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

ALEX QUINTINO BARBI

**A informação mútua como medida de dependência não linear na estrutura
de rede do mercado brasileiro de ações**

Orientador: Prof. Dr. Gilberto Aparecido Prativiera

RIBEIRÃO PRETO
2017

Prof. Dr. Marco Antônio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Dante Pinheiro Martinelli
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Marcio Mattos Borges de Oliveira
Chefe do Departamento de Administração

ALEX QUINTINO BARBI

A informação mútua como medida de dependência não linear na estrutura de rede do mercado brasileiro de ações

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Organizações da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto - Universidade de São Paulo, como requisito para o título de Mestre em Ciências - Área: Administração de Organizações. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP.

Orientador: Prof. Dr. Gilberto
Aparecido Prata

RIBEIRÃO PRETO

2017

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.

Barbi, Alex Quintino

A informação mútua como medida de dependência não linear na estrutura de rede do mercado brasileiro de ações / Alex Quintino Barbi – Ribeirão Preto, 2017.

147p.: il.; 30 cm

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Organizações da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto - Universidade de São Paulo, como requisito para o título de Mestre em Ciências - Área: Administração de Organizações. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP. – Universidade de São Paulo

Orientador: Prata, Gilberto Aparecido

1. Mercado de Ações. 2. Modelos de Dependência. 3. Informação Mútua. 4. Análise de Redes Complexas. .

ALEX QUINTINO BARBI

A informação mútua como medida de dependência não linear na estrutura de rede do mercado brasileiro de ações

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Organizações da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto - Universidade de São Paulo, como requisito para o título de Mestre em Ciências - Área: Administração de Organizações. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP.

Área de Concentração: Administração de Organizações

Data de Aprovação: 15/12/2017

Banca Examinadora:

Prof. Dr. Gilberto Aparecido Prata
Orientador

Prof. Dr. Evandro Marcos Saidel Ribeiro

Prof. Dr. Francisco Aparecido Rodrigues

Prof. Dr. Marcos César de Oliveira

AGRADECIMENTOS

Primeiramente, a meu orientador, Prof. Dr. Gilberto Aparecido Prata, pelo entusiasmo com este trabalho e pela extrema dedicação com a qual conduz seus ensinamentos.

Ao Prof. Joaquim Quintino Filho por suas sábias ideias e pela revisão gramatical do texto.

Aos professores que também dedicaram seu tempo ao projeto: prof. Dr. Evandro Saidel Ribeiro e prof. Dr. Carlos Alberto Gabrielli Barreto Campello.

Aos funcionários da FEA-RP, especialmente ao André Luiz Martins Pignata, por sua maestria na depuração dos códigos deste trabalho.

Não poderia deixar de agradecer ao colega Matheus Albino por ter disponibilizado este template em $\text{\LaTeX}$ para a comunidade FEA-RP.

Por fim, este autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

“Some things benefit from shocks; they thrive and grow when exposed to volatility, randomness, disorder, and stressors and love adventure, risk, and uncertainty.”

Nassim Nicholas Taleb, 2014.

RESUMO

Título: A informação mútua como medida de dependência não linear na estrutura de rede do mercado brasileiro de ações.

Mercados financeiros são sistemas complexos com estrutura e comportamento extremamente dependentes das interrelações entre os seus componentes. Em particular, a teoria de redes tem contribuído para caracterizar e compreender o comportamento e as interdependências entre vários componentes do mercado financeiro, em especial, o mercado de ações. Pesquisas nessa área indicam que a estrutura de rede gerada do mercado pode conter informações úteis para um melhor entendimento do mercado como um todo e até mesmo prever a ocorrência de eventos extremos, como, por exemplo, uma crise financeira. Em geral, os estudos consideram apenas dependências lineares entre os objetos da rede baseados no coeficiente de correlação linear de Pearson, e nesse sentido, a proposta deste projeto é a aplicação de conceitos e métodos de teoria de redes e de teoria da informação para caracterizar e explorar o efeito de dependências não lineares na estrutura de rede do mercado brasileiro de ações. Para tal, a informação mútua foi usada como medida de dependência não linear para gerar a estrutura de redes que foi comparada com a obtida a partir da correlação linear de Pearson. Por fim, investigou-se como a estrutura da rede e suas métricas poderiam ajudar a caracterizar e a entender o comportamento dos mercados financeiros, analisando-se dois períodos, o primeiro sob gestão da Presidente Dilma Rousseff, com um retorno do índice de ações de -42%, e o segundo sob gestão do Presidente Michel Temer, com um retorno deste índice de 50%. Para tal fim, foram utilizados dados de alta frequência, sendo uma cotação a cada 15 minutos. Em suma, concluiu-se que os retornos dos ativos no segundo período parecem ter maior dependência não-linear quando comparados aos retornos do período anterior. A rede para este período é a que se mostra mais arriscada em termos de estrutura de 'transmissão de volatilidades', tanto pela análise do coeficiente de robustez da rede, quanto pela estimativa do parâmetro da lei de potência. Encontrou-se evidência da relação entre estrutura das redes e desempenho das ações. Além disso, vimos a grande importância do setor financeiro nas redes. Finalmente, tecemos comentários quanto a aplicação destas redes para diversos fins.

Palavras-chave: Mercado de Ações, Teoria da Informação, Redes Complexas

SHORT ABSTRACT

Title: The mutual information as a nonlinear dependence measure in Brazilian network financial assets structure.

This paper has the purpose to apply concepts and methods from network and information theory to characterize and to explore the role of nonlinear dependencies over the Brazilian network stock market structure. In particular, the minimum spanning tree network structure generated from the mutual information as a measure of nonlinear dependence was compared with the one obtained by Pearson's correlation coefficient. We analyzed two periods, the first under the management of President Dilma Rousseff, with an index return of -42%, and the second one, under the management of President Michel Temer, with an index return of 50%. For this purpose, high frequency data of fifteen minutes interval was used. Our analysis suggest that the assets returns of Temer's presidential term seem to have greater nonlinear dependence when compared to the returns of the previous period. Also, the network's robustness coefficient and power law parameter suggests that the network for the second period is the most risky in terms of volatility transmission structure. Also, we find evidence of network structure and stock performance relationship. Finally, we have also seen the great importance of financial sector within Brazilian's stock network

Keywords: Stock Market, Information Theory, Complex Networks

LISTA DE ILUSTRAÇÕES

2.1	Rede de doenças humanas	22
2.2	Rede da internet	22
2.3	Ilustração das pontes da cidade de Königsberg.	23
2.4	Ilustração pensada por Euler para o problema das sete pontes de Königsberg.	24
2.5	Exemplo de representação de grafos pela matriz de adjacência	25
2.6	Tipos de centralidade	30
2.7	Ilustração de um componente gigante	32
2.8	Exemplo de distribuição de Poisson	33
2.9	Distribuição de Poisson e Empírica para três redes	33
2.10	Distribuição de Poisson e lei de potência	35
2.11	Hubs em redes livre de escala	37
2.12	Regimes do modelo Barabási-Albert	40
2.13	Níveis de energia de Bose-Einstein	43
2.14	Exemplo de simulação de percolação em grades	44
2.15	Ciranda: um componente gigante se forma	45
2.16	Falhas e ataques em redes livre de escala	46
2.17	Redes assortativas e dissortativas	49
2.18	Comunidades Francesas e Alemãs na Bélgica	51
2.19	Modularidade em partições	53
3.1	Diagramas de Anscombe	61
3.2	Coeficiente de correlação de Pearson para três relações funcionais	62
3.3	Exemplo de distribuição normal bivariada	63
3.4	Exemplo de famílias de cópulas	67
3.5	Spearman vs. Pearson	69
3.6	Dependência de cauda para cópulas	71
4.1	Entropia e probabilidade	75
4.2	Relação entre entropia e informação mútua	76
4.3	Exemplos de histogramas para cada largura de janela	80
4.4	Exemplos de histogramas para origens diferentes	81
4.5	Exemplos de kernels para janelas diferentes	82
4.6	Estimativa por kernel - Soma de Densidades	85
4.7	Estimativa de densidade pelo kernel normal bivariado	85
4.8	Estimativa de densidade empírica	87
4.9	Estimativa de densidade pelo kernel normal bivariado para ações brasileiras	87
5.1	Distribuições dos retornos de 100 ações americanas	90

5.2	Árvore geradora mínima para ações baseada em distância ultramétrica	92
5.3	Arranjos: Redes de ações reais vs. redes gaussianas aleatórias	93
5.4	AGM em comparação com a PMF	94
5.5	Correlação linear média e volatilidade do período	95
5.6	Assortatividade e dissortatividade em redes de ações	96
5.7	Ações centrais vs. ações periféricas	97
5.8	Distribuição empírica da informação mútua para alguns países	98
5.9	Árvore geradora para três períodos ao redor da crise de 2008	99
5.10	Correlação e informação mútua: matrizes de calor	101
5.11	Redes de informação mútua e suas centralidades.	103
5.12	Redes de dependência de caudas inferiores e superiores	104
7.1	Árvores geradora mínima (AGM) para o período de Dilma Rousseff.	113
7.2	Árvores geradora mínima (AGM) para o período de Michel Temer.	114
7.3	Distribuição dos valores absolutos da correlação linear de Pearson e do coeficiente global de correlação entre os períodos	115
7.4	Matriz simétrica cujos valores são dados por $ \lambda_{ij} - \rho_{ij} $	116
7.5	Função de distribuição cumulativa complementar baseada no grau ponderado . . .	117
A.1	Fronteira eficiente do portfólio de ações para o período DR (Dilma).	138
A.2	Fronteira eficiente do portfólio de ações para o período MT (Temer).	138

LISTA DE TABELAS

1.1	Objetivos propostos para a pesquisa	20
2.1	Exemplos de avalanche para algumas redes reais	48
6.1	Ações selecionadas para compor a análise	107
7.1	Relação entre centralidade e desempenho das ações.	118
7.2	Resultados da Regressão - período Dilma	119
7.3	Resultados da regressão - período Temer	120
7.4	Principais resultados	122
7.5	Descrição dos setores acionários com maiores e menores centralidades	123

Sumário

1	INTRODUÇÃO	15
1.1	Relevância deste estudo	17
1.2	Propostas e objetivos	19
1.3	Organização do trabalho	20
2	REDES COMPLEXAS	21
2.1	Conceitos básicos de redes	23
2.1.1	Nós e ligações	23
2.1.2	O grau de um nó	24
2.1.3	Matriz de adjacência	25
2.1.4	Distribuição de grau	26
2.1.5	Distâncias e caminhos	26
2.1.6	Árvores	27
2.1.7	Transitividade	28
2.1.8	Outras medidas de nós	28
2.2	Redes randômicas	30
2.2.1	Ligações em redes randômicas	30
2.2.2	Regimes em redes randômicas	31
2.2.3	Distribuição de grau para redes randômicas	32
2.2.4	Fenômeno de mundos pequenos em redes randômicas	33
2.2.5	Coeficiente de agrupamento para redes randômicas	34
2.2.6	Redes reais não são randômicas	34
2.3	Redes livre de escala	34
2.3.1	Lei de potência	35
2.3.2	Nodos centrais (Hubs)	36
2.3.3	Significado de livre de escala	36
2.3.4	Ultra mundos pequenos	38
2.3.5	Redes reais são livre de escala	38
2.3.6	Modelo Barabási-Albert	39
2.3.7	Modelo Bianconi-Barabási	40
2.4	Robustez em redes	42
2.4.1	Teoria de percolação	43
2.4.2	Robustez em redes livre de escala	44
2.4.3	Ataques em redes	45

2.4.4	Ajustando a robustez	47
2.4.5	Efeitos cascata	47
2.5	Aspectos mesoscópicos de redes	48
2.5.1	Arquiteturas de ligações: assortatividade e dissortatividade	48
2.5.2	Comunidades em redes	50
2.5.3	Caracterizando comunidades	53
2.6	Softwares para redes complexas	54
2.7	Como construir redes do mercado de ações?	55
3	MEDIDAS DE DEPENDÊNCIA	57
3.1	Independência e dependência	57
3.1.1	Examinando dependências	59
3.2	Dependências lineares	59
3.2.1	Popularidade da distribuição normal em finanças	62
3.3	Propriedades da dependência	64
3.4	Cópuas	64
3.5	Correlação por ranqueamento ou concordância	66
3.6	Dependência de cauda	70
3.7	Métricas da teoria da informação	71
4	TEORIA DA INFORMAÇÃO	73
4.1	Medidas de informação	74
4.1.1	Entropia conjunta e condicional	75
4.1.2	Informação mútua	76
4.2	Entropia de distribuições contínuas de probabilidade	77
4.3	Estimando entropia e informação	78
4.3.1	Estimando densidades empiricamente	79
4.3.2	Outros métodos	84
4.4	Informação mútua entre ações	86
5	REDES NO MERCADO DE AÇÕES	89
5.1	Complexidade em mercados financeiros	89
5.2	Redes no mercado de ações	91
5.2.1	Redes de ações com estruturas lineares	91
5.2.2	Redes de ações com estruturas não lineares	98
6	METODOLOGIA	105
6.1	Base de dados	105
6.1.1	Comentários sobre os períodos	106

6.2	A estrutura de rede	109
6.3	Código-fonte final	110
7	RESULTADOS E DISCUSSÕES	111
7.1	Comentários	121
8	CONCLUSÕES	125
8.1	Limitações e pesquisas futuras	126
8.2	Antifrágeis	127
	REFERÊNCIAS	129
A	ANEXO A - O ÍNDICE DE SHARPE	137
A	APÊNDICE A - MATERIAL SUPLEMENTAR	139
A.1	Código em R para entropia e informação mútua	139
A.2	Código em R para teste do código do Apêndice A.1	141
A.3	Transformando matrizes de correlação em grafos (redes)	142
A.4	Código final para cálculo da informação mútua para n vetores	142
A.5	Testes de bandas	144
A.6	Base de dados	146

CAPÍTULO 1

INTRODUÇÃO

Uma rede é um conjunto de elementos discretos (vértices) e um conjunto de conexões (arestas) que ligam os elementos aos pares (NEWMAN, 2010; BARABASI, 2016). Os elementos e suas conexões podem ser diversos objetos - pessoas e amigos (RAPOPORT; HORVATH, 1961); páginas web (BARABASI; ALBERT, 1999), doenças infecciosas (BARRAT; BARTHELEMY; VESPIGNANI, 2008), ou mesmo ações no mercado financeiro (MANTEGNA, 1999).

O estudo da estrutura de redes ganhou desenvolvimento nas décadas de 1940, 50 e 60. Dentre alguns trabalhos importantes, destaca-se o de Solomonoff e Rapoport (1951), no qual os autores apresentam um primeiro estudo sistemático de grafos aleatórios, e o de Erdős e Rényi (1960), considerados os pais da teoria moderna de grafos aleatórios. Desde então o estudo de redes tornou-se um campo ativo de pesquisa, tanto do ponto de vista de ciência básica (TRAVERS; MILGRAM, 1969; WATTS; STROGATZ, 1998; BARABASI; ALBERT, 1999; NEWMAN, 2003), como em aplicações em redes reais (MANTEGNA, 1999; CALDARELLI; VESPIGNANI, 2007; ROSVALL; BERGSTROM, 2007; BARRAT; BARTHELEMY; VESPIGNANI, 2008; BARABASI, 2016; SILVA; ZHAO, 2016). Várias medidas, tais como grau, centralidade de proximidade, centralidade de intermediação, modularidade e pageRank foram introduzidas para caracterizar o comportamento de redes e utilizadas no estudo de seus mais diversos tipos (NEWMAN, 2003).

O primeiro passo no estudo de redes é a regra que gera a estrutura de rede e que diz como objetos são ou não conectados, ou seja, como eles interagem. O conceito básico é o de matriz de adjacência, com elementos zero ou um, representando a ligação ou não ligação entre dois objetos, respectivamente (NEWMAN, 2010). Alternativas ponderadas da matriz de adjacência também são possíveis, substituindo-se o conceito de ligado/não ligado pelo de intensidade da

ligação (NEWMAN, 2010). Uma alternativa interessante para a matriz de adjacência é baseada no coeficiente de correlação de Pearson como medida do grau de interação entre objetos e tem sido muito utilizado no estudo de redes financeiras (MANTEGNA; STANLEY, 1999; TABAK; SERRA; CAJUEIRO, 2010). A partir do coeficiente de correlação, pode-se definir uma medida de distância que permite obter a chamada árvore geradora mínima da rede (KRUSKAL, 1956; MANTEGNA, 1999). A árvore geradora mínima fornece uma estrutura hierárquica que é então utilizada para representar as maiores correlações. Esta ideia foi proposta pela primeira vez por Mantegna (1999) para o estudo de correlações de ativos com o objetivo de extrair uma sub-rede com as conexões mais relevantes. A árvore geradora mínima reduz a complexidade de uma rede, pois permite a visualização das conexões não redundantes mais importantes. A partir do trabalho de Mantegna (1999), vários trabalhos considerando correlações e a árvore geradora mínima foram sendo realizados (MICCICHÈ et al., 2003; COELHO et al., 2007; TABAK; SERRA; CAJUEIRO, 2010; GILMORE; LUCEY; BOSCIA, 2010; ZHANG et al., 2011; SENSOY; YUKSEL; ERTURK, 2013; SENSOY; TABAK, 2014). Entretanto, o coeficiente de correlação captura apenas a dependência linear entre variáveis. Assim, torna-se importante introduzir, no estudo de redes, medidas de interação que indiquem a dependência não linear entre objetos que as constituem.

A teoria da informação (SHANNON, 1948; COVER; THOMAS, 2005) é uma área que pode fornecer conceitos e métodos interessantes no estudo de redes. O conceito básico da teoria da informação, e que tem sido aplicado nas mais diversas áreas, é a medida de entropia. Esta é uma medida do grau de incerteza de uma distribuição de probabilidades (COVER; THOMAS, 2005). Trabalhos recentes mostram que métodos baseados em medidas de informação podem ser muito úteis em análise de dados (GULKO, 1999; DARBELLAY; WUERTZ, 2000; DIONISIO; MENEZES; MENDES, 2004; ROSVALL; BERGSTROM, 2007; KRASKOV; GRASSBERGER, 2009; PERON; COSTA; RODRIGUES, 2012; RIBEIRO; PRATAVIERA, 2014; BEKIROS et al., 2016). Em particular, a informação mútua, que é uma medida do grau de dependência entre variáveis, pode ser uma alternativa ao coeficiente de correlação como medida do grau de associação entre objetos, pois mede tanto estruturas lineares quanto não lineares (FRASER; SWINNEY, 1986; JOE, 1997; KRASKOV; STÖGBAUER; GRASSBERGER, 2004; DIONISIO; MENEZES; MENDES, 2004; COVER; THOMAS, 2005; KINNEY; ATWAL, 2014). Estudos nessa direção indicam que a dependência não linear pode ter papel importante na estrutura de certas redes (ROSVALL; BERGSTROM, 2007; YANG et al., 2014). A fusão de métodos de teoria da informação e teoria de redes é um tema interessante a ser explorado, particularmente útil em pesquisas sobre o mercado financeiro (DIONISIO; MENEZES; MENDES, 2004; YANG et al., 2014; FIEDOR, 2014).

Os mercados financeiros têm sido muito estudados do ponto de vista de processos aleatórios (MANTEGNA, 1999). Todavia, a presença de memória nas séries temporais de retornos, a detecção de informação econômica pela matriz de correlação e o comportamento em forma de cascata durante eventos extremos fizeram a discussão tomar rumo diferente, como o proposto

pela teoria de sistemas complexos (MANTEGNA; STANLEY, 1999; BONANNO; LILLO; MANTEGNA, 2001). De fato, “deve haver uma regra mais profunda que controle as propriedades estatísticas do sistema global tanto em dias típicos como em eventos extremos” (BONANNO; LILLO; MANTEGNA, 2001, p. 26). Um dos caminhos foi procurar arranjos topológicos de redes que pudessem estar relacionados aos movimentos dos preços dos ativos. Segundo o estudo de Mantegna (1999), tal topologia resulta do fato de que as séries temporais carregam valiosas informações econômicas. Além das informações contidas nos retornos dos ativos, Brida e Risso (2008) obtiveram evidências de que as ligações entre os ativos estavam longe de ser árvores aleatórias. Esses estudos contribuíram para despertar grande interesse de pesquisas em redes no mercado de ações. Fiedor (2014), utilizando uma metodologia de redes de expansão mínima, comparou as diferenças entre as topologias de rede pela correlação linear de Pearson e pela informação mútua. O resultado foi que a segunda produziu mais agrupamentos que a primeira, modificando a estrutura da rede. You, Fiedor e Hořda (2015), utilizando metodologia semelhante, concluem que a rede formada pela informação mútua é caracterizada por caudas grossas, representando uma distribuição de rede livre de escala, pelo fato de que as variações nos preços são fortemente influenciadas por relativamente pequenos números de ações. Yang et al. (2014) verificaram que a estrutura de rede se alterou significativamente durante a crise de 2008 para o mercado norte americano, sugerindo que, como o desvio padrão aumenta nos períodos de crise, a informação mútua entre ativos deve aumentar, pois efeitos não lineares devem ser importantes nesses períodos. Esta estrutura muda de uma compacta forma de estrela para uma em forma de cadeia longa, evidenciando a topologia como uma característica física de risco. Zhang et al. (2011) concluem que a investigação das propriedades dessas estruturas é importante principalmente em países com economias em crise e de alta volatilidade.

Para o mercado brasileiro, os poucos estudos baseados em redes são restritos a dependências lineares mediante o coeficiente de correlação linear de Pearson (TABAK; SERRA; CAJUEIRO, 2010). Portanto, o estudo da estrutura do mercado de ações usando teorias de redes, juntamente com teoria de informação, neste caso para incluir dependências não lineares, pode revelar padrões importantes e gerar novas perspectivas para caracterizar e entender o comportamento do mercado brasileiro de ações. Além disso, há uma ausência de trabalhos comparando de forma sistemática as redes geradas por correlação de Pearson com as geradas pela Informação Mútua.

1.1 Relevância deste estudo

A importância do estudo de relações não lineares entre objetos, e em particular o mercado financeiro, pode ser apreciada no recente artigo de MARTI et al. (2017). Alguns pontos que consideramos importantes para esta dissertação, incluem:

- a instabilidade das ligações na rede são causadas, em parte, pelo coeficiente de correlação de Pearson, pois este é muito sensível a *outliers*;
- talvez seja possível que resultados mais satisfatórios possam ser obtidos por métodos que são livres de distribuição dado que o coeficiente de correlação de Pearson tem boa confiabilidade apenas quando uma dada distribuição conjunta é Gaussiana. Nós discutiremos mais sobre este aspecto ao decorrer deste trabalho;
- completa-se que um estudo mais adequado seria utilizando-se ligações baseadas em medidas de informação, tal como a informação mútua, métrica que contempla também relações não-lineares;
- simulações empíricas, principalmente nos métodos livres de distribuição (também chamados de não-paramétricos), tendem a ser mais confiáveis quando utilizam-se dados em alta frequência (o número de observações tem grande impacto nestes métodos). Isto é, dados intradiários (de horas ou até mesmo minutos) dos retornos das ações;
- outros desafios concernem na dinâmica da estrutura das redes, da dinâmica dos grupos formados e das aplicações no âmbito das Finanças, tais como otimização de portfólio, medidas comparativas de valor em risco etc. Discutiremos este tópico ao longo deste trabalho.

Retornos não-lineares no mercado de ações são atualmente motivo de frequentes estudos em econofísica (MANTEGNA; STANLEY, 1999; SORNETTE; MALEVERGNE, 2006), de tal forma influenciados pelo artigo dos economistas Scheinkman e LeBaron (1989). Estes autores concluíram que uma parte substancial da variação dos retornos das ações proviam de não-linearidades, em oposição à aleatoriedade previamente defendida por Fama (1970). Concluem os autores que a dependência às condições iniciais que faz a trajetória de alguns sistemas não-lineares determinísticos parecerem aleatórios, também faz com que a compreensão da lei do movimento dessas trajetórias seja extremamente difícil. Essa dependência no movimento tem como consequência o aumento da volatilidade aparente em tais conjuntos não-lineares sujeitos a choques aleatórios (SCHEINKMAN; LEBARON, 1989). Esta conclusão é suficiente para que estudiosos se atenham ao objetivo de se aprofundarem nas pesquisas de tais dependências, pois é bem possível que tais sistemas não-lineares (no caso aqui, os retornos das ações no mercado) apresentem alta volatilidade, tanto para baixo quanto para cima, e isto é de suma importância para investidores interessados nestes ativos.

1.2 Propostas e objetivos

A proposta deste projeto é a aplicação de conceitos e métodos de redes complexas juntamente com técnicas da teoria da informação, em especial a informação mútua, a fim de quantificar as inter-relações não lineares no mercado brasileiro de ações. O objetivo da pesquisa é o estudo da estrutura e métricas da rede financeira quando é utilizada a informação mútua como medida de dependência não linear em comparação com a estrutura gerada a partir do coeficiente de correlação linear de Pearson, a partir das séries de retornos das ações no mercado de ações brasileiro. Espera-se que a estrutura e métricas da rede a partir de medidas de informação, em especial a informação mútua, possam ser úteis e mais confiáveis como medida de avaliação empírica desses ativos do mercado brasileiro de ações do que aquelas baseadas em redes geradas a partir das correlações lineares.

Especificamente, nesta comparação de redes, serão investigadas:

1. as mudanças ocorridas na estrutura global destas redes, como por exemplo sua distribuição de grau, sua estrutura de grupos e de suas ligações internas, e na estrutura local, como as medidas de centralidade de cada ação em redes obtidas a partir do coeficiente de correlação de Pearson e da Informação mútua (IM).
2. as mudanças em relação a diferentes períodos. Neste trabalho, verificaremos como são as mudanças nas estruturas de rede via correlação de Pearson e Informação Mútua no período político de transição da gestão do segundo mandato da presidente Dilma Rousseff para a do presidente Michel Temer.
3. se as medidas de rede baseadas em informação mútua fornecem uma indicação de fatores de risco mais realistas em relação às obtidas com a correlação de Pearson.

Em particular, o conjunto de dados para a pesquisa será as séries temporais de alta frequência dos retornos das ações do índice Bovespa (Ibovespa), que compõe uma carteira teórica das principais ações brasileiras. Séries de alta frequência (séries de intervalos menores que um dia) dos preços das ações do mercado brasileiro estão disponíveis no site da Bovespa, e podem ser manipuladas por um pacote do R chamado 'GetHFDData'. Nós utilizaremos, neste trabalho, dados de alta frequência de intervalo de 15 minutos. Se fossem utilizadas janelas menores, incorreríamos em muitos dados faltantes. Ainda assim, conseguimos um tamanho amostral satisfatório da ordem de 4000 retornos.

A tabela 1.1 sumariza as propostas desta pesquisa, bem como as análises subsequentes.

1.3 Organização do trabalho

O trabalho está dividido da seguinte maneira: o capítulo 2 apresenta os principais conceitos sobre a teoria das redes complexas. O capítulo 3, introduz o conceito de dependência entre variáveis aleatórias, bem como várias medidas para sua mensuração. O capítulo 4 discorre sobre o conceito de entropia e informação mútua, bem como suas estimativas. O capítulo 5 apresenta uma revisão dos principais trabalhos e métodos desenvolvidos no estudo de redes em mercados de ações e que motivaram o desenvolvimento deste trabalho. O capítulo 6 discute o método adotado pra realização da pesquisa. Já o capítulo 7 descreve e analisa os resultados obtidos nesta pesquisa. Finalmente, o capítulo 8 traz as considerações finais, bem como sugestões para possíveis trabalhos futuros. Por fim, o Apêndice A apresenta os códigos dos algoritmos utilizados e demais derivações importantes.

Tabela 1.1 – Objetivos propostos para a pesquisa

PROPOSTAS	MEDIDAS DA REDE (ANÁLISE DE RESULTADOS)
COMPARAÇÃO DE REDES Verificar diferenças na rede via informação mútua e pela correlação de Pearson.	ESTRUTURA GLOBAL DA REDE Analisar a distribuição de grau das redes de ações, verificando sua estrutura a partir da árvore geradora mínima.
	ESTRUTURA DE GRUPOS (MESO) Verificar mudanças nos principais agrupamentos econômicos que surgem.
	ESTRUTURA LOCAL Evidenciar mudanças nas medidas de centralidade.
COMPARAÇÃO DE PERÍODOS A informação mútua capta dependências que são particularmente úteis em períodos diferentes?	DIFERENÇAS DE INFORMAÇÃO Medir a informação mútua e compará-la ao coeficiente de Pearson em diferentes períodos, particularmente de interesse aos investidores.
AValiação Avaliar quais medidas da rede podem ser úteis quando tratamos de risco geral e específico no mercado de ações.	MUDANÇAS ESTRUTURAIS As medidas de centralidade e de caminhos da rede podem ser úteis na estimativa do risco empírico das ações?
RETORNO E INFORMAÇÃO	'HUBS' DE INFORMAÇÃO Avaliar como a informação mútua pode ajudar a encontrar ações que sofrem (sofreram) alta volatilidade, como em processos de arbitragem e especulação, e que dessa forma, sugerem um apoio aos estudos de Scheinkman e LeBaron (1989).

Fonte: Elaboração Própria.

CAPÍTULO 2

REDES COMPLEXAS

Consideremos dois sistemas, inicialmente distintos, as células de câncer e as redes de eletricidade. Na evolução do câncer, por exemplo, pode-se chegar ao ponto de todas nossas células serem infectadas num típico movimento de falha em cascata. Em outros casos, o câncer pode ser rapidamente controlado. Caso semelhante ocorre nas redes elétricas. Qual deve ser o limiar para que esta entre em colapso, deixando milhares ou mesmo milhões no escuro? O fator predominante em ambos os casos parece ser a forma com que esses sistemas exibem interconexões. Barabasi (2016) chama esse fenômeno de vulnerabilidade por interconectividade. Apesar de tais sistemas, a princípio parecerem totalmente randômicos e imprevisíveis, eles geralmente seguem leis que podem ser quantificadas e até preditas usando as ferramentas da teoria de redes complexas (NEWMAN, 2010; BARABASI, 2016).

Uma importante descoberta da ciência de redes complexas é que as arquiteturas de redes em vários domínios do conhecimento são similares uma com as outras, uma consequência de serem governadas pelos mesmos princípios e leis. Consequentemente, podemos então usar um ferramental matemático comum que explore tais sistemas (BARABASI; ALBERT, 1999). Esse ferramental, baseado em estudos empíricos, modelos matemáticos e estatísticos e algoritmos computacionais, foi importante para o desenvolvimento da ciência de redes, pois somente dessa forma pode-se tratar estruturas de redes reais. As figuras 2.1 e 2.2 são exemplos do tamanho e da complexidade que exibem algumas dessas estruturas.

Nesta seção apresentaremos os conceitos básicos para o entendimento de redes complexas.

Figura 2.1 – Rede de doenças humanas: interligadas se compartilham o mesmo gene. Nota: versão colorida na opção eletrônica.

Fonte: Barabasi (2016)

Figura 2.2 – Rede da internet: cada cor remete a um domínio diferente. Neste caso os nós da rede são roteadores e as ligações são conexões de internet. Nota: versão colorida na opção eletrônica.

Fonte: Newman (2003)

2.1 Conceitos básicos de redes

Os moradores da cidade de Königsberg, na antiga Prússia, tinham um peculiar problema matemático no século XVIII. Dadas as sete pontes da cidade, cinco nas quais a interligavam com a ilha de Kneiphof, e duas que cruzavam braços de rios, os curiosos se perguntavam se era possível cruzar todas as sete pontes sem que se cruzasse uma mesma mais de uma vez, de acordo com o esquema da ilustração 2.3. Esse problema ficou sem solução até 1735, quando Leonard Euler ofereceu uma prova que tal caminho único não existia. Euler representou as pontes de acordo com o esquema da figura 2.4. A figura representando nós onde ocorrem ligações é chamado de grafo. Observando o agora chamado grafo da figura 2.4, ele observou que se houvesse um caminho único entre as pontes, então os nós com número de ligações ímpar deveriam ser o caminho de chegada e também o de saída. Dessa maneira, não poderia existir este caminho se existissem mais do que dois nós com um número ímpar de ligações. Pela figura 2.4, pode-se observar que a construção de uma ponte entre os nós A e C faria com que ficássemos com dois nós com ligações ímpares. Assim, podemos encontrar o caminho desejado.

A prova de Euler é importante porque nos mostra que é mais simples e tratável representar um problema por meio de um grafo. Se quisermos entender as várias maneiras que as redes afetam as propriedades de um sistema, precisamos entender sobre alguns conceitos básicos em teoria de redes (NEWMAN, 2010; BARABASI, 2016).

Figura 2.3 – Ilustração das pontes da cidade de Königsberg.

Fonte: Barabasi (2016)

2.1.1 Nós e ligações

Uma rede é um conjunto de elementos discretos (vértices, nós) e um conjunto de conexões (arestas, ligações) que ligam os elementos aos pares (NEWMAN, 2010; BARABASI, 2016). Uma rede é definida pelo número N de nós e o número de vértices L . Por exemplo, na figura 2.4, temos o exemplo de um grafo com $N = 4$ e $L = 7$. Além disso, essas ligações podem ser unidirecionais ou bidirecionais e apresentarem ou não pesos diferentes (grafos ponderados).

Figura 2.4 – Ilustração pensada por Euler para o problema das sete pontes de Königsberg.
Fonte: Elaboração própria.

2.1.2 O grau de um nó

Um conceito importante em teoria de grafos é o grau de um nó, o número de ligações daquele nó, denotado por k . Por exemplo, na figura 2.4, temos $k_A = 3$, $k_B = 5$, $k_C = 3$ e $k_D = 3$. Podemos expressar L , o total de ligações de uma rede em termos de k_i . Para redes unidirecionais

$$L = \frac{1}{2} \sum_{i=1}^n k_i. \quad (2.1)$$

O fator $\frac{1}{2}$ é introduzido em (2.1) para não contar-se as ligações duas vezes.

Já o grau médio de um grafo unidirecional é dado por

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2L}{N}. \quad (2.2)$$

Para grafos bidirecionais, temos k ligações que chegam, e k ligações que saem do nó. Pode-se verificar que a equação 2.2 torna-se $\frac{L}{N}$, já que $L = \sum k_{chegam} = \sum k_{saem}$ (BARABASI, 2016).

2.1.3 Matriz de adjacência

Uma maneira de representar grafos é por meio de uma matriz de adjacência, $\mathbf{A}$, com elementos $A_{ij} = 1$, caso haja ligação entre i e j , ou 0, caso contrário. A matriz de adjacência pode ter valores diferentes de um, por exemplo, em grafos ponderados, assim $A_{ij} = w_{ij}$, onde w_{ij} é o peso da ligação. Podemos calcular o grau k_i de um nó i através da matriz de adjacência. Para um grafo unidirecional, k_i é dado pela soma das colunas ou das linhas da matriz $\mathbf{A}$. Assim, podemos expressar

$$k_i = \sum_{j=1}^N A_{ij} \quad (2.3)$$

Um exemplo para grafo unidirecional e bidirecional é dado pela figura 2.5

Figura 2.5 – Exemplo de representação de grafos pela matriz de adjacência: notar que no cálculo de L para grafos unidirecionais, deve-se dividir a soma para cada i de (2.3) por 2. Porém, o mesmo não se aplica para o caso bidirecional, onde L é diretamente dado pela soma para cada i da equação (2.3).

Fonte: Adaptado de Barrat, Barthelemy e Vespignani (2008).

A partir de (2.1) ou (2.2), encontrar L é de certa forma interessante, pois podemos compará-la com L_{max} , ou seja o número total de ligações presentes num grafo completo (unidirecional e bidirecional), onde cada nó é conectado por cada outro nó. Assim

$$L_{max} = \frac{N(N-1)}{2}. \quad (2.4)$$

Podemos chamar de um grafo esparso aquele em que $L \ll L_{max}$. Por exemplo, a rede da figura 2.2 tem 192.244 nós e 1,5 milhão de ligações (NEWMAN, 2010). Pela equação (2.4), podemos verificar que $L_{max} \sim 1,8 \times 10^{10}$ ligações. Essa conclusão é válida para diversos tipos de redes reais (BARABASI, 2016) e pode nos indicar a fragmentação de redes e a importância relativa de certos nós em relação aos demais.

2.1.4 Distribuição de grau

A distribuição de grau, p_k , denota a probabilidade que um nó aleatoriamente selecionado tenha grau k . Para um grafo com N nós, a distribuição segue

$$p_k = \frac{N_k}{N}, \quad (2.5)$$

onde N_k é o número de nós com grau k . A distribuição de grau tem papel muito importante na análise de redes. Segundo Barabasi (2016), a forma funcional de p_k determina muitos fenômenos que ocorrem em redes, por exemplo, na composição de grupos e na robustez de um sistema. Voltaremos a falar sobre a distribuição de grau quando mostrarmos as redes randômicas.

2.1.5 Distâncias e caminhos

Em grafos, a distância entre dois nós é dada pela chamado 'comprimento de caminho'. Um caminho é uma rota que liga os elementos de um grafo. Podemos definir o caminho mais curto em um grafo como o de menor número de ligações entre dois nós quaisquer (NEWMAN, 2010). O caminho mais curto é também chamado de caminho geodésico. De modo semelhante, podemos definir o caminho mais longo como o diâmetro do grafo. Se este caminho conter o mesmo começo e fim é chamado de ciclo. Além dessas medidas, outra de grande interesse é o comprimento de caminho médio, $\langle d \rangle$. Ele pode ser calculado como a média de todos os menores caminhos dentre todos os pares de nós. O número de menores caminhos, N_{ij} , pode ser calculado diretamente da matriz de adjacência, A_{ij} . Dado que $d_{ij} = d$ se há um caminho de tamanho d entre i e j , então $A_{ik} \dots A_{lj} = 1$. Assim, o número de caminhos de tamanho d entre i e j é (BARABASI, 2016)

$$N_{ij}^d = A_{ij}^d. \quad (2.6)$$

Um grafo é dito conectado quando todos seus pares de nós estão conectados. Se, em pelo menos um nó, $d_{ij} = \infty$, o componente é chamado de desconectado. Sua importância está na descoberta de nós do tipo 'pontes', aqueles que, se retirados do grafo, desconectam seus componentes. Para grandes grafos, o cálculo de N_{ij} e do número de componentes conectados

é difícil, e algoritmos eficazes, como o *breadth-first-search*, foram desenvolvidos para este fim (NEWMAN, 2010).

2.1.6 Árvores

Uma árvore é um grafo conectado, unidirecional, que não contém ciclos fechados. Segundo Newman (2010), uma das propriedades mais importantes de árvores é que existe exatamente um caminho entre qualquer par de nós, pois não há ciclos fechados. Dessa forma, o cálculo de várias propriedades do grafo torna-se mais simples, pois o número de ligações é exatamente $n - 1$, caso contrário violaria a propriedade de árvores não conterem ciclos fechados.

Podemos também chamar essas árvores de árvores de expansão. Um interessante modelo de árvore de expansão é a chamada Árvore Geradora Mínima, do inglês *minimum-spanning tree*. Ela é definida como a árvore de expansão quando a soma de distância das ligações é mínima, ou seja, minimizando a função peso (CORMEN et al., 2001)

$$w(t) = \sum_{ij \in T} w(i, j). \quad (2.7)$$

Existem dois principais algoritmos de busca de árvore geradora mínima, o algoritmo de Kruskal (KRUSKAL, 1956) e o de Prim. Por exemplo, o algoritmo de Kruskal adiciona novas ligações de menor peso até que não hajam mais opções de novas adições sem que um ciclo seja formado, e que $k_i \geq 1$. Assim, temos $L = n - 1$ ligações. Neste processo podem existir mais de uma árvore mínima por grafo, porém todas com o mesmo peso (CORMEN et al., 2001). A seguir apresentamos um pseudo-algoritmo para encontrar uma árvore geradora mínima, seguindo os passos de Kruskal (1956).

Algorithm 1 Algoritmo de Kruskal simplificado

```

1: procedure KRUSKAL
2:  $A \leftarrow \emptyset$ 
3: para cada vértice  $v \in V$ 
4:   faça um conjunto ( $v$ ) ▷ criam-se  $V$  árvores, cada uma contendo um vértice  $v$ .
5:   filtre as ligações de  $V$  em ordem crescente por peso  $w$ 
6: para cada ligação  $(u, v) \in V$ 
7:   faça se conjunto ( $u$ )  $\neq$  conjunto ( $v$ ) ▷ se os pontos finais de  $u$  e  $v$  pertencerem à mesma
   árvore  $V$ , estes não podem ser ligados, pois se caso fossem, criariam um ciclo. Assim, esta
   ligação é descartada.
8:   então  $A \leftarrow A \cup (u, v)$  ▷ Caso respeitem condição acima, os vértices são adicionados à
   A.
9:   una ( $u, v$ ) ▷ ligam-se os vértices.
10: retorne  $A$ 

```

2.1.7 Transitividade

A transitividade de um nó pode ser calculada a partir do coeficiente de agrupamentos. Essa medida captura a probabilidade de que um dado nó i se ligue com outro nó j , e é definida como (BARABASI, 2016)

$$C_i = \frac{2L_i}{k_i(k_i - 1)}, \quad (2.8)$$

onde L_i representa o número de ligações entre k_i vizinhos do nó i . Dessa forma, C_i varia entre 0 até 1 e representa a probabilidade de que dois vizinhos tenham uma ligação entre si. Em outras palavras, C_i mede a densidade local do grafo: mais perto de 1, maior sua densidade local. Somando cada C_i e dividindo por N , temos o coeficiente de agrupamento médio, $\langle C \rangle$, do grafo.

2.1.8 Outras medidas de nós

Até agora, discutimos que o grau de conectividade do nó k era definido assumindo-se o número de ligações n que este tinha. Algumas vezes essa definição pode não ser suficiente para classificarmos a importância do nó dentro da estrutura do grupo. Precisamos então da definição de centralidade (C) de um nó. Essa definição depende da medida a ser utilizada. Por exemplo, uma medida bastante utilizada é a centralidade de intermediação, dada por (NEWMAN, 2010)

$$x(i) = \sum_{j < k} \frac{d_{jk}(i)}{d_{jk}}, \quad (2.9)$$

onde d_{jk} é o número de caminhos mais curtos entre j e k e $d_{jk}(i)$ é o número de caminhos mais curtos entre j e k que passam por i . Dessa forma, para a medida (2.9), centralidade é estar sempre nos caminhos mais curtos entre dois pares de nós i e j .

Uma medida de centralidade bastante similar é a chamada centralidade de proximidade (NEWMAN, 2010). Intuitivamente, é dada pelo inverso da distância média entre todos os outros nós. Matematicamente,

$$x(i) = \left(\sum_{j=1}^N d(i, j) \right)^{-1}. \quad (2.10)$$

Neste caso, centralidade é estar o mais próximo entre todos os demais nós.

Outra medida bastante interessante é a chamada centralidade baseada em autovetores. Para esta medida, um nó é importante se ele está conectado com outros nós que também são

importantes. Pode ser definida somando-se a centralidade de todos os vizinhos do nó i (NEWMAN, 2010)

$$x'_i = \sum_j A_{ij}x_j, \quad (2.11)$$

onde A_{ij} é um elemento da matriz de adjacência. Note que pode-se escrever essa expressão em forma de notação matricial $\mathbf{x}' = \mathbf{A}\mathbf{x}$, onde $\mathbf{x}$ é um vetor de elementos x_i . Tomando-se o conceito do autovalor da matriz $\mathbf{A}$, podemos então reescrever a equação (2.11) como

$$x'_i = \kappa_1^{-1} \sum_j A_{ij}x_j, \quad (2.12)$$

onde κ_1 é o maior autovalor de $\mathbf{A}$.

Dessa forma, um vértice (nó) pode ter alto valor de centralidade baseado em autovetor se este tiver muitos vizinhos, ou se estes vizinhos (mesmo que poucos) forem muito importantes.

Outras medidas de centralidade são derivadas da equação (2.12). Por exemplo, podemos utilizar um termo constante multiplicando $\mathbf{A}$, tornando agora a centralidade de Katz (NEWMAN, 2010). Quando lidando com grafos bidirecionais, uma usual medida é a *PageRank*, que utiliza medidas de grau de entrada e saída, k_{entra} e k_{sai} . Essas medidas costumam caracterizar bastante adequadamente redes de sites de internet, pois ajustam a centralidade dividindo-a por k_{sai} , uma forma de fazer com que páginas de busca não tenham uma centralidade tão desproporcional.

Um exemplo interessante que mostra as diferenças entre as centralidades é ilustrado na figura 2.6.

Em suma, a definição de centralidade varia de acordo com o contexto e proposta. Podemos definir o grau k_i como sendo uma medida estritamente local da centralidade de um nó i . De mesmo modo, podemos defini-la relativa ao resto da grafo, ou seja, utilizando medidas dos vizinhos (centralidade baseada em autovetores) ou mesmo mensurando os caminhos de todo o grafo (centralidade de intermediação).

As medidas que mostramos nessa seção podem nos ajudar a diagnosticar várias propriedades centrais no estudo de redes reais. Enquanto a teoria de grafos, um ramo da matemática, estuda o comportamento de pequenos grafos, a ciência de redes tem como objetivo caracterizar, em geral gigantescas redes reais, de modo bastante empírico. Nas próximas seções serão apresentados alguns modelos de redes que permitem entender sua estrutura e dinâmica.

Figura 2.6 – Tipos de centralidade: Quem é mais central? Se considerarmos a medida tradicional de grau k , temos que Dan é o mais importante, pois tem $k = 4$. Agora, se considerarmos a medida em (2.9), Gus se encontra em vários caminhos mais curtos entre vários pares de nós. Dessa maneira, agora Gus é o mais importante. Se considerarmos a equação (2.10), agora Fay tem a menor distância média entre todos os demais nós, portanto agora é o mais central, pois funciona como uma ponte de ligação entre todos os demais. Finalmente, se consideramos a mensuração em (2.12), Edy é o mais central, pois se conecta ao nó de maior grau k , Dan.

Fonte: Elaboração própria com nomes fictícios.

2.2 Redes randômicas

A partir dessa seção, serão discutidos alguns modelos de redes randômicas que explicam alguns padrões de estrutura de redes. Erdős e Rényi (1960) definem uma rede randômica onde N nós são conectados com L ligações atribuídas aleatoriamente. Para gerar uma rede aleatória, podemos seguir os seguintes passos:

- selecionar um par de nós e gerar um número aleatório entre 0 – 1. Se esse número exceder certo p , conectar os nós, caso contrário deixá-los desconectados.
- repetir o passo anterior para cada $\frac{N(N-1)}{2}$ pares de nós.

A rede assim gerada é chamada de rede de Erdős-Rényi, porém Rapoport e Horvath (1961) também merecem crédito pelo estudo inicial de tais redes.

2.2.1 Ligações em redes randômicas

A probabilidade de uma rede randômica ter exatamente L ligações é o produto de três termos: a probabilidade p que dado par de nó se conecte, $1 - p$, a probabilidade que dado par

de nó não se conecte, e um fator combinacional que conte o número de diferentes maneiras que podemos colocar L ligações entre $\frac{N(N-1)}{2}$ pares de nós. Assim, temos que

$$P_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}. \quad (2.13)$$

Dado que a equação (2.13) segue uma distribuição binomial, o número de ligações esperadas em uma rede randômica é (BARABASI, 2016)

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L P_L = p \frac{N(N-1)}{2}. \quad (2.14)$$

A equação (2.14) mostra que $\langle L \rangle$ é o produto da probabilidade p de ligação entre dois nós quaisquer dentre todos os pares de nós que desejamos ligar, que é dado por (2.4).

Usando $\langle L \rangle$ em (2.2), pode-se definir o grau médio de uma rede randômica, $\langle k \rangle$, dado por

$$\langle k \rangle = p \frac{2\langle L \rangle}{N} = p(N-1), \quad (2.15)$$

onde p é a probabilidade que dois nós se liguem e $(N-1)$, o número máximo de ligações que um nó pode ter numa rede unidirecional de tamanho N .

Em suma, se aumentarmos p , a rede fica mais densa: o número médio de ligações aumenta linearmente de $\langle L \rangle = 0$ para L_{max} , enquanto o grau médio de um nó aumenta linearmente de $\langle k \rangle = 0$ para $\langle k \rangle = N-1$.

2.2.2 Regimes em redes randômicas

Podemos destacar quatro regimes diferentes para as redes randômicas. Para $\langle k \rangle < 1$, o tamanho do maior grupo é dado por $N_g \sim \ln N$. Quando $\langle k \rangle = 1$, é chamado de ponto crítico, e o tamanho do maior grupo cresce para $N_g \sim N^{\frac{2}{3}}$. Para $\langle k \rangle > 1$, há a formação de um componente gigante, ou seja nós que são todos conectados entre si, e neste caso, $N_g \sim (p-p_c)N$, onde p_c é igual à configuração randômica $\frac{1}{N}$. Caso $\langle k \rangle \sim \ln N$, o componente gigante se torna único e $N_g \sim N$. Um exemplo de componente gigante é ilustrado pela figura 2.7.

O modelo descrito é importante, pois podemos comparar o grau médio (2.2) de uma rede real com (2.15), verificando se a rede real tem ou não características da rede de Erdős-Rényi, e além disso, mostrar em qual regime se encontra.

Figura 2.7 – Ilustração de um componente gigante: quando o grau médio de ligação $\langle k \rangle \rightarrow \ln N$, encontramos um componente gigante, um clique.

Fonte: Barabasi (2016).

2.2.3 Distribuição de grau para redes randômicas

Vimos que a equação (2.13) nos dá a forma da distribuição de grau de k . No limite em que $N \rightarrow \infty$ e $p \rightarrow 0$, podemos aproximar a forma da distribuição binomial pela distribuição de Poisson pela equação (2.16). As figuras 2.8 e 2.9 mostram respectivamente, um exemplo para a distribuição de Poisson e três distribuições de grau de redes reais, comparando a distribuição empírica com a dada pela distribuição de Poisson. A distribuição de Poisson para o grau k pode ser dada por

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}. \quad (2.16)$$

Figura 2.8 – Exemplo de distribuição de Poisson: notar que a aproximação da binomial pela Poisson se torna indistinguível para redes grandes.

Fonte: Barabasi (2016).

Figura 2.9 – Pode-se comparar a distribuição de Poisson (pontilhado), com dados de distribuições empíricas, da esquerda para a direita, da internet, da rede de colaboração na ciência e na interação entre proteínas. O modelo de rede randômica subestima a ocorrência de nós com k muito pequeno ou então k muito grande.

Fonte: Barabasi (2016).

2.2.4 Fenômeno de mundos pequenos em redes randômicas

Em linguagem de redes, o fenômeno de mundos pequenos implica que a distância entre dois nós é sempre pequena. Esse fenômeno é tipicamente definido pela equação (2.17). Nota-se que 'pequena' significa que o tamanho médio do caminho, $\langle d \rangle$, depende logaritmicamente do tamanho da rede N , e não linear a N ou alguma potência de N (BARABASI, 2016). Esse fato

foi bastante estudado por Travers e Milgram (1969), no aclamado *Seis Graus de Separação*, e posteriormente por Watts e Strogatz (1998), no modelo Watts-Strogatz. Por exemplo, para redes sociais, os valores para (2.17) variam entre 3 e 6 (WATTS; STROGATZ, 1998). Assim, estamos de três até seis apertos de mão de qualquer pessoa na Terra.

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}. \quad (2.17)$$

2.2.5 Coeficiente de agrupamento para redes randômicas

Vimos que um parâmetro importante para caracterizar uma rede é o coeficiente de agrupamento, definido pela eq. (2.8). Para as redes randômicas, a equação (2.8) torna-se p , pois é a probabilidade de que dois nós quaisquer se liguem. Dessa forma, o coeficiente de agrupamento é dado por

$$C_i = p = \frac{\langle d \rangle}{N}. \quad (2.18)$$

Pesquisas posteriores (WATTS; STROGATZ, 1998; NEWMAN, 2010) verificaram que em redes reais o coeficiente de agrupamento era relativamente alto, independente de N e variava com k , contradizendo a equação (2.18). Neste caso, o modelo Watts e Strogatz (1998) mostra redes de mundos pequenos com alto coeficiente de agrupamento.

2.2.6 Redes reais não são randômicas

Para Barabasi (2016, p. 98), redes não poderiam ser randômicas, pois deveria haver uma razão mais profunda que pudesse descrever a arquitetura do sistema, produzindo os desvios encontrados quando comparados com a configuração de rede randômica. Em resumo, se o modelo randômico estiver presente no sistema, este terá propriedades aleatórias, porém caso não esteja, este certamente vai requerer uma explicação mais profunda. Apesar de a teoria de redes randômicas ser um importante passo inicial na teoria de redes, ainda havia muito o que se investigar.

2.3 Redes livre de escala

Na seção anterior, vimos que no modelo randômico não há chances de termos nós com k muito pequeno ao mesmo tempo com outros com k muito alto. A figura 2.9 mostra alguns exemplos dessa anomalia em redes reais. De fato, as redes reais exibem propriedades que fazem

com que a distribuição de grau p_k não seja bem explicada por uma distribuição de Poisson. Redes com variação de k muito além de $\langle k \rangle$ são representadas pela chamada rede livre de escala.

2.3.1 Lei de potência

As distribuições de grau da figura 2.9 sugerem um comportamento linear quando apresentadas em um gráfico na escala log-log. Assim, estas distribuições podem ser aproximadas por uma lei de potência (BARABASI, 2016)

$$p_k \sim k^{-\gamma}, \quad (2.19)$$

ou numa escala log-log

$$\ln p_k \sim -\gamma \ln k. \quad (2.20)$$

A equação (2.20) nos mostra que $\ln p_k$ depende linearmente de $\ln k$, com coeficiente angular γ , que corresponde ao expoente da lei de potência. Dessa maneira, para redes nas quais sua distribuição segue uma lei de potência, esta é chamada livre de escala (NEWMAN, 2010; BARABASI, 2016). A figura 2.10 apresenta um exemplo comparando as redes estudadas até então. Para compararmos, a 2.9 mostrou três exemplos de distribuições de redes reais que seguem uma lei de potência (nota-se que a função de Poisson subestima a ocorrência de p_k quando $k \gg \langle k \rangle$).

Figura 2.10 – O exemplo compara redes randômicas, que seguem uma distribuição exponencial, à esquerda, com as livre de escala, que seguem uma lei de potência, à direita. Notar que a rede livre de escala possui uma linha reta no gráfico log-log.

Fonte: Barabasi e Albert (1999).

2.3.2 *Nodos centrais (Hubs)*

Vimos na subseção anterior que a diferença entre a rede randômica e a livre de escala está justamente nas caudas da distribuição de p_k . Reanalizando a figura 2.9, podemos verificar três condições:

- Para k pequeno, a lei de potência produz mais nós do que pela função Poisson nessa região.
- Para k na vizinhança de $\langle k \rangle$, a função Poisson superestima a quantidade de nós nessa região.
- Para $k \gg \langle k \rangle$, a função Poisson nos mostra muito menos nós do que a lei de potência nessa região. Esse último caso é de suma importância para as redes livre de escala. Nós em que $k \gg \langle k \rangle$ são chamados de nodos centrais (tradução do inglês *hubs*). Por questões de comparabilidade, denotaremos estes nodos centrais por hubs ao longo do texto. Pode-se definir um hub como um ponto central, foco principal de atividade de uma rede. Este tipo de nó costuma receber boa parte das ligações de uma rede.

Para uma rede livre de escala, o maior hub esperado é dado por (BARABASI, 2016)

$$k_{max} = k_{min} N^{\frac{1}{\gamma-1}}. \quad (2.21)$$

Podemos verificar pela equação (2.21) que k_{max} depende de uma potência de N . Esta dependência de k_{max} em N implica ordens de magnitude de diferença entre k_{min} e k_{max} . Esse é um dos sintomas da falta de escala que exhibe esse tipo de rede.

A figura 2.11 mostra um exemplo, comparando redes randômicas e redes livres de escala. A presença de hubs torna a rede bastante diferente da randômica, diminuindo ainda mais a média de caminhos mais curtos.

2.3.3 *Significado de livre de escala*

A distribuição de probabilidades de uma variável aleatória X permite obter os chamados momentos da distribuição. Os momentos de uma distribuição são definidos pelo valor esperado de X^n . O n -ésimo momento de uma variável aleatória contínua X , cuja função densidade de probabilidade é dada por $f_X(x)$, é definido por (MEYER, 1969)

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx, \quad (2.22)$$

onde $n = 1$ define o primeiro momento, correspondente à média de X . Para $n = 2$, obtemos o segundo momento, que tem relação com o desvio padrão de X . Já quando $n = 3$, temos o terceiro momento, relacionado à assimetria da distribuição de X , enquanto que o quarto momento é relacionado à curtose da distribuição de X . Assim, a partir da distribuição de probabilidade p_K do grau k dos nós de uma rede, podemos obter o grau médio $\langle k \rangle$, e momentos de ordem mais altas, $\langle k^2 \rangle$, $\langle k^3 \rangle$ etc.

Para a rede livre de escala, se $n - \gamma + 1 \leq 0$, então todos os momentos que satisfazem $n \leq \gamma - 1$ são finitos. Caso contrário, então $\langle k^n \rangle \rightarrow \infty$. Nesse caso, todos os momentos que excedem $\gamma - 1$ divergem.

Figura 2.11 – A comparação da rede randômica, que segue uma distribuição de Poisson, com a livre de escala, que segue uma lei de potência, nos mostra que a presença de hubs é uma forma natural da apresentação das redes livre de escala

Fonte: Barabasi (2016).

Pesquisas (BARRAT; BARTHELEMY; VESPIGNANI, 2008; NEWMAN, 2010; BARABASI, 2016) apontam que, na maioria das redes reais estudadas, γ está entre 2 e 3. Dessa forma, o segundo e demais momentos superiores da distribuição divergem. Quando γ é menor que 2, a distribuição não possui escala alguma, pois todos os momentos divergem. Esse fenômeno da falta de escala não é de estranhar, visto que as redes livre de escala possuem nós com $k \gg \langle k \rangle$. Neste caso, $\langle k \rangle$ não serve como escala para k . Daí o nome livre de escala.

Muitas redes reais são livre de escala. Newman (2010) e Barabasi (2016) sugerem que a rede de internet, a de colaboração na ciência, a de atores e a de interação de proteínas sejam livres de escala. Já a rede elétrica se adapta melhor a um modelo aproximadamente randômico.

2.3.4 *Ultra mundos pequenos*

A figura 2.11 nos mostra que, ao menos visualmente, as distâncias em redes livres de escala são bastante reduzidas se comparadas ao modelo randômico. A dependência da distância média $\langle d \rangle$ no tamanho da rede N e do coeficiente γ é dada por (BARABASI, 2016)

$$\langle d \rangle \sim \begin{cases} \text{const.}, & \gamma = 2 \\ \ln \ln N, & 2 < \gamma < 3 \\ \frac{\ln N}{\ln \ln N}, & \gamma = 3 \\ \ln N, & \gamma > 3 \end{cases} \quad (2.23)$$

Neste caso, temos quatro regimes diferentes (BARABASI, 2016):

- regime anômalo: Quando $\gamma = 2$, de acordo com a eq. (2.21), o maior hub cresce linearmente com N . Isso força todos os nós a se ligarem a um só *hub* central.
- ultra-mundo-pequeno: Quando γ está entre 2 e 3, (2.23) cresce com $\ln \ln N$, uma velocidade bem menor que a esperada para redes randômicas, $\ln N$. Dessa maneira, os hubs da rede fazem com que a distância entre caminhos seja bem menor do que a encontrada em redes de mundos pequenos do modelo randômico.
- ponto crítico: No ponto em que $\gamma = 3$, podemos verificar que o momento de segunda ordem da distribuição já não diverge mais. Esse ponto é chamado de ponto crítico, um ponto entre a configuração randômica, quando $\gamma > 3$, e da configuração livre de escala, quando $2 < \gamma < 3$.
- mundo-pequeno: Aqui, voltamos aos mundos-pequenos da configuração de redes randômicas. Neste caso, $d = \ln N$. Os hubs encontrados em redes quando $\gamma > 3$ não são suficientemente grandes para terem impacto na arquitetura da rede. Podemos verificar tal situação examinando a equação (2.21).

2.3.5 *Redes reais são livre de escala*

Muitas das redes de exemplos anteriores seguiam uma forma livre de escala. Podemos nos questionar a origem dessa configuração. Dois principais modelos emergem nessa questão: o modelo Barabási-Albert (BARABASI; ALBERT, 1999) e o modelo Bianconi-Barabási (BIANCONI; BARABASI, 2001). Discutiremos rapidamente esses modelos nas próximas subseções.

2.3.6 Modelo Barabási-Albert

A grande maioria das redes não são estáticas, podem crescer ou decrescer ao longo do tempo. O modelo Barabási Albert (BA) (BARABASI; ALBERT, 1999) é baseado na hipótese de que a rede cresce pela adição de novos nós, e suas conexões dependem da regra de ligação em que se baseavam os nós mais antigos. O modelo prevê que as redes randômicas se diferenciam da livre de escala por duas importantes características:

- crescimento de nós em um processo contínuo no tempo: a cada tempo t , um novo nó é adicionado à rede, distribuindo suas m novas ligações.
- ligação preferencial: enquanto que nas redes aleatórias temos a probabilidade p aleatória de ligação entre dois nós, agora se constata que os nós preferem se conectar com outros que tenham mais ligações. A ligação preferencial ficou conhecida como a relação em que 'o rico sempre fica cada vez mais rico'.

Podemos dizer que a probabilidade de um nó se conectar a outro é proporcional à k , dado que

$$pk_i = \frac{k_i}{\sum_j k_j}. \quad (2.24)$$

Desse modo, por exemplo, verificamos que a probabilidade de um nó ligar-se a um nó com $k = 2$ é a metade de se ligar a outro com $k = 4$. Esse fenômeno concede a nós de maior ligação, os hubs, a prioridade no recebimento de novas ligações. Esse efeito também é conhecido como a vantagem do primeiro jogador, pois quanto mais cedo um dado nó entra na rede, maior é seu grau k num dado momento t .

Para calcular a distribuição de grau do modelo BA, pode-se calcular o número de nós com grau menor que k , isto é $k(t) < k$. Utilizando equações de taxa, Barabasi e Albert (1999) chegaram à sua exata distribuição, dada por

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}. \quad (2.25)$$

A dedução da equação (2.25), pode ser vista em (BARABASI, 2016, p. 194-196). Para o caso em que $k, m \rightarrow \infty$, (2.25) torna-se

$$p_k \sim 2m^2 k^{-3}. \quad (2.26)$$

Em resumo, os cálculos predizem que o modelo BA gera uma rede livre de escala com $\gamma = 3$. Além disso, não há dependência em relação ao tempo t , ou seja, a distribuição é

estacionária, o que em parte explica porque redes de diferentes tamanhos, histórias e idades se desenvolvem com uma topologia muito semelhante entre si.

Pela equação (2.24), se o expoente da função de decaimento da distribuição pk é α , temos que

$$pk \sim k^\alpha. \quad (2.27)$$

Neste caso, o modelo BA prediz $\alpha = 1$, assim $pk \sim k_i$. Podemos chamá-lo agora de modelo de ligação preferencial linear.

Apesar de nem sempre o coeficiente α da forma funcional de pk ser linear no decaimento em k , ou seja, nem sempre teremos $\alpha = 1$ (ver os modelos de ligação não linear da figura 2.12), a mensagem mais importante do modelo BA é que a estrutura e evolução da rede são inseparáveis (CALDARELLI; VESPIGNANI, 2007). Assim, para entendermos a topologia de sistemas complexos, como os que foram ilustrados até agora, precisa-se descrever como eles se desenvolveram até ali. Barabasi (2016) chama isso da 'captura da gênese do sistema'.

Figura 2.12 – Regimes do modelo BA: quando $\alpha = 1$, temos o modelo BA com $\gamma = 3$. Quando α se aproxima de $\frac{1}{2}$, recuperamos a função randômica dos modelos exponenciais (sublinear). Já quando α se aproxima de $\frac{3}{2}$, temos um modelo do tipo 'Regime Anômalo' (neste caso, superlinear) discutido anteriormente, quando há um hub muito grande. Quanto mais α aumenta, mais completa fica a ligação da rede até se tornar um clique.

Fonte: Barabasi (2016).

2.3.7 Modelo Bianconi-Barabási

Se o modelo BA pudesse prever o surgimento de grandes sites de internet, o 'novato' Google provavelmente seria até então uma ferramenta de busca bem menor de que realmente é,

se comparado ao Alta Vista, por exemplo. Isso é devido à 'vantagem do primeiro jogador', visto anteriormente. Assim, podemos dizer que o Google ganhou uma série de visitantes que, uma vez que o utilizou, sempre o fez novamente, à revelia dos demais mais antigos. Essa característica que coloca alguns à frente dos outros é chamada de função aptidão, ou do inglês *fitness*, surgindo assim o modelo Bianconi-Barabási (BB) (BIANCONI; BARABASI, 2001). Há algumas mudanças em relação ao modelo BA:

- crescimento de nós, em um processo contínuo no tempo: a cada tempo t , um novo nó com m novas ligações e η_j *fitness*, onde η_j é escolhido de uma distribuição *fitness* $p(\eta)$.
- ligação preferencial: agora a probabilidade da ligação de um novo nó é proporcional ao produto de seu grau k_j e sua *fitness* η_j , no qual nos fornece (2.28).

$$p(\eta_j) = \frac{\eta_j k_j}{\sum_i \eta_i k_i}. \quad (2.28)$$

A equação (2.28) nos mostra que nós mais novos em relação aos demais podem adquirir ligações mais rapidamente, caso houvesse uma maior *fitness* associada. De fato, a precisa forma de $p(\eta)$ afeta toda a estrutura da rede. Caso a forma da distribuição *fitness* fique longe de uma distribuição uniforme, a rede pode adquirir propriedades do regime 'superlinear', fenômeno descrito anteriormente. Dessa forma, a topologia da rede torna-se uma grande forma de estrela, com um hub muito grande ao centro e demais nós ligados a ele. A essa mudança para a forma de estrela pode ser associada uma analogia com um fenômeno físico chamado condensação de Bose-Einstein (previsto por Bose e Einstein em 1925, e produzido mais de 70 anos depois). Na condensação de Bose-Einstein, um conjunto de partículas de gás abaixo de uma certa temperatura passam a ocupar o mesmo estado de energia. Analogamente, podemos dizer que nessa forma estrela, muitas redes se submetem a uma condensação do tipo Bose-Einstein, no sentido de que todos as partículas (nós da rede) se concentram no nível mais baixo de energia, deixando os demais livres. Neste caso, o nó que tiver mais função *fitness* dominaria todos os demais. A figura 2.13 mostra um exemplo dessa configuração. O leitor pode verificar mais detalhes sobre a condensação Bose-Einstein e sua relação com redes complexas em Bianconi e Barabasi (2001).

O modelo BB assume algumas outras características que diferem do modelo BA, por exemplo (BIANCONI; BARABASI, 2001; BARABASI, 2016):

- presença de atratividade inicial: pelo fato de que nós isolados não poderiam ganhar ligações, adicionamos uma constante A na função dada por (2.27), que se chamará atratividade inicial do nó. Dessa maneira, o coeficiente da lei de potência γ se reduz para $3 + \frac{A}{m}$, diminuindo os efeitos das redes livres de escala.

- ligações internas (n): chamamos de n o número de novas ligações internas entre os nós já existentes. A ligação interna pode seguir uma função aleatória ou seguir (2.24). Caso seguir uma função aleatória, o expoente γ aumenta para $3 + \frac{2n}{m}$. Caso seguir a função do modelo BA, γ se reduz para $2 + \frac{m}{m+2n}$, o que nos dá γ entre 2 e 3. Isso nos indica que o processo de adicionar ligações entre os nós já existentes mimica o processo para redes randômicas. Já quando seguem a função da equação (2.24), Barabasi (2016) chama de Dupla Ligação Preferencial. Neste caso, a rede torna-se mais heterogênea, aumentando o tamanho dos hubs.
- eliminação de nós: neste caso, a uma taxa igual a r , eliminamos nós do modelo BA. Essa eliminação aumenta γ de 3 para $3 + \frac{2r}{1-r}$, melhor homogeneizando a rede. Quando a taxa de entrada e saída de nós é igual, ou seja, $r = 1$, a rede perde sua natureza livre de escala, assumindo uma distribuição de Poisson (BARABASI, 2016). De fato, neste caso $\gamma \rightarrow \infty$.
- idade (τ): nós podem ter tempo de vida limitado. É o caso, por exemplo, de redes de empresa ou atores. Pode-se estimar que a probabilidade de ligação torna-se $k\tau^{-\nu}$ (BIANCONI; BARABASI, 2001), onde ν é um parâmetro de escala para o parâmetro de idade τ . Para ν positivo, nós são encorajados a se ligarem a nós mais novos. Essa maneira de ligação suaviza a propriedade livre de escala, tornando $\gamma > 3$. Para ν negativo, os nós são encorajados a se ligarem a nós mais antigos na rede, aumentando o tamanho dos hubs, aprimorando a propriedade livre de escala, diminuindo assim o expoente γ para valores abaixo de 3. Quanto mais ν se aproxima de 1, mais a rede torna-se equivalente à randômica (BARABASI, 2016).

Em resumo, seguindo Bianconi e Barabasi (2001), podemos chegar a uma conclusão baseada nesta seção: se quisermos entender a estrutura da rede, devemos entender sua dinâmica. A topologia é sua consequência.

As próximas seções são destinadas a ilustrar algumas características que afetam a topologia das redes randômicas e livres de escala, acarretando nestas, consequências bastante diferentes entre si. Discutiremos brevemente a robustez desses tipos de redes e apresentaremos as arquiteturas de ligações entre nós que permitem o aparecimento de grupos (*clusters*) nas redes.

2.4 Robustez em redes

Robustez é uma questão central para sociologistas, que tentam prever efeitos de cunho social, como o efeito de opinião (BARRAT; BARTHELEMY; VESPIGNANI, 2008); economistas que buscam entender a instabilidade do mercado financeiro (MANTEGNA; STANLEY, 1999); médicos

Figura 2.13 – Em (A), um exemplo de partículas de gás que foram atraídas para o nível mais baixo de energia, C_0 . Essa característica produz configurações de redes do tipo mostrada em (B), em que o aeroporto de Denver é o hub central, onde vários outros se ligam através dele.

Fonte: Bianconi e Barabasi (2001) e Google Imagens.

que pesquisam o tratamento de células de câncer (BARABASI; GULBAHCE; LOSCALZO, 2011); engenheiros e físicos, que se ocupam do design de veículos, sistemas de comunicação, entre outros, que evidentemente não podem falhar (SCCELLATO et al., 2011).

2.4.1 Teoria de percolação

A Teoria de percolação (BARABASI, 2016) nos oferece respostas a perguntas sobre o número de nós que podemos excluir da rede sem que esta se fragmente em numerosos componentes. De fato, esta teoria nos diz que há um valor crítico p_c em que podemos observar uma transição de fase entre pequenos grupos isolados para a formação de grupo gigante ou componente gigante. A figura 2.14 mostra um exemplo de como atua a percolação no exemplo de grades, simulando uma floresta. Fato semelhante ocorre, por exemplo, na transição da água no estado líquido para o gasoso. Neste caso, a percolação de p_c é representada por 100°C . Pode-se já ter ideia de que tal p_c deve diferir entre redes randômicas e livres de escala. Os próximos tópicos discutem essa afirmação, verificando a robustez de redes contra falha e ataques deliberados.

Figura 2.14 – Simulação de fogo em floresta por meio de uma grade 250 x 250. Pode-se verificar que, quando $p < p_c$, ou seja, $p = 0,55$, o agrupamento de fogo (pontos escuros) é quase imperceptível. No limiar $p_c = 0,593$, o grupo de fogo se alastra (percolação), consumindo a floresta. Quanto maior p em relação a p_c , maior a percolação exercida pelo fogo, até que se consome toda a floresta. Este é um exemplo de como pode ser abrupta a mudança de fase, em confrontação com o que se esperava: um movimento suave e proporcional.

Fonte: Barabasi (2016).

2.4.2 Robustez em redes livre de escala

A teoria de percolação tem foco em grades regulares, como na figura 2.14 ou redes randômicas. Para o caso de redes livres de escala, simulações indicam que uma rede desse tipo perde seu componente gigante apenas quando p_c se aproxima de 0 (BARABASI, 2016). Assim, precisaríamos retirar quase a totalidade dos nós da rede livre de escala para que esta se fragmentasse em pequenos grupos isolados. Este é um fato marcante na robustez contra falhas nas redes livre de escala.

Podemos observar que para um componente gigante existir, cada nó que estiver conectado a ele, ao menos, deve estar conectado com outros dois. Para ilustrar isso, podemos pensar numa ciranda, tradicional brincadeira infantil, onde cada participante segura a mão de outros dois (figura 2.15). Dessa forma, todos são conectados. De modo semelhante, a rede tem componente gigante se

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} > 2. \quad (2.29)$$

Esse critério é chamado de Mollow-Reed (BARABASI, 2016). Utilizando a equação (2.29) em redes randômicas, podemos verificar que

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle (1 + \langle k \rangle)}{\langle k \rangle} = 1 + \langle k \rangle > 2, \quad (2.30)$$

Figura 2.15 – Nesta ilustração de uma ciranda, podemos verificar que quando cada indivíduo segura a mão de outros dois, temos aí um componente gigante.

Fonte: Google Imagens.

que implica que $\langle k \rangle > 1$. De fato, em redes randômicas, é necessário somente que o número médio de ligações supere uma ligação por nó para que surja um componente gigante.

Para redes livres de escala, podemos utilizar (2.29) para chegar ao parâmetro limiar inverso de robustez f_c (BARABASI, 2016)

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}, \quad (2.31)$$

onde f_c representa a fração de nós que precisamos retirar para que um componente gigante se desfça.

De fato, verificamos em seções anteriores que $\langle k^2 \rangle$ diverge para $\gamma < 3$. No limite em que $\langle k^2 \rangle \rightarrow \infty$, pela equação (2.31), podemos verificar que $f_c \sim 1$. Em outras palavras, para fragmentar uma rede livre de escala, com base em falhas aleatórias, precisamos retirar quase todos seus nós. A presença dos nodos centrais (*hubs*) é responsável por tal fato, pois a propriedade da lei de potência nos mostra que muitos são os nós com grau k pequeno, sendo igualmente diminutas as chances de encontrarmos um hub. Dessa forma, a remoção desses nós de k pequeno não altera a estrutura da rede.

Em geral, uma rede terá robustez realçada toda vez que f_c exceder ao parâmetro de robustez para redes randômicas, dado por (BARABASI, 2016)

$$f_c^R = 1 - \frac{1}{\langle k \rangle}. \quad (2.32)$$

2.4.3 Ataques em redes

Até agora, vimos os efeitos de falhas aleatórias em redes, restando abordar a questão dos ataques deliberados a tais estruturas. Sob ataque, uma rede livre de escala pode rapidamente se desintegrar, visto que agora seus hubs terão a maior probabilidade de ataque, ou seja, possui f_c

bastante baixo. A figura 2.16 mostra f_c para uma rede livre de escala sob ataques e sob falhas aleatórias.

Figura 2.16 – A probabilidade f_c de que a rede se fragmente quando em ataques e falhas aleatórias. Nota-se que a rede se fragmenta com a retirada de cerca de 25% de seus nós quando sob ataques. Já para falhas aleatórias, a rede torna-se bastante robusta, atingindo o patamar de 75%. Essa ilustração nos mostra a extrema fragilidade de redes livre de escala quando sob ataques. A rede inicial tem $\gamma = 2,5$, $k_{min} = 2$ e $N = 10.000$.

Fonte: Barabasi (2016).

Encontrar f_c depende do parâmetro da função de potência γ (BARABASI, 2016)

- f_c diminui para grande γ e cresce para pequeno γ . Isso está em linha com o fato de que, com pequenos valores de γ , temos maior concentração de um único hub, até que a rede se torna um clique.
- para valores altos de γ , a rede livre de escala atua como a rede randômica, isto é, sob ataque ou sob falhas aleatórias, tem mesmo valor de f_c . Isso está em linha com o fato de que f_k para redes randômicas é uniforme.

Voltemos à figura 2.13. Poderíamos indagar o que aconteceria se retirássemos o aeroporto de Denver de operação. Provavelmente o tráfego aéreo norte-americano viraria um caos em pouco tempo. O mesmo não aconteceria caso retirássemos, por exemplo, o aeroporto de Atlanta. Apesar da má notícia para os aeroportos, ataques em redes podem ter benefícios, por exemplo, em medicina, quando a retirada de um hub da rede de proteína de uma bactéria causa a rápida interrupção de uma infecção.

2.4.4 Ajustando a robustez

Ainda resta saber qual deveria ser a configuração da rede de maior robustez possível. Isto é, contra ataques e falhas ao mesmo tempo, ou seja

$$f_c^{total} = f_c^{ataques} + f_c^{falhas}. \quad (2.33)$$

Os autores Paul et al. (2004) descobriram que a rede com o máximo de f_c (ou o mínimo de p_c) tem um único hub com grau k_{max} e o resto dos nós com k_{min} . A função k_{max} depende de N , e é dada por (PAUL et al., 2004)

$$k_{max} \sim N^{\frac{2}{3}}. \quad (2.34)$$

Podemos verificar que essa topologia de hub gigante é robusta contra falhas aleatórias, dado que a chance de retirarmos o hub é quase nula para $N \rightarrow \infty$. Além disso, é robusta contra ataques, pois os nós formam um componente gigante. Assim, todos os nós k_{min} são conectados com o hub e com pelo menos outros dois nós com k_{min} . Quando retiramos o nó central, os restantes ainda continuam robustos contra ataques subsequentes (BARABASI, 2016).

2.4.5 Efeitos cascata

A falha de um nó pode causar o lapso de seus vizinhos e assim sucessivamente. Eventualmente, esta falha inicial não se localiza apenas no nó inicial, mas se propaga pelas ligações dos nós, atingindo aqueles relativamente longínquos (BARRAT; BARTHELEMY; VESPIGNANI, 2008). Podemos considerar alguns exemplos de ocorrência, apagões em redes elétricas, bloqueio de serviço em redes de internet, crises financeiras etc.

Para descobrir os padrões que governavam alguns desses efeitos, foi feita uma aproximação da sua distribuição, dada por (DOBSON et al., 2007)

$$p_s \sim s^{-a}, \quad (2.35)$$

onde s é o tamanho da avalanche, e o expoente de avalanche a é dado empiricamente. Essa configuração em lei de potência prediz que a maioria dos efeitos cascata serão pequenos ou nulos. Porém, coexistirão alguns que ocasionalmente produzirão grandes efeitos de avalanche quando, por exemplo, milhões de pessoas ficariam no escuro. Resultados empíricos mostram que (DOBSON et al., 2007; BARABASI, 2016)

$$a = \begin{cases} \frac{3}{2}, & \gamma \geq 3 \\ \frac{\gamma}{\gamma-1}, & 2 < \gamma < 3. \end{cases} \quad (2.36)$$

Observando a equação (2.36), podemos inferir que os maiores efeitos de tal avalanche seriam vistos quando uma rede possuísse $2 < \gamma < 3$. Felizmente, a maioria das ocorrências serão relativamente de pequena magnitude, e neste caso $\alpha \sim \frac{3}{2}$. A tabela 2.1 mostra alguns expoentes para algumas fontes. Notar que dentre os países com fonte de avalanche na rede elétrica, os Estados Unidos (EUA) apresentam a maior probabilidade de ocorrência de eventos extremos em forma de cascata, onde neste caso, $\gamma = 2$.

Tabela 2.1 – A tabela mostra os coeficientes de avalanche para vários países e fontes para a distribuição de lei de potência.

Fonte	α
Elétrica (EUA)	2,0
Elétrica (Suécia)	1,6
Elétrica (Noruega)	1,7
Elétrica (Nova Zelândia)	1,6
Elétrica (China)	1,8
Twitter	1,75
Terremoto	1,67

Fonte: Dobson et al. (2007).

Além dos efeitos de robustez e avalanche, as propriedades que regem as ligações entre os nós e suas estruturas de comunidades são de extrema importância para a ciência de redes. O último tópico deste capítulo trata então de arquiteturas de ligação e comunidades em redes.

2.5 Aspectos mesoscópicos de redes

No começo deste capítulo, iniciamos o estudo microscópico da rede, basicamente basicamente mediante as métricas relacionadas aos nós. Depois, nos atentamos às propriedades topológicas da rede e seus modelos teóricos, quando mudamos nosso foco para um olhar macroscópico da rede. Nós agora mudamos nosso foco do olhar macroscópico da rede para um olhar mesoscópico. Esse foco nos permite conhecer a estrutura de grupos e subgrupos, a partir da estrutura de conexões existentes entre os nós. De fato, o estudo das arquiteturas de ligação e comunidades em redes permeia entre os estudos microscópicos e macroscópicos da rede.

2.5.1 Arquiteturas de ligações: assortatividade e dissortatividade

Podemos ver que na sociedade as pessoas têm uma forte tendência de se associarem a outras de quem elas percebem que são similares de alguma maneira. De mesma forma, podemos

dizer que em redes os hubs gostariam de se associar a outros de seus semelhantes de igual importância. Chamamos uma rede que tem essa tendência de assortativa (NEWMAN, 2010). De modo contrário, há sistemas em que os hubs tendem a se isolar dos outros nós centrais, conectando-se a pequenos nós, formando componentes. A este caso, damos o nome de redes dissortativas (NEWMAN, 2010). A figura 2.17 mostra um exemplo dessas duas redes.

Figura 2.17 – Representação gráfica de uma rede assortativa (A) e uma rede dissortativa (B). Notar como a estrutura de comunidades que podemos observar na rede muda drasticamente.

Fonte: Newman (2003).

Se assumimos que esse processo é aleatório, ou seja, para uma rede neutra, temos que a probabilidade de que dois nós se associem com grau k e k' é (BARABASI, 2016)

$$p_{k,k'} = \frac{kk'}{2L}. \quad (2.37)$$

Inspecionando a equação (2.37), podemos perceber que se os hubs têm maior chance de se conectarem entre si, quanto maior é kk' , tanto maior será $p_{k,k'}$. Se os hubs se conectarem com probabilidade maior que a esperada por (2.37), esta rede é assortativa. Caso se conectem com probabilidade menor que a esperada por (2.37), esta rede é dissortativa. Podemos dizer, então, que as redes exibem correlação de grau. Essa característica pode ser observada pela matriz de correlação entre os graus dos nós.

É de certa forma difícil verificar as correlações pela matriz de correlação. Um método bastante eficiente foi derivado por Barrat, Barthélemy e Vespignani (2008), sob cuja

ótica a aproximação da função da correlação de grau é dada por

$$k_{nn}(k) \sim k^\mu. \quad (2.38)$$

A equação (2.38) nos mostra que a correlação de grau é dada pelo sinal do expoente de correlação μ . Para:

- redes assortativas: $\mu > 0$
- redes neutras: $\mu = 0$
- redes dissortativas: $\mu < 0$

Em suma, a correlação de grau nos permite capturar a presença ou não de correlações em redes reais, utilizando o parâmetro de escala μ para medir a magnitude dessa relação. Essa propriedade de redes muda sua topologia desde a transição de fase, quanto a sua robustez diante de falhas e ataques, ambos elucidados anteriormente. Além disso, acaba por modificar as estruturas de comunidades encontradas nas redes.

2.5.2 Comunidades em redes

Em ciência de redes, podemos dizer que as comunidades são um grupo de nós que têm alta preferência para se ligarem a outros de mesma comunidade do que de outras. A figura 2.18 mostra o trabalho dos autores Blondel et al. (2008) no estudo de comunidades sociais.

Formalmente, uma comunidade é um subgrafo local densamente conectado (ou um clique) em uma rede (BARABASI, 2016). Cliques são raros em redes, pois dependem de um subgrafo completo. Dessa maneira, pode-se relaxar essa restrição, definindo dois tipos de comunidades, a forte, onde cada nó da comunidade (C) tem mais ligações dentro de C do que fora de C , e a fraca, onde a soma dos graus internos de C é maior do que a soma dos graus externos à C (RADICCHI et al., 2004).

Para encontrarmos o número de comunidades em uma rede, precisamos de um algoritmo de partição, a fim de encontrarmos, ao melhor das hipóteses, n subgrafos não sobrepostos. Examinado todas as partições do grafo, seleciona-se a que melhor satisfaça nossa definição de comunidade. Essa não é uma tarefa plausível, pois a complexidade computacional em inspecionar todas as partições cresce exponencialmente com N (RADICCHI et al., 2004). Dessa maneira, necessitamos algoritmos que não dependam dessa restrição.

No caso do agrupamento do tipo hierárquico, inicia-se por construir uma matriz de similaridade para a rede. Então, o algoritmo hierárquico escolhido identifica os grupos de nós com a maior similaridade. Pode-se dividir os algoritmos hierárquicos em dois, sendo:

Figura 2.18 – Representação gráfica de comunidades extraídas de dois milhões de cidadãos da Bélgica acerca da língua falada, obtida por contatos telefônicos: cor mais escura (vermelho) para francês e mais clara (verde) para alemão. Nota-se também como os nós na intersecção dos grupos também contêm várias subcomunidades. Este exemplo nos mostra a importância dos agrupamentos na topologia das redes.

Fonte: Blondel et al. (2008).

- aglomerativos: juntam nós com a maior similaridade entre si.
- divisivos: divide as comunidades, retirando as ligações que conectam nós com baixa similaridade.

A definição da similaridade entre um nó i e outro j não é trivial. Porém, pode-se utilizar o algoritmo de Ravasz (RAVASZ; BARABÁSI, 2003) segundo o qual nós que se conectam diretamente um com o outro e compartilham múltiplos vizinhos têm mais chance de pertencer a uma mesma comunidade, sugerindo, dessa maneira, maior potencial de similaridade. No método aglomerativo, depois de aplicado o algoritmo, deve-se decidir o procedimento de aglomeração. Pode-se decidir entre vários algoritmos tradicionais, como o de ligação simples, ligação completa, ligação média e método de Ward's. O algoritmo de Ravasz (RAVASZ; BARABÁSI, 2003)

avalia a similaridade entre n_{ij} para cada outro par da rede, até que se encontre aquele com maior similaridade entre si, ligando o par. O processo de cálculo da nova similaridade continua, até que todos os nós estejam ligados numa única comunidade. Processo semelhante ocorre com os algoritmos de ligação simples, completa e média listados anteriormente. Após esse processo, deve-se proceder o corte no dendrograma formado, encontrando os grupos da rede. Um problema é que o algoritmo não nos fornece este corte.

Já no método divisivo removem-se sistematicamente todas as ligações que se conectam a diferentes comunidades. Define-se uma medida de centralidade entre as ligações e aplica-se um método hierárquico, já visto. Recordemos que ligações com alta centralidade por serem 'pontes' nas redes, devem ser removidas. Dessa maneira, ligações com baixa centralidade estão dentro de comunidades, enquanto ligações com alta centralidade estão intercomunidades (entre elas). O algoritmo de Girvan-Newman (NEWMAN, 2003) utiliza a centralidade *betweenness*, x_{ij} , e remove sistematicamente as ligações com maior x_{ij} . O processo de recálculo das centralidades continua até que todas as ligações sejam removidas. Agora, podemos utilizar a medida da modularidade M para obter um corte otimizado do dendrograma.

A modularidade (M) é uma medida associada a uma partição. Ela mede a diferença entre o diagrama real da rede, A_{ij} , e o número esperado de ligações entre i e j , p_{ij} , caso esse diagrama fosse aleatório. Assim

$$M_c = \frac{1}{2L} \sum_{i,j \in c} (A_{ij} - p_{ij}) \quad (2.39)$$

Dessa maneira, se M_c é positivo, então um subgrafo C_c tem mais ligações do que esperado por (2.37). Assim, representa uma potencial comunidade. Caso $M_c = 0$, então a conectividade dos nós é explicada por um processo aleatório. Finalmente, caso $M_c < 0$, os nós de C_c não formam nenhuma comunidade. A figura 2.19 ilustra esse conceito.

Utilizando (2.37), e somando para todas C_c comunidades, temos que a modularidade M_c é igual

$$M_c = \sum_{c=1}^{n_c} \left\{ \frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right\}, \quad (2.40)$$

onde L_c é o número de ligações totais dentro de uma comunidade C_c , e k_c é o número total de grau dos nós que estão dentro dessa comunidade.

Dessa maneira, a partição que tiver a máxima modularidade M para uma dada rede, oferece a estrutura ótima de comunidades (NEWMAN, 2003). O algoritmo 'Greedy' busca a

Figura 2.19 – A ilustração acima verifica a medida de modularidade M para várias partições da rede. Em (d), temos $M < 0$, neste caso cada nó é uma comunidade. Em (c), temos $M = 0$, assim, temos uma comunidade somente. Em (b) observamos que $M > 0$, assim, temos duas comunidades, porém ainda não é seu valor máximo, o qual se apresenta em (a), sendo então $M = 0,41$ a partição ótima encontrada para a rede. Nota: versão colorida na opção eletrônica.

Fonte: Barabasi (2016).

maior modularidade M , ligando novos pares de nós que apresentem $\Delta M > 0$ (NEWMAN, 2003).

Apesar de ser um dos algoritmos mais utilizados em comunidades em redes, a modularidade também tem suas limitações, como (NEWMAN, 2010; BARABASI, 2016):

- a maximização de modularidade não consegue detectar comunidades cuja soma de grau de todos os seus nós seja igual ou menor que $\sqrt{2L}$. Dessa maneira, pequenas comunidades, normalmente presentes em redes reais, ficarão forçadas a se juntarem a outras maiores.
- em certos casos, fica difícil distinguir a partição ótima de numerosas outras subótimas partições nas quais a modularidade é muito próxima. Segundo Barabasi (2016), isto decorre do fato de que a função modularidade não tem um pico ao redor de um simples valor, mas tem um 'platô' com valores muito próximos à M_{max} .

2.5.3 Caracterizando comunidades

Em redes sociais, podemos verificar que, quanto mais tempo dois indivíduos passarem juntos, maior a chance de pertencerem à mesma comunidade. Dessa maneira, as comunidades

tendem a se formar ao redor de grandes nós centrais. Do mesmo modo, as ligações intercomunidades tendem a ser fracas. Já em redes de transporte, podemos correlacionar os pesos das ligações à sua medida de intermediação (*betweenness*), por onde deve passar grande carga. Dessa maneira, os grandes (*hubs*) estão entre as comunidades, acarretando acentuado enfraquecimento das ligações dentro delas. O fenômeno dos pesos das ligações, dessa maneira, potencializa a eficácia dos algoritmos baseados em modularidade.

A evolução de comunidades também nos fornece informações sobre a dinâmica da rede (PALLA; BARABÁSI; VICSEK, 2007). Para o crescimento de uma comunidade, dependemos do número de ligações internas entre os nós dessas comunidades. Inversamente, a contração, ou mesmo desaparecimento de uma rede depende do número de ligações entre elementos de diferentes comunidades que seus nós apresentam. Outra característica importante é a idade das comunidades, a qual parece aumentar sua resiliência, modificando a estabilidade das redes.

Em suma, apesar do sucesso dos algoritmos de identificação de comunidades em redes, ainda há questões abertas, como (PALLA; BARABÁSI; VICSEK, 2007; NEWMAN, 2010; BARABASI, 2016):

- temos mesmo comunidades nas redes ou é um artefato computacional?
- devem todos os nós de uma comunidade estarem mesmo nela?
- as comunidades importam?

A teoria de redes para comunidades se desenvolve rapidamente, e novas respostas e perguntas certamente surgirão.

2.6 Softwares para redes complexas

Muito do conteúdo exposto está implementado nos principais softwares livres de análise de redes complexas. Pode-se destacar os seguintes softwares:

- *igraph*: o pacote para análise de redes livre '*igraph*' está disponível para o software estatístico R e para python. Um livro está disponível com todas as utilidades do pacote (KOLACZYK; CSÁRDI, 2014). É um dos mais completos em termos de algoritmos.
- *Gephi*: o popular software livre de visualização de redes contém vários pacotes adicionados pelos usuários (JACOMY et al., 2014).
- *Cytoscape*: o software livre *cytoscape* é bastante utilizado no meio biológico, porém tem todas as análises que seu irmão mais famoso, o *gephi*, tem, incluindo outras características visuais e de modelagem (FRANZ et al., 2016).

Algumas vezes, não é possível, somente com os softwares livres indicados acima proceder com a análise de redes. Alguma programação em python ou JavaScript pode ser necessária, por exemplo, para importação e preparação de dados ou mesmo uma visualização interativa na web.

2.7 Como construir redes do mercado de ações?

Neste capítulo, identificamos três estruturas básicas da teoria de redes complexas: seus aspectos locais, meso e macroscópicos. Agora, podemos começar a verificar aspectos sobre redes no mercado de ações.

Na introdução, já foram comentados alguns aspectos de redes para o mercado de ações. Porém, quando pensamos nestas, surgem as seguintes questões:

- o que conecta uma ação x_i a outra x_j ?
- se tal métrica existe, como obtê-la?
- o que tais métricas podem ajudar a compreender sobre a estrutura e dinâmica do mercado de ações?

Os capítulos 3 e 4 tratam sobre tais questões.

Página intencionalmente deixada em branco.

CAPÍTULO 3

MEDIDAS DE DEPENDÊNCIA

Os preços de todas as ações do mercado estão constantemente mudando em resposta a novas notícias, numa tentativa de antecipar o desempenho futuro de determinada empresa titular de ação submetida a esta ou aquela oscilação. Numa economia interconectada, não é surpresa que os movimentos dos preços das ações sejam também interrelacionados, ou seja, devem exibir algum tipo de dependência entre si. Destarte, a estrutura de dependência entre todas as ações é a chave determinante para estimar modelos de risco para o mercado (ENGLE, 2009). Estimar essa estrutura não é tarefa trivial, pois depende de várias estimativas acerca da métrica de similaridade entre as ações a ser utilizada, além do fato de essas estruturas de similaridades poderem não ser lineares (MANTEGNA, 1999; SORNETTE; MALEVERGNE, 2006). Além disso, dado que as empresas mudam com o tempo, é provável que suas ações também exibam dependências que são temporalmente-dependentes (ENGLE, 2009; SORNETTE; MALEVERGNE, 2006). O estudo de dependências é importante para entendermos a formação de redes financeiras.

3.1 Independência e dependência

A probabilidade de ocorrência de uma variável aleatória X , dada $P(X = x)$, pode ser modificada depois que informações acerca de outra variável aleatória Y , dadas por $P(Y = y)$ forem conhecidas. Esta revisão sobre as probabilidades de X dado Y é chamada de probabilidade condicional e denotada por $P(X = x | Y = y)$, onde (MARI; KOTZ, 2001)

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \text{ se } P(Y = y) > 0, \quad (3.1)$$

em que $P(X = x, Y = y)$ é a probabilidade conjunta das variáveis aleatórias X e Y .

Equivalentemente, podemos escrever

$$P(X = x, Y = y) = P(Y = y) P(X = x | Y = y), \quad (3.2)$$

que é então chamada de Lei da Multiplicação em probabilidade. Esta lei infere que a probabilidade condicional de uma variável aleatória multiplicada pela probabilidade da variável aleatória condicionante nos dá a probabilidade da intersecção destas variáveis, denotada por $P(X = x, Y = y)$.

Uma situação especial se verifica quando a probabilidade de X não é afetada pelo conhecimento de Y , ou seja,

$$P(X = x | Y = y) = P(X = x), \quad (3.3)$$

e neste caso dizemos que X e Y são independentes. Na condição de independência, pode-se observar que a equação (3.2) torna-se

$$P(X = x, Y = y) = P(X = x) P(Y = y), \quad \forall x, \forall y \quad (3.4)$$

Já no caso de variáveis contínuas, podemos verificar independência pela função densidade de probabilidade conjunta X, Y . Neste caso, a equação (3.4) pode ser reescrita como (MARI; KOTZ, 2001)

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad \forall x, \forall y \in \mathbb{R}, \quad (3.5)$$

onde $f_{X,Y}(x, y)$ é chamada de função densidade de probabilidade conjunta de X, Y , e $f_X(x)$, $f_Y(y)$ são denominadas funções densidade de probabilidade marginais de X e Y , respectivamente.

A equação (3.3) mostra que duas variáveis aleatórias são independentes se o conhecimento da informação de uma das variáveis aleatórias não altera a estimativa da probabilidade de ocorrência da outra. Podemos ver também que, caso algum par (x, y) viole a equação (3.5), então X e Y serão ditos dependentes. De maneira formal, quando podemos escrever a função conjunta $P(X = x, Y = y)$ como o produto das funções marginais $P(X = x)$ e $P(Y = y)$, como em (3.4) ou (3.5), conclui-se que as variáveis aleatórias são independentes (MEYER, 1969).

Das definições apresentadas, torna-se evidente que a função densidade de probabilidade conjunta (fdpc) determina, univocamente, as funções de densidade de probabilidade marginais (fdpm). O conhecimento da fdpc $f_{X,Y}(x, y)$ nos fornece as fdpm $f_X(x)$ e $f_Y(y)$. A recíproca não é verdadeira, infelizmente. Só poderemos obter a fdpc através das fdpm se, e somente se,

X e Y forem independentes (MEYER, 1969), ou seja, caso as equações (3.4) ou (3.5) sejam verdadeiras.

3.1.1 *Examinando dependências*

A dependência entre variáveis aleatórias surge como um campo de notório estudo na estatística e probabilidade, pois até que se conheça a estrutura de dependência dos dados, nenhum modelo estatístico pode ser construído (JOE, 1997; MARI; KOTZ, 2001). Francis Galton (1822-1911) e Karl Pearson (1857-1936) são reconhecidos como os pioneiros nos conceitos de dependência e em sua mensuração. Contribuíram substancialmente também o estatístico Corrado Gini (1884-1965) e o matemático Maurice Fréchet (1878-1973).

Uma inspeção gráfica pode ajudar a encontrar o tipo de dependência nos dados. Um método bastante eficiente neste sentido é o chamado diagrama de dispersão. Ainda que este tipo de análise não nos confirme uma dependência ou ausência dela nos dados, poderia ser a primeira tarefa de um pesquisador que gostaria de verificar como seus dados se comportam, para que possa, por exemplo, utilizar algum pressuposto para construir um modelo estatístico para os dados. Quantitativamente, o coeficiente de correlação de Pearson é um dos mais usados parâmetros de dependência. Entretanto, ele só indica uma relação linear entre as variáveis. A seguir, apresentaremos várias medidas de dependência.

3.2 Dependências lineares

Uma maneira simples de verificar uma dependência linear é pela medida de covariância, denotada por $Cov(X, Y)$. Essa medida utiliza o desvio esperado das respectivas médias de cada variável aleatória. Se $Cov(X, Y) = 0$, então as variáveis aleatórias são linearmente independentes. Caso $Cov(X, Y) \neq 0$, então as variáveis aleatórias são linearmente dependentes. Essa dependência pode ser positiva ou negativa, de acordo com o sinal da covariância. Ressalte-se que se $Cov(X, Y) = 0$, não podemos dizer apenas que as variáveis são independentes, pois a covariância apenas capta a dependência linear. Outros tipos de dependência, como formas quadráticas, não são levadas em consideração (RACHEV et al., 2010). Pode-se dizer apenas que se $Cov(X, Y) = 0$, as variáveis são não correlacionadas. Caso $Cov(X, Y) \neq 0$, as variáveis são correlacionadas.

Matematicamente, a covariância para variáveis discretas é dada por

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = \sum_x \sum_y (x - E(X))(y - E(Y))P(X = x, Y = y), \quad (3.6)$$

onde $E(X)$ é a média de X e $E(Y)$ a média de Y . A equação (3.6) computa a soma dos desvios conjuntos, multiplicada pela probabilidade de todas as combinações entre x e y . Para o caso contínuo, trocam-se as duplas somas pela integral dupla.

Um dos problemas da covariância é que ela é sensível a transformações de escala. Por exemplo, se multiplicarmos X por uma constante a' e Y por outra b' , a covariância muda de $Cov(X, Y)$ para $a'b'Cov(X, Y)$. Outro problema, é que ela não tem limite inferior nem superior. Pode assumir qualquer valor, ficando, desse modo, difícil em comparar covariâncias entre vários pares de variáveis medidas em diferentes escalas. No caso de ações, se trocarmos as séries temporais dos retornos das ações da janela diária para semanal, esta troca afetará a medida de covariância. Então, precisamos de uma medida que seja invariante de escala e padronizada: o coeficiente de correlação linear de Pearson, definido como ρ de Pearson.

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}, \quad (3.7)$$

onde $-1 \leq \rho \leq 1$. Dessa forma, se $\rho = 0$, então as variáveis X e Y não são correlacionadas. Caso $-1 \leq \rho < 0$, as variáveis X e Y são negativamente correlacionadas. Finalmente, caso $0 < \rho \leq 1$, as variáveis X e Y são positivamente correlacionadas. No caso em que $\rho \pm 1$, podemos dizer que Y é uma função linear de X (com probabilidade 1). Desse modo, o coeficiente de correlação de Pearson é uma medida do grau de linearidade entre X e Y . Valores positivos de ρ mostram que Y tende a crescer com o crescimento de X , enquanto valores negativos de ρ mostram que Y tende a decrescer com valores crescentes de X . Para o caso de n variáveis, podemos montar uma matriz de correlação cujos elementos correspondem às correlações entre os pares de variáveis.

Um teorema que decorre da função de correlação é (MEYER, 1969): Se X e Y forem independentes, então $\rho = 0$. Isto ocorre porque $E(XY) = E(X)E(Y)$, ou seja, a fdpc fatora.

A recíproca do Teorema 1, em geral, não é verdadeira. Assim, podemos ter $\rho = 0$, e no entanto X e Y não precisam ser independentes (MEYER, 1969; JOE, 1997). Neste caso, apenas diremos que as variáveis não são correlacionadas. Portanto, correlação e independência, em geral, não são equivalentes. Por exemplo, um valor de ρ próximo a zero indica apenas a ausência de relação linear entre X e Y , porém não elimina a possibilidade de alguma relação não linear. Mostraremos alguns exemplos acerca deste comentário mais adiante. Além do mais, elucidaremos mais tarde neste capítulo uma exceção a essa regra, quando tratarmos da distribuição bivariada normal.

Apesar de a correlação linear de Pearson ser muito usada para medir dependências em finanças (MANTEGNA; STANLEY, 1999; TABAK; SERRA; CAJUEIRO, 2010), ρ é muito sensível a pares de valores extremos, além de não conseguir capturar funções não lineares. Além do mais:

- o coeficiente de correlação de Pearson é apenas definido quando as expectativas da variância são finitas;
- o coeficiente de correlação de Pearson depende das distribuições marginais de X e Y (MARI; KOTZ, 2001);
- conforme já comentado, geralmente zero correlação não implica em independência.

A figura 3.1 mostra um conhecido diagrama de dispersão elaborado por Anscombe (ANScombe, 1973). Em cada diagrama de dispersão, a média de x e y é igual (9; 7,5) e suas variâncias iguais a (11; 4,122), respectivamente. Além disso, todos possuem o coeficiente de correlação linear de Pearson igual a 0,816, encaixados numa reta do tipo $y = 3 + 0,5x$. Podemos ver claramente como a estrutura linear de Pearson não se mostra adequada para dados 'pouco comportados', isto é, com presença de valores extremos ou funções não lineares.

Figura 3.1 – À esquerda superior, são mostrados alguns dados vindos de distribuições normais, bem acomodados por uma reta. Já à direita superior, vemos uma clara relação não linear, onde uma reta não seria adequada. Abaixo, à esquerda, temos 10 pontos perfeitamente lineares, porém um valor fora desta reta muda a estrutura linear. Já à direita, abaixo, observamos 10 pontos na vertical e um ponto extremo. A reta passa por este último ponto, ignorando toda a estrutura da maioria dos dados.

Fonte: Anscombe (1973).

Já a figura 3.2 mostra algumas funções, bem como o coeficiente de Pearson estimado. Inspeccionando esta figura, obtemos $\rho = 0$ para a função (não linear) $Y = \cos(4\pi X)$, apesar de

claramente X ser uma função de Y , embora não linear.

Figura 3.2 – Para uma função tal como em (C), $Y = \cos(4\pi X)$, o coeficiente de correlação de Pearson é igual a zero, em contraste com o valor que assume para a relação linear em (A), um.

Fonte: Kinney e Atwal (2014).

Outrora verificamos que o fato de o coeficiente de correlação de Pearson ser zero não implicá independência (ver comentários do Teorema 1 e figura 3.2). Mas essa conclusão é verdadeira, por exemplo, se considerarmos um tipo muito conhecido de distribuição, a bivariada normal, representada por suas duas variáveis aleatórias, X_1 e X_2 como $N_1(\mu_1, \sigma_1)$ e $N_2(\mu_2, \sigma_2)$, onde μ_1 e μ_2 são as médias das variáveis aleatórias, e σ_1, σ_2 são os desvios padrão das variáveis aleatórias. Acrescente-se que temos um vetor de médias μ e um vetor de desvios-padrão, este agora chamado de matriz de covariância Σ . De fato, podemos mostrar o Teorema a seguir:

Se X e Y têm uma distribuição bivariada (duas dimensões) normal e a covariância ou correlação entre elas é zero, então as duas variáveis são independentes (MEYER, 1969; RACHEV et al., 2010).

A figura 3.3 ilustra um exemplo de função densidade de probabilidade conjunta para uma distribuição normal bivariada com correlação ρ igual a zero.

A função densidade de probabilidade conjunta (fdpc ou somente fdp) $P(X = x_1, Y = y_1)$ da distribuição bivariada normal (figura 3.3) é dada por (MEYER, 1969)

$$P(x_1, x_2) = \frac{\exp\left(\frac{1}{2}\left(x_1\left(-\left(\frac{x_1}{1-\rho^2} - \frac{\rho x_2}{1-\rho^2}\right)\right) + (-x_2)\left(\frac{x_2}{1-\rho^2} - \frac{\rho x_1}{1-\rho^2}\right)\right)\right)}{2\pi\sqrt{1-\rho^2}} \quad (3.8)$$

Notar que para este caso, se $\rho = 0$, então as variáveis aleatórias são independentes.

3.2.1 Popularidade da distribuição normal em finanças

Pela fdp da equação (3.8), podemos observar que, para a distribuição normal, a correlação ρ define toda sua estrutura. Também podemos dizer que a facilidade no cálculo do

Figura 3.3 – Exemplo de distribuição normal bivariada em 3 dimensões.

Fonte: Elaboração Própria.

coeficiente linear de Pearson é uma das principais razões para a grande popularidade da distribuição normal utilizada nos modelos econômicos e sociais, especialmente em finanças. Alguns desses modelos que utilizam o requisito de que seus dados vieram de uma distribuição normal são:

- O modelo *Markowitz*, usado na seleção de carteiras de investimentos eficientes.
- O modelo *CAPM* (do inglês, *Capital Asset Pricing Model*) para avaliação do risco individual de cada ação.
- o índice de *Sharp*, derivado do modelo *CAPM*, na avaliação do risco de um portfólio.
- O modelo de *Black-Scholes* para precificação de opções de ações.
- O modelo *VaR*, do inglês *Value at Risk*, modelagem que determina a máxima perda esperada em um portfólio de investimentos.

- O modelo de *Baumol*, para gerenciamento da variação de caixa.

Assim, boa parte dos modelos tradicionais em finanças utilizam procedimentos baseados em variância e no coeficiente de correlação de Pearson, embora este apresente limitações para uma ampla gama de dados (ver seção 3.2). A seguir, apresentaremos algumas propriedades que definem uma medida de dependência mais robusta.

3.3 Propriedades da dependência

Podemos definir algumas propriedades que fazem de uma medida de dependência, D_{ij} , uma boa estimativa da similaridade entre as variáveis aleatórias (SORNETTE; MALEVERGNE, 2006; ENGLE, 2009; RACHEV et al., 2010):

- 1. D_{ij} é definida para qualquer i e j ;
- 2. $D_{ij} = D_{ji}$;
- 3. D_{ij} é invariante sob transformações lineares;
- 4. $0 \leq D_{ij} \leq 1$, para métricas que mensurem somente a magnitude da dependência;
- 5. $D_{ij} = 0$ se, e apenas se, i e j sejam independentes;
- 6. $D_{ij} = 1$, se caso i e j sejam uma função estritamente monótona uma da outra;
- 7. D_{ij} deve captar tanto funções lineares quanto não lineares, ou seja, deve ser independente da função densidade de probabilidade f_{ij} ;
- 8. D_{ij} não deve ser afetada por valores extremos.

Como o coeficiente de correlação de Pearson viola as propriedades 5, 7 e 8, faz-se importante buscar métricas mais robustas, que consigam não só ser mais independentes de valores extremos, mas também captar estruturas não lineares.

3.4 Cópulas

Pesquisas em finanças (MANTEGNA, 1999; SORNETTE; MALEVERGNE, 2006; RACHEV et al., 2010; ENGLE, 2009) descobriram que os retornos das ações do mercado financeiro exibiam caudas mais 'grossas' nas distribuições, uma tendência que têm os ativos financeiros de assumir valores extremos. Uma das maneiras de resolver alguns desses problemas foi utilizando formas

de fdp que independam das funções de densidade marginais, dando acesso à estrutura das distribuições conjuntas, as então chamadas cópulas (RACHEV et al., 2010).

A estrutura de dependência entre variáveis aleatórias é determinada pela distribuição conjunta entre elas. Seja $X = (X_1, X_2, \dots, X_d)$ um vetor aleatório, cuja distribuição conjunta F seja expressa como (JOE, 1997)

$$\begin{aligned} F(x_1, x_2, \dots, x_d) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d) = \\ &= C(P(X_1 \leq x_1), P(X_2 \leq x_2), \dots, P(X_d \leq x_d)) = \\ &= C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \end{aligned} \quad (3.9)$$

onde F_i são funções marginais de distribuição, e C é a função no espaço d -dimensional numa unidade cúbica, chamada de cópula da distribuição conjunta F . As novas variáveis aleatórias $U_i = F_i(X_i)$ são distribuídas uniformemente no intervalo entre $[0 - 1]$. Neste caso, conhecido como o Teorema de Sklar's, a função cópula (SORNETTE; MALEVERGNE, 2006)

$$C : [0, 1]^d \rightarrow [0, 1] \quad (3.10)$$

Podemos verificar que a distribuição multivariada F pode ser escrita como uma função de C no intervalo real entre os números $[0 - 1]$ (3.10). Isso significa que a função conjunta de um vetor $\mathbf{X}$ é determinado pelas funções marginais de seus componentes - assumindo valores de $[0 - 1]$ - acoplados em C (SORNETTE; MALEVERGNE, 2006). Desse modo, podemos garantir, com apoio nesse teorema, que qualquer distribuição conjunta possa ser decomposta em distribuições marginais e em suas estruturas de dependência. O inverso nem sempre será verdadeiro.

Embora as densidades individuais (marginais) $f_X(x)$ e $f_Y(y)$ possam sempre ser calculadas a partir da densidade conjunta $f_{X,Y}(x, y)$ apenas para variáveis aleatórias independentes, pode-se obter a densidade conjunta a partir das densidades marginais pelo Teorema de Sklar's. Pela notação em (3.9), é imediato perceber que a função C é uma distribuição (densidade) conjunta para $(F_1(x_1), F_2(x_2))$. Por exemplo, a cópula gaussiana transforma as densidades marginais para $(\phi^{-1}(F_1(x_1)), \phi^{-1}(F_2(x_2)))$, onde ϕ^{-1} denota a função distribuição univariada normal padronizada. Para qualquer par de distribuição marginal $(F_1(x_1), F_2(x_2))$, podemos construir uma distribuição conjunta C com uma função cópula, sendo a função mais comum a gaussiana. Dessa maneira, podemos entender que a estrutura de dependência nos dados é dada pela junção das densidades marginais mediante uma dada cópula. Em casos mais complexos, podemos ter, por exemplo, densidades marginais normais, porém densidade bivariada conjunta não definida (KOWALSKI, 1973).

Suponhamos que gostaríamos de representar a estrutura de dependência, a dizer, da saúde pessoal (y) com o consumo de cigarro (x). Uma maneira óbvia é gerando variáveis alea-

tórias correlacionadas a partir de uma distribuição multivariada gaussiana. Porém, não podemos assumir por enquanto que estas variáveis tenham uma distribuição marginal gaussiana. Destarte, podemos continuar a utilizar as distribuições marginais $f_X(x)$ e $f_Y(y)$ destas variáveis, porém de forma a considerar sua estrutura de dependência conjunta. Para isso, podemos gerar pares (a, b) de uma distribuição gaussiana com correlação ρ e transformá-las utilizando a distribuição gaussiana cumulativa. Assim, $u = \Phi(a)$, $v = \Phi(b)$. Temos agora (u, v) com distribuições marginais uniformes. Finalmente, transformamos as variáveis novamente para: $x = F^{-1}(u)$, $y = G^{-1}(v)$, onde mantemos as densidades marginais que queremos e ainda preservamos a estrutura original de dependência dos dados. Neste caso, acabamos de usar a chamada cópula gaussiana.

As cópulas exibem algumas propriedades que são (RACHEV et al., 2010)

- densidade: muitas cópulas têm funções densidade que podem ser tratadas de forma analítica, como são as arquimedianas. Porém, nem sempre uma forma de solução fechada poderá ser encontrada, como é o caso de algumas cópulas (p.e., a cópula-T).
- Limites: usualmente, padronizam-se as funções de densidade marginais para valores entre 0 e 1. Dessa forma, toda a estrutura de dependência é representada pela função de densidade conjunta, ou seja, pela cópula.
- Invariância sob transformações de escala: este é um dos requisitos já discutidos para uma boa medida de dependência. Pode ser útil quando trabalhando com variáveis padronizadas.

Algumas famílias de cópulas mais populares incluem a gaussiana (normal), a T-student, cujas características dependem do grau de liberdade desejado, e algumas especiais, como a Gumbel, a Clayton e a Frank, que pertencem a classes de cópulas arquimedianas. A figura 3.4 ilustra algumas dessas famílias (funções) de cópulas, bem como seu diagrama de dispersão.

Para aplicações práticas, esses modelos têm a função de separar um conjunto univariado de distribuição de probabilidade, chamado de marginais, e um outro conjunto, que carrega a dependência entre os componentes, então chamado de cópula, pela transformação da escala das marginais para $[0, 1]$. Algumas medidas de dependência, como as de concordância, derivam da cópula estimada, e dessa maneira não dependem das densidades marginais (em confronto com a correlação de Pearson, medida na qual depende destas densidades).

3.5 Correlação por ranqueamento ou concordância

Uma forma natural de verificar e quantificar correlações, por exemplo, das ações no mercado financeiro, é comparar a probabilidade de que duas ações subam ou caiam ao mesmo tempo com a probabilidade de que uma suba e a outra caia ou vice-versa (SORNETTE; MALEVERGNE,

Figura 3.4 – A cópula normal é bem distribuída ao longo das marginais, enquanto que a T depende do parâmetro do grau de liberdade. Já as cópulas arquimedianas são bastante assimétricas. No caso da Gumbel há bastante dependência na cauda positiva da distribuição. Já na Clayton, a maior dependência é na cauda negativa da distribuição, indicando dependência em eventos extremos negativos (como a queda generalizada do mercado financeiro). Finalmente, a Frank tem diagrama de espalhamento bastante disperso nas marginais, indicando que a média passa a ser um parâmetro pouco confiável para representar as variáveis aleatórias. As três cópulas mais acima na figura são da família arquimediana, enquanto as outras duas mais abaixo são da família elíptica (possuem contornos elípticos).

Fonte: Naimy (2012).

2006). Sejam duas realizações (X_1, Y_1) e (X_2, Y_2) do mesmo par de variáveis aleatórias (X, Y) . Então:

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (3.11)$$

Dessa maneira, na equação (3.11), a parte mais esquerda do segundo termo nos dá a probabilidade de concordância, ou a probabilidade de que X e Y 'caminhem' juntos. Já a parte mais a direita nos dá a probabilidade de que X e Y 'caminhem' em lados opostos. Podemos chamar essa maneira de correlacionar eventos de correlação por postos. A expressão (3.11) quantifica a chamada Kendall's τ (ENGLE, 2009). A medida de Kendall é invariante em escala e

dessa forma depende apenas da cópula $C(X, Y)$ na forma de (SORNETTE; MALEVERGNE, 2006)

$$\tau(C) = 4 \iint C(u, v) dC(u, v) - 1. \quad (3.12)$$

Assim, a medida de concordância τ é a probabilidade de que um par de observações sejam concordantes menos a probabilidade que sejam discordantes (não concordantes). τ pode ser calculado para várias famílias de cópulas. A relação entre ρ e τ para qualquer cópula elíptica pode ser dada por (SORNETTE; MALEVERGNE, 2006)

$$\tau = \frac{2}{\pi} \arcsin \rho. \quad (3.13)$$

A equação (3.13) é interessante quando a distribuição não admite um segundo momento. Dessa forma, Kendall's τ tem a vantagem de sempre existir e ser fácil de ser estimado. Essa estimativa de τ é mais robusta do que a estimativa de ρ (SORNETTE; MALEVERGNE, 2006), especialmente quando as distribuições têm caudas pesadas e não exponenciais.

Para verificar a similaridade entre qualquer cópula C_1 e outra referência C_2 , podemos utilizar a medida de concordância de Spearman ρ^s , métrica muito semelhante à que foi vista pela de Kendall's τ (ver Sornette e Malevergne (2006) para uma comparação entre τ e ρ^s). O ranqueamento dos valores da distribuição marginal F_X e F_Y nos fornece a medida de Spearman, ρ^s (RACHEV et al., 2010). Em termos de cópula, ρ^s é dado por (SORNETTE; MALEVERGNE, 2006).

$$\tau(C) = 12 \iint C(u, v) dudv - 3. \quad (3.14)$$

Do mesmo modo que τ , ρ^s pode encontrado a partir de ρ para qualquer tipo de cópula elíptica, por (RACHEV et al., 2010)

$$\rho^s(X, Y) = \rho(F_X(X), F_Y(Y)), \quad (3.15)$$

onde F_X e F_Y são as respectivas funções marginais de X e Y .

Uma maneira fácil para se estimar ρ^s é pelo ranqueamento dos valores X e Y . Neste caso,

$$\rho^s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (3.16)$$

onde d_i é a diferença entre os dois ranqueamentos para cada par de observação e n é o número de observações.

Os índices de correlação por ranqueamento são especialmente interessantes quando a relação entre duas variáveis é estritamente uma função monotônica (função que sempre preserva a ordem, crescente ou decrescente) (SORNETTE; MALEVERGNE, 2006). Ademais, são mais robustos contra extremas realizações de pares de observações (outliers) do que o coeficiente de correlação de Pearson, por utilizarem o ranqueamento das variáveis, e não seu valor, ou seja, independem das distribuições marginais. Na figura 3.5, podemos verificar a robustez de ρ^s para uma função monotônica não linear.

Figura 3.5 – O coeficiente de Spearman vale 1 quando as variáveis são perfeitamente monotônicas, mesmo nesta relação não linear. Em contraste, o coeficiente de correlação de Pearson não consegue medir uma relação perfeita (0,88). Esta relação monotônica entre as variáveis é importante na estimativa de um robusto valor de ρ^s . Nota: $Y = 4 \left(\log \left(\frac{x}{1-x} \right) \right)$.

Fonte: Elaboração própria.

A interpretação dos valores medidos para Kendall τ e Spearman ρ^s é idêntica à correlação de Pearson ρ . Ou seja, seus valores são mensurados entre $[-1, 1]$ e chegam a esses limites quando X e Y são funções contramonotônica e comonotônica (funções perfeitamente decrescentes e crescentes, respectivamente), e iguais a zero para variáveis independentes. Além disso, também são simétricos e definidos para qualquer par de variáveis contínuas.

3.6 Dependência de cauda

Outra medida relativamente importante em finanças é a chamada dependência de cauda. Ela se foca nas partes extremas da fdp para variáveis aleatórias, quer dizer, mede a ocorrência de valores muito extremos (nas caudas da fdp) tanto positivos quanto negativos. A chamada dependência de cauda inferior expressa a probabilidade de um dos componentes assumir valores bem pequenos dado que o outro componente já está nesse estado (RACHEV et al., 2010). Essa medida pode ser interessante, por exemplo, ao se verificar a probabilidade de efeito cascata negativo em mercado de ações. A medida de dependência de cauda superior é análoga à de dependência de cauda inferior. Neste caso, exemplificando, um investidor pode aplicar em ações que tenham essa característica, esperando um efeito cascata positivo. Podemos apresentar de forma matemática (SORNETTE; MALEVERGNE, 2006)

$$\begin{aligned}\lambda_s(X, Y) &= \lim_{u \rightarrow 1} P[X > F_X^{-1}(u) \mid Y > F_Y^{-1}(u)] \quad \text{para cauda superior} \\ \lambda_i(X, Y) &= \lim_{u \rightarrow 0} P[X < F_X^{-1}(u) \mid Y < F_Y^{-1}(u)] \quad \text{para cauda inferior}\end{aligned}\tag{3.17}$$

Se λ_i ou $\lambda_s > 0$, então eventos extremos podem acontecer em diversos sistemas e afetar simultaneamente vários de seus componentes. Se λ_i ou $\lambda_s = 0$, a cópula não tem dependência de cauda, e X e Y são ditos assintoticamente independentes. A exemplo das demais métricas apresentadas nesta seção, o coeficiente de cauda λ_i ou λ_s tem formas analíticas para diversas famílias de cópulas (SORNETTE; MALEVERGNE, 2006). Para a gaussiana, não existe dependência de cauda, exceto quando $\rho = 1$ (SORNETTE; MALEVERGNE, 2006; RACHEV et al., 2010). Dessa forma, utilizar um modelo gaussiano para modelar dependências no mercado financeiro, especialmente em ações, pode ser perigoso, pois negligencia a mensuração da dependência $\lambda_i(X, Y)$.

Uma melhor alternativa é utilizar a cópula da distribuição-T de Student. Se duas variáveis aleatórias estão correlacionadas, mas não perfeitamente, então devem exibir algum tipo de dependência de cauda inferior, dado que o coeficiente de dependência de cauda inferior para a cópula-T é (RACHEV et al., 2010)

$$\lambda_i(X, Y) = 2t_{v+1} \left(-\sqrt{\frac{(v+1)(1+p)}{1+p}} \right)\tag{3.18}$$

A figura 3.6 mostra novamente algumas famílias de cópulas, porém agora expondo mais especificamente sua dependência de cauda.

O software estatístico **R** tem uma série de pacotes para ajudar na estimativa de tais medidas de dependência. O leitor pode consultar os pacotes do R 'Hmisc' e 'copula' para mais

Figura 3.6 – As cópulas arquimedianas são bastante assimétricas. No caso da Gumbel (A) há bastante dependência na cauda positiva da distribuição, indicando a dependência em eventos extremos positivos. Já em Clayton (C), a maior dependência é na cauda negativa da distribuição, indicando dependência em eventos extremos negativos (como a queda generalizada do mercado financeiro). Destaca-se também que para a gaussiana (B) não há dependência de cauda.

Fonte: Mai e Scherer (2014)

informações.

3.7 Métricas da teoria da informação

Ao final desta última seção, já temos algumas definições sobre as perguntas que fizemos ao fim do capítulo 2. À pergunta 'o que conecta as ações?', já temos a resposta de que é uma medida de dependência. Apresentamos também como medir as mais diversas formas de dependência bem como o que cada uma representa.

A próxima medida de dependência que iremos abordar advém da teoria da informação.

A chamada informação mútua é vista como uma das mais robustas em estimar dependências, sem a necessidade de depender de linearidade, ou mesmo de funções monotônicas das variáveis, exemplificadas pela figura 3.5 (GULKO, 1999; DARBELLAY; VAJDA, 1999; SORNETTE; MALE-VERGNE, 2006; KRASKOV; GRASSBERGER, 2009). De fato, a informação mútua pode capturar relações não lineares nos dados (GULKO, 1999; DARBELLAY; VAJDA, 1999; SORNETTE; MALE-VERGNE, 2006; KRASKOV; GRASSBERGER, 2009; FIEDOR, 2014; BEKIROS et al., 2016). Diante de sua importância para essa pesquisa, a informação mútua será apresentada no capítulo 4.

CAPÍTULO 4

TEORIA DA INFORMAÇÃO

Neste capítulo apresentamos alguns conceitos da teoria da informação que podem ser úteis no estudo de dependências entre variáveis. A teoria da informação estuda a transmissão, o processamento, a utilização e a extração de informação. Estes conceitos abstratos foram aplicados em 1948 por Claude Shannon em sua tese *A Mathematical Theory of Communication*, cujo objetivo era permitir que o receptor de mensagens pudesse reconstruir o conteúdo delas, mesmo que estas tivessem sido enviadas através de um canal com ruído (SHANNON, 1948). Para isso, esta teoria se baseia em probabilidade e estatística e se preocupa com as medidas de informação de distribuições associadas a variáveis aleatórias, tais como a entropia e a informação mútua.

Aplicações das medidas de informação têm efeito na (COVER; THOMAS, 2005):

- matemática, na distribuição de funções, caracterizando longas sequências de variáveis aleatórias;
- computação, na ideia da complexidade de Kolmogorov, caracterizando a mínima sequência binária de um código;
- física: na explicação do comportamento termodinâmico macroscópico a partir da estrutura microscópica da matéria;
- economia: nos índices de desigualdade de distribuição de renda, ou na incerteza de séries temporais em finanças.

4.1 Medidas de informação

A ideia inicial da teoria da informação residia na área de compressão e transmissão de dados. Para isto tornou-se necessário uma medida da quantidade de informação. Para medi-la, Shannon (SHANNON, 1948) introduziu o conceito de entropia. A entropia de uma variável aleatória discreta X com a função de probabilidade $p(x)$ é definida por (SHANNON, 1948)

$$H(X) = - \sum_x p(x) \log_2 p(x). \quad (4.1)$$

Podemos dizer que a entropia é o número de bits, em média, requerido para descrever uma variável aleatória. Em outras palavras, é a medida da incerteza média em uma variável aleatória. Observa-se que a entropia é função da distribuição de x , dada por $p(x)$. Não depende, portanto, dos valores de x , mas da probabilidade de ocorrência destes.

Para uma variável aleatória que tenha uma distribuição uniforme, em que por exemplo $n = 32$, é suficiente dizer que precisamos de 5 bits para descrevê-la, pois

$$H(X) = \sum_{i=1}^{32} p(i) \log p(i) = - \sum_{i=1}^{32} \frac{1}{32} \log \frac{1}{32} = 5 \text{ bits}. \quad (4.2)$$

Para uma distribuição não uniforme, por exemplo, a probabilidade de oito aviões terem falha mecânica, dada por $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$, a entropia da distribuição da probabilidade de falhas será

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{16} \log \frac{1}{16} - 4 \frac{1}{64} \log \frac{1}{64} = 2 \text{ bits}. \quad (4.3)$$

Caso tivéssemos uma distribuição uniforme dessas mesmas falhas, teríamos $H(X) = 3$ bits.

Os dois exemplos ilustram a associação da entropia como medida da incerteza de uma distribuição. Para o caso da distribuição de probabilidades ser uniforme, (4.2), a entropia é máxima, pois a informação sobre os eventos é mínima. Do mesmo modo, quando temos mais informações sobre estes, (4.3), temos uma menor entropia, portanto sabemos mais sobre as probabilidades de ocorrência de cada evento. A entropia torna-se zero para o caso em que um único evento tenha probabilidade 1. A entropia máxima para qualquer distribuição não pode ser maior que $\log |X|$, onde $|X|$ denota o número de elementos de X . A figura 4.1 ilustra um caso especial para uma distribuição de Bernoulli. Observa-se que a entropia é máxima quando os eventos são igualmente prováveis, em que $p = 0,50$.

Figura 4.1 – A ilustração mostra a função cônica da entropia máxima e mínima para cada valor esperado de probabilidade p . Observa-se que para eventos igualmente prováveis, $p = 0,5$, a entropia é máxima.

Fonte: Cover e Thomas (2005)

4.1.1 Entropia conjunta e condicional

A extensão da entropia para um caso bivariado é a chamada entropia conjunta, formulada pela extensão da equação (4.1), por

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y). \quad (4.4)$$

No caso bivariado podemos obter a chamada entropia condicional $H(Y | X)$, é definida por (COVER; THOMAS, 2005)

$$\begin{aligned} H(Y | X) &= \sum_{x \in X} p(x) H(Y | X = x) \\ &= - \sum_{x \in X} p(x) \sum_{y \in Y} p(y | x) \log p(y | x) \\ &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y | x). \end{aligned} \quad (4.5)$$

Pode-se verificar que a entropia conjunta pode ser escrita como Cover e Thomas (2005, p. 17):

$$H(X, Y) = H(X) + H(Y | X). \quad (4.6)$$

No caso em que as variáveis são independentes, a entropia conjunta é a soma da entropia de cada variáveis, ou seja, $H(X, Y) = H(X) + H(Y)$.

4.1.2 Informação mútua

Para duas variáveis aleatórias X e Y com distribuição conjunta de probabilidade $p(x, y)$ e distribuições marginais de probabilidade $p(x)$ e $p(y)$, dizemos que a informação mútua $I(X, Y)$ é definida por (COVER; THOMAS, 2005):

$$I(X, Y) = \sum \sum p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (4.7)$$

Em termos de entropia, a informação mútua pode ser escrita como $I(X, Y) = H(X) - H(X | Y)$. Dado que $H(X, Y) = H(X) + H(Y | X)$, tem-se que

$$I(X, Y) = H(X) + H(Y) - H(X, Y). \quad (4.8)$$

Caso $X = Y$, temos que $I(X, X) = H(X) - H(X | X) = H(X)$. Essa é a razão para que a entropia também seja chamada de informação própria (COVER; THOMAS, 2005). A relação entre a entropia e a IM pode ser visualizada no diagrama de Venn (figura 4.2).

Figura 4.2 – Diagrama de Venn ilustrando as várias medidas de informação de duas variáveis aleatórias.

Fonte: Cover e Thomas (2005)

Podemos verificar duas propriedades da informação mútua:

- $I(X, Y) \geq 0$ (não negatividade) da IM: a informação mútua é nula se X e Y forem independentes.
- $H(X | Y) \leq H(X)$. Este condicionamento reduz a entropia e aumenta a informação mútua: Há uma igualdade apenas se X e Y forem independentes.

4.2 Entropia de distribuições contínuas de probabilidade

Na sessão anterior discutimos a teoria da informação para dados discretos. Para dados contínuos, as definições básicas são semelhantes, exceto que trocamos agora as somas por integrais. Seja agora X uma variável aleatória contínua com função densidade de probabilidade (fdp) $f(x)$. A entropia é dada por (SHANNON, 1948)

$$H(X) = - \int f(x) \log f(x) dx. \quad (4.9)$$

Se (X, Y) for um par de variáveis aleatórias com fdpc $f(x, y)$ e fdp marginais $f(x)$ e $f(y)$, então a entropia conjunta será (SHANNON, 1948)

$$H(X, Y) = - \int \int f(x, y) \log (f(x, y)) dx dy, \quad (4.10)$$

onde $H(X, Y)$ é máximo quando X e Y são independentes (SHANNON, 1948).

A entropia é definida para várias distribuições, em particular, a entropia para uma variável X que segue uma distribuição gaussiana com desvio padrão σ é dada por (SHANNON, 1948; COVER; THOMAS, 2005)

$$H(X) = \frac{1}{2} \log 2\pi e \sigma^2. \quad (4.11)$$

Para uma distribuição multivariada normal com média μ e matriz de covariância S , temos que (COVER; THOMAS, 2005, p. 250)

$$H(X_1, X_2, \dots, X_n) = H(N(\mu, S)) = \frac{1}{2} \log((2\pi e)^n |S|), \quad (4.12)$$

em que $|S|$ denota o determinante de S .

Para o caso contínuo, informação mútua (IM) é definida agora por (SHANNON, 1948)

$$I(X, Y) = \int \int f(x, y) \log \left(\frac{f(x, y)}{f(x)f(y)} \right) dx dy. \quad (4.13)$$

Caso (X, Y) forem independentes, então $I(X, Y) = 0$.

No caso de distribuições gaussianas bivariadas, utilizando as equações (4.8), (4.11), (4.12), podemos verificar que

$$I(X, Y) = -\frac{1}{2} \log(1 - \rho^2). \quad (4.14)$$

Para este caso, se o coeficiente de correlação de Pearson ρ for zero, a informação mútua $I(X, Y)$ também será zero. Caso $\rho = \pm 1$, X e Y são perfeitamente correlacionados e $I(X, Y) \rightarrow \infty$. Isso está de acordo com o fato de que para a distribuição normal bivariada a correlação implica em dependência entre as variáveis.

Uma comparação entre a informação mútua e o coeficiente de correlação de Pearson não é direta, pois o primeiro assume valores no intervalo $[0, \infty]$, enquanto que o segundo tem valores possíveis no intervalo $[-1, 1]$. Para comparar o grau de dependência entre as variáveis medido pela correlação e pela informação mútua é conveniente um parâmetro que assuma valores no mesmo intervalo. Para isso, podemos usar o resultado da equação (4.14) para a distribuição normal bivariada. Invertendo-a e generalizando o conceito, pode-se definir o coeficiente de dependência por (JOE, 1989)

$$\lambda = \sqrt{1 - e^{-2I(X, Y)}}, \quad (4.15)$$

onde λ assume valores no intervalo de $[0, 1]$, e é igual ao módulo do coeficiente linear de Pearson $|\rho|$ para a distribuição normal bivariada. O coeficiente λ permite a comparação do grau de dependência via correlação de Pearson e informação mútua, e é às vezes chamado de coeficiente global de correlação (GRANGER; LIN, 1994; DIONISIO; MENEZES; MENDES, 2004).

4.3 Estimando entropia e informação

Para obtermos a entropia de uma variável aleatória, precisamos conhecer sua distribuição de probabilidades. É necessário um modelo teórico ou uma estimativa empírica das probabilidades ou da função densidade de probabilidade.

Os métodos para estimação de densidade de probabilidades são classificados em paramétricos, quando assumem formas funcionais predefinidas, e não paramétricos, quando não dependem de uma forma funcional já conhecida. Os métodos paramétricos incluem os estimadores *bayesianos* e estimadores de máxima verossimilhança, enquanto os não paramétricos incluem histogramas, estimadores de densidades por kernels (funções núcleo), estimadores de *k*-vizinhos mais próximos e estimadores de *wavelets* (KRASKOV; GRASSBERGER, 2009; WALTERS-WILLIAMS; LI, 2009).

Na próxima seção, discutiremos dois métodos não paramétricos de estimação de distribuições de probabilidade, os histogramas e os estimadores de densidades por kernels (funções núcleo).

4.3.1 Estimando densidades empiricamente

Um dos métodos não paramétricos mais utilizados para estimar a densidade é o histograma, obtido a partir de uma partição de X em intervalos de tamanho finito. Assim, pela discretização dos valores de X , pode-se inferir a distribuição de probabilidades a partir dela, simplesmente contando os valores distribuídos em cada intervalo, podendo ser estimada a significância da independência dos intervalos em termos do teste de qui-quadrado (FRASER; SWINNEY, 1986). O problema é estimar o número e largura dos intervalos sem incorrer em erros sistemáticos (KRASKOV; GRASSBERGER, 2009). Segundo Moon, Rajagopalan e Lall (1995), a desvantagem do histograma é que a função estimada é descontínua e se modifica com a escolha da origem e da largura do intervalo.

Uma maneira de estimar a função densidade f num ponto x é (HOLLANDER; WOLFE; CHICKEN, 2015)

$$\hat{f}(x) = \frac{F_n(x - \frac{h}{2}) - F_n(x + \frac{h}{2})}{nh} = \frac{\#X_i \text{ no intervalo } (x - \frac{h}{2}, x + \frac{h}{2})}{nh}. \quad (4.16)$$

A função histograma remove o requisito de a média de qualquer intervalo ser x , utilizando, ao invés disso, um número de valores de centro dentro de um intervalo. Os valores da janela de tamanho h são escolhidos para que os intervalos $I_j = (x - \frac{h}{2}, x + \frac{h}{2})$ não se sobreponham. A esses intervalos da eq. (4.16), damos o nome de largura da janela.

Podemos simplificar a equação (4.16) como uma divisão entre

$$\hat{f}(x) = \frac{\#X_i \text{ no intervalo } I_j}{nh} = \frac{n_j}{nh}, \quad x \text{ em } I_j. \quad (4.17)$$

Importante ressaltar que, na construção do histograma, precisamos escolher uma origem x e um tamanho de janela h , que primariamente controla quanto de suavização teremos no procedimento. Exemplificando, na figura 4.3, vemos distintas estimativas de densidades, de acordo com os valores de largura da janela h .

A modificação do tamanho de h tem um efeito bastante significativo na estimação de densidade pelo histograma. É crucial, portanto, a boa escolha da largura h do intervalo. Se forem escolhidos valores pequenos, finas estruturas espúrias podem ser vistas. Caso sejam escolhido valores altos, será difícil perceber qualquer formação de padrão.

Figura 4.3 – Em (a), a distribuição estimada parece superestimar a verdadeira densidade pelo pouco 'detalhamento' nos intervalos ($h = 0,244$). Já em (b), a distribuição estimada parece subestimar a verdadeira densidade pelo muito detalhamento dos intervalos, ocasionando em regiões com zero probabilidade e aparecimento de uma densidade bimodal ($h = 0,037$). Neste caso, temos uma estimativa mais realista da densidade em (c), cujo $h = 0.129$.

Fonte: Hollander, Wolfe e Chicken (2015)

Problemas com histogramas - Alguns dos problemas com a estimativa de densidade pelo método do histograma são (SILVERMAN, 1986):

- **descontinuidade:** esta causa extrema dificuldade se uma derivada tiver que ser estimada.
- **origem:** a escolha de diferentes origens afeta de modo importante a densidade estimada.
- **dimensão:** histogramas para ordens maiores que 1 apresentam sérias dificuldades. A dependência não fica apenas na escolha da origem, mas também das coordenadas na direção dos intervalos.

A figura 4.4 ilustra o problema da descontinuidade e da origem.

Outro método muito usual para estimar uma densidade de probabilidade a partir dos dados observados é o chamado estimador de densidade por kernel (SILVERMAN, 1986). Para Moon, Rajagopalan e Lall (1995), os métodos de kernels são superiores ao histograma pois:

Figura 4.4 – Exemplos de histogramas para para os mesmos dados, porém origens diferentes: observe-se como há o aparecimento de estruturas de densidade diferentes para cada escolha de origem. Podemos verificar também a descontinuidade em vários pontos dos histogramas, como no dia 500.

Fonte: Silverman (1986)

- têm uma melhor taxa de convergência de erro quadrático quando da estimação da densidade de probabilidade.
- não dependem da escolha da origem.
- permitem especificar formas de intervalos (também chamados de janelas), definindo-os por meio de uma função K , mais sofisticada, assim, do que o padrão retangular de intervalo usado no histograma. A escolha da janela afeta muito mais a estimativa de densidade do que a escolha do tipo de função de kernel (SILVERMAN, 1986; SCOTT, 2015). A figura 4.5 ilustra esse conceito da dependência de h .

A função kernel K deve satisfazer a condição tal que

$$\int_{-\infty}^{+\infty} K(x) dx = 1. \quad (4.18)$$

Figura 4.5 – Exemplos de densidades obtidas por kernels para os mesmos dados, porém com janelas diferentes: observe-se como há o aparecimento de estruturas de densidade diferentes para cada escolha de janela h , (a) 0,1; (b) 0,3; (c) 0,6. Algo semelhante ocorre com o histograma da figura 4.3.

Fonte: Silverman (1986)

Usualmente, K será uma função simétrica como, por exemplo, a normal ou a triangular. Por analogia com o método do histograma, tem-se que o kernel é definido por (SILVERMAN, 1986)

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (4.19)$$

onde X_i é o centro do kernel, K , a função kernel e h , o tamanho da janela utilizada, também chamado de parâmetro de suavização ou largura da banda.

Para um conjunto de variáveis d dimensionais $\mathbf{x} = (x_1, \dots, x_d)^T$ e um dado conjunto de dados $\{x_1, \dots, x_n\}$, cuja densidade deve ser estimada, o estimador de densidade kernel multivariado K e largura de janela $h_1, \dots, h_d$ é definido por (SILVERMAN, 1986)

$$f(\mathbf{x}) = \frac{1}{nh_1 \cdots h_d} \sum_{i=1}^n \left[\prod_{j=1}^d K\left(\frac{x_i - x_{ij}}{h_j}\right) \right], \quad (4.20)$$

onde a função kernel satisfaz $\int_d K(\mathbf{x}) d\mathbf{x} = 1$.

O kernel gaussiano é uma das variantes mais populares, e sua utilização é generalizada para uma distribuição normal multivariada padrão radialmente simétrica como (SILVERMAN, 1986)

$$K(\mathbf{x}) = \frac{e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}}{(2\pi)^{d/2}}. \quad (4.21)$$

Em suma, a função estimada por (4.20) é uma média ponderada local das frequências relativas das observações ao redor do ponto de estimação. O objetivo da função kernel é fazer inferências sobre a densidade de probabilidade em todo o espaço, incluindo aquele onde não há dados observáveis. Na estimação por kernel, a contribuição de cada ponto observado é suavizada em uma região dependente do valor de h (que faz o papel do tamanho dessa região). Finalmente, agregando as contribuições de cada ponto, podemos estimar a estrutura da função de densidade. Intuitivamente, a estimativa por kernel é uma soma de ‘saltos’ com uma altura e uma extensão em torno de cada valor amostral. O salto é representado pela função escolhida como kernel e centrada em cada observação X_i e a sua extensão representa a probabilidade associada aos valores vizinhos. Assim, os pontos que estão próximos a X_i recebem um peso maior. Para uma boa estimativa da densidade, é importante ter-se uma boa estimativa do tamanho do intervalo h . Procedimentos para encontrar um valor ótimo de h ficam dependentes da própria densidade a ser estimada, cujo valor ótimo é encontrado para um kernel multivariado gaussiano por (SCOTT, 2015; MOON; RAJAGOPALAN; LALL, 1995)

$$h_i = \left(\frac{4}{(d+2)} \right)^{\frac{1}{d+4}} \hat{\sigma}_i n^{-\frac{1}{(d+4)}}, \quad (4.22)$$

onde $\hat{\sigma}_i$ é o desvio padrão amostral da variável i , e $d = 2$ para um kernel bivariado.

Para outras bandas além da referência gaussiana o tamanho da largura h deve sofrer uma redução. Alguns exemplos são discutidos em (SCOTT, 2015), como o método de validação cruzada e o de Sheather e Jones (SHEATHER, 2004), baseados nos dados empíricos, que utilizam estimativas para a derivada segunda da densidade, cujo propósito é minimizar o erro quadrático assintótico médio, *AMISE*. Cabe observar que o objetivo de tais métodos é estimar uma

derivada segunda de algo que ainda não conhecemos (a verdadeira densidade $f(x)$). O método de Sheather e Jones ainda continua a ser um dos mais poderosos, principalmente em se tratando de densidades suavizadas (SHEATHER, 2004; HOLLANDER; WOLFE; CHICKEN, 2015).

Em suma, vimos que quando a aproximação da referência normal, via (4.22), não for realmente plausível, deve-se optar por uma redução no parâmetro h , a fim de conseguirmos captar mais detalhes nos dados, especialmente se estes não tiverem uma densidade mais suavizada. Dessa maneira, o método de Sheather e Jones é uma boa alternativa quando da minimização deste parâmetro (SHEATHER, 2004). Não obstante, ao diminuirmos a largura da banda h estaremos incorrendo em maiores chances de erros sistemáticos (dependências espúrias) quando da estimativa da densidade. Para tanto, fica evidente que a 'Regra de Scott' é um bom começo (um modo conservador) para se testarem os possíveis valores de h . O autor Chiu (1991) faz uma boa análise das diversas metodologias de seleção de bandas h . Em suma, se acharmos razoável estimar h a depender de uma densidade referência, podemos implementar os métodos da 'Regra de Scott' ou método de Sheather e Jones. Particularmente, este último método não costuma mensurar bandas muito pequenas nem superestimar a verdadeira densidade (CHIU, 1991). De modo alternativo, se estimarmos densidades que apresentem características únicas, como 'caudas grossas', não há uma clara indicação de qual método utilizarmos, sendo necessários maiores cuidado na estimação de h . Neste trabalho, implementamos densidades que apresentam largura de banda h baseada na 'Regra de Scott' (SCOTT, 2015) e no método de Sheather e Jones (SHEATHER, 2004). Em nossas simulações, o primeiro método foi mais conservador ao estimar bandas maiores, algumas vezes subestimando a verdadeira densidade, enquanto que o segundo, em várias oportunidades, a superestimou. Nós providenciamos no Apêndice A.5 os resultados da simulação quanto a essas bandas. No software R, esses métodos estão implementados no pacote *MASS* e no *ks*.

Outros métodos de minimização incluem bandas h variáveis, como os estimadores de balão (HOLLANDER; WOLFE; CHICKEN, 2015), que utilizam uma banda h local para cada observação X_i . Este autor recomenda os livros de Silverman (1986), Scott (2015) para uma discussão aprofundada sobre esses métodos.

As figuras 4.6 e 4.7 mostram uma visualização da densidade estimada pelo método do kernel para dados univariados e bivariados. A função densidade resultante (linha contínua superior das figuras) é obtida somando-se a densidade em cada ponto, que no caso são funções gaussianas.

4.3.2 Outros métodos

Nós não iremos discutir com profundidade outros métodos de estimação de densidade, mas alguns deles são também importantes. Citando como exemplo, um outro método não

Figura 4.6 – Ilustração do conceito de estimação da densidade de probabilidade por kernel para uma amostra de 7 elementos. O kernel usado é uma função gaussiana. A curva superior contínua é a estimativa de $f(x)$, que obviamente não é gaussiana.

Fonte: Silverman (1986)

Figura 4.7 – Ilustração do conceito de estimação da densidade de probabilidade por um kernel normal bivariado para uma amostra com 100 elementos e $h = 1, 2$. A superfície superior contínua é a estimativa de $f(x, y)$.

Fonte: Silverman (1986)

paramétrico para estimar a densidade de probabilidade é o k -vizinhos mais próximos, uma metodologia supervisionada na qual a ideia básica é estimar a entropia de X pela média das distâncias para outros k -vizinhos mais próximos, tipicamente com k entre 2 e 4 (KRASKOV; GRASSBERGER, 2009). Segundo Walters-Williams e Li (2009), estimadores baseados no k -vizinhos mais próximos são mais estáveis, porém dependem de valor adequado do número de vizinhos k . Já o método de wavelets é uma outra alternativa de estimação de densidade. As *wavelets* (ou ondoletas) são funções localizadas tanto no tempo quanto na frequência e não forçam uma estimativa de alisamento para a distribuição, como é feito no método de kernel (WALTERS-WILLIAMS; LI, 2009). Para Kraskov e Grassberger (2009), os métodos de kernels tendem a ser inferiores aos de estimação por ondoletas, entretanto, estes são de implementação mais complexa em termos computacionais. Há também algumas derivações do modelo original de kernel, como o kernel variável, cujo parâmetro de escala h sempre varia de um ponto a outro.

Em suma, deve-se escolher qual método utilizar de acordo com o problema prático a ser enfrentado, além de adotar os parâmetros adequados de que as funções de densidade necessitam para ser aplicadas.

4.4 Informação mútua entre ações

Estimadas as densidades $f(x)$, $f(y)$ e $f(x, y)$, pode-se obter a informação mútua $I(x, y)$ via equações (4.9), (4.10), inserindo-as na equação (4.8), ou diretamente pela expressão dada em (4.13). Assim a $I(x, y)$ será a medida de dependência não linear entre, por exemplo, duas ações x_i e x_j .

Para ilustrar o caso bivariado, a figura 4.8 mostra as curvas de nível da função densidade estimada por kernel de uma amostra de pares (x, y) gerados de uma distribuição normal bivariada, com médias $\mu = (5; 10)$, desvios-padrão, $\sigma = (2, 24; 1, 41)$ e correlação de Pearson, $\rho = 0,95$. As curvas de nível têm forma elíptica, como esperado para uma distribuição normal bivariada com correlação não nula (SORNETTE; MALEVERGNE, 2006). No caso de correlação nula, estas curvas podem assumir forma elíptica com eixo maior horizontal ou degenerar para a forma de círculos. Para uma distribuição normal, toda dependência entre as variáveis está contida no coeficiente de correlação, como já elucidado.

Para comparar com o caso anterior, na figura 4.9 temos as curvas de nível para densidade conjunta estimada por kernel para dois ativos do Ibovespa (índice brasileiro de ações), vale dizer, os preços das ações da VALE5 (Companhia Vale do Rio Doce) e SUZB5 (Suzano Papel e Celulose), duas empresas do setor de exportação de commodities com coeficiente de correlação praticamente nulo ($\rho = -0,07$). Nota-se que as curvas de nível diferem significativamente da forma elíptica esperada para uma distribuição normal. Assim, a aproximação do grau de dependência destas ações por uma função linear não é adequada e, neste caso, pode subestimar de forma comprometedor a real estrutura de dependência entre elas. Neste caso, a informação mútua seria uma medida interessante de dependência não linear (global) entre as ações, pois capta a estrutura de densidade observada, medindo tanto dependências lineares quanto não lineares (DARBELLAY; VAJDA, 1999; KRASKOV; GRASSBERGER, 2009).

Vamos testar o código do Apêndice A.1 para o caso de uma distribuição conjunta bivariada normal com correlação de Pearson igual a 0 e variâncias marginais iguais a 4. Utilizando o código 'teste' em R do apêndice A.2, temos que a entropia para esta distribuição é igual a 4,2225 (ver primeiro resultado do Apêndice A.2). Utilizando a entropia teórica para uma normal multivariada, pela equação (4.12), encontramos o valor de 4,2241 (ver segundo resultado do Apêndice A.2). Uma boa estimativa, admita-se.

Utilizando o coeficiente global de correlação, definido pela equação (4.15), podemos

Figura 4.8 – Curvas de nível da função densidade estimada a partir de valores amostrais de uma distribuição normal bvariada com coeficiente de correlação $\rho = 0,95$.
Nota: versão colorida na opção eletrônica.

Fonte: Elaboração Própria

Figura 4.9 – Curvas de nível da função densidade conjunta estimada por kernel a partir de valores amostrais dos preços de duas ações do Índice Bovespa citadas, VALE5 e SUZB5. O coeficiente de dependência global, mensurado pela função kernel K normal e h pela regra de Scott, atinge $\lambda \cong 0,83$, enquanto o coeficiente de correlação linear é $\rho = -0,07$. O Apêndice A.1 contém o código fonte utilizado para cálculo da IM e do gráfico apresentado. Nota: versão colorida na opção eletrônica.

Fonte: Elaboração Própria

definir uma distância entre uma ação x_i e outra x_j , como

$$d_{ij}(\text{global}) = 1 - \lambda_{ij}, \quad (4.23)$$

e compará-la diretamente com a equação resultante da definição de dependência sob uma distribuição normal, uma distância assim definida por

$$d_{ij}(\textit{linear}) = 1 - |\rho_{ij}|. \quad (4.24)$$

Definidas as medidas de distância entre os objetos (as ações, digamos), estas podem ser usadas para a formação da árvore geradora mínima (tradução do inglês para *minimum spanning tree*, *MST*) de uma rede. O Apêndice A.3 mostra um código em R para transformar uma matriz simétrica em grafo.

No próximo capítulo, discutiremos alguns desses modelos de redes de ações baseados nos conceitos apresentados neste e em capítulos anteriores.

CAPÍTULO 5

REDES NO MERCADO DE AÇÕES

No capítulo 2, apresentamos vários conceitos de redes complexas, evidenciando sua importância para a ciência atualmente. No capítulo 3, mostramos o conceito de dependência e suas mensurações. Já no capítulo 4, apresentamos a teoria de informação e suas aplicações, enfatizando a medida de informação mútua como uma boa medida de dependência global das ações do mercado financeiro. O intuito desses capítulos foi atualizar o leitor a compreender melhor os estudos anteriores em redes de ações que farão parte deste capítulo. Após chegarmos às equações (4.23) e (4.24), que são as unidades de distância das ações em uma rede financeira, podemos agora verificar como diversos autores tratam dessa temática.

5.1 Complexidade em mercados financeiros

Os mercados financeiros estão entre os sistemas mais complexos que conhecemos até então. Isso é devido principalmente à difícil estimação dos fatores internos e externos que afetam o mercado, além do fato de esses fatores estarem fortemente interrelacionados por uma rede de conexões desconhecida (KWAPIEN; DROZDZ, 2012; MANTEGNA; STANLEY, 1999; SORNETTE; MALEVERGNE, 2006). Essas características têm como consequência a auto-organização dos mercados de maneira muito rápida, aumentando seus níveis de complexidade (KWAPIEN; DROZDZ, 2012). De fato, resultados de pesquisas mostram que o mercado vai progressivamente se estruturando em função do horizonte de tempo, sugerindo também que essa estrutura ocorra primeiramente conectando ações de mesmo subsetor, para então, conectar os papéis que representam o mesmo setor econômico (TUMMINELLO et al., 2007; MATTEO; POZZI; ASTE, 2010). Porém, além dos fatores que contribuem para essa auto-organização, esse tipo de sistema exhibe também muito

ruído, acarretando bastante dificuldade na simulação de tais mercados.

Uma característica fundamental nessa modelagem é a forma da distribuição das flutuações de preço. O estudo dos autores Kwapien e Drozd (2012) comparou a distribuição dos retornos de 100 companhias americanas. Na figura 5.1, verifica-se que esses retornos são, de fato, pobremente estimados por uma distribuição gaussiana, os quais, no entanto, são modelados de forma satisfatória por uma distribuição mais leptocúrtica, como pode ser vista por uma lei de potência (MANTEGNA; STANLEY, 1999; SORNETTE; MALEVERGNE, 2006; ENGLE, 2009).

Figura 5.1 – As análises das distribuições dos retornos das 100 maiores ações americanas (em vermelho) mostram que estão longe de serem modeladas a partir de uma distribuição normal (mostrada pela curva exponencial $N(0;1)$). Porém são bastante próximas a distribuições em forma de lei de potência com parâmetro de escala entre $[3 - 4]$.

Fonte: Kwapien e Drozd (2012)

Uma das maneiras mais novas e promissoras para tratar de sistemas complexos e suas dependências (lineares ou não) veio justamente com a ciência de redes complexas. Seus mapas transformam qualquer modelo de sistema em um modelo universal, permitindo investigar regularidades e leis que governam a estrutura e evolução desses sistemas.

5.2 Redes no mercado de ações

5.2.1 Redes de ações com estruturas lineares

Conforme visto na Introdução, um dos caminhos para tratar de redes financeiras foi a busca por arranjos hierárquicos que pudessem estar relacionados aos movimentos das ações, à época em que Mantegna (1999) revelou o fato de que as séries temporais carregavam valiosas informações econômicas, que de fato poderiam ser tratadas pela metodologia de redes. Se há uma regra que controle as propriedades estatísticas do sistema financeiro global tanto em dias típicos quanto em extremos (BONANNO; LILLO; MANTEGNA, 2001), então é provável que a estruturação do mapa em forma de rede desse sistema possa fornecer um melhor conhecimento de sua estrutura e dinâmica (MANTEGNA; STANLEY, 1999; BONANNO; LILLO; MANTEGNA, 2001; BRIDA; RISSO, 2008; TUMMINELLO et al., 2007; TABAK; SERRA; CAJUEIRO, 2010; YANG et al., 2014; FIEDOR, 2014).

Na construção de arranjos financeiros, podemos obter uma distância métrica e, extraída dela, uma árvore geradora mínima, revelando a estrutura geométrica e taxonômica presente na dependência entre os pares de ações do mercado financeiro. Segundo Mantegna (1999), a estrutura presente nos arranjos em rede das ações é útil na procura de fatores econômicos comuns que afetam um determinado grupo de papéis e podem ser evidenciados a partir das séries temporais dos retornos deles.

A mais utilizada medida de distância entre duas ações é definida pelo coeficiente de correlação de Pearson, dada por (MANTEGNA, 1999)

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \quad (5.1)$$

Tem-se que a distância na equação (5.1) se distribui entre $0 \leq d_{ij} \leq 2$. Isso significa que ações com correlações lineares negativas possuem distâncias maiores que àquelas com correlações positivas. A equação dada por (5.1) satisfaz as propriedades de uma distância, ou seja (MANTEGNA; STANLEY, 1999):

- $d(x, y) \geq 0$ para $\forall x, y$
- $d(x, y) = 0$ para $x = y$
- $d(x, y) = d(y, x), \forall (x, y)$
- $d(x, z) \leq (d(x, y) + d(y, z)), \forall (x, y, z)$.

Os trabalhos de Onnela, Kaski e Kertesz (2004), Bonanno et al. (2004), Brida e Risso (2008), Coelho et al. (2007) mostraram que a estrutura de redes do mercado de ações não é

aleatória, mas altera-se de acordo com as métricas e janelas temporais utilizadas. Assim, estudiosos na área de economia, física, computação e finanças desenvolveram novos trabalhos em redes para ações, no intuito de encontrarem novas conclusões empíricas acerca do funcionamento desse mercado, notadamente sua estrutura topológica e dinâmica, e suas implicações no gerenciamento de risco nos investimentos.

A figura 5.2 ilustra uma estrutura com árvore geradora mínima e distância entre ações baseada na equação (5.1). Já a figura 5.3 compara uma rede de ações real com uma artificialmente criada a partir de realizações gaussianas aleatórias.

Figura 5.2 – Árvore geradora mínima para trinta ações norte-americanas, baseada na distância ultramétrica, definida em (5.1). Nota-se como, a partir do arranjo, podemos evidenciar estruturas de grupos e ações mais centrais para o mercado.

Fonte: Mantegna e Stanley (1999)

Trabalhos posteriores também utilizaram a medida de distância da equação (5.1) na montagem das redes. Boginski, Butenko e Pardalos (2005), Tse, Liu e Lau (2010) criaram uma rede de correlações com filtro baseado em um parâmetro limite θ . Ambos estudos evidenciaram uma distribuição de grau em forma de lei de potência, com os setores financeiros tendo maior centralidade na rede. Dada a forma subjetiva de controle de θ , outros métodos foram sendo testados, como as redes baseadas em árvores geradoras mínimas (AGM) e as planares de máximo filtro (PMF) (TUMMINELLO; LILLO; MANTEGNA, 2010).

A diferença básica entre a árvore geradora mínima (ver seção 2.1.6) e as planares de máximo filtro reside no fato de que a primeira não permite ciclos, ao passo que a segunda não possui essa restrição, permitindo o aparecimento de cliques (componentes conectados). A

Figura 5.3 – As redes de ações reais, (A), apresentam vários agrupamentos, nós com alto grau de centralidade k , além de pequenos caminhos entre si. Já em redes baseadas a partir de realizações gaussianas aleatórias não correlacionadas, (B), poucos nós têm grau k alto, e boa parte destes apresentam longos caminhos entre si. A conclusão é que redes financeiras não são aleatórias e apresentam propriedades topológicas únicas.

Fonte: Bonanno et al. (2004)

única restrição da PMF é as ligações ficarem num mesmo plano, não permitindo ligações entre mais de um plano. Além disso, pela sua restrição topológica, só são permitidos cliques de 3 ou 4 elementos. Dessa forma, o número de ligações da AGM é $N - 1$, enquanto da PMF é $3(N - 2)$, além de que, segundo Tumminello, Lillo e Mantegna (2010), a PMF sempre contém a AGM. Estudos de autores como os três acima e Matteo, Pozzi e Aste (2010) não encontraram evidências das diferenças na análise quando elaborada com uma topologia ou outra. A figura 5.4 ilustra a comparação de uma rede de ações para essas duas classes topológicas de redução de complexidade de redes.

Já se viu que a maioria dos trabalhos recentes em redes de ações utiliza a medida de correlação e a medida de distância definida em (5.1) como métrica para obter as topologias citadas anteriormente. Vejamos agora como alguns autores analisam esse tipo de rede.

Os trabalhos de Zhang et al. (2011), Sensoy, Yuksel e Erturk (2013), Yang et al. (2014), Heiberger (2014) mostraram que a estrutura das redes alterava-se de acordo com a volatilidade do período em análise (ver figura 5.5). Para períodos de crise, foi visto que as medidas do tamanho médio dos caminhos (*average path length*) alteravam-se de acordo com a estrutura topológica utilizada. Para redes baseadas em expansão mínima (AGM), esses caminhos apresentavam características inversamente proporcionais às volatilidades atuais. Por exemplo, em tempos de crise, como a de 2008, o tamanho médio dos caminhos da rede se reduzia de forma brusca, pois aumentavam-se as ligações em distâncias menores (em virtude das maiores correlações entre os ativos). As redes que apresentavam características de caminhos bem curtos foram chamadas de estrelas, e as que apresentavam caminhos muito longos, chamadas de redes

Figura 5.4 – A comparação visual da AGM, (A), com a PMF, (B), evidencia que a segunda é mais rica em detalhes e ainda contém a AGM. Ressalte-se que a AGM é uma árvore, enquanto que a PMF forma um plano.

Fonte: Tumminello, Lillo e Mantegna (2010)

em cadeia. Em suma, a medida do tamanho médio dos caminhos parecia bem representar a real dinâmica dos mercados de ações.

Conforme visto no capítulo 2, na discussão sobre redes assortativas e dissortativas, outra importante descoberta em redes de ações foi a mudança do tipo de arquitetura de ligação entre as ações de um período para outro. A pesquisa de Heiberger (2014) (a partir de correlações cruzadas) encontrou uma associação entre esses tipos de redes e a volatilidade do período em análise. Segundo Heiberger (2014), em períodos normais, as ações tendem a se ligar com aquelas do seu mesmo setor ou grupo, podendo, neste caso, a rede ser vista como assortativa. Já em períodos de crise, as ações perdem essa característica, apresentando uma topologia dissortativa. A figura 5.6 mostra a ilustração do estudo de Heiberger (2014).

Outra configuração bastante importante no estudo de redes de ações foi a análise da centralidade e perifricidade. O estudo de Matteo, Pozzi e Aste (2010) encontrou evidências de dois principais tipos de ações, segundo sua centralidade ou perifricidade. As ações mais centrais, no estudo, eram as de maiores valores de grau e de centralidade de intermediação (discutidas no capítulo 2). Por sua vez, as ações mais periféricas possuíam excentricidade maior. Entende-se que, para um nó i , a excentricidade designa o máximo tamanho dos caminhos mais curtos que

Figura 5.5 – Períodos mais voláteis fazem as ações exibirem correlações cruzadas maiores. Observe-se que, no período entre 2008-2009, a correlação média entre as ações norte-americanas foi a maior, chegando a valores maiores que 0,50.

Fonte: Heiberger (2014)

conectam aquele nó a outro nó j . Assim, quanto maior esse valor de excentricidade, tanto mais periférica está a ação dentro da rede. Matteo, Pozzi e Aste (2010) apresentam esta classificação de ações:

- conectadas e centrais.
- conectadas, porém periféricas.
- ações pouco conectadas, porém centrais.
- ações pouco conectadas e periféricas.

Anos mais tarde, em novo estudo destes autores, mostrou-se que ações centrais e periféricas tinham comportamentos bastante distintos entre si (POZZI; MATTEO; ASTE, 2013). Verificou-se que o desempenho das ações mais periféricas era diferente daquele das mais centrais: carteiras de ativos periféricos tinham menor risco e melhores retornos que aquelas baseadas em ações mais centrais. O baixo desempenho destas últimas podia ser consequência do efeito cascata durante períodos de alta perturbação (volatilidade), pelo fato de que empresas (representadas por seus papéis) 'centrais' pudessem ser mais acometidas por decisões irracionais de investidores, como o efeito de compra/venda de 'manada' (POZZI; MATTEO; ASTE, 2013). A figura 5.7 ilustra

Figura 5.6 – Assortatividade e dissortatividade em redes de ações: em (a), vemos uma rede em um período normal, tendo característica de fato assortativa, uma vez que ações se ligam por meio de outros papéis de mesmo setor. Por sua vez, em (b), vemos um período de crise, tornando a rede bastante interligada e dissortativa, uma vez que estas mesmas ações se ligam a diferentes setores. Nota: para melhor interpretação, verificar versão colorida disponibilizada na opção eletrônica.

Fonte: Heiberger (2014)

a rede de ações centrais e periféricas. Na figura, quanto maior o tamanho do nó, maior a relação retorno/risco da ação considerada ¹.

Finalmente, temos um estudo para o mercado brasileiro (TABAK; SERRA; CAJUEIRO, 2010), também baseado na distância em (5.1) e pela obtenção da AGM. Pela análise da centralidade de proximidade e da dominância (ponderação do peso do nó i pelo do total da rede), pôde-se chegar a algumas conclusões para o mercado local:

- sugere-se que os setores financeiro, material e de energia sejam os mais centrais na rede, influenciando os demais. Essa visão é compartilhada por outros estudos em locais diferen-

¹ Apesar de não discutido neste trabalho, o premiado modelo de Markowitz para carteiras de ativos pode ser usado a partir das relações de risco/retorno estabelecidas a partir da análise das ações centrais e periféricas.

Figura 5.7 – Ações centrais vs. ações periféricas: na figura, quanto maior o tamanho do nó, maior a relação retorno/risco daquela ação. Aqui, 'P' representa as ações periféricas e 'C', as mais centrais.

Fonte: Matteo, Pozzi e Aste (2010)

tes, como em Matteo, Pozzi e Aste (2010), em relação à bolsa norte-americana.

- as medidas de centralidade da rede não são muito robustas, nem têm desempenho padrão em relação ao tempo, enquanto aquelas baseadas em 'caminhos' pareciam ser as mais robustas (TABAK; SERRA; CAJUEIRO, 2010).

Novas formas de modelagem de AGM foram propostas utilizando-se algumas abordagens dos modelos econométricos, como os que contemplam atrasos no cálculo de correlações. Segundo Sensoy, Yuksel e Erturk (2013), relação de atraso é uma característica possível em séries temporais financeiras, pois os ativos podem responder com certo atraso às variações de outros. Um desses modelos que visam tratar deste tipo de dinâmica é conhecido como ARMA-GARCH-DCC, estrutura utilizada pelos autores Sensoy e Tabak (2014). De fato, as mesmas conclusões são evidenciadas: as medidas de redes complexas, bem como sua estrutura e dinâmica, podem evidenciar as dependências entre as ações ou mercados como um todo e ajudar na previsão de novas movimentações bruscas, efeitos cascata ou mesmo na mensuração da robustez dos mercados.

5.2.2 Redes de ações com estruturas não lineares

As redes de correlação de Pearson tinham seu ponto forte: bastante simples de ser elaboradas, e ao mesmo tempo podiam reter boa parte da informação entre as ações (desde que evidentemente essas informações fossem relativamente lineares). Com o avanço da pesquisa em redes financeiras, especialmente no mercado de ações, percebeu-se que a correlação de Pearson, enfim, não era uma boa medida de dependência entre ações, como comentado no início desse capítulo. Os trabalhos que vieram a seguir iniciaram o estudo de redes de ações via estruturas de dependência não lineares, especialmente mediante o uso da informação mútua (IM).

Yang et al. (2014) utilizam a IM como intensidade da ligação entre os mercados de ações, sendo este o primeiro trabalho a mostrar a distribuição da informação mútua para a análise de alguns mercados. A figura 5.8 ilustra o fato de que a IM média nos mercados de ações dos países aumentou consideravelmente em períodos próximos a grandes crises, como a de 2008-9.

Figura 5.8 – Distribuição empírica da informação mútua para alguns países: observa-se que a informação tem distribuição assimétrica à esquerda para o período de maior volatilidade (em quadrados amarelos, nos EUA, em círculos vermelhos para Inglaterra, triângulos lilases para China e Hong-Kong. Nota: para melhor interpretação, verificar versão colorida disponibilizada na opção eletrônica.

Fonte: Yang et al. (2014)

Na figura 5.9, Yang et al. (2014) mostram como as redes de IM se alteram em períodos de crise, passando de uma rede em forma de estrela para uma em forma de cadeia longa, em sintonia com os estudos de Zhang et al. (2011), Sensoy, Yuksel e Erturk (2013), Tumminello et al. (2007) para redes de correlação linear.

Figura 5.9 – Árvore Geradora para três períodos, (A), 2005, (B), 2008-2009, e (C), 2010, utilizando informação mútua como medida de dependência entre as ações. A topologia em 2005 é em forma de estrela, a mostrar baixa dependência entre as ações. Por seu turno, a rede próxima à crise de 2008-2009 mostra-se bastante alongada, em forma de cadeia, evidenciando alta dependência entre as ações. Novamente, em 2010, a rede volta a apresentar uma estrutura em forma estrelar e revela recuperação dos ativos. Nota: neste trabalho, utiliza-se a árvore geradora máxima ao invés da tradicional árvore geradora mínima.

Fonte: Yang et al. (2014)

No mesmo ano, o trabalho de Fiedor (2014) mostrou AGM baseada em taxa de informação mútua. Da mesma forma que a taxa de entropia representa a entropia por unidade de tempo, a taxa de informação mútua (TIM) representa a IM trocada entre duas variáveis dinâmicas ao longo do tempo (FIEDOR, 2014). Para duas sequências de variáveis dinâmicas X e Y , temos que a TIM é dada por

$$\text{TIM} = \lim_{n \rightarrow \infty} \frac{\text{IM}(n)}{n}, \quad (5.2)$$

onde IM representa a informação mútua entre duas sequências de tamanho n .

De modo inverso ao apresentado no capítulo 4, o estudo de Fiedor (2014) utilizou uma abordagem paramétrica na estimativa da IM e da TIM, elaborada a partir da distribuição de Dirichelet. Além disso, encontraram-se os valores para IM e TIM por discretização em janelas de tamanho h pelo método do histograma. A métrica de distância utilizada a partir das medidas de informação foi (FIEDOR, 2014)

$$d(X, Y) = H(X) - H(Y) - 2I(x, y), \quad (5.3)$$

que satisfaz a desigualdade triangular, é positiva definida e simétrica. Para não depender do

tamanho n , utilizou-se uma normalização

$$D(X, Y) = 1 - \frac{d(X, Y)}{H(X, Y)}. \quad (5.4)$$

Apesar de não mostrar a comparação com a rede de correlação linear, Fiedor (2014) registra que foram observados mais agrupamentos em redes baseadas em IM do que as baseadas em correlação linear ρ . Nos setores centrais observados, chegou-se à conclusão semelhante àquela observada por Tabak, Serra e Cajueiro (2010), Sensoy e Tabak (2014), ou seja, não se pode desvincular o setor financeiro às maiores centralidades na rede observada. Um ano mais tarde, o estudo de You, Fiedor e Hořda (2015), de metodologia similar ao trabalho anterior, chegou à mesma conclusão de seus antecessores sobre os agrupamentos. Além disso, foi calculada a correlação entre as medidas de distância baseadas em AGM de correlações lineares e nas AGM de IM, chegando-se ao valor de 0,80. Assim, verificou-se que a IM poderia trazer novas e melhores formas de caracterizar o mercado de ações, em detrimento da tradicional correlação de Pearson (YOU; FIEDOR; HOŘDA, 2015). Em suma, não se pode deixar de lado as dependências não lineares em se tratando de ações do mercado financeiro.

De fato, ainda não estava claro o potencial da IM para o mercado de ações. Talvez a IM pudesse considerar outros fatores até então inacessíveis à correlação. E foi partindo do pressuposto de que seria provável que as ações respondessem com períodos de atrasos às variações de suas similares, que Junior, Mullokandov e Kenett (2015) estudaram essa dinâmica para matrizes baseadas em correlação linear e em IM. Neste estudo foram encontradas diferenças bastante significativas entre elas. A figura 5.10 ilustra essa diferença na análise da matriz de dependência de 83 bolsas internacionais.

Além do fato da diferença na matriz de dependência, os autores verificaram também diferentes formações de grupos quando do uso da medida de informação em vez da correlação linear. Pode-se conferir uma análise dos autores para o setor de produção de petróleo, a mostrar que a correlação falha em evidenciar os principais grupos de países influenciadores, enquanto que as medidas baseadas em informação, notadamente a informação mútua, capturam de forma mais adequada essa relação (JUNIOR; MULLOKANDOV; KENETT, 2015, p. 253-254).

O mais recente de que temos conhecimento a utilizar medidas de informação em redes de ações é o estudo de Kaya (2015), divulgado pela nova revista sobre o tema: o *Jornal de Teoria de Redes em Finanças* (*Journal of Network Theory in Finance*), o qual usa a distância definida em (4.23), baseada na equação (4.15), para montar uma AGM em redes de ações americanas entre 1996-2013. O uso da equação baseada em (4.15) se mostra interessante, pois essa distância $1 - \lambda$ pode ser facilmente comparada com a obtida em (4.24), $1 - |\rho|$, baseada na correlação linear de Pearson. Apesar de o estudo de Kaya (2015) não proceder dessa maneira, ele evidencia que a IM pode ser uma boa forma de investigação de regularidades e leis que governam a estrutura e

Figura 5.10 – Informação mútua (A) vs. correlação (B): matrizes de calor - Observe-se como a estrutura de dependência baseada em IM (A, primeiro quadrante) é bastante dissimilar à estrutura de correlação linear, (B), com relação à dependência temporal com atraso de um dia (a correlação não é capaz de identificá-la). Nota: áreas mais escuras correspondem a valores de dependência menores.

Fonte: Junior, Mullokandov e Kenett (2015)

evolução dessas redes. As figuras 5.11a e 5.11b detalham as AGM para dois períodos (2008 e 2013) e a mensuração de centralidade baseada em excentricidade, já comentada anteriormente, para vários setores da economia norte-americana.

Para finalizar essa seção, cabe mencionar o trabalho de Wang e Xie (2016) publicado com o tema de AGM para redes de ações com estruturas de dependência de cauda inferiores e superiores, apresentadas no capítulo 3. Relembremos que essas medidas visam a estimar a dependência de extremas flutuações de uma ação i tanto de alta (cauda superior) como de baixa

(a) (A) e (B) mostram redes baseadas na eq. (4.23), para os períodos de 2008-2009 e 2012-2013, respectivamente. Podemos observar que as distâncias médias em A, rede estrela, são bem menores que as distâncias médias em B, rede em cadeia longa, em virtude de no primeiro período em análise os valores de dependência serem relativamente mais altos em relação aos do segundo período.

(b) Por outro lado, em (C) mostra-se a medida de centralidade baseada em excentricidade para vários setores da economia americana: observe-se que podemos ver um aumento em boa parte das centralidades dos setores analisados perto de 2008 (os setores financeiro e elétrico parecem ser notórias exceções), junto à queda brusca do S&P 500, índice de ações norte-americano, ilustrado em tom mais escuro (marrom). Essa é uma previsão para a formação das redes do tipo em (A).

Figura 5.11 – Redes de informação mútua e suas centralidades.

Fonte: Kaya (2015).

(cauda inferior), em relação à mesma flutuação extrema de outra ação j . Nesse trabalho, Wang e Xie (2016) sugerem que a estrutura de dependência de cauda baseada em cópulas (utilizou-se uma cópula SJC, chamada de 'Simétrica-Joe-Clayton', variante dinâmica das cópulas de Joe e Clayton (WANG; XIE, 2016)) pode evidenciar propriedades não mensuradas pela correlação

linear de Pearson e pelas métricas tradicionais baseadas em ranqueamento (Spearman e Kendall) em todo o espaço xy . Mostrou-se também que as características de dependência são diferentes nas caudas inferior e superior, contendo diferentes agrupamentos e ligações internas. A figura 5.12 mostra como as estruturas topológicas da AGM para redes de ações norte-americanas são diferentes quando mensuradas as dependências dessas caudas.

Figura 5.12 – A topologia da AGM para redes de ações norte-americanas são evidentemente diferentes para (A), AGM com estrutura de dependência baseada em cauda inferior e (B), superior. Em (A), podemos ver uma rede em forma de estrela, enquanto em (B) uma rede mais alongada, em forma de cadeia. Isso revela o maior risco em redes do tipo em (A) para extremas flutuações na queda generalizada dos preços de ações.

Fonte: Wang e Xie (2016)

Os trabalhos apresentados indicam que estudo de redes considerando-se dependências não lineares entre seus componentes, via medidas da teoria da informação, é um tema promissor de pesquisa com muitas possibilidades a serem exploradas. Verifica-se uma ausência de trabalhos que comparem de modo mais assertivo as diferenças nas redes geradas por correlação de Pearson com as geradas através de IM. Além disso, as análises são feitas, de maneira geral, incompletas, não integrando as várias medidas de redes. Nesse sentido, analisando diversas propriedades, como as medidas mais microscópicas (caminhos, centralidades), além das mesoscópicas (formação de agrupamentos) e macroscópicas (modelos de redes, arquiteturas de ligação), podemos ter uma visão mais aprofundada sobre a estrutura de rede que estamos a analisar. Finalmente, vale ressaltar a escassez de trabalhos nesse tema para o mercado brasileiro.

CAPÍTULO 6

METODOLOGIA

Nesta seção apresentaremos a base de dados estudada e as técnicas de análise utilizadas.

6.1 Base de dados

O conjunto de dados para a pesquisa será o das ações do Índice Bovespa (Ibovespa), que compõe uma carteira teórica dos principais ativos (papéis) brasileiros, mais importante indicador do desempenho médio das cotações do mercado de ações brasileiro, a relevância do Ibovespa advém do fato de ele retratar o comportamento dos principais papéis negociados na Bolsa de Valores de São Paulo e também de sua tradição, pois mantém a integridade de sua série histórica e não sofre modificações metodológicas desde sua implementação em 1968.

A participação de cada ação na carteira do Ibovespa tem relação direta com a representatividade desse título no mercado à vista - em termos de número de negócios e volume financeiro - ajustada ao tamanho amostral. Isso é captado pelo índice de negociabilidade da ação (IN), calculado pela seguinte fórmula

$$IN = \sqrt{\frac{n_i}{N} \times \frac{v_i}{V}}, \quad (6.1)$$

em que n_i é o número de negócios com a ação i , N , o número total de negócios, v_i , o volume financeiro gerado pelas negociações com a ação i , e V , o volume total negociado, considerando-se o mercado à vista para todas as variáveis.

Seguindo a metodologia de representatividade de cada ação, calculamos o IN a partir de

uma janela de 5 anos (2012-2016) e selecionamos as primeiras 93 ações de maior negociabilidade. Isso significa que aquelas pouco representativas no mercado, seja por baixo número de negócios ou por baixo volume negociado, ou que ainda existam dados faltantes, ficarão de fora da análise por representarem mais ruído do que informação. A amostra representa 20% do total de ações do mercado brasileiro e quase 80% do total do IN, obtidos a partir do Economatica (ECONOMATICA, 2016). A tabela 6.1 mostra as ações que entraram para a análise, bem como seu índice de negociabilidade. Por exemplo, a ação da Petrobrás preferencial, PETR4, representa em torno de 8% do total dos negócios/volume do mercado de ações à vista brasileiro. Apesar do fato de nossa base de dados se expandir de 2012 a 2016, nossa análise será restrita aos dados de dois períodos que consideramos relevantes:

1. Período DR, sob gestão da Presidente Dilma Rousseff (01/06/2015 - 26/01/2016) - 159 pregões; retorno do índice no período: queda de 42%. Foram coletados 3888 retornos (dados a cada 15 minutos).
2. Período MT, sob gestão do Presidente Michel Temer (27/01/2016 - 08/09/2016) - 160 pregões; retorno do índice no período: alta de 50%. Foram coletados 3969 retornos (dados a cada 15 minutos).

O Apêndice A.6 mostra o código fonte para importar, reorganizar e exportar dados em alta frequência do site da Bovespa pelo pacote GetHFData no R.

6.1.1 Comentários sobre os períodos

Em meados de 2015, o Brasil passava por uma série de dificuldades econômicas e políticas, que afetaram diretamente o desempenho das ações do mercado. Nessa época, o índice da Bovespa caiu 42% entre o período de junho de 2015 a janeiro de 2016. No início de 2016, iniciava-se o processo de impeachment da presidente Dilma Rousseff, por indícios de irregularidades em contratos da Petrobrás e diversas manobras contábeis irregulares. A iminente queda da presidente Dilma animou investidores, fazendo com que as ações comesçassem a subir já a partir de fevereiro de 2016.

Em maio de 2016, assumiu a presidência Michel Temer, substituindo o mandato da presidente afastada Dilma Rousseff. As ações continuaram a subir até meados de setembro de 2016, enquanto eram propostas diversas reformas no país pelo atual presidente. Nessa época, o índice da Bovespa subiu 50% entre o período de fevereiro a setembro de 2016. Porém, a situação acabou se deteriorando com a recessão da economia, o alto desemprego e denúncias de corrupção dentro do núcleo do governo, culminando na prisão do presidente da Câmara Federal, Eduardo Cunha, pela 23ª fase da Operação Lava-Jato, em setembro de 2016.

Tabela 6.1 – Ações selecionadas para compor a análise: para o período DR (período de gestão da Presidente Dilma Rousseff), tivemos que excluir da análise a ação FLRY3, LREN3 e RUMO3 por dados inconsistentes e ou falta de dados; já para o período MT (período de gestão do Presidente Michel Temer), tivemos que excluir as ações OIBR3 e MDIA3 inconsistentes e ou falta de dados.

Nome	Classe	Bolsa	Código	IN (%)
Oi	ON	Bovespa	OIBR3	0.1113
Iochp-Maxion	ON	Bovespa	MYPK3	0.1175
Taesa	UNT N2	Bovespa	TAE11	0.1177
Qgep Part. SA	ON	Bovespa	QGEP3	0.1194
Valid	ON	Bovespa	VLID3	0.1235
Minerva	ON	Bovespa	BEEF3	0.1238
Eztec	ON	Bovespa	EZTC3	0.1251
Sao Martinho SA	ON	Bovespa	SMT03	0.1253
Iguatemi	ON	Bovespa	IGTA3	0.1276
Copasa	ON	Bovespa	CSMG3	0.1559
Qualicorp	ON	Bovespa	QUAL3	0.1662
Fleury SA	ON	Bovespa	FLRY3	0.1691
Sul America	UNT N2	Bovespa	SULA11	0.1774
Equatorial	ON	Bovespa	EQTL3	0.1841
Ecorodovias	ON	Bovespa	ECOR3	0.1884
Aliansce SA	ON	Bovespa	ALSC3	0.1917
Tran Paulist	PN	Bovespa	TRPL4	0.1938
BBSeguridade	ON	Bovespa	BBSE3	0.1971
Randon Part	PN	Bovespa	RAPT4	0.2044
Banrisul	PNB	Bovespa	BRSR6	0.2069
BR Propert	ON	Bovespa	BRPR3	0.2126
Ultrapar	ON	Bovespa	UGPA3	0.2199
Weg	ON	Bovespa	WEGE3	0.2203
Totvs	ON	Bovespa	TOTS3	0.2282
Odontoprev	ON	Bovespa	ODPV3	0.2327
Grendene SA	ON	Bovespa	GRND3	0.2386
Multiplan	ON	Bovespa	MULT3	0.2402
Porto Seguro	ON	Bovespa	PSSA3	0.2493
Multiplus SA	ON	Bovespa	MPLU3	0.2496
Kroton	ON	Bovespa	KROT3	0.2509
Even	ON	Bovespa	EVEN3	0.2585

RaiaDrogasil	ON	Bovespa	RADL3	0.2641
Marcopolo	PN	Bovespa	POMO4	0.2728
B2W Digital	ON	Bovespa	BTOW3	0.2854
Duratex	ON	Bovespa	DTEX3	0.3016
Alpargatas SA	PN	Bovespa	ALPA4	0.3084
Cetip	ON	Bovespa	CTIP3	0.3128
Dasa	ON	Bovespa	DASA3	0.3134
Estacio Part	ON	Bovespa	ESTC3	0.3410
Light S/A	ON	Bovespa	LIGT3	0.3495
Rumo Log	ON	Bovespa	RUMO3	0.3756
Marfrig	ON	Bovespa	MRFG3	0.3915
M. Dias Branco SA	ON	Bovespa	MDIA3	0.3921
Santander BR	UNT N2	Bovespa	SANB11	0.3923
Fibria	ON	Bovespa	FIBR3	0.3970
Eletropaulo	PN	Bovespa	ELPL4	0.4142
Copel	PNB	Bovespa	CPLE6	0.4401
CPFL Energia	ON	Bovespa	CPFE3	0.4477
Energias BR	ON	Bovespa	ENBR3	0.4487
Cia Hering	ON	Bovespa	HGTX3	0.4523
Via Varejo SA	UNT2	Bovespa	VVAR11	0.4538
Bradesco	ON	Bovespa	BBDC3	0.4544
Sabesp	ON	Bovespa	SBSP3	0.4562
Localiza	ON	Bovespa	RENT3	0.4736
Cesp	PNB	Bovespa	CESP6	0.5195
Eletrobras	PNB	Bovespa	ELET6	0.5460
Gol	PN	Bovespa	GOLL4	0.5499
Eletrobras	ON	Bovespa	ELET3	0.5571
Telef Brasil	PN	Bovespa	VIVT4	0.5881
Braskem	PNA	Bovespa	BRKM5	0.6172
Gerdau Met	PN	Bovespa	GOAU4	0.6189
Hypermarcas	ON	Bovespa	HYPE3	0.6341
Embraer	ON	Bovespa	EMBR3	0.6590
Tim Part S/A	ON	Bovespa	TIMP3	0.6593
Cosan	ON	Bovespa	CSAN3	0.6636
BR Malls Par	ON	Bovespa	BRML3	0.6729
P.Acucar-Cbd	PN	Bovespa	PCAR4	0.7033
Natura	ON	Bovespa	NATU3	0.7099

Oi	PN	Bovespa	OIBR4	0.7240
MRV	ON	Bovespa	MRVE3	0.7323
Bradespar	PN	Bovespa	BRAP4	0.7402
Lojas Americ	PN	Bovespa	LAME4	0.7516
Lojas Renner	ON	Bovespa	LREN3	0.7610
Suzano Papel	PNA	Bovespa	SUZB5	0.7647
JBS	ON	Bovespa	JBSS3	0.8158
Cielo	ON	Bovespa	CIEL3	0.8237
Gafisa	ON	Bovespa	GFSA3	0.8499
Cyrela Realt	ON	Bovespa	CYRE3	1.0614
Ambev S/A	ON	Bovespa	ABEV3	1.0639
CCR SA	ON	Bovespa	CCRO3	1.1115
Cemig	PN	Bovespa	CMIG4	1.1591
BRF SA	ON	Bovespa	BRFS3	1.1729
Sid Nacional	ON	Bovespa	CSNA3	1.4951
BmfBovespa	ON	Bovespa	BVMF3	1.7284
Usiminas	PNA	Bovespa	USIM5	1.7625
Gerdau	PN	Bovespa	GGBR4	2.0078
Itausa	PN	Bovespa	ITSA4	2.1858
Vale	ON	Bovespa	VALE3	2.1892
Brasil	ON	Bovespa	BBAS3	2.2292
Petrobras	ON	Bovespa	PETR3	2.6206
Bradesco	PN	Bovespa	BBDC4	2.9727
ItauUnibanco	PN	Bovespa	ITUB4	3.7831
Vale	PNA	Bovespa	VALE5	6.7624
Petrobras	PN	Bovespa	PETR4	7.9724

6.2 A estrutura de rede

A estrutura de redes do mercado será obtida a partir das matrizes de correlação e de informação mútua entre os retornos dos preços das ações. O retorno, R_T do preço de uma ação é definido como

$$R_T = \ln \left(\frac{p_T}{p_{T-1}} \right), \quad (6.2)$$

onde p_T e p_{T-1} são os fechamentos dos preços das ações no tempo T e $T - 1$, respectivamente.

A correlação de Pearson pode ser calculada para qualquer par de ações pela equação (3.7). Já, para estimarmos a informação mútua, iremos recorrer à abordagem não paramétrica, via densidade kernel, dada pela equação (4.20), e discutida no capítulo 4. Neste trabalho, utilizamos o kernel Gaussiano, dado pela Eq. (4.21) e, para estimativa da janela h , recorreremos à 'Regra de Scott', dado pela equação (4.22). Quanto à elaboração da Árvore Geradora Mínima (AGM), iremos utilizar o algoritmo de Kruskal (KRUSKAL, 1956), descrito no capítulo 2.

Nossa proposta é a aplicação de estruturas e métricas de redes complexas na comparação de topologias no mercado de ações, baseadas na correlação linear de Pearson e em informação mútua, a fim de quantificar as interrelações lineares e não lineares no mercado brasileiro de ações. Assim, pode-se verificar como as estruturas não lineares afetam a topologia e métricas das redes e suas possíveis implicações econômicas. Para isso, iremos utilizar a medida definida em (4.23), baseada na equação (4.15), com o objetivo de montar uma árvore geradora mínima em redes de ações do mercado brasileiro, justificando-se essa escolha pela falta de estudos sobre o mercado local. O uso da equação baseada em (4.15), como já visto, mostra-se interessante, pois essa distância $1 - \lambda$ pode ser facilmente comparada com a distância obtida em (4.24), $1 - |\rho|$, baseada em Pearson. As análises seguirão a seguinte estrutura:

- Para cada um dos períodos de análise (ver seção 6.1), iremos montar as AGM para as medidas de correlação linear e informação mútua (IM), gerando quatro redes diferentes.
- estas redes serão comparadas, ao nível macro e mesoscópicos, mediante sua topologia e distribuição de grau, centralidade e distâncias médias, robustez e formação de agrupamentos. No nível microscópico, buscaremos verificar quais as ações que desempenham papel mais importante por meio das medidas de centralidade. Esse tipo de verificação microscópica pode ser útil na estimativa de risco empírico das ações do mercado brasileiro. As redes e suas análises serão elaboradas pelos softwares Gephi e R. A Tabela 1.1 do capítulo 1 sumariza as principais propostas deste trabalho.

6.3 Código-fonte final

O código para cálculo da informação mútua mediante n -vetores encontra-se na seção A.4 do Apêndice A. Podemos utilizar uma tabela de vetores dos resultados da equação (6.2) de qualquer tamanho n , iterando cada par de vetores através de um comando 'for'. Temos, ao final deste processo, uma matriz baseada nos resultados da equação (4.15). Vale notar que estamos a usar a Regra de Scott (SCOTT, 2015) para obtenção do tamanho da largura da banda h . Finalmente, para a transformação em grafo, utilizaremos o código da seção A.3 do Apêndice A.

CAPÍTULO 7

RESULTADOS E DISCUSSÕES

Nesta seção apresentaremos os resultados obtidos comparando-se as redes obtidas a partir da correlação de Pearson e da Informação Mútua para dois períodos. O primeiro período, DR, corresponde ao do governo Dilma Rousseff, enquanto que o segundo período, MT, corresponde ao do governo Michel Temer.

As figuras 7.1a e 7.1b mostram a estrutura da rede obtidas a partir das distâncias baseadas na correlação linear e na informação mútua, respectivamente, para o período DR (presidente Dilma). As ações são destacadas pela estatística de centralidade. Dessa forma, quanto maior sua centralidade, maior seu tamanho na figura. Além disso, a tonalidade da cor de cada nó mostra a variação daquele ativo no período, vermelho para variação negativa, até o azul, para variações positivas. Além disso, foram destacados os ativos que comporiam a carteira que maximiza o chamado índice de Sharpe, através da técnica de seleção de carteiras eficientes, a chamada fronteira eficiente (mais detalhes no Anexo A). Este índice seleciona carteiras com a melhor relação retorno-risco ao longo desta fronteira. As ações circuladas foram selecionadas a partir do modelo de seleção de carteiras de Markowitz, pela otimização do índice de Sharpe, quando representarem ao menos 3% da carteira. Na comparação visual das figuras, observam-se caminhos mais longos em (a, rede correlação) em relação à (b, rede informação mútua), ou seja, há uma estrutura de dependência mais visível entre as ações em (b). Além disso, podemos ver que, em (a) os maiores quocientes retorno e desvio padrão podem ser obtidos tanto por ativos ao redor do centro, quanto mais os mais periféricos. Já em (b), estes podem ser obtidos mais pela periferia de modo mais homogêneo do que pelo que foi visto na rede formada a partir da correlação linear. Já as Figuras 7.2a e 7.2b mostram respectivamente, a mesma estrutura da rede formada a partir da correlação linear e da informação mútua para o período Michel Temer

(MT). Na Fig. 7.2a, averiguamos que os maiores quocientes retorno e desvio padrão podem ser obtidos pelos ativos mais periféricos. Já na Fig. 7.2b, podemos ver que estes são agora obtidos pelos ativos mais centrais, uma mudança que ainda não tinha sido verificada nas análises para o período DR. Finalmente, na Fig. 7.2a-b vemos uma maior aproximação entre os ativos em comparação com a da Fig. 7.1a-b. Isto indica que a rede obtida a partir da informação mútua durante o período MT tem mais fortes interdependências em uma estrutura de transmissão de volatilidade mais interconectada. Durante o período de DR, esta conclusão é mais fraca, porém as dependências baseadas na rede via informação mútua sempre estão acima daquelas baseadas em correlação linear.

Para enfatizar o papel das dependências não lineares, plotamos na Fig. 7.3 as distribuições do valor absoluto da correlação linear e o coeficiente de dependência global. Neste histograma, a correlação tem grande parte de seus valores centrados entre 0 e 0,1, enquanto que a informação mútua tem pico entre 0,2 e 0,4. Não há variações entre a correlação e a informação mútua a partir de 0,50. Se considerarmos a distribuição do período MT, na Fig. 7.3b, vemos que cerca de 82% da distribuição de dependência global está concentrada de 0,25 a 0,45, enquanto que para o período DR, na Fig. 7.3a, este valor é de cerca de 77%. Além disso, há uma maior proporção de valores de correlação linear entre 0 e 0,1 no período MT do que no período DR, o que significa que a diferença geral de informação mútua e correlação linear é maior no período MT do que no período DR. Esta conclusão é reforçada pelas Figs. 7.4a-b, onde mostramos uma representação pictórica da matriz simétrica cujas entradas são dadas pela diferença do valor absoluto entre o coeficiente global de dependência e da correlação linear para ambos os períodos. Estas matrizes são representadas pela intensidade de cor, de modo que os pontos escuros correspondem a uma diferença de pelo menos 0,30. A maior concentração de pontos escuros na Fig. 7.4b mostra que este período é caracterizado por uma estrutura geral de dependência não-linear mais interconectada. De fato, a estrutura de correlação linear no período MT tem uma dependência geral mais baixa do que no período DR, como também mostrado na Fig. 7.3. Isso indica que uma análise de risco baseada na correlação linear para o período MT é ainda menos apropriada.

(a) AGM a partir da distância baseada na correlação de Pearson.

(b) AGM a partir da distância baseada na informação mútua.

Figura 7.1 – Árvores geradora mínima (AGM) para o período de Dilma Rousseff. (a) AGM com base na correlação linear; (b) AGM com base em informações mútuas. O tamanho dos nós está relacionado à centralidade de autovetor, já o grau de matiz de cor com a variação das ações no período, vermelho para retornos negativos, até o azul escuro, para os positivos. As ações são destacadas dentro de um quadrado se compõem um mínimo de 3% do portfólio que maximiza o chamado índice de Sharpe. Distância média da AGM: (a) 0,64; (b) 0,56.

(a) AGM a partir da distância baseada na correlação de Pearson.

(b) AGM a partir da distância baseada na informação mútua.

Figura 7.2 – O mesmo que na Fig. 7.1 para o período de Michel Temer. (a) AGM com base na correlação linear; (b) AGM com base em informação mútua. As ações são destacadas dentro de um quadrado se compõem um mínimo de 3 % do portfólio que maximiza o índice de Sharpe. Distância média da AGM: (a) 0,66; (b) 0,54.

Figura 7.3 – Distribuição da dependência linear (azul) e global (amarelo) em valores absolutos. Sua interseção está em rosa. (a) Período de Dilma Rousseff; (b) período de Michel Temer.

As Figs. 7.5a - 7.5d mostram a distribuição cumulativa complementar de grau ponderado para a rede de correlação linear e aquela obtida a partir do coeficiente de dependência global, onde destaca-se em verde o ajuste para um modelo de lei de potência para todos os períodos. Para comparar a decomposição das caudas, também incluímos um ajuste de densidade log-normal (linha preta). Ao comparar as Figs. 7.5a e 7.5c com 7.5b e 7.5d, vemos que as redes de

Figura 7.4 – Representação pictórica da matriz simétrica cujos valores são dados por $|\lambda_{ij} - \rho_{ij}|$. (a) Período Dilma Rousseff; (b) Período Michel Temer. Pontos escuros mostram diferenças de no mínimo 0,3.

correlação seguem mais de perto a distribuição log-normal, enquanto as obtidas via informação mútua têm dependência de cauda próximas a uma lei de potência. Ainda mais, a Fig. 7.5d, que corresponde ao período com mais dependências não-lineares, mostra um comportamento mais próximo de uma lei de potência do que o que mostra o da Fig. 7.5b. Essa relação linear indica que há uma maior probabilidade de encontrarmos nós com grau bastante alto no período MT do que a esperada para o período DR. Além disso, sugere-se uma estrutura de dependência mais forte nesse período em relação ao primeiro. Nas análises foi utilizado o método descrito em Gillespie (2015).

(a) Distribuição de grau ponderado baseado em correlação linear para o período DR.

(b) Distribuição grau ponderado baseado em informação mútua para o período DR.

(c) Distribuição de grau ponderado baseado em correlação linear para o período MT.

(d) Distribuição grau ponderado baseado em informação mútua para o período MT.

Figura 7.5 – Função de distribuição cumulativa complementar baseada no grau ponderado para cada período em uma escala log-log. O ajuste de lei de potência é representado pela linha verde. Parâmetro de escala estimado: (a) $\alpha = 3.03$; (b) $\alpha = 2.44$; (c) $\alpha = 4.19$; (d) $\alpha = 2.39$. O ajuste da densidade lognormal na linha preta foi incluído para fins comparativos.

Na Tabela 7.1, separamos as ações da rede em dois grupos. Parte (a) contém os retornos dos 20 ativos com as maiores e menores centralidade baseada em autovetor, enquanto a parte (b) mostra o grau ponderado dos 20 ativos com os maiores e menores retornos para as ambas as redes baseadas na dependência linear e global, e onde μ e σ representam a média e o desvio

padrão para cada subgrupo. Nesta análise, estamos interessados em estabelecer a relação entre a estrutura do grau médio e os retornos de ativos. Na Tabela 7.1 (a), apontamos que, nas redes via informação mútua, os ativos mais centrais tiveram desempenho melhores do que os menos centrais para o período MT, e exatamente o contrário aconteceu durante o período DR, ou seja, os ativos mais centrais tiveram realmente os piores desempenhos. Por outro lado, as redes baseadas na correlação linear não apresentam tal conclusão. De fato, os ativos centrais e periféricos tiveram resultados muito similares nestas redes. Já na parte (b), mostramos que, nas redes via informação mútua as ações com um desempenho (retorno) muito positivo têm um grau médio ponderado que está bem acima do grupo de desempenho mais fraco, especialmente para período MT, e muito além do que foi verificado pela análise da rede via correlação linear. Tal conclusão sugere, em primeiro lugar, que as ações centrais são as que realmente 'movem' o mercado e, em segundo lugar, que as redes via informação mútua possuem mais informações úteis em relação à dualidade entre centralidade e retorno das ações do que aquelas via correlação linear. Por exemplo, em (a), um investidor, utilizando as redes via IM, aplicou nos 20 ativos com as maiores centralidades, no período MT ganhou 40% a mais do que se ele tivesse utilizado somente redes via correlação linear. Para o período DR, se utilizasse as análises advindas das redes via informação mútua, ele teria perdido quase 20% a menos se utilizasse uma estratégia de aplicar nos ativos menos centrais. De fato, em (b) vemos que os melhores retornos são relacionados com os maiores graus ponderados no período MT, e com os piores no período DR.

Tabela 7.1 – Relação entre centralidade e desempenho das ações; (a) retorno médio dos 20 ativos com as maiores e menores centralidade baseada em autovetor; (b) grau ponderado médio dos 20 ativos com os maiores e menores retornos para as ambas as redes baseadas na dependência linear e global. μ_r e σ_r são a média e o desvio padrão para os retornos, e μ_k e σ_k são a média e o desvio padrão para o grau médio.

<i>Período</i>	<i>Dilma Rousseff</i>				<i>Michel Temer</i>			
<i>(a) Retornos dos 20 ativos com as maiores e menores centralidade baseada em autovetor</i>								
	<i>mais central</i>		<i>menos central</i>		<i>mais central</i>		<i>menos central</i>	
	μ_r	σ_r	μ_r	σ_r	μ_r	σ_r	μ_r	σ_r
<i>Correlação linear</i>	-20.4%	14.7%	-27.7%	33.1%	64.3%	63.3%	64.4%	67.9%
<i>Dependência global</i>	-32.7%	19.1%	-8.3%	24.9%	105.5%	126.7%	46.5%	63.8%
<i>(b) Grau ponderado dos 20 ativos com os maiores e menores retornos</i>								
	<i>melhores retornos</i>		<i>piores retornos</i>		<i>melhores retornos</i>		<i>piores retornos</i>	
	μ_k	σ_k	μ_k	σ_k	μ_k	σ_k	μ_k	σ_k
<i>Correlação linear</i>	1.45	0.83	1.90	1.21	2.2	1.4	1.5	0.76
<i>Dependência global</i>	1.75	1.58	2.55	2.14	3	3	1.3	0.8

Finalmente, na Tabela 7.2 apresentamos alguns resultados da regressão linear elaborada, analisando em mais detalhes a relação entre as variáveis independentes, grau e centralidade, com a variável dependente do retorno das ações para o período DR. Neste caso, confirmamos não haver grandes diferenças entre o modelo 1 (rede de correlação linear) em relação ao 2 (rede via informação mútua), ou seja, não encontramos genuínas relações lineares entre os retornos e as variáveis de centralidade e grau na rede, na medida em que o R^2 é muito próximo a zero para ambos os modelos do período DR. Já a Tabela 7.3 apresenta os resultados da regressão linear para o período MT. Agora, vê-se alguma diferença entre o modelo 1 (rede de correlação linear) em relação ao 2 (rede via informação mútua). Nesse caso há uma certa dependência linear entre as variáveis retorno e grau ponderado, até porque o R^2 já pode ser considerado mais relevante para este segundo modelo.

Tabela 7.2 – Resultados da Regressão - período Dilma Rousseff (DR)

	<i>Variável dependente:</i>	
	Retorno (Linear)	(Global)
grau ponderado (rede de correlação)	-0.074*** (0.026)	
centralidade autovetor (rede de correlação)	1.033*** (0.360)	
grau ponderado (rede de inf. mútua)		-0.063** (0.027)
centralidade autovetor (rede de inf. mútua)		0.291 (0.281)
Constante	-0.221*** (0.043)	-0.170*** (0.046)
Observ.	90	90
R^2	0.092	0.073
R^2 Ajustado	0.071	0.052
Erro padrão resid. (df = 87)	0.288	0.291
Estatística F (df = 2; 87)	4.427**	3.422**
Nota:	*p<0.1; **p<0.05; ***p<0.01	

Estes resultados mostram que, apesar das redes de informação mútua dos períodos DR e MT não serem muito diferentes entre si, vimos que algumas características da estrutura da rede são diferentes, a começar pela relação entre centralidade e retorno das ações, mais presente no período MT. Verifiquemos agora outras diferenças.

A Tabela 7.4 mostra os principais resultados deste trabalho. A parte (a) da tabela 7.4 contém algumas métricas das redes. Vemos que, para as redes de informação mútua, a distância

Tabela 7.3 – Resultados da regressão - período Temer (MT).

	<i>Variável dependente:</i>	
	Retorno	
	(Linear)	(Global)
grau ponderado (rede de correlação)	0.129 (0.092)	
centralidade autovetor (rede de correlação)	0.860 (0.922)	
grau ponderado (rede de inf. mútua)		0.247*** (0.067)
centralidade autovetor		1.375 (0.955)
Constante	0.610*** (0.142)	0.481*** (0.118)
Observ.	91	91
R ²	0.023	0.198
R ² Ajustado	0.001	0.180
Erro padrão resid. (df = 88)	0.848	0.769
Estatística F (df = 2; 88)	1.032	10.848***
Nota: *p<0.1; **p<0.05; ***p<0.01		

média da AGM diminuiu 18% para o período MT e apenas 12,5% para o período DR. Enquanto que o grau médio permaneceu o mesmo nestas redes, o coeficiente de robustez, calculado a partir de Equação (2.31), aumentou em 27% durante o período MT (durante DR, eleva-se em 5%), enquanto que o valor estimado do parâmetro α da lei de potência caiu de 4,19 a 2,39 durante o período MT (para o período DR diminuiu de 3,03 para 2,44), uma mudança significativa. Mesmo que ambas as redes baseadas em informação mútua mostrem um bom ajuste à lei de potência, ao observarmos o período DR, alguns pontos na cauda da distribuição diminuem mais rápido do que o comportamento linear, enquanto esta queda no período MT é muito mais lenta. Por outro lado, as redes via correlação linear mostram um bom ajuste pela distribuição log-normal. Podemos dizer que as redes obtidas através da métrica de informação mútua são mais parecidas com uma rede livre de escala do que as formadas a partir da correlação linear, para ambos os períodos.

Já a parte (b) da Tabela 7.4 contém algumas características qualitativas das AGMs que são relevantes, por exemplo, a relação entre centralidade e o retorno das ações. Ambas as redes baseadas em correlação linear e informação mútua (dependência global) sugerem que os retornos dos ativos são dependentes da centralidade, como se vê na Tabela 7.1. Esse comportamento fica ainda mais evidente para redes baseadas em informação mútua. Além disso, como mostrado

nas Figs. 7.1 e 7.2, para verificarmos as mudanças individuais dentro das redes, utilizamos o índice de Sharpe para descrever os melhores ativos em relação à sua taxa de retorno pela sua variabilidade. Enquanto em redes via correlação linear todos os melhores ativos do índice de Sharpe são difusos ao redor da árvore, quando analisamos as redes via informação mútua, observamos uma grande mudança. Durante o período de DR, os ativos com os maiores índices de Sharpe aparecem na periferia do gráfico, enquanto isso no período de MT, essas mesmas ações agora se mostram na parte central. Durante o período DR, estes ativos centrais ‘moveram’ o mercado para um retorno negativo de 42%, enquanto aqueles com os melhores índices de Sharpe permaneceram na periferia. Durante o período MT, os ativos centrais agora empurraram o mercado para um retorno positivo de 50%. Neste período, os melhores índices de Sharpe se concentraram na parte central do gráfico. Isto realmente sugere que os ativos centrais são os que realmente ‘movem’ o mercado, ou seja, têm maior risco durante uma transmissão de volatilidade. Por transmissão de volatilidade, entendemos que uma ação muito central (Hubs) tem dependência geral maior do que outros ativos menos centrais, assim, é de se supor que uma vez aquele Hub atingido por certa volatilidade, este a espalhe por suas ligações, causando um efeito cascata.

A parte (c) da Tabela 7.4 mostra a importância dos setores econômicos dentro das redes usando suas centralidades. Confirmamos a estabilidade dos bancos nas análises, sendo o setor com maior centralidade em todas as redes, enquanto os setores da indústria, construção e energia estão muito agrupados na periferia da rede, indicando que eles não são muito importantes em termos de transmissão de volatilidade.

Por fim, a Tabela 7.5 apresenta um resumo da análise em relação aos setores mais evidenciados de cada rede elaborada, onde destacamos a importância do setor financeiro e de serviços como essenciais ‘transmissores de crise’, uma vez que apresentam altas centralidades em todas as redes. Já o setor elétrico e da indústria parecem estar à revelia dessas transmissões, atuando mais como setores independentes.

7.1 Comentários

Podemos concluir que, tanto no período MT quanto no período DR, as redes sofreram mudanças estruturais quando elaboradas a partir do coeficiente global, derivado da informação mútua. Mesmo que ambas as redes via IM sejam de certa forma similares entre si, ocorre que a intensidade da diferença nas redes para o período MT foi muito maior pelo fato da estrutura de dependência linear estar bastante fraca do que aquela para o período DR. Isto pudemos comprovar através dos esquemas das Figura 7.3. Em suma, conclui-se que os retornos dos ativos no período entre o afastamento da presidente Dilma e à gestão Temer parecem ter maior dependência não-linear quando comparados aos retornos do período anterior. Dessa maneira, a

Tabela 7.4 – Principais resultados

<i>Período</i>	<i>Dilma Rousseff (DR)</i>		<i>Michel Temer (MT)</i>	
<i>Dependência</i>	<i>Linear</i>	<i>Global</i>	<i>Linear</i>	<i>Global</i>
(a) Métricas das Árvores				
<i>Grau médio</i>	1.98	1.98	1.99	1.99
<i>Distância média da AGM</i>	0.64	0.56	0.66	0.54
<i>Parâmetro α estimado</i>	3.03	2.44	4.19	2.39
<i>Parâmetro de robustez</i>	0.62	0.65	0.56	0.71
(b) Características das Árvores				
<i>Centralidade vs. retornos</i>	Alta centralidade indica baixos retornos, relação esta mais clara nas redes via dependência global.		Alta centralidade indica bons retornos, relação esta mais clara nas redes via dependência global.	
<i>Ajuste da distribuição de grau</i>	Próxima a uma log-normal	Cauda próxima a uma lei de potência	Próxima a uma log-normal	Cauda próxima a uma lei de potência
<i>Dispersão do índice de Sharpe</i>	Difuso na rede	Periférico na rede	Difuso na rede	Central na rede
(c) Importantes setores econômicos				
<i>Maiores centralidades (o maior hub)</i>	Bancos (Bancos Bradesco)	Bancos (Banco Itaú)	Bancos (Banco Itaú)	Bancos (Banco Itaú)
<i>Menores centralidades (o mais periférico)</i>	Indústria (EZTEC S.A.)	Indústria (WEG S.A.)	Indústria (Gafisa S.A.)	Energia (Equatorial S.A.)

análise do coeficiente de robustez mostra que a rede para o período MT é a que se mostra mais arriscada em termos de estrutura de 'transmissão de volatilidades'. Finalmente, a análise da rede via informação mútua parece trazer benefícios aos investidores quanto à tomada de decisão de investimento, notadamente no período MT, quando confrontamos a estrutura de centralidade da rede e o desempenho das ações.

Conforme vimos no Capítulo 5, vários autores têm salientado a importância de dependências não lineares no estudo das interrelações no mercado de ações. Nossos resultados corroboram estudos recentes (MARTI et al., 2017), indicando que medidas de dependência mais gerais, como a informação mútua, deveriam ser mais utilizadas nos estudos dos mercados financeiros. Essa importância vem do fato de que, apesar da dificuldade na compreensão da lei do movimento dessas dependências nas trajetórias das séries temporais do mercado, estas sujeições não lineares fazem com que o sistema exiba aumento de volatilidade (SCHEINKMAN; LEBARON, 1989). Para a situação estudada nesta pesquisa, verificamos que os ativos que sofreram alta volatilidade (ex. OIBR3 e OIBR4 no período DR, e GOLL4, RUMO3 e ELET3 no período MT) apresentaram diferenças significativas entre o coeficiente linear e o coeficiente global de dependência, exibindo

Tabela 7.5 – Descrição dos setores acionários com maiores e menores centralidades (auto-vetor) para cada rede explorada. Destaque para o setor financeiro e serviços, bem como energia elétrica e indústria pesada, respectivamente com as maiores e menores centralidades em ambas metodologias adotadas.

	Dilma				Temer			
	Correlação	Info. mútua	nº	%	Correlação	Info. mútua	nº	%
Maiores centralidades								
Energia Elétrica	1	-	1	3%	-	-	0	0%
Financeiro	3	6	9	30%	6	5	11	37%
Concessionárias (outras)	1	-	1	3%	1	2	3	10%
Construção Civil	3	1	4	13%	1	-	1	3%
Varejo	2	-	2	7%	-	-	0	0%
Frigoríficos	-	-	0	0%	-	1	1	3%
Indústria pesada	1	3	4	13%	1	1	2	7%
Logística (serviços)	-	-	0	0%	-	2	2	7%
Mineração	-	1	1	3%	-	-	0	0%
Petróleo	-	-	0	0%	2	2	4	13%
Serviços (outros)	4	4	8	27%	4	2	6	20%
Menores centralidades								
Energia Elétrica	1	3	4	13%	2	5	7	23%
Financeiro	-	1	1	3%	-	-	0	0%
Concessionárias (outras)	-	1	1	3%	1	-	1	3%
Construção Civil	3	3	6	20%	4	1	5	17%
Varejo	1	-	1	3%	2	1	3	10%
Frigoríficos	1	1	2	7%	-	1	1	3%
Indústria pesada	3	4	7	23%	3	4	7	23%
Petróleo	2	1	3	10%	-	-	0	0%
Serviços (outros)	4	1	5	17%	2	3	5	17%
TI	-	-	0	0%	1	-	1	3%

maior centralidade nas redes de dependências não lineares via informação mútua.

Além disso, verificamos uma grande mudança na estrutura geral da rede no período de transição (MT). Este período exibiu maior diferença entre correlação linear e informação mútua, além de características de distribuição de grau próxima a uma lei de potência nas caudas, como estudado por Kwapień e Drozd (2012), e uma relação mais clara entre centralidade e desempenho das ações, conforme estudos de Matteo, Pozzi e Aste (2010).

Os resultados apresentados mostram também uma relação entre ações da carteira ótima e estrutura da rede via dependência não linear. Enquanto que para períodos de desempenho negativo, as ações da carteira ótima estão nas periferias da estrutura, para períodos de desempenho positivo, estas mesmas ações serão mais centrais na rede. Resultados parcialmente semelhantes foram encontrados em Matteo, Pozzi e Aste (2010), uma vez que estes autores se limitaram a redes via correlação linear.

Para complementar, a Tabela 7.5 apresentou uma classificação dos setores acionários com as maiores e menores centralidades para cada rede explorada nos períodos DR e MT. Notou-

se a grande dominância do setor financeiro e o alto nível de agrupamento periférico nos setores industriais e elétrico. Um resultado semelhante foi obtido em estudos anteriores no mercado brasileiro (TABAK; SERRA; CAJUEIRO, 2010), porém apenas baseado em correlação linear. Pudemos analisar que pequenas quantidades de altas correlações lineares são suficientes para organizarem as ligações da rede, separando os setores do mercado, especialmente o bancário, o industrial e o elétrico. Neste caso, altas correlações lineares serão acompanhadas, também, por altos valores de informação mútua. Assim, estas estruturas setoriais não sofrem grandes alterações quando utilizado um modelo via informação mútua. Ou seja, para estes casos, a dependência será majoritariamente linear.

CAPÍTULO 8

CONCLUSÕES

Neste trabalho, comparamos a estrutura de redes do mercado brasileiro de ações obtida a partir de duas medidas de dependência, o coeficiente de correlação de Pearson e o coeficiente global de dependência, definido a partir da Informação Mútua. O objetivo foi verificar a presença de dependências não lineares no mercado brasileiro de ações, já que estudos baseados apenas na correlação de Pearson capturam apenas as dependências lineares. Nos capítulos iniciais, estudamos um pouco da Teoria de Redes Complexas, para que pudéssemos analisar as redes financeiras, a partir daí, nos enveredamos para a análise de dependência de modo mais formal, incluindo as diversas métricas para medi-la, e finalmente nos aprofundamos em uma destas, a informação mútua. Após esta primeira parte, para compor a matriz de dependência, e finalmente as estruturas em rede, analisamos dados de alta frequência de 91 ações do Ibovespa para dois períodos diferentes, um no período do exercício da presidência Dilma Rousseff (DR), e outro a partir da conturbada troca de governo, quando assumi o presidente, então interino, Michel Temer (MT). No estudo destas redes vimos

- que a transição de presidentes e a conturbada crise econômico-política pareceu ter feito com que as ações exibissem certos tipos de volatilidades acompanhadas de relações não lineares entre elas.
- o conceito de 'transmissão de volatilidades', aspecto bastante importante na análise dessas estruturas para o mercado de ações, pois fortes dependências podem induzir a uma maior propagação do risco.
- diferenças em suas estruturas gerais, quando comparadas as redes via correlação linear e

informação mútua, mais notadamente no período MT. Conseguimos extrair essa característica a partir da estrutura da árvore geradora mínima e de suas métricas, como centralidade, coeficiente de robustez e do parâmetro da lei de potência. Também utilizamos o método de seleção de carteiras eficientes para demonstrar como as estruturas se modificam.

- alguma relação entre centralidade e desempenho da ação, notadamente no período MT, quando utilizamos dados da rede baseada na informação mútua.
- a identificação dos setores que desempenham características que os tornam muito importantes para o mercado financeiro brasileiro, visíveis tanto nas redes via correlação linear quanto nas via informação mútua.

Conclui-se que a análise da rede no mercado financeiro via informação mútua parece poder trazer benefícios aos investidores quanto à tomada de decisão de investimento, particularmente útil na análise da estrutura da rede e desempenho das ações.

8.1 Limitações e pesquisas futuras

Algumas limitações dessa pesquisa podem ser destacadas:

1. o método de estimação da informação mútua é de certa forma sujeito a vieses de estimação. Quando da estimação não paramétrica, notadamente pelo método do kernel, a calibração da janela que faz a função de suavização da função deve ser feita caso a caso. Isto não foi possível em nosso caso, pois tínhamos de iterar milhares de casos para a formação da matriz de dependência. Dessa forma, tivemos que usar a 'regra de bolso' da janela gaussiana. Dentre todos os métodos que testamos, esta foi a que tivemos mais confiança nos resultados da simulação.
2. não calculamos erros de estimativas e intervalo de confiança.
3. tivemos que nos ater a um período muito recente da base de dados das ações. Isto porque estes dados de alta frequência teve início somente em 2013 no site da Bovespa.
4. a topologia da rede via Árvore Geradora Mínima costuma restringir algumas análises. Por exemplo, podemos perder ligações que realmente eram importantes, pois o algoritmo de Kruskal é do tipo 'guloso' ¹.

¹ Algoritmo guloso é uma técnica de projeto de algoritmos que tenta resolver o problema fazendo a escolha localmente ótima em cada fase com a esperança de encontrar um ótimo global.

Por fim, nossas considerações sobre as pesquisas futuras recaem sobre nossas próprias limitações. Em relação ao primeiro item, verifica-se a necessidade de testar outras funções kernel, bem como outros tipos de janelas (talvez baseadas em outros métodos empíricos). Também pensamos ser importante introduzir estimativas de erros e intervalo de confiança para as métricas de dependência. Outra forma de análise seria replicar este estudo para outros dados, e talvez para outros mercados, como o de moedas. Uma última menção seria tornar este estudo em tempo-real e de maneira interativa, de modo que investidores pudessem acompanhar a evolução das estruturas de rede. Por enquanto ainda é um desafio, dada a complexidade computacional exigida para o cálculo da informação mútua.

A análise de redes complexas é uma ciência que se desenvolve rapidamente, e não seria de se estranhar que tenhamos métodos muito mais rápidos, flexíveis e interativos de análise e visualização de dados em pouco tempo. Esperamos também uma maior dedicação da área de ciências sociais aplicadas ao desenvolvimento de tais pesquisas.

8.2 Antifrágeis

Uma das frases que talvez possa definir o propósito desse trabalho é a que está na epígrafe. Se o leitor não a viu, reescrevo-a:

“Some things benefit from shocks; they thrive and grow when exposed to volatility, randomness, disorder, and stressors and love adventure, risk, and uncertainty”².

Esta colocação é sobre a definição do título do livro de Nassim Nicholas Taleb, denominado "Antifrágil"(2014). Pelo que mostramos ao longo desse estudo, podemos caracterizar o mercado financeiro, em especial o acionário, por esta lista de substantivos, tais como volatilidade, aleatoriedade, desordem e incerteza. Esses termos 'fogem' da linearidade e do controle de riscos gaussiano. Este 'gosto pela aventura' deve ser estudado de forma mais realística, e um dos métodos para tratá-lo é através de métricas como Entropia e Informação Mútua, de tal forma aliadas a estruturas mais complexas de visualização, algumas das quais estudadas pela Teoria de Redes Complexas e seus mapas de relacionamentos. Se a chave determinante para estimar modelos de risco para o mercado é a estrutura de dependência entre todos seus ativos, parece que ter em mãos um mais acurado mapa já é um ótimo começo.

² Algumas coisas se beneficiam de choques; eles prosperam e crescem quando expostos à volatilidade, aleatoriedade, desordem e estressores, e amam aventura, risco e incerteza - tradução do autor

Página intencionalmente deixada em branco.

REFERÊNCIAS

ANSCOMBE, F. J. Graphs in statistical analysis. **The American Statistician**, [American Statistical Association, Taylor Francis, Ltd.], v. 27, n. 1, p. 17–21, 1973. ISSN 00031305. Disponível em: <http://www.jstor.org/stable/2682899>.

BARABASI, AL; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n. 5439, p. 509–512, oct 1999. Disponível em: <http://dx.doi.org/10.1126/science.286.5439.509>.

BARABASI, Albert-László. **Network Science**. Cambridge CB2 8BS, United Kingdom: Cambridge University Press, 2016. 456 p. ISBN 978-1-107-07626-6.

BARABASI, Albert-László; GULBAHCE, Natali; LOSCALZO, Joseph. Network medicine: a network-based approach to human disease. **Nat Rev Genet**, v. 12, n. 1, p. 56–68, jan 2011. Disponível em: <http://dx.doi.org/10.1038/nrg2918>.

BARRAT, Alain; BARTHELEMY, Marc; VESPIGNANI, Alessandro. **Dynamical Processes on Complex Networks**. Cambridge: Cambridge University Press, 2008. ISBN 9780511791383. Disponível em: <http://ebooks.cambridge.org/ref/id/CBO9780511791383>.

BEKIROU, Stelios et al. Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. **European Journal of Operational Research**, jul 2016. ISSN 03772217. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0377221716304921>.

BIANCONI, G; BARABASI, A L. Bose-einstein condensation in complex networks. **Phys Rev Lett**, v. 86, n. 24, p. 5632–5635, jun 2001. Disponível em: <http://dx.doi.org/10.1103/PhysRevLett.86.5632>.

BLONDEL, Vincent D et al. Fast unfolding of communities in large networks. **J. Stat. Mech.**, v. 2008, n. 10, p. P10008, oct 2008. ISSN 1742-5468. Disponível em: <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52>.

BOGINSKI, Vladimir; BUTENKO, Sergiy; PARDALOS, Panos M. Statistical analysis of financial networks. **Comput Stat Data Anal**, v. 48, n. 2, p. 431–443, feb 2005. ISSN 01679473. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0167947304000258>.

BONANNO, G. et al. Networks of equities in financial markets. **The European Physical Journal B - Condensed Matter**, v. 38, n. 2, p. 363–371, mar 2004. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1140/epjb/e2004-00129-6>.

BONANNO, Giovanni; LILLO, Fabrizio; MANTEGNA, Rosario N. Levels of complexity in financial markets. **Physica A: Statistical Mechanics and its Applications**, v. 299, n. 1-2, p. 16–27, oct 2001. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437101002795>.

BRIDA, Juan Gabriel; RISSO, Wiston Adrián. Multidimensional minimal spanning tree: The dow jones case. **Physica A: Statistical Mechanics and its Applications**, v. 387, n. 21, p. 5205–5210, sep 2008. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437108004299>.

CALDARELLI, Guido; VESPIGNANI, Alessandro. **Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science**. WORLD SCIENTIFIC, 2007. v. 2. (Complex Systems and Interdisciplinary Science, v. 2). ISSN 1793-4540. ISBN 978-981-270-664-5. Disponível em: <http://www.worldscientific.com/worldscibooks/10.1142/6455>.

CHIU, Shean-Tsong. Bandwidth selection for kernel density estimation. **Ann. Statist.**, v. 19, n. 4, p. 1883–1905, dec 1991. ISSN 0090-5364. Disponível em: <http://projecteuclid.org/euclid.aos/1176348376>.

COELHO, Ricardo et al. The evolution of interdependence in world equity markets from minimum spanning trees. **Physica A: Statistical Mechanics and its Applications**, v. 376, p. 455–466, mar 2007. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437106010624>.

CORMEN, Thomas et al. **Introduction to Algorithms, Second Edition**. [S.l.]: Cambridge, Massachusetts London, England, 2001.

COVER, Thomas M.; THOMAS, Joy A. **Elements of Information Theory**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005. ISBN 9780471241959. Disponível em: <http://doi.wiley.com/10.1002/047174882X>.

DARBELLAY, G.A.; VAJDA, I. Estimation of the information by an adaptive partitioning of the observation space. **IEEE Trans. Inform. Theory**, v. 45, n. 4, p. 1315–1321, may 1999. ISSN 00189448. Disponível em: <http://ieeexplore.ieee.org/document/761290/>.

DARBELLAY, Georges A; WUERTZ, Diethelm. The entropy as a tool for analysing statistical dependences in financial time series. **Physica A: Statistical Mechanics and its Applications**, v. 287, n. 3-4, p. 429–439, dec 2000. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437100003824>.

DIONISIO, Andreia; MENEZES, Rui; MENDES, Diana A. Mutual information: a measure of dependency for nonlinear time series. **Physica A: Statistical Mechanics and its Applications**, v. 344, n. 1-2, p. 326–329, dec 2004. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437104009598>.

- DOBSON, Ian et al. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. **Chaos**, v. 17, n. 2, jun 2007. Disponível em: <<http://dx.doi.org/10.1063/1.2737822>>.
- ECONOMATICA. 2016. <<https://economica.com/>>. Acesso: nov, 2016.
- ENGLE, Robert. **Anticipating Correlations: A New Paradigm for Risk Management**. Princeton: Princeton University Press, 2009. ISBN 9781400830190. Disponível em: <<http://www.degruyter.com/view/books/9781400830190/9781400830190/9781400830190.xml>>.
- ERDŐS, Paul; RÉNYI, Alfréd. On the evolution of random graphs . **Publications of the Mathematical Institute of the Hungarian Academy Science**, in print, n. 5, 1960.
- FAMA, Eugene F. Efficient capital markets: A review of theory and empirical work. **J Finance**, v. 25, n. 2, p. 383, may 1970. ISSN 00221082. Disponível em: <<http://www.jstor.org/stable/2325486?origin=crossref>>.
- FIEDOR, Paweł. Networks in financial markets based on the mutual information rate. **Phys Rev E Stat Nonlin Soft Matter Phys**, v. 89, n. 5, p. 052801, may 2014. Disponível em: <<http://dx.doi.org/10.1103/PhysRevE.89.052801>>.
- FRANZ, Max et al. Cytoscape.js: a graph theory library for visualisation and analysis. **Bioinformatics**, v. 32, n. 2, p. 309–311, jan 2016. Disponível em: <<http://dx.doi.org/10.1093/bioinformatics/btv557>>.
- FRASER, AM; SWINNEY, HL. Independent coordinates for strange attractors from mutual information. **Phys Rev A Gen Phys**, v. 33, n. 2, p. 1134–1140, feb 1986. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9896728>>.
- GILLESPIE, Colin S. Fitting heavy tailed distributions: The powerLaw package. **Journal of Statistical Software**, v. 64, n. 2, p. 1–16, 2015. Disponível em: <<http://www.jstatsoft.org/v64/i02/>>.
- GILMORE, Claire G.; LUCEY, Brian M.; BOSCIA, Marian W. Comovements in government bond markets: A minimum spanning tree analysis. **Physica A: Statistical Mechanics and its Applications**, v. 389, n. 21, p. 4875–4886, nov 2010. ISSN 03784371. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0378437110006059>>.
- GRANGER, Clive; LIN, Jin-Lung. Using the mutual information coefficient to identify lags in nonlinear models. **J Time Series Analysis**, v. 15, n. 4, p. 371–384, jul 1994. ISSN 0143-9782. Disponível em: <<http://doi.wiley.com/10.1111/j.1467-9892.1994.tb00200.x>>.
- GULKO, LES. The entropy theory of stock option pricing. **Int. J. Theor. Appl. Finan.**, v. 02, n. 03, p. 331–355, jul 1999. ISSN 0219-0249. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S0219024999000182>>.
- HEIBERGER, Raphael H. Stock network stability in times of crisis. **Physica A: Statistical Mechanics and its Applications**, v. 393, p. 376–381, jan 2014. ISSN 03784371. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0378437113008030>>.
-

HOLLANDER, Myles; WOLFE, Douglas A.; CHICKEN, Eric. **Nonparametric Statistical Methods**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2015. ISBN 9780470387375. Disponível em: <<http://doi.wiley.com/10.1002/9781119196037>>.

JACOMY, Mathieu et al. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. **PLoS ONE**, v. 9, n. 6, p. e98679, jun 2014. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0098679>>.

JOE, Harry. Relative entropy measures of multivariate dependence. **J Am Stat Assoc**, v. 84, n. 405, p. 157–164, mar 1989.

_____. **Multivariate Models and Multivariate Dependence Concepts**. Chapman and Hall/CRC, 1997. v. 19970691. (C&H/CRC Monographs on Statistics & Applied Probability, v. 19970691). ISSN 2154-3038. ISBN 978-0-412-07331-1. Disponível em: <<http://www.crcnetbase.com/doi/book/10.1201/b13150>>.

JUNIOR, Leonidas; MULLOKANDOV, Asher; KENETT, Dror. Dependency relations among international stock market indices. **JRFM**, v. 8, n. 2, p. 227–265, may 2015. ISSN 1911-8074. Disponível em: <<http://www.mdpi.com/1911-8074/8/2/227/>>.

KAYA, Hakan. Eccentricity in asset management. **Journal of Network Theory in Finance**, v. 1, p. 1–31, 2015.

KINNEY, Justin B; ATWAL, Gurinder S. Equitability, mutual information, and the maximal information coefficient. **Proc Natl Acad Sci U S A**, v. 111, n. 9, p. 3354–3359, mar 2014. Disponível em: <<http://dx.doi.org/10.1073/pnas.1309933111>>.

KOLACZYK, Eric D.; CSÁRDI, Gábor. BOOK. **Statistical Analysis of Network Data with R**. New York, NY: Springer New York, 2014. v. 65. (Use R!, v. 65). ISSN 2197-5736. ISBN 978-1-4939-0982-7. Disponível em: <<http://link.springer.com/10.1007/978-1-4939-0983-4>>.

KOWALSKI, Charles J. Non-normal bivariate distributions with normal marginals. **The American Statistician**, [American Statistical Association, Taylor Francis, Ltd.], v. 27, n. 3, p. 103–106, 1973. ISSN 00031305. Disponível em: <<http://www.jstor.org/stable/2683630>>.

KRASKOV, Alexander; GRASSBERGER, Peter. Mic: Mutual information based hierarchical clustering. In: EMMERT-STREIB, Frank; DEHMER, Matthias (Ed.). **Information Theory and Statistical Learning**. Boston, MA: Springer US, 2009. p. 101–123. ISBN 978-0-387-84815-0. Disponível em: <http://link.springer.com/10.1007/978-0-387-84816-7_5>.

KRASKOV, Alexander; STÖGBAUER, Harald; GRASSBERGER, Peter. Estimating mutual information. **Phys Rev E Stat Nonlin Soft Matter Phys**, v. 69, n. 6 Pt 2, p. 066138, jun 2004. Disponível em: <<http://dx.doi.org/10.1103/PhysRevE.69.066138>>.

KRUSKAL, Joseph B. On the shortest spanning subtree of a graph and the traveling salesman problem. **Proc. Amer. Math. Soc.**, v. 7, n. 1, p. 48–48, jan 1956. ISSN 0002-9939. Disponível em: <<http://www.ams.org/jourcgi/jour-getitem?pii=S0002-9939-1956-0078686-7>>.

KWAPIEN, Jarosław; DROZDZ, Stanisław. Physical approach to complex systems. **Physics Reports**, v. 515, n. 3-4, p. 115–226, jun 2012. ISSN 03701573. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0370157312000166>>.

MAI, Jan-Frederik; SCHERER, Matthias. **Financial Engineering with Copulas Explained**. London: Palgrave Macmillan UK, 2014. ISBN 978-1-137-34630-8. Disponível em: <http://link.springer.com/10.1057/9781137346315>.

MANTEGNA, R.N. Hierarchical structure in financial markets. **Eur. Phys. J. B**, v. 11, n. 1, p. 193–197, sep 1999. ISSN 1434-6028. Disponível em: <http://link.springer.com/10.1007/s100510050929>.

MANTEGNA, Rosario N.; STANLEY, H. Eugene. **Introduction to Econophysics: Correlations and Complexity in Finance**. Cambridge: Cambridge University Press, 1999. ISBN 9780511755767. Disponível em: <http://ebooks.cambridge.org/ref/id/CBO9780511755767>.

MARI, Dominique Drouet; KOTZ, Samuel. **Correlation and Dependence**. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., 2001. ISBN 978-1-86094-264-8. Disponível em: <http://www.worldscientific.com/worldscibooks/10.1142/p226>.

MARTI, G. et al. A review of two decades of correlations, hierarchies, networks and clustering in financial markets. **ArXiv e-prints**, mar. 2017.

MATTEO, T. Di; POZZI, F.; ASTE, T. The use of dynamical networks to detect the hierarchical organization of financial market sectors. **Eur. Phys. J. B**, v. 73, n. 1, p. 3–11, jan 2010. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/index/10.1140/epjb/e2009-00286-0>.

MEYER, Paul. **Introductory Probability and Statistical Applications**. PUBLISHED BY ADDISON-WESLEY PUBLISHING COMPANY, INC., 1969. Disponível em: <http://www.worldscientific.com/worldscibooks/10.1142/p226>.

MICCICHÈ, Salvatore et al. Degree stability of a minimum spanning tree of price return and volatility. **Physica A: Statistical Mechanics and its Applications**, v. 324, n. 1-2, p. 66–73, jun 2003. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437103000025>.

MOON, YI; RAJAGOPALAN, B; LALL, U. Estimation of mutual information using kernel density estimators. **Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics**, v. 52, n. 3, p. 2318–2321, sep 1995. Disponível em: <http://dx.doi.org/10.1103/PhysRevE.52.2318>.

NAIMY, Viviane Y. Gaussian copula vs. loans loss assessment: A simplified and easy-to-use model. **JBCS**, v. 8, n. 5, p. 533, sep 2012. ISSN 2157-8826. Disponível em: <http://cluteinstitute.com/ojs/index.php/JBCS/article/view/7252>.

NEWMAN, Mark. **Networks**. Oxford University Press, 2010. ISBN 9780199206650. Disponível em: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650>.

NEWMAN, M. E. J. The structure and function of complex networks. **SIAM Rev.**, v. 45, n. 2, p. 167–256, jan 2003. ISSN 0036-1445. Disponível em: <http://epubs.siam.org/doi/abs/10.1137/S003614450342480>.

- ONNELA, J.-P.; KASKI, K.; KERTSZ, J. Clustering and information in correlation based financial networks. **The European Physical Journal B - Condensed Matter**, v. 38, n. 2, p. 353–362, mar 2004. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1140/epjb/e2004-00128-7>.
- PALLA, Gergely; BARABÁSI, Albert-László; VICSEK, Tamás. Quantifying social group evolution. **Nature**, v. 446, n. 7136, p. 664–667, apr 2007. Disponível em: <http://dx.doi.org/10.1038/nature05670>.
- PAUL, G. et al. Optimization of robustness of complex networks. **Eur. Phys. J. B**, v. 38, n. 2, p. 187–191, mar 2004. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/index/10.1140/epjb/e2004-00112-3>.
- PERON, Thomas Kaue Dal'Maso; COSTA, Luciano da Fontoura; RODRIGUES, Francisco A. The structure and resilience of financial market networks. **Chaos**, v. 22, n. 1, p. 013117, mar 2012. Disponível em: <http://dx.doi.org/10.1063/1.3683467>.
- POZZI, F.; MATTEO, T Di; ASTE, T. Spread of risk across financial markets: better to invest in the peripheries. **Sci Rep**, v. 3, p. 1665, 2013. Disponível em: <http://dx.doi.org/10.1038/srep01665>.
- RACHEV, Svetlozar T. et al. **Probability and Statistics for Finance**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010. ISBN 9780470400937. Disponível em: <http://doi.wiley.com/10.1002/9781118267912>.
- RADICCHI, Filippo et al. Defining and identifying communities in networks. **Proc Natl Acad Sci U S A**, v. 101, n. 9, p. 2658–2663, mar 2004. ISSN 0027-8424. Disponível em: <http://dx.doi.org/10.1073/pnas.0400054101>.
- RAPOPORT, Anatol; HORVATH, William J. A study of a large sociogram. **Syst. Res.**, v. 6, n. 4, p. 279–291, jan 1961. ISSN 00057940. Disponível em: <http://doi.wiley.com/10.1002/bs.3830060402>.
- RAVASZ, Erzsébet; BARABÁSI, Albert-László. Hierarchical organization in complex networks. **Phys Rev E Stat Nonlin Soft Matter Phys**, v. 67, n. 2 Pt 2, p. 026112, feb 2003. ISSN 1063-651X. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.67.026112>.
- RIBEIRO, E.M.S.; PRATAVIERA, G.A. Information theoretic approach for accounting classification. **Physica A: Statistical Mechanics and its Applications**, v. 416, p. 651–660, dec 2014. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437114007729>.
- ROSVALL, Martin; BERGSTROM, Carl T. An information-theoretic framework for resolving community structure in complex networks. **Proc Natl Acad Sci U S A**, v. 104, n. 18, p. 7327–7331, may 2007. Disponível em: <http://dx.doi.org/10.1073/pnas.0611034104>.
- SCELLATO, Salvatore et al. Understanding robustness of mobile networks through temporal network measures. In: **2011 Proceedings IEEE INFOCOM**. IEEE, 2011. p. 1–5. ISBN 978-1-4244-9919-9. Disponível em: <http://ieeexplore.ieee.org/document/5935006/>.
-

SCHEINKMAN, Jose A.; LEBARON, Blake. Nonlinear dynamics and stock returns. **J Bus**, v. 62, n. 3, p. 311, jan 1989. ISSN 0021-9398. Disponível em: <http://www.jstor.org/stable/2353350>.

SCOTT, David W. **Multivariate Density Estimation: Theory, Practice, and Visualization**. Hoboken, NJ: John Wiley & Sons, Inc, 2015. (Wiley Series in Probability and Statistics). ISSN 19406347. ISBN 9780471697558. Disponível em: <http://doi.wiley.com/10.1002/9781118575574>.

SENSOY, Ahmet; TABAK, Benjamin M. Dynamic spanning trees in stock market networks: The case of asia-pacific. **Physica A: Statistical Mechanics and its Applications**, v. 414, p. 387–402, nov 2014. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437114006517>.

SENSOY, A.; YUKSEL, S.; ERTURK, M. Analysis of cross-correlations between financial markets after the 2008 crisis. **Physica A: Statistical Mechanics and its Applications**, v. 392, n. 20, p. 5027–5045, oct 2013. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437113005645>.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, n. 4, p. 623–656, oct 1948. ISSN 00058580. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773067>.

SHEATHER, Simon J. Density estimation. **Stat Sci**, v. 19, n. 4, p. 588–597, nov 2004. ISSN 0883-4237. Disponível em: <http://projecteuclid.org/Dienst/getRecord?id=euclid.ss/1113832723>.

SILVA, Thiago Christiano; ZHAO, Liang. **Machine Learning in Complex Networks**. Cham: Springer International Publishing, 2016. ISBN 978-3-319-17289-7. Disponível em: <http://link.springer.com/10.1007/978-3-319-17290-3>.

SILVERMAN, B. W. **Density Estimation for Statistics and Data Analysis**. Boston, MA: Springer US, 1986. ISBN 978-0-412-24620-3. Disponível em: <http://link.springer.com/10.1007/978-1-4899-3324-9>.

SOLOMONOFF, Ray; RAPOPORT, Anatol. Connectivity of random nets. **Bull Math Biophys**, v. 13, n. 2, p. 107–117, jun 1951. ISSN 0007-4985. Disponível em: <http://link.springer.com/10.1007/BF02478357>.

SORNETTE, Didier; MALEVERGNE, Yannick. **Extreme Financial Risks - From Dependence to Risk Management**. Berlin/Heidelberg: Springer-Verlag, 2006. ISBN 3-540-27264-X. Disponível em: <http://link.springer.com/10.1007/b138841>.

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**, v. 389, n. 16, p. 3240–3249, aug 2010. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437110002992>.

TRAVERS, Jeffrey; MILGRAM, Stanley. An experimental study of the small world problem. **Sociometry**, v. 32, n. 4, p. 425, dec 1969. ISSN 00380431. Disponível em: <http://www.jstor.org/stable/2786545?origin=crossref>.

TSE, Chi K.; LIU, Jing; LAU, Francis C.M. A network perspective of the stock market. **Journal of Empirical Finance**, v. 17, n. 4, p. 659–667, sep 2010. ISSN 09275398. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0927539810000368>.

TUMMINELLO, Michele; LILLO, Fabrizio; MANTEGNA, Rosario N. Correlation, hierarchies, and networks in financial markets. **J Econ Behav Organ**, v. 75, n. 1, p. 40–58, jul 2010. ISSN 01672681. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0167268110000077>.

TUMMINELLO, M. et al. Correlation based networks of equity returns sampled at different time horizons. **Eur. Phys. J. B**, v. 55, n. 2, p. 209–217, jan 2007. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/index/10.1140/epjb/e2006-00414-4>.

WALTERS-WILLIAMS, Janett; LI, Yan. Estimation of mutual information: A survey. In: WEN, Peng et al. (Ed.). **Rough Sets and Knowledge Technology**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, (Lecture Notes in Computer Science, v. 5589). p. 389–396. ISBN 978-3-642-02961-5. Disponível em: http://link.springer.com/10.1007/978-3-642-02962-2_49.

WANG, Gang-Jin; XIE, Chi. Tail dependence structure of the foreign exchange market: A network view. **Expert Syst Appl**, v. 46, p. 164–179, mar 2016. ISSN 09574174. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0957417415007411>.

WATTS, D J; STROGATZ, S H. Collective dynamics of 'small-world' networks. **Nature**, v. 393, n. 6684, p. 440–442, jun 1998. Disponível em: <http://dx.doi.org/10.1038/30918>.

YANG, Chunxia et al. Research on the evolution of stock correlation based on maximal spanning trees. **Physica A: Statistical Mechanics and its Applications**, v. 415, p. 1–18, dec 2014. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437114006554>.

YOU, Tao; FIEDOR, Paweł; HOIDA, Artur. Network analysis of the shanghai stock exchange based on partial mutual information. **JRFM**, v. 8, n. 4, p. 266–284, jun 2015. ISSN 1911-8074. Disponível em: <http://www.mdpi.com/1911-8074/8/2/266>.

ZHANG, Yiting et al. Will the us economy recover in 2010? a minimal spanning tree study. **Physica A: Statistical Mechanics and its Applications**, v. 390, n. 11, p. 2020–2050, jun 2011. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437111000847>.

ANEXO A

O ÍNDICE DE SHARPE E A FRONTEIRA EFICIENTE

No Cap. 7, utilizamos conceitos da moderna teoria de carteiras, porém não demos muitos detalhes. Pois bem, a fronteira eficiente (ou fronteira da carteira) é uma carteira de investimentos que ocupa as partes "eficientes" do espectro risco-retorno. Formalmente, é o conjunto de carteiras que satisfazem a condição de que nenhum outro portfólio existe com um retorno esperado mais alto, mas com o mesmo desvio padrão de retorno. O limite superior positivamente inclinado desta região é uma porção de uma parábola e é chamada de "fronteira eficiente", formulada pela primeira vez por Harry Markowitz em 1952. Matrizes são utilizadas nestes cálculos. A fronteira eficiente é encontrada minimizando-se a expressão

$$w^T \sum w - q * R^T w$$

onde w é um vetor peso do portfólio, e $\sum_i w = 1$, $\sum$ é a matriz de covariância para os retornos dos ativos do portfólio, $q \geq 0$ é um fator de risco, onde 0 resulta na carteira de mínima variância, e ∞ nos limites da fronteira, R é um vetor de retornos esperados, $w^T \sum w$, a variância dos retornos do portfólio, e $R^T w$ é o retorno esperado do portfólio.

Podemos definir o índice de Sharpe por

$$S = \frac{E(R_a)}{\sqrt{\text{var}(R_a)}},$$

onde $E(R_a)$ é a expectativa do retorno da carteira, e $\sqrt{\text{var}(R_a)}$ é o desvio padrão dos retornos dessa carteira. Então, quer-se um ponto x com uma maior inclinação na fronteira eficiente. Em

outras palavras, um investidor procura a carteira que maximiza esta expressão.

Por exemplo, as figuras a seguir mostram a fronteira eficiente para os períodos analisados nesse trabalho (DR, presidente Dilma, e MT, presidente Temer). Podemos ver que consegue-se um melhor retorno com variância bem menor no período MT do que no período DR (em MT, o ponto vermelho desloca-se para um ponto mais inclinado na fronteira, ou seja, não precisamos incorrer em grandes riscos para termos um bom retorno esperado; o inverso ocorre para o período DR). O leitor pode consultar o pacote do R 'fPortfolio' para mais informações.

Figura A.1 – Fronteira eficiente do portfólio de ações para o período DR (Dilma).

Figura A.2 – Fronteira eficiente do portfólio de ações para o período MT (Temer).

APÊNDICE A

MATERIAL SUPLEMENTAR

A.1 Código em R para entropia e informação mútua

Código em **R** para cálculo da entropia e informação mútua do exemplo da figura 4.9. Sejam x = vetor série temporal 1, e y = vetor série temporal 2. Assim,

```
library(MASS)
library(ks)
library(rgl)
library(iplots)
library(RColorBrewer)
library(kedd)
library(ggplot2)
library(stats)

xlim<-range(x)
ylim<-range(y)

## PARA H (X,Y)
denz<-kde2d(x,y, h=c(width.SJ(x), width.SJ(y)), n=512, lims=c(xlim,ylim))
# Kernel gaussiano e h pelo método de Sheather e Jones (2004)
z<-denz$z #extrai a densidade
z<-z[z!=0] #Seja log0 indefinido, excluimos os locais com densidade zero
cell_sizez<-(diff(xlim)/512) * (diff(ylim)/512)
#estima o tamanho de cada grade retangular. #A densidade expande em grades
#2D (neste caso com 512 partições), com densidade z em cada uma
#A integração numérica torna-se direta, quando agregadas todas as grades
```



```

normz<-sum(z)*cell_sizez #Soma da fdp da entropia, aproximando-se de 1
integrandz<--(z*log(z)) #Integrando (entropia)
entropyz<-sum(integrandz)*cell_sizez #Integral numérica por dupla soma
entropyz<-entropyz/normz #Normalizando

##PARA H(X) - Kernel gaussiano e h pelo método de Sheather e Jones (2004)
denx<-kde(x=x, h=c(width.SJ(x), gridsize = 512, xmin=xlim[1], xmax = xlim[2])
zx<-denx$estimate
zx <- zx[zx!=0]
cell_sizecx<-(diff(xlim)/512)
normx<-sum(zx)*cell_sizecx
integrandx<--(zx*log(zx))
entropyx<-sum(integrandx)*cell_sizecx
entropyx<-entropyx/normx

##PARA H(Y) - Kernel gaussiano e h pelo método de Sheather e Jones (2004)
deny<-kde(x=y, h=c(width.SJ(x), gridsize = 512, xmin=ylim[1], xmax = ylim[2])
zy<-deny$estimate
zy <- zy[zy!=0]
cell_sizecy<-(diff(ylim)/512)
normy<-sum(zy)*cell_sizecy
integrandy<--(zy*log(zy))
entropyy<-sum(integrandy)*cell_sizecy
entropyy<-entropyy/normy

##PARA IM(X,Y)
iMI <- entropyx+entropyy-entropyz #Cálculo da Informação Mútua.
CD<-sqrt(1-exp(-2*iMI)) #Cálculo do coeficiente global de Correlação.
pearson<-cor(x,y) #Correlação de Pearson.
Diff<-CD-abs(pearson) #Diferença entre CD e Pearson.

##CORES PARA PLOTS
colors <- c("blue", "red")
coll<- rainbow(length(denz$z)) [rank(denz$z)] #degradê
k <- 11
my.cols <- rev(brewer.pal(k, "RdYlBu")) #paletas de cores

##Gráficos de dispersão e densidades
plot(x, y, pch=19, main="Scatter Plot with all Observations")
iplot(x,y) #Dispersão
#Densidades
filled.contour(denz,
  color.palette=colorRampPalette(c('white','blue','yellow','red','darkred')))
contour(denz$x,denz$y,denz$z) #Contorno de densidade

```

A.2 Código em R para teste do código do Apêndice A.1

Teste do código acima para entropia de uma distribuição gaussiana bivariada não correlacionada.

```
set.seed(123); x<-rnorm(1000,0,2) #variância marginal: 4
set.seed(456); y<-rnorm(1000,0,2) #variância marginal: 4

library(MASS)
library(rgl)
library(RColorBrewer)
library(stats)

xlim<-range(x)
ylim<-range(y)

## Densidade (X,Y)
denz<-kde2d(x,y, h=c(width.SJ(x), width.SJ(y)), n=512, lims=c(xlim,ylim))
#Kernel gaussiano e h pelo método de Sheather e Jones (2004)
z<-denz$z #extraí a densidade
z<-z[z!=0] #Seja log0 indefinido, excluimos os locais com densidade zero
cell_sizez<-(diff(xlim)/512) * (diff(ylim)/512)
#estima o tamanho de cada grade retangular. #A densidade expande em grades
#2D (neste caso com 512 partições), com densidade z em cada uma
#A integração numérica torna-se direta, quando agregadas todas as grades
normz<-sum(z)*cell_sizez #Soma da fdp da entropia, aproximando-se de 1
integrandz<--(z*log(z)) #Integrando (entropia)
entropyz<-sum(integrandz)*cell_sizez #Integral numérica por dupla soma
entropyz<-entropyz/normz #Normalizando
entropyz

## [1] 4.222565

#entropia para distrib. normal bivariada: 1/2*log((2*pi*e)^2*det(sigma))
1/2*log(((2*pi*exp(1))^2)*16)

## [1] 4.224171
```

```
#Gráfico de densidade conjunta
filled.contour(denz,
  color.palette=colorRampPalette(c('white','blue','yellow','red','darkred')))
```

A.3 Transformando matrizes de correlação em grafos (redes)

O código em R abaixo mostra como transformar matrizes de correlação em redes. Neste exemplo, utilizamos a medida de dissimilaridade baseada no coeficiente linear de Pearson, $1 - |\rho|$. Vale ressaltar que podemos trocar a correlação pelo coeficiente de dependência, $1 - \lambda$.

```
library(igraph)
library(MASS)

corrdata<-cor(ativos_retornos) #Correlação de Pearson
cor_mat<-as.matrix(corrdata) #Transforma em matriz
#Transforma a matriz em grafo
graph <- graph.adjacency(1-abs(cor_mat), weighted=TRUE, mode="lower")
#Exporta grafo no formato universal .graphml
write.graph(graph, file="acoes.graphml", format="graphml")
#Plot de teste
plot(graph, layout=layout_with_kk, vertex.color="green")
#Árvore Geradora Mínima teste
mst<-mst(graph)
plot(mst)
```

A.4 Código final para cálculo da informação mútua para n vetores

Este é o código, baseado no kernel gaussiano e largura da banda h pela referência normal (Regra de Scott), que será utilizado para a análise dos resultados.

```
library(MASS)
library(ks)
library(kedd)
library(stats)

MI = function(x,y){ #cria função MI, representada por dois vetores x e y.

xlim<-range(x)
ylim<-range(y)

library("gplm")
#largura da janela h baseada na regra de Scott.
hx<-bandwidth.scott(x, kernel = "gaussian")
hx<-hx*1.06 #constante para a referência normal quando d=1.
hy<-bandwidth.scott(y, kernel = "gaussian")
hy<-hy*1.06 #constante para a referência normal quando d=1.
```

```

detach("package:gplm", unload=TRUE)

library(MASS)
library(stats)
library(ks)

## Densidade (X,Y)
denz<-kde2d(x,y, h = c(bandwidth.nrd(x), bandwidth.nrd(y)),
n=128, lims=c(xlim,ylim))
#kernel gaussiano e h pelo método da referência normal (Regra de Scott)
z<-denz$z #extraí a densidade
z<-z[z!=0] #Seja log0 indefinido, excluimos os locais com densidade zero
cell_sizez<-(diff(xlim)/128) * (diff(ylim)/128)
#estima o tamanho de cada grade retangular. #A densidade expande em grades
#2D (neste caso com 128 partições), com densidade z em cada uma
#A integração numérica torna-se direta, quando agregadas todas as grades
normz<-sum(z)*cell_sizez #Soma da fdp da entropia, aproximando-se de 1
integrandz<--(z*log(z)) #Integrando (entropia)
entropyz<-sum(integrandz)*cell_sizez #Integral numérica por dupla soma
entropyz<-entropyz/normz #Normalizando

#kernel gaussiano e h pela referência normal.
denx<-kde(x=x,h=hx,gridsize = 128, xmin=xlim[1], xmax = xlim[2])
zx<-denx$estimate
zx <- zx[zx!=0]
cell_sizex<-(diff(xlim)/128)
normx<-sum(zx)*cell_sizex
integrandx<- -zx*log(zx)
entropyx<-sum(integrandx)*cell_sizex
entropyx<-entropyx/normx

#kernel gaussiano e h pela referência normal.
deny<-kde(x=y,h=hy,gridsize = 128, xmin=ylim[1], xmax = ylim[2])
zy<-deny$estimate
zy <- zy[zy!=0]
cell_sizey<-(diff(ylim)/128)
normy<-sum(zy)*cell_sizey
integrandy<- -zy*log(zy)
entropyy<-sum(integrandy)*cell_sizey
entropyy<-entropyy/normy

iMI<-entropyx+entropyy-entropyz #Retorna a informação mútua, iMI.

if (iMI < 0) #Não existe iMI < 0
  return(0) #Se ocorrerem erros de aprox., a função retorna 0.

```

```

    else #de modo contrário...
      return(sqrt(1-exp(-2*iMI))) #retorna o cálculo correto.
  }

  i <- 1
  j <- 2
  x <- ativos_retornos[,i] #i vetor da matriz de retornos das ações.
  y <- ativos_retornos[,j] #j vetor da matriz de retornos das ações.
  mRet <- diag(100) #cria matriz nula
  for (i in 1:ncol(ativos_retornos)) { #comando for para iteração da
    #função MI entre cada vetor da matriz de retornos das ações.
    iIni <- i+1 #início do for para i.
    if (iIni <= 100) { #neste caso, temos 100 vetores.
      for(j in (i+1):ncol(ativos_retornos)) { #início do for para j.
        if (i != j) { #comando para não calcular a matriz completa.
          mRet[i,j] <- MI(ativos_retornos[,i],ativos_retornos[,j])
          #retorna resultado da função MI para dois pares de vetores i e j.
          print(mRet) #retorna a matriz baseada na função MI.
        }
      }
    }
  }
}

```

A.5 Testes de bandas

Aqui, testamos alguns cenários. Os dois primeiros cenários são elaborados a partir de uma função não linear com algum ruído aleatório, descrita a seguir em código R. Podemos verificar que o primeiro caso (h pela regra de Scott), obtivemos um coeficiente de dependência próximo a 0,86, enquanto que no segundo caso (h pelo método de Sheather e Jones), um valor próximo a 0,98. Para este caso, a banda pelo segundo método pareceu superavaliar a real estrutura de dependência, enquanto que o primeiro método a mensurou de modo mais conservador. Nota: a correlação de Pearson é 0,00269.

```

x = runif(5000, -1, 1)
x = sin(x*pi) + rnorm(5000, 0, 1/8)
y = cos(x*pi) + rnorm(5000, 0, 1/8)
plot(cbind(x,y), xlim = c(-1.5, 1.5), ylim = c(-1.5, 1.5))

```

```
## [1] 0.002696478 (Pearson)
## [1] 0.8556761 (Coeficiente de dependência)
```

```
x = runif(5000, -1, 1)
x = sin(x*pi) + rnorm(5000, 0, 1/8)
y = cos(x*pi) + rnorm(5000, 0, 1/8)
plot(cbind(x,y), xlim = c(-1.5, 1.5), ylim = c(-1.5, 1.5))
```

```
## [1] 0.002944974 (Pearson)
## [1] 0.9823322 (Coeficiente de dependência)
```

Os dois cenários finais retratam a estimação para densidade multivariada gaussiana com variâncias marginais iguais a 4 e correlação 0,25: no primeiro caso (h pelo método de Scott), temos um valor de dependência bastante próximo ao verdadeiro valor: 0,25 contra 0,2647 da simulação, enquanto que no segundo caso (h pelo método de Sheather e Jones), a estimativa superestimou a real dependência, onde neste caso, CD é próximo a 0,55.

```
library(mvtnorm)
sigma <- matrix(c(4,1,1,4), ncol=2)
xn <- rmvnorm(n=5000, mean=c(1,2), sigma=sigma)
```

```
cor(xn)
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.2410966
## [2,] 0.2410966 1.0000000
```

```
x<-xn[,1]
```

```
y<-xn[,2]
```

```
## [1] 0.2410966 (Pearson)
```

```
## [1] 0.2647679 (Coeficiente de dependência)
```

```
library(mvtnorm)
```

```
sigma <- matrix(c(4,1,1,4), ncol=2)
```

```
xn <- rmvnorm(n=5000, mean=c(1,2), sigma=sigma)
```

```
cor(xn)
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.2669493
## [2,] 0.2669493 1.0000000
```

```
x<-xn[,1]
```

```
y<-xn[,2]
```

```
## [1] 0.2669493 (Pearson)
```

```
## [1] 0.5460185 (Coeficiente de dependência)
```

A.6 Base de dados

Neste último apêndice, mostramos como buscar dados em alta frequência das ações do Bovespa a partir do pacote GetHFDData, bem como organizar os dados em forma de tabela estruturada. Finalmente, a última parte do código mostra como exportar ao Excel, sem incorrer em falta de memória RAM.

```
install.packages("GetHFData")
library(GetHFData)

first.time <- '10:15:00'
last.time <- '16:45:00'
first.date <- '2016-01-27'
last.date <- '2016-09-08'
type.output <- 'agg'
agg.diff <- '15 min'
my.assets <- c("OIBR3", "BBRK3", "MYPK3", "TAE11", "PRIO3" (...))
type.market = 'equity'

df.out <- ghfd_get_HF_data(my.assets=my.assets,
                          type.market = type.market,
                          first.date = first.date,
                          last.date = last.date,
                          first.time = first.time,
                          last.time = last.time,
                          type.output = type.output,
                          agg.diff = agg.diff)
dados <- df.out[order(df.out$InstrumentSymbol), ]

library(tidyr)
df<-spread(dados, key = "InstrumentSymbol", value = last.price)
df[] <- lapply(df, function(x) `length<-`(na.omit(x), nrow(df)))
drops <- c("SessionDate", "TradeDateTime", "n.trades",
          "weighted.price", "period.ret", "period.ret.volat",
          "sum.qtd", "sum.vol", "n.buys", "n.sells", "Tradetime")
dft<-df[ , !(names(df) %in% drops)]
dft<-dft[1:7500,]

options(java.parameters = "-Xmx4086m")
library(xlsx)
write.xlsx2(x = dft, file = "x.xlsx", sheetName = "TestSheet",
            row.names = FALSE, showNA=F)
```