autora.

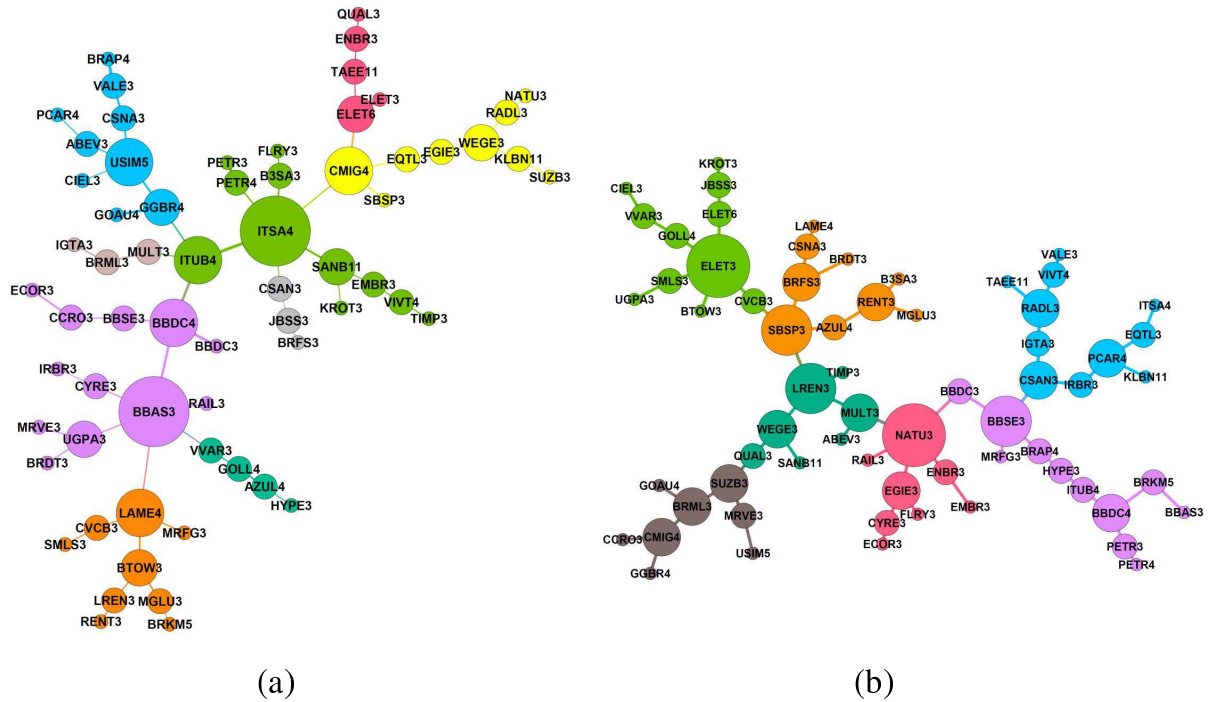


Figura 15 – Árvore geradora mínima para o mês de agosto de 2019. (a) Rede baseada em correlação. (b) Rede baseada em distribuição de probabilidade.

Fonte: Elaborada pela autora.

Note que as árvores geradoras mínimas representam um resumo da interação entre os vértices de uma rede na qual foram selecionadas as interações mais importantes. As intensidades das ligações são dadas pelo valor absoluto das correlações para as redes obtidas a partir de correlações e pela intensidade dada pela equação 3.31 para as redes obtidas a partir de distâncias entre distribuições. Além disso, a intensidade de ligação entre vértices é uma variável contínua definida no intervalo  $[0, 1]$ . Para mostrar as diferenças entre as intensidade de ligações nos dois tipos de redes vamos apresentar a distribuição de intensidades por meio de um diagrama de caixa (boxplot) (JOHNSON; BHATTACHARYYA, 2009). O diagrama de caixa contém o valor mínimo, o 1º quartil, a mediana (2º quartil), o 3º quartil e o valor máximo de um conjunto de valores. Ele divide os dados ordenados em 4 intervalos. Cada intervalo do diagrama representa 25% dos valores que acumulam naquele intervalo. A caixa central do diagrama que vai do 1º ao 3º quartil contém a 50% dos valores centrais dos dados. A distância entre o 1º e 3º é chamada de intervalo interquartil e é uma medida da variabilidade dos dados. A simetria do box-plot indica a simetria da distribuição de frequências da variável considerada, no caso o grau ponderado das ações.

A figura 16 mostra o diagrama de caixa das intensidades das ligações entre as ações nas árvores geradoras mínimas de correlação e de distribuição de probabilidade para os meses de maio, junho, julho e agosto de 2019. As figuras mostram que as intensidades das ligações nas redes baseadas em distribuição de probabilidade são sempre próximas da intensidade máxima 1, variam num intervalo pequeno de valores e são sempre maiores que as intensidades das

redes baseadas em correlação. Isto mostra que as distribuições de probabilidade das ações são muito similares em cada período e essa característica praticamente não muda no decorrer dos períodos. Por outro lado, a distribuição de intensidade das correlações varia num intervalo maior de valores em cada período e também no decorrer dos períodos.

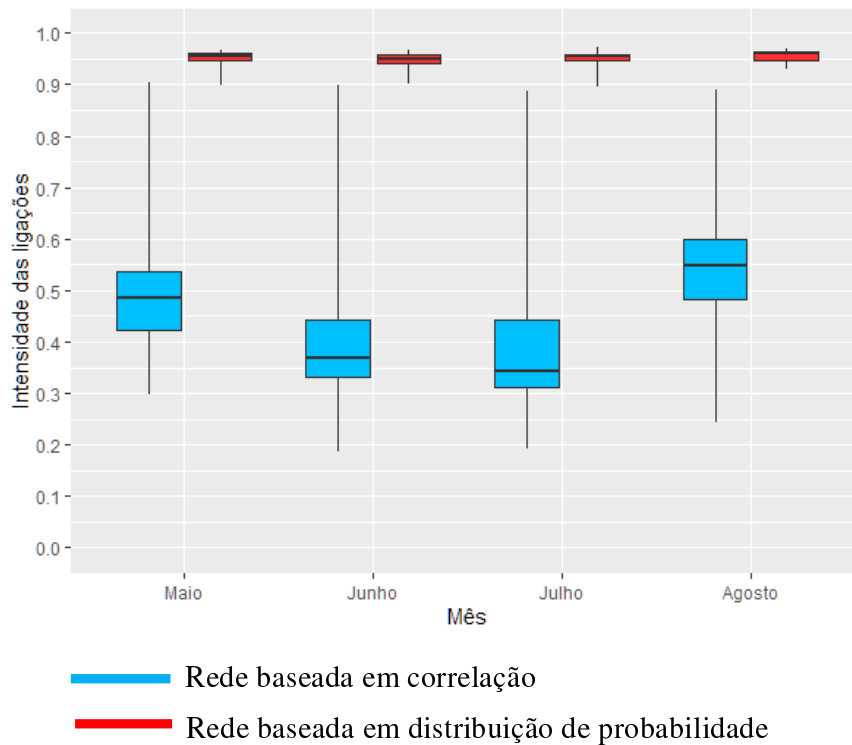


Figura 16 – Diagrama de caixa da intensidade das ligações dos elementos das árvores geradoras mínimas

Fonte: Elaborada pela autora.

## 5.1 Árvores geradoras mínimas com ligações simples

Nesta seção estudaremos algumas estatísticas baseadas na distribuição de grau simples das redes. Assim, consideramos que a rede mínima é uma rede com ligações binárias e a cada vértice é atribuído uma intensidade de valor 1. Neste caso o grau médio é o mesmo para todas as redes e igual a  $\bar{k} = 1,9692$ . Os gráficos na figura 17 mostram as distribuições de grau simples para cada mês. Os gráficos da esquerda correspondem as distribuição de grau para as redes baseadas em correlações e os da direita as distribuições de grau das redes baseadas em distribuição de probabilidade. Inspeccionando as distribuições de grau observa-se que as redes baseadas em correlação podem apresentar graus mais altos do que as redes baseadas em distribuição de probabilidades.

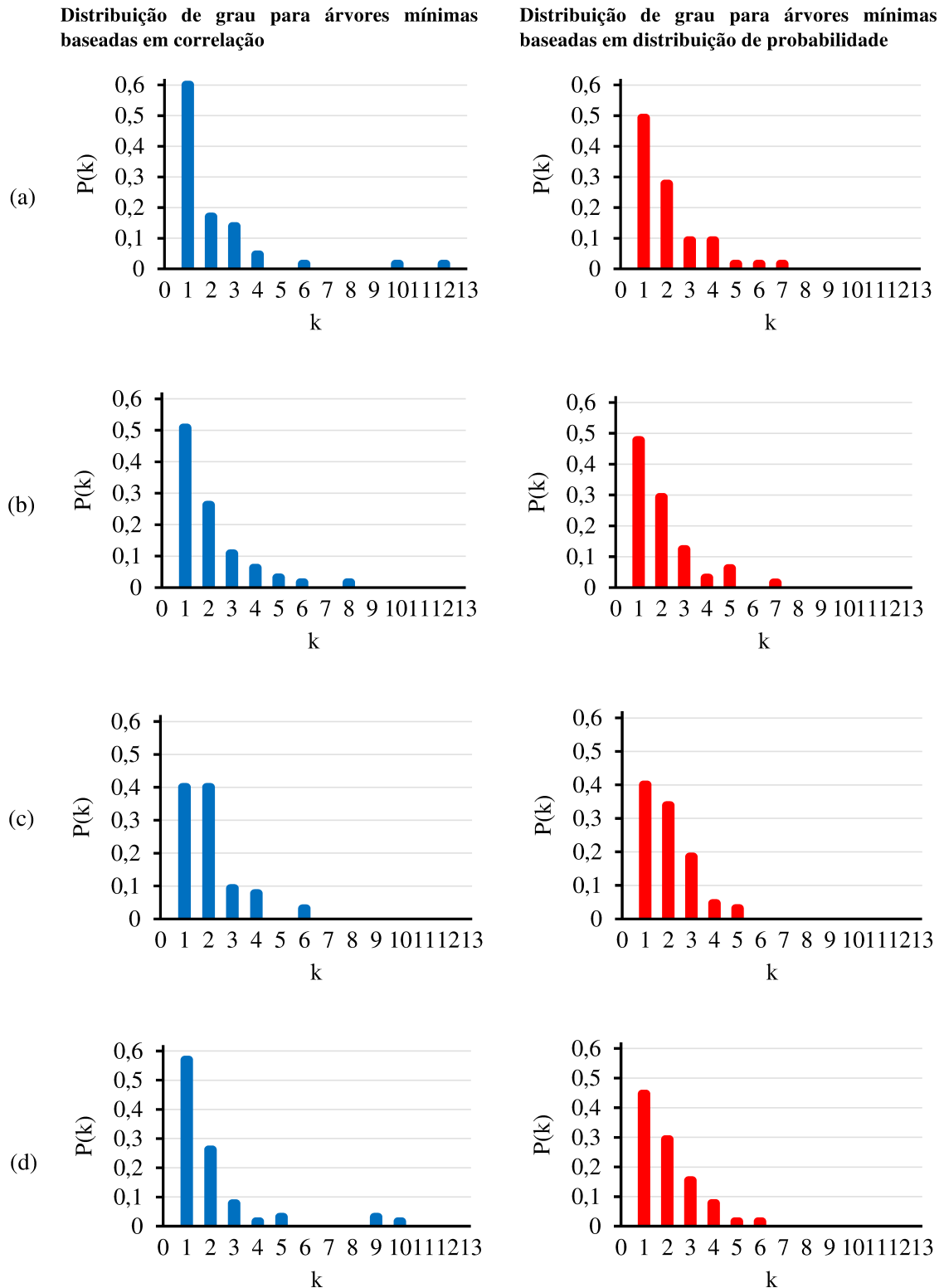


Figura 17 – Distribuição de grau simples das árvores geradoras mínimas para os meses de (a) maio, (b) junho, (c) julho e (d) agosto. Lado esquerdo: distribuições de grau das árvores mínimas baseadas em correlações. Lado direito: distribuições de grau das árvores mínimas baseadas em distribuições de probabilidade.

Fonte: Elaborada pela autora.

Para caracterizar a dispersão das distribuições a figura 18 mostra o desvio padrão da distribuição de grau para cada mês. Observa-se que a dispersão dos graus das redes baseadas em distribuição de probabilidade é sempre menor do que a dispersão das redes baseadas em correlação. Além disso, em função do tempo, a dispersão dos graus nas redes de correlação apresentam mais variação quando comparadas às redes de distribuição de probabilidade.

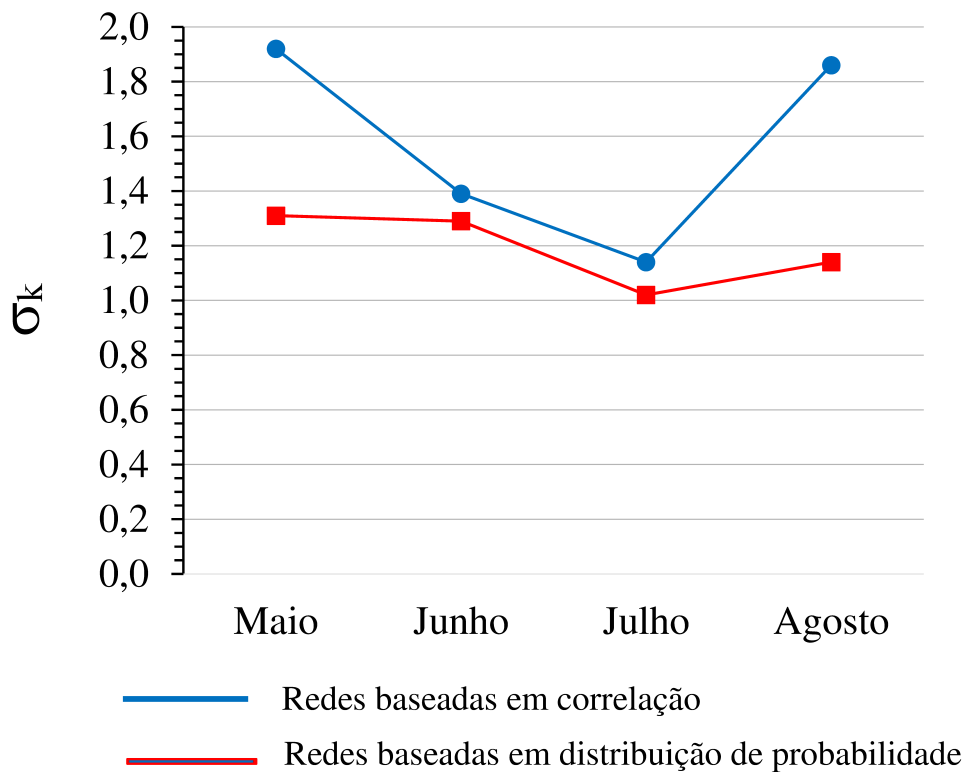


Figura 18 – Desvio padrão das distribuições de grau das árvores geradoras mínimas. O grau médio  $\bar{k} = 1,9692$  é o mesmo para todas as redes.

Fonte: Elaborada pela autora.

Outra métrica que usamos para caracterizar as redes mínimas foi o parâmetro de robustez dado pela equação 2.13. A figura 19 mostra a robustez das redes mínimas baseadas em correlação e distribuição de probabilidade para cada mês. Podemos observar que todas as redes baseadas em correlação tem uma rigidez maior do que as redes baseadas em distribuição de probabilidade.

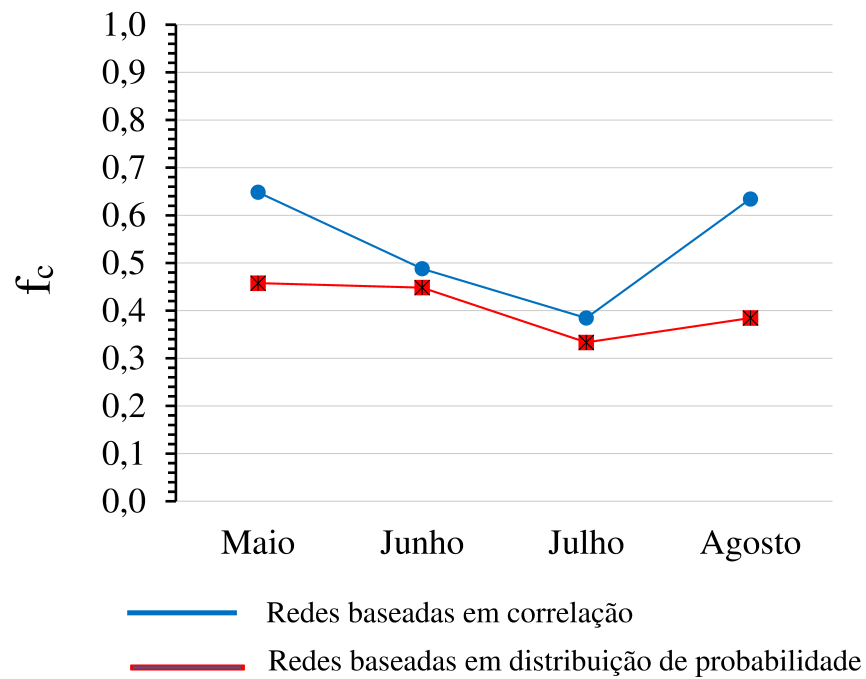


Figura 19 – Robustez das árvores geradoras mínimas.

Fonte: Elaborada pela autora.

As figuras 20, 21, 22 e 23 mostram as classificações das ações de acordo com o grau para cada mês. As ações não apresentam necessariamente o mesmo grau quando comparamos as redes baseadas em distribuição de probabilidade com as redes baseadas em correlação. Além disso, para os períodos considerados as ações mais periféricas das redes (com grau 1) são em menor número nas redes baseadas em distribuição de probabilidade, com exceção do mês de julho em que o número de ações na periferia é o mesmo para ambos os tipos e redes.

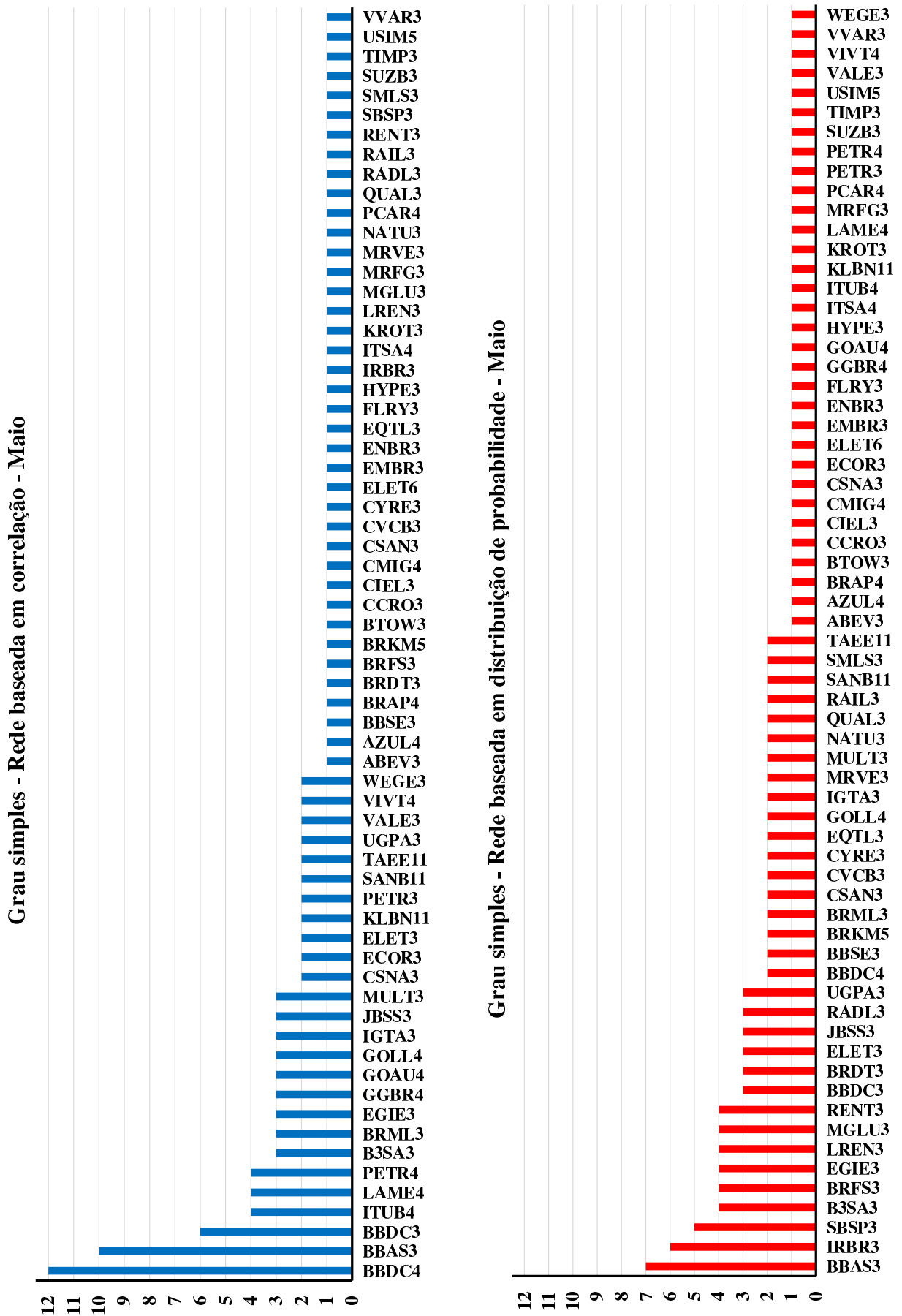


Figura 20 – Grau das ações para o mês de maio de 2019.

Fonte: Elaborada pela autora.



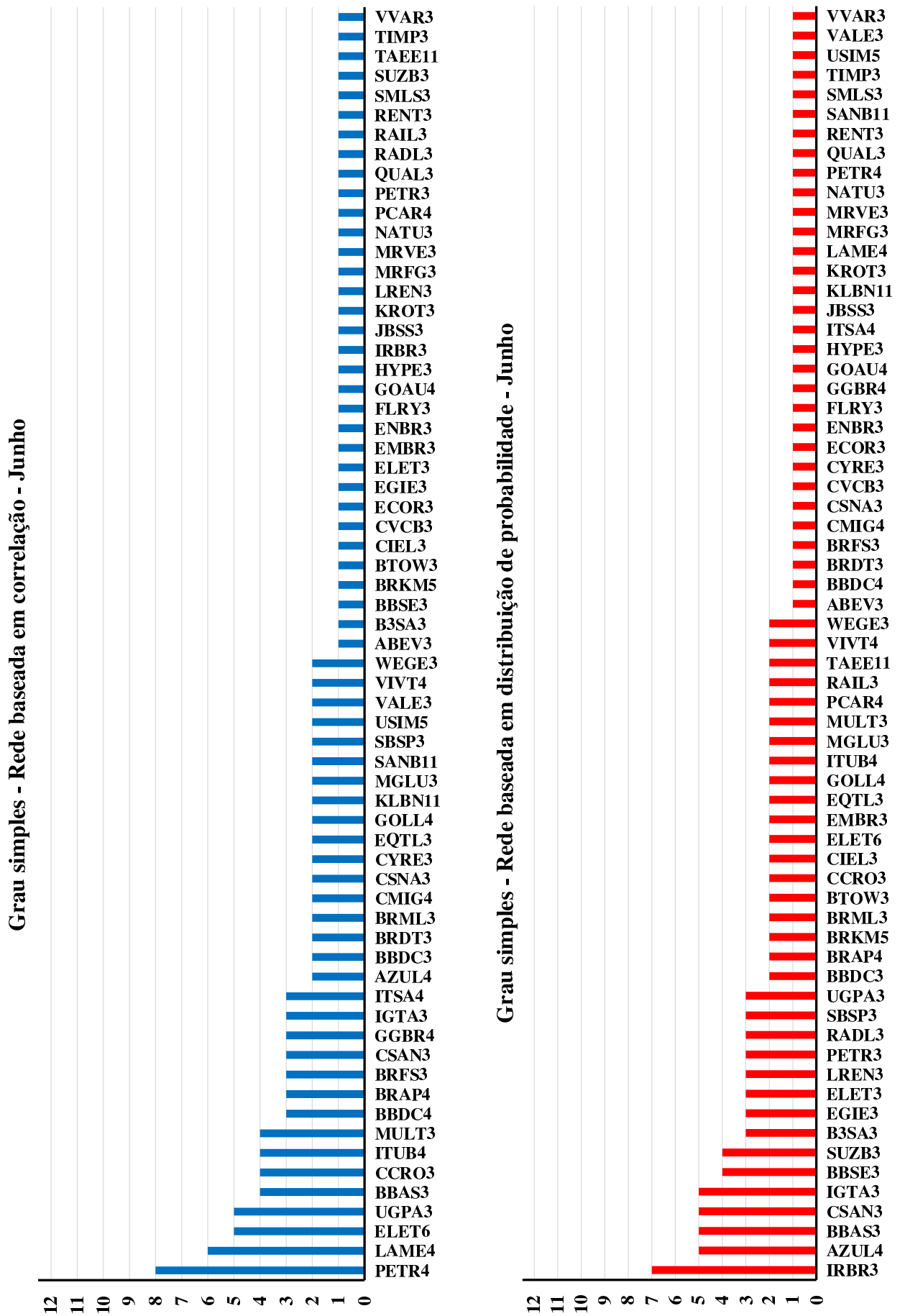


Figura 21 – Grau das ações para o mês de junho de 2019.

Fonte: Elaborada pela autora.

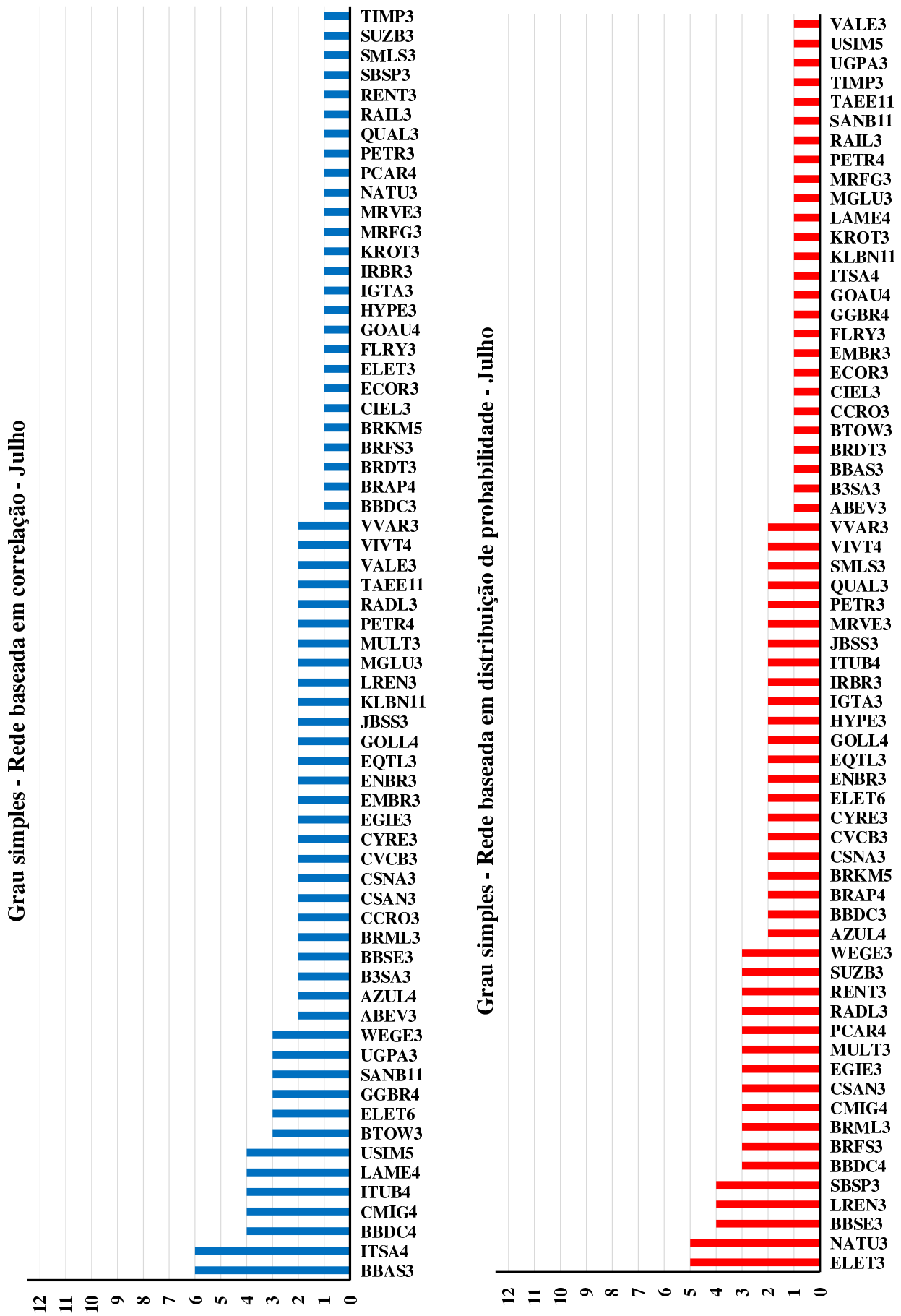


Figura 22 – Grau das ações para o mês de julho de 2019.  
 Fonte: Elaborada pela autora.

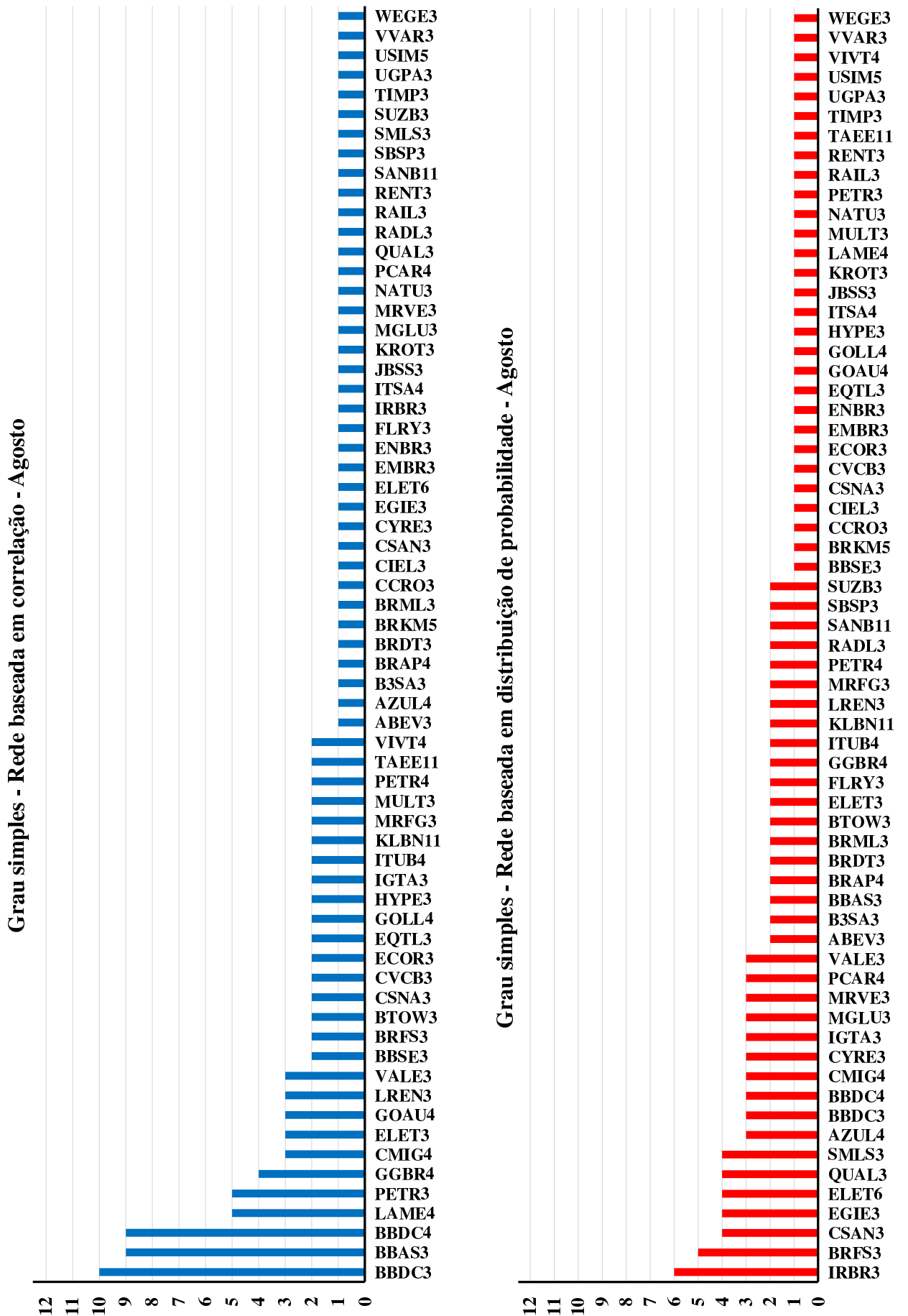


Figura 23 – Grau das ações para o mês de agosto de 2019.

Fonte: Elaborada pela autora.

## 5.2 Árvores geradoras mínimas ponderadas

Nesta seção apresentamos algumas características das redes mínimas considerando que as ligações entre os vértices são ponderadas pela intensidade das correlações no caso de redes de correlação e pela intensidade definida pela equação 3.31 no caso de redes de distribuição de probabilidade. Para obter o grau ponderado de um vértice somam-se os graus simples de cada aresta incidente multiplicado pela intensidade da ligação. Note que o grau ponderado é uma variável contínua. Para visualizar a distribuição de grau ponderado usaremos novamente o diagrama de caixa (em inglês: Boxplot). A figura 24 mostra o diagrama de caixa (boxplot)(JOHNSON; BHATTACHARYYA, 2009) para a distribuição de grau ponderado das árvores mínimas obtidas para os meses de maio, junho, julho e agosto de 2019. Os diagramas indicam que as distribuições de grau são assimétricas para ambos os tipos de redes. O valor mínimo de grau ponderado e a mediana de grau ponderado é sempre maior nas redes baseadas em distribuição de probabilidade.

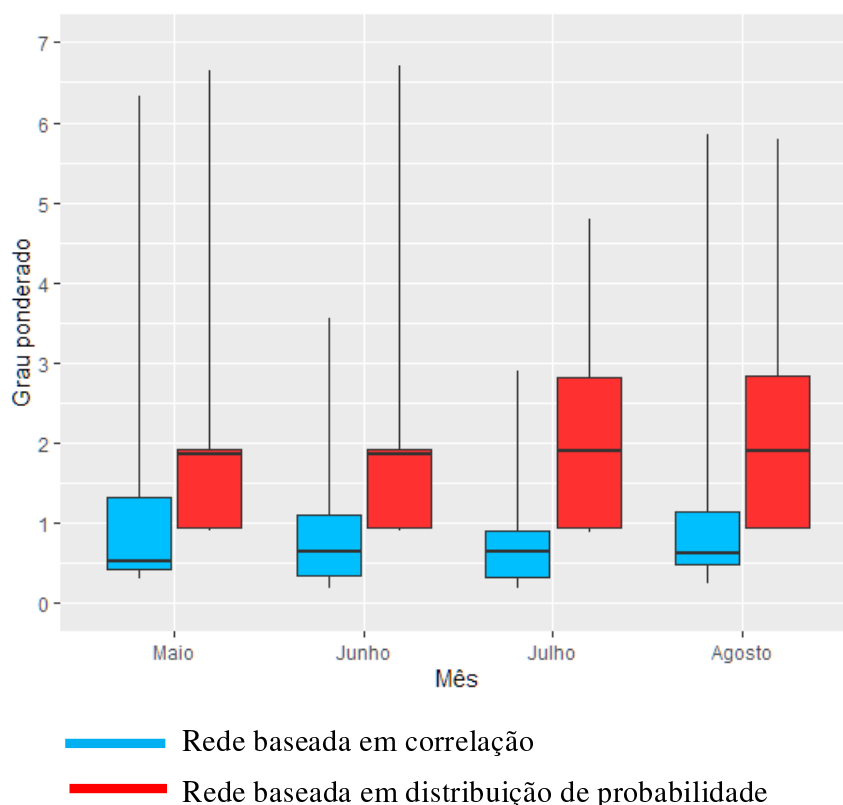


Figura 24 – Diagrama de caixa da distribuição de grau ponderado para as árvores geradoras mínimas.

Fonte: Elaborada pela autora.

Para complementar, as figuras 25(a) e 25(b) mostram respectivamente os valores do grau médio ponderado e do desvio padrão de grau ponderado das redes mínimas para cada mês. Para as redes baseadas em distribuição de probabilidade o grau médio ponderado é praticamente constante ao longo dos períodos e sempre maior do que o grau médio das redes de correlação.

Isso indica que as redes baseadas em distribuição de probabilidade refletem a característica de que as intensidades de ligação são muito próximas do valor máximo 1, isto é, a semelhança entre os ativos em relação as suas distribuições de probabilidade não sofrem grandes alterações no decorrer dos períodos. Além disso, o desvio padrão de grau ponderado das redes baseadas em distribuição de probabilidade também variam menos ao longo dos períodos quando comparado ao das redes baseadas em correlação.

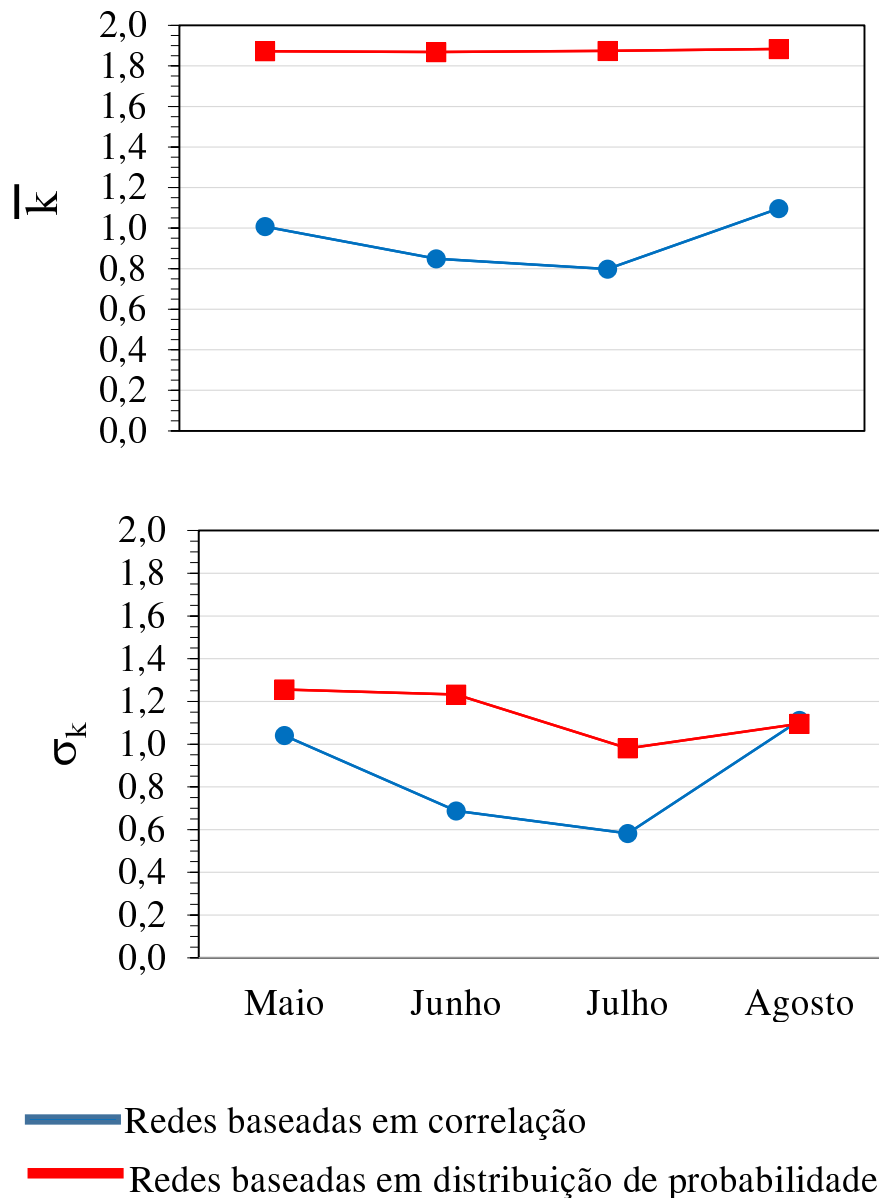


Figura 25 – Grau médio e desvio padrão da distribuição de grau ponderado das árvores geradoras mínimas.

Fonte: Elaborada pela autora.

As figuras 26, 27, 28 e 29 mostram as classificações das ações segundo o grau ponderado para cada mês. Inspeccionando as figuras observa-se que os grupos de ações com mesmo grau simples apresentam pouca variação de grau ponderado nas redes baseadas em distribuição de probabilidade, se comparadas com as redes baseadas em correlação. Novamente, isso decorre do fato que a rede mínima seleciona as intensidades segundo a similaridade entre as distribuições e estas são próximas de 1. Por outro lado, a intensidade da correlação tem maior variabilidade num grupo de ações com mesmo grau simples. Nesse sentido, os graus ponderados das redes baseadas em distribuição de probabilidades são menos adequados para identificar diferenças em um grupo de ações com mesmo grau simples.

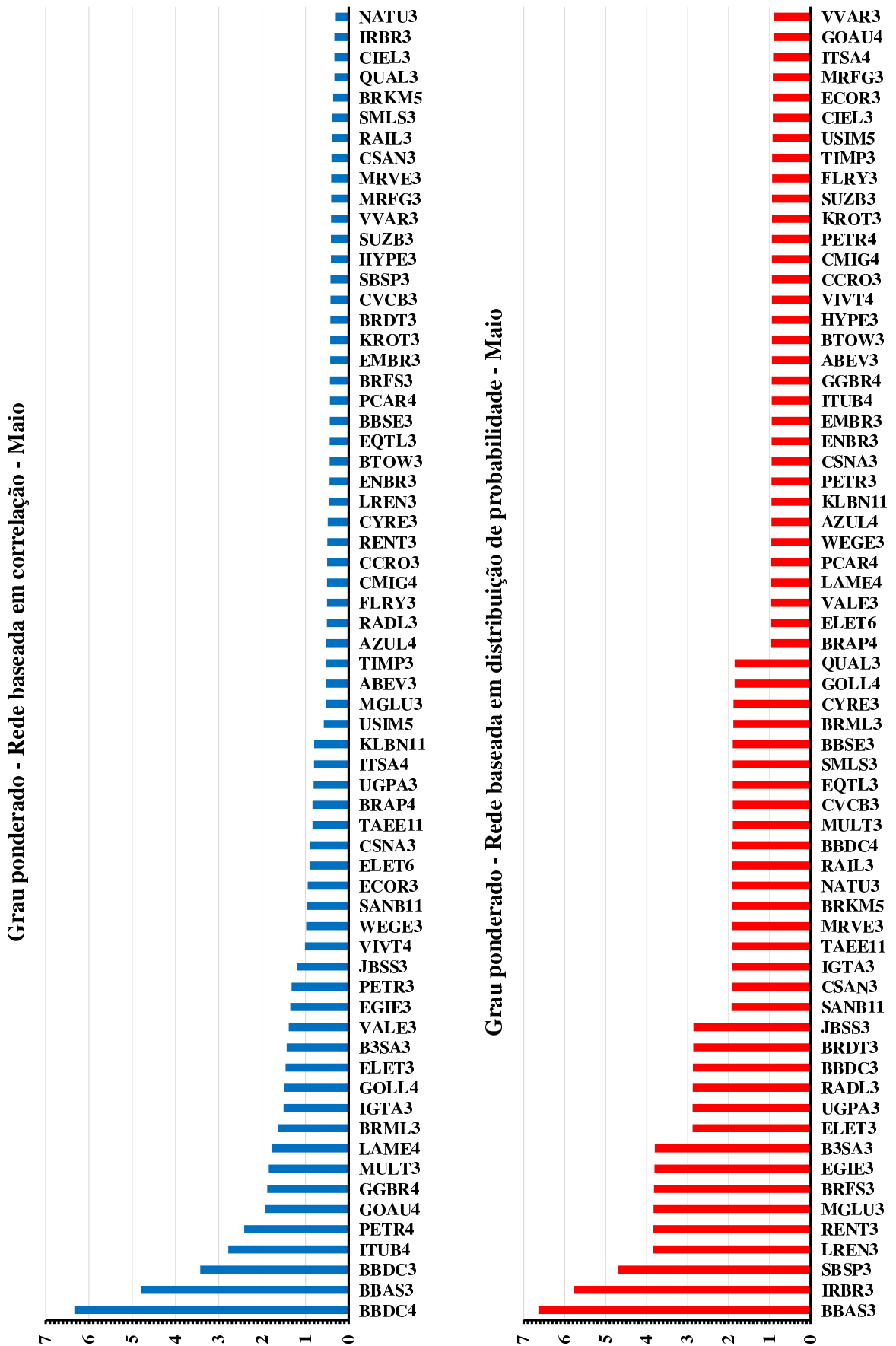


Figura 26 – Grau ponderado das ações para o mês de maio de 2019.

Fonte: Elaborada pela autora.

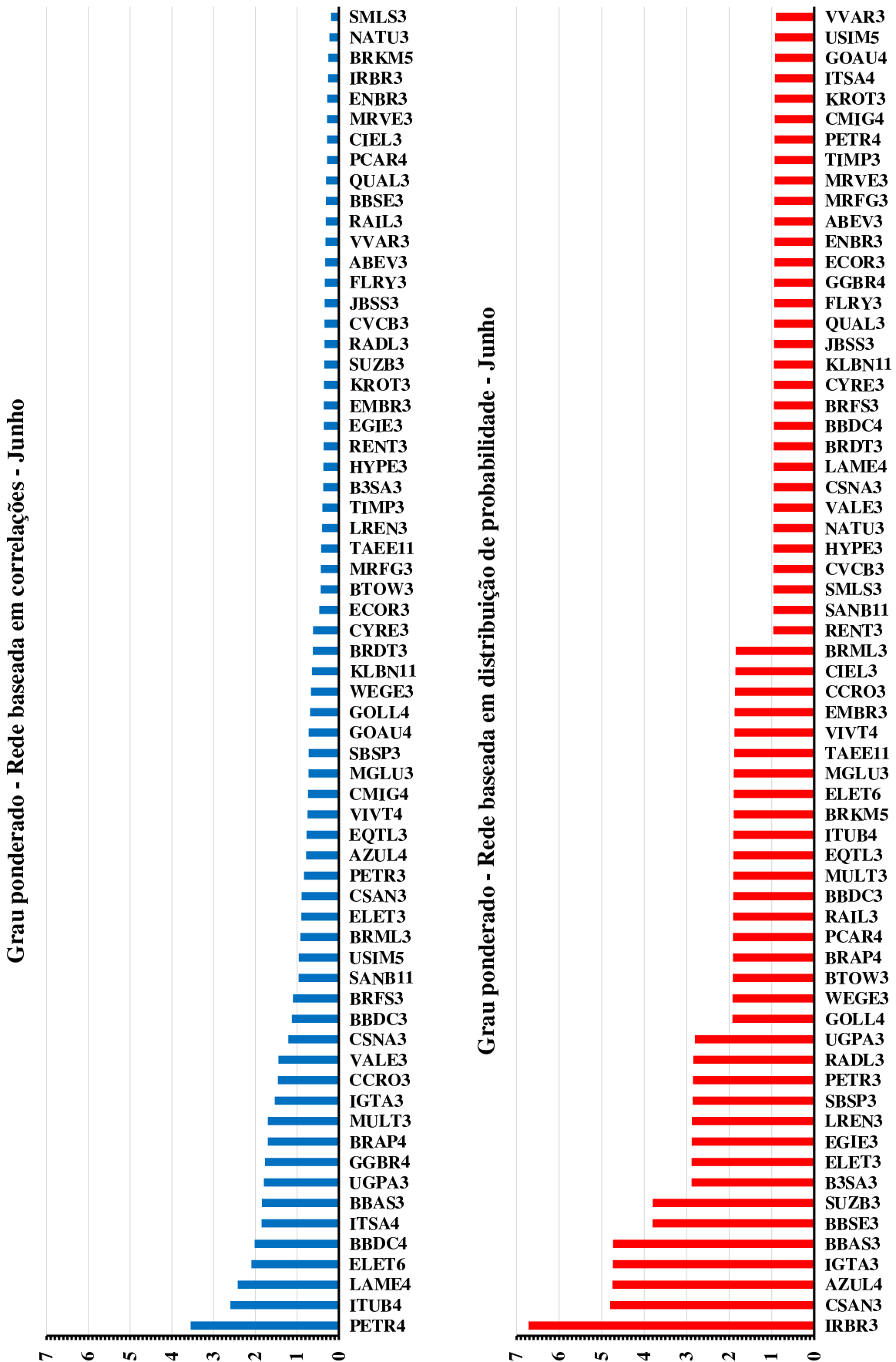


Figura 27 – Grau ponderado das ações para o mês de junho de 2019.

Fonte: Elaborada pela autora.



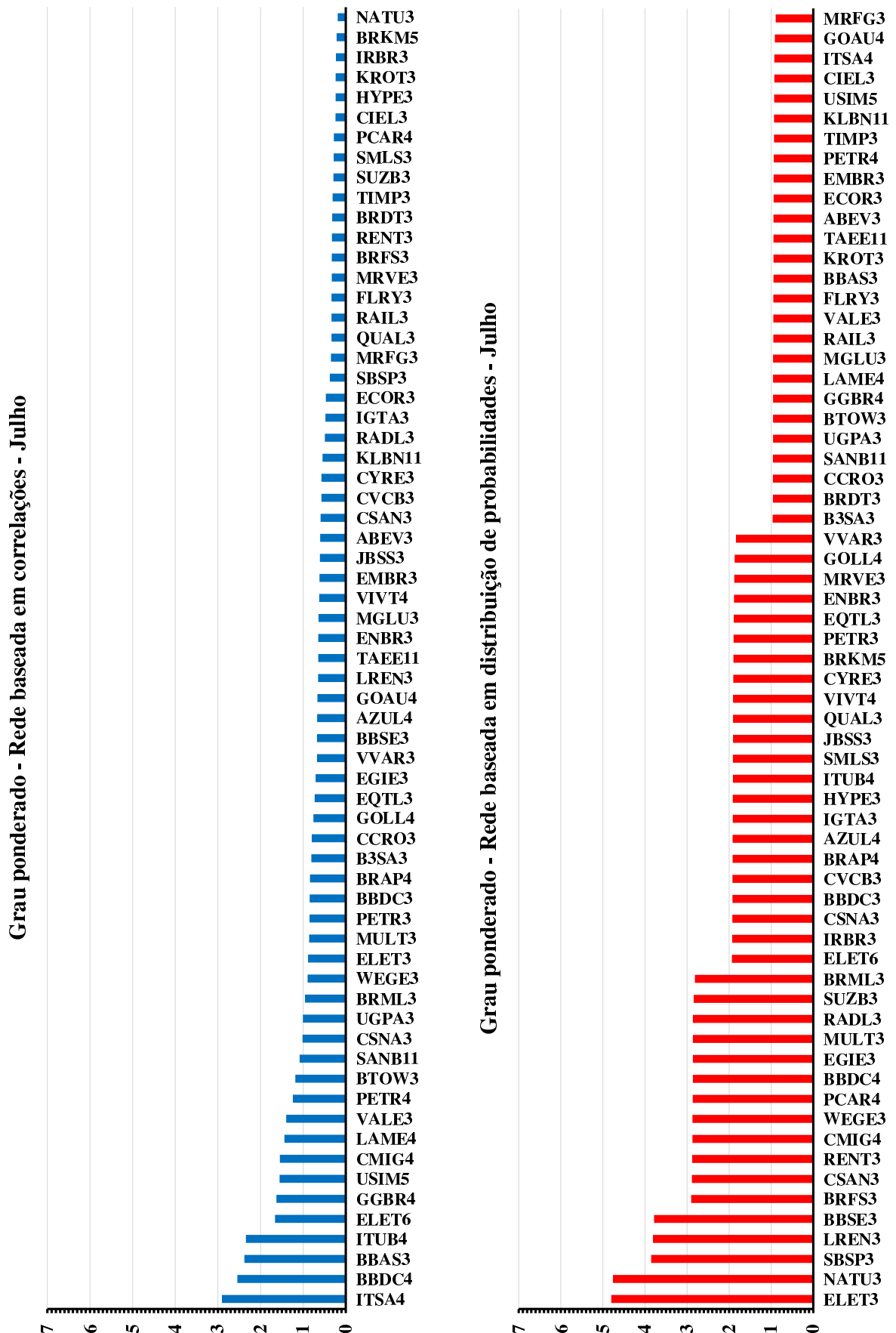


Figura 28 – Grau ponderado das ações para o mês de julho de 2019.

Fonte: Elaborada pela autora.

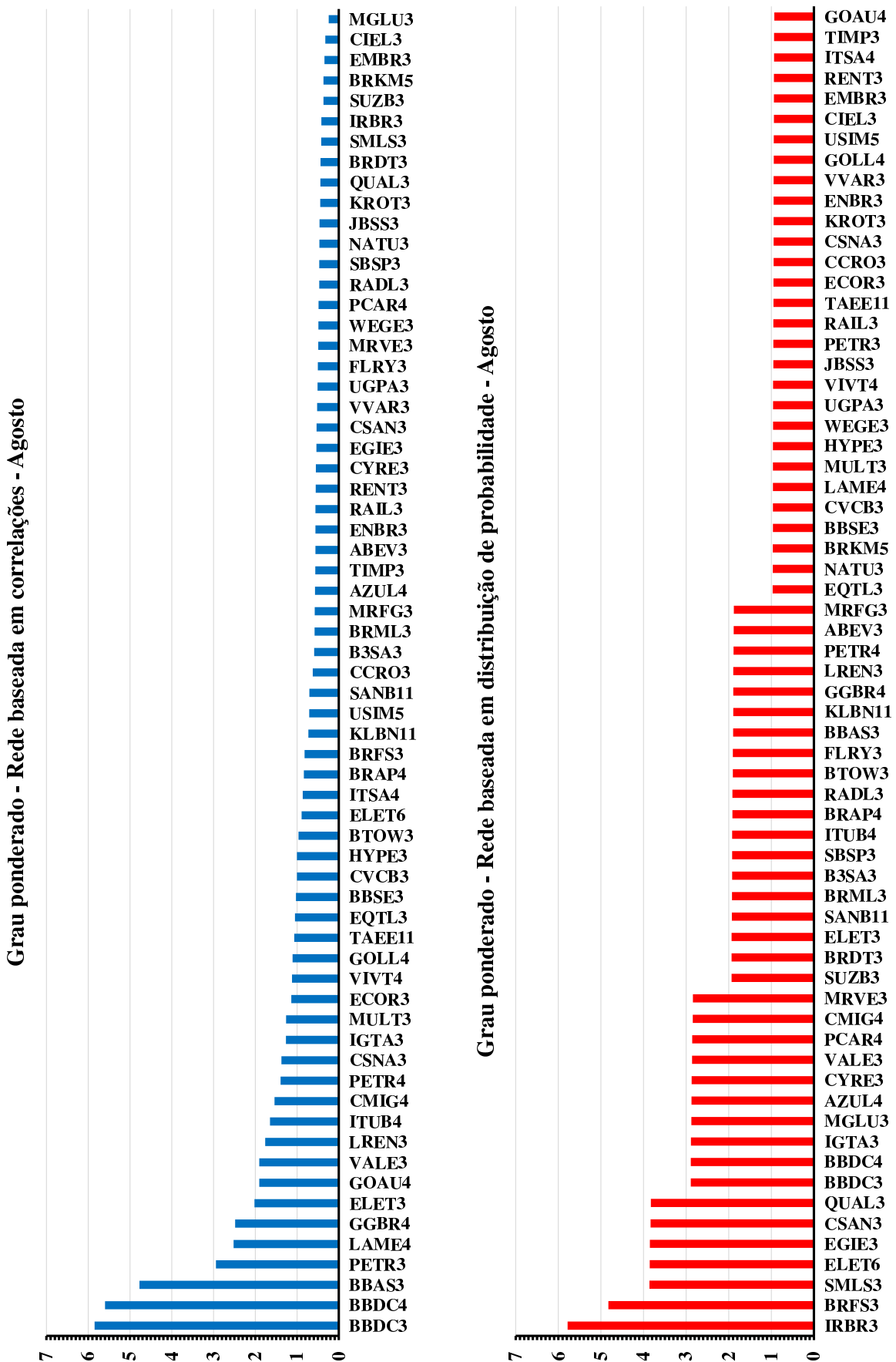


Figura 29 – Grau ponderado das ações para o mês de agosto de 2019.

Fonte: Elaborada pela autora.

## 6 CONCLUSÕES

Nesta dissertação propomos e exploramos uma distância entre distribuições de probabilidade de retornos de ações para obtenção da árvore mínima geradora do mercado financeiro. A proposta foi aplicada para obtenção de árvores geradoras mínimas de uma carteira de ações do IBOVSPA para os meses de maio, junho, julho e agosto de 2019. As árvores mínimas baseadas em distribuições de probabilidade foram comparadas com árvores mínimas obtidas a partir de uma distância baseada em correlação.

Apresentamos as distribuições de intensidade de ligações entre os elementos das redes geradoras mínimas usando um diagrama de caixa. As distribuições de intensidades de ligação das redes mínimas obtidas a partir de distribuições de probabilidade variam em um intervalo de valores menores e mais próximos da intensidade máxima quando comparadas com as distribuições das redes mínimas baseadas em correlação.

Fizemos um estudo considerando as redes mínimas com ligações simples, isto é, que a matriz de adjacência tem valores 1 quando existe ligação entre dois elementos da rede mínima. As distribuições de grau simples apresentam dispersão menor para as redes baseadas em distribuição de probabilidade. As redes mínimas baseadas em distribuição de probabilidade também apresentam valores de rigidez menores do que as das redes baseadas em correlação. As redes baseadas em distribuição de probabilidade apresentam um número menor de ações na periferia da rede (ações com grau 1) quando comparadas as redes baseadas em correlação.

Também estudamos as redes mínimas ponderadas, isto é, com elementos da matriz de adjacência igual a intensidade das ligações definida pela equação 3.31 no caso de redes mínimas baseadas em distribuição de probabilidade e igual ao valor absoluto da correlação no caso de redes mínimas baseadas em correlação. A distribuição de grau das redes ponderadas baseadas em distribuição de probabilidade apresentaram grau médio ponderado praticamente constante ao longo dos meses e com valores sempre acima do grau médio ponderado das redes baseadas em correlação. A classificação das ações nos graus ponderados das redes baseadas em distribuição de probabilidade tem pouca variabilidade se consideradas ações com mesmo grau simples quando comparadas com as redes baseadas em correlação. Nesse sentido, os graus ponderados das ações nas redes baseadas em distribuição de probabilidades são menos adequadas do que os das redes baseadas em correlação para identificar diferenças em um grupo de ações com mesmo grau simples.

Concluindo, as redes mínimas baseadas na distância (ou similaridade) entre as distribuições de probabilidade dos retornos de ativos do mercado financeiro representam um resumo das conexões mais importantes em termos de similaridade entre distribuições de probabilidade que por sua vez são relacionadas ao Valor em Risco dos ativos (VAR). Assim, em trabalhos futuros seria interessante explorar as redes mínimas baseadas em distribuição de probabilidade entre ativos do mercado financeiro como uma ferramenta alternativa na identificação do risco sistêmico em mercados financeiros.

## REFERÊNCIAS

- BARABÁSI, Albert-László. *Network science*. [S.l.]: Cambridge university press, 2016.
- BARBI, AQ; PRATAVIERA, GA. Nonlinear dependencies on brazilian equity network from mutual information minimum spanning trees. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 523, p. 876–885, 2019.
- BARBI, Alex Quintino. *A informação mútua como medida de dependência não linear na estrutura de rede do mercado brasileiro de ações*. Tese (Doutorado) — Universidade de São Paulo, 2017.
- BOLLOBÁS, Béla. *Modern graph theory*. [S.l.]: Springer Science & Business Media, 2013. v. 184.
- BONANNO, Giovanni et al. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E, APS*, v. 68, n. 4, p. 046130, 2003.
- BONANNO, Giovanni; VANDEWALLE, Nicolas; MANTEGNA, Rosario N. Taxonomy of stock market indices. *Physical Review E, APS*, v. 62, n. 6, p. R7615, 2000.
- CAJUEIRO, Daniel O; TABAK, Benjamin M. The role of banks in the brazilian interbank market: does bank type matter? *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 387, n. 27, p. 6825–6836, 2008.
- CARRINGTON, Peter J; SCOTT, John; WASSERMAN, Stanley. *Models and methods in social network analysis*. [S.l.]: Cambridge university press, 2005. v. 28.
- COELHO, Ricardo et al. The evolution of interdependence in world equity markets: Evidence from minimum spanning trees. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 376, p. 455–466, 2007.
- DEGROOT, Morris H; SCHERVISH, Mark J. *Probability and statistics*. [S.l.]: Pearson Education, 2012.
- EASLEY, David; KLEINBERG, Jon. *Networks, crowds, and markets: Reasoning about a highly connected world*. [S.l.]: Cambridge University Press, 2010.
- ERDÖS, Paul. Graph theory and probability. *Canadian Journal of Mathematics*, Cambridge University Press, v. 11, p. 34–38, 1959.
- EULER, Leonhard. Leonhard euler and the königsberg bridges. *Scientific American, JSTOR*, v. 189, n. 1, p. 66–72, 1953.
- GARAS, Antonios; ARGYRAKIS, Panos. Correlation study of the athens stock exchange. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 380, p. 399–410, 2007.
- GOWER, John C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, Oxford University Press, v. 53, n. 3-4, p. 325–338, 1966.
- JACKSON, Matthew O. *Social and economic networks*. [S.l.]: Princeton university press, 2010.
- JOHNSON, Richard A.; BHATTACHARYYA, Gouri K. *Statistics: principles and methods*. [S.l.]: John Wiley & Sons, 2009.

JOHNSON, Richard A; WICHERN, Dean. *Multivariate analysis*. [S.l.]: Wiley Online Library, 2002.

JORION, Philippe. *Value at risk*. McGraw-Hill Professional Publishing, 2000.

JUNG, Woo-Sung et al. Characteristics of the Korean stock market correlations. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 361, n. 1, p. 263–271, 2006.

KREYSZIG, Erwin. *Advanced Engineering Mathematics, 8-th edition*. [S.l.]: John Wiley & Sons, 1999.

KRUSKAL, Joseph B. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, JSTOR, v. 7, n. 1, p. 48–50, 1956.

KWAPIEN, Jaroslaw; GWOREK, Sylwia; DROZDZ, Stanislaw. Structure and evolution of the foreign exchange networks. *arXiv preprint arXiv:0901.4793*, 2009.

MALEVERGNE, Yannick; SORNETTE, Didier. *Extreme Financial Risks: From Dependence to Risk Management*. [S.l.]: Springer, 2006.

MANTEGNA, Rosario N. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, Springer, v. 11, n. 1, p. 193–197, 1999.

MANTEGNA, Rosario N; STANLEY, H Eugene. *Introduction to econophysics: correlations and complexity in finance*. [S.l.]: Cambridge university press, 1999.

MARKOWITZ, Harry. Portfolio selection. *The journal of finance*, Wiley Online Library, v. 7, n. 1, p. 77–91, 1952.

MATTEO, Tiziana Di et al. Interest rates hierarchical structure. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 355, n. 1, p. 21–33, 2005.

MICCICHÈ, Salvatore et al. Degree stability of a minimum spanning tree of price return and volatility. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 324, n. 1-2, p. 66–73, 2003.

MOLLOY, Michael; REED, Bruce. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, Wiley Online Library, v. 6, n. 2-3, p. 161–180, 1995.

NAYLOR, Michael J; ROSE, Lawrence C; MOYLE, Brendan J. Topology of foreign exchange markets using hierarchical structure methods. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 382, n. 1, p. 199–208, 2007.

NEWMAN, Mark. *Networks: an introduction*. [S.l.]: Oxford university press, 2010.

ONNELA, J-P et al. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, APS, v. 68, n. 5, p. 056110, 2003.

ORIGUELA, Leticia Aparecida. *Estudo da influência de eventos sobre a estrutura do mercado brasileiro de ações a partir de redes ponderadas por correlações de Pearson, Spearman e Kendall*. Tese (Doutorado) — Universidade de São Paulo, 2018.

ORTEGA, Guillermo J; MATESANZ, David. Cross-country hierarchical structure and currency crises. *International Journal of Modern Physics C*, World Scientific, v. 17, n. 03, p. 333–341, 2006.

- PAPADIMITRIOU, Theophilos; GOGAS, Periklis; TABAK, Benjamin M. Complex networks and banking systems supervision. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 392, n. 19, p. 4429–4434, 2013.
- PERALTA, Gustavo; ZAREEI, Abalfazl. A network approach to portfolio selection. *Journal of Empirical Finance*, Elsevier, v. 38, p. 157–180, 2016.
- POZZI, Francesco; MATTEO, Tiziana Di; ASTE, Tomaso. Centrality and peripherality in filtered graphs from dynamical financial correlations. *Advances in Complex Systems*, World Scientific, v. 11, n. 06, p. 927–950, 2008.
- SIECZKA, Paweł; HOŁYST, Janusz A. Correlations in commodity markets. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 388, n. 8, p. 1621–1630, 2009.
- SINHA, Sitabhra et al. *Econophysics: an introduction*. [S.l.]: John Wiley & Sons, 2010.
- SORAMAKI, Kimmo; COOK, Samantha; LAUBSCH, Alan. A network-based method for visual identification of systemic risks. *Journal of Network Theory in Finance*, v. 2, n. 1, p. 67–101, 2016.
- TABAK, Benjamin M; CAJUEIRO, Daniel O; SERRA, Thiago R. Topological properties of bank networks: the case of brazil. *International Journal of Modern Physics C*, World Scientific, v. 20, n. 08, p. 1121–1143, 2009.
- TABAK, Benjamin M; LUDUVICE, André Victor D; CAJUEIRO, Daniel O. Modeling default probabilities: The case of brazil. *Journal of International Financial Markets, Institutions and Money*, Elsevier, v. 21, n. 4, p. 513–534, 2011.
- TABAK, Benjamin Miranda et al. *Conectividade e risco sistêmico no sistema de pagamentos brasileiro*. [S.l.], 2012.
- TABAK, Benjamin M; SERRA, Thiago R; CAJUEIRO, Daniel O. Topological properties of stock market networks: The case of brazil. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 389, n. 16, p. 3240–3249, 2010.
- VANDEWALLE, N et al. Non-random topology of stock markets. *Quantitative Finance*, Taylor & Francis, v. 1, n. 3, p. 372–374, 2001.