

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

FELIPE RODRIGUES DE MENEZES TOMAZ

ESTUDO DE UTILIZAÇÃO DE UMA *MINIMUM SPANNING TREE* DE
CORRELAÇÕES COMO SELETORA DE AÇÕES EM UMA ESTRATÉGIA DE
COINTEGRAÇÃO NO MERCADO BRASILEIRO

SÃO PAULO
2016

FELIPE RODRIGUES DE MENEZES TOMAZ

ESTUDO DE UTILIZAÇÃO DE UMA *MINIMUM SPANNING TREE* DE
CORRELAÇÕES COMO SELETORA DE AÇÕES EM UMA ESTRATÉGIA DE
COINTEGRAÇÃO NO MERCADO BRASILEIRO

Dissertação apresentada à Escola de
Economia de São Paulo da Fundação
Getulio Vargas como requisito para
obtenção do título de Mestre em
Economia.

Campo do Conhecimento:
Economia e Finanças.

Orientador:
Prof. Dr. Ricardo R. Rochman

SÃO PAULO
2016

Tomaz, Felipe Rodrigues de Menezes.

Estudo de utilização de uma minimum spanning tree de correlações como seletora de ações em uma estratégia de cointegração no mercado brasileiro / Felipe Rodrigues de Menezes Tomaz. - 2016.

41 f.

Orientador: Ricardo Ratner Rochman

Dissertação (MPFE) - Escola de Economia de São Paulo.

1. Ações (Finanças). 2. Operações com pares (Finanças). 3. Bolsa de valores - Brasil. 4. Mercado de capitais. I. Rochman, Ricardo Ratner. II. Dissertação (MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 336.763.21(81)

FELIPE RODRIGUES DE MENEZES TOMAZ

ESTUDO DE UTILIZAÇÃO DE UMA *MINIMUM SPANNING TREE* DE
CORRELAÇÕES COMO SELETORA DE AÇÕES EM UMA ESTRATÉGIA DE
COINTEGRAÇÃO NO MERCADO BRASILEIRO

Dissertação apresentada à Escola de
Economia de São Paulo da Fundação
Getulio Vargas como requisito para
obtenção do título de Mestre em
Economia.

Campo do Conhecimento:
Economia e Finanças.

Data da Aprovação:

__ / __ / ____

Banca Examinadora:

Prof. Dr. Ricardo Ratner Rochman
(Orientador). FGV – EESP

Prof. Dr. Clemens Vinícius de Azevedo
Nunes. FGV - EESP

Prof. Dr. Arthur Ridolfo Neto. FGV -
EESP

DEDICATÓRIA

Dedico esta dissertação à minha esposa Patrícia Tomaz por toda paciência e companheirismo em todo o desenvolvimento deste trabalho. Seu apoio incondicional foi imprescindível para a conclusão deste trabalho.

AGRADECIMENTOS

Agradeço ao meu orientador e coordenador do curso de mestrado profissional de economia da FGV, Prof. Dr. Ricardo Rochman por ajudar a direcionar este trabalho. Agradeço à FGV por me proporcionar um ensino de qualidade, contribuindo para meu crescimento intelectual e profissional. Agradeço aos meus familiares e amigos pelo suporte e apoio para que pudesse alcançar mais este objetivo pessoal e acadêmico.

RESUMO

Essa dissertação teve como objetivo avaliar a possibilidade de utilização de uma *minimum spanning tree (MST)* de correlação entre ativos como instrumento de seleção em uma estratégia de *pairs trading* que teve como base a cointegração. A partir de uma seleção de ativos do índice Bovespa, foram feitos testes de cointegração entre pares de ativos e, posteriormente, partindo de resultados positivos, foram executados *backtestings* que foram filtrados por informações provenientes de uma MST.

Para determinado período, ao se encontrar uma cointegração entre um par dos ativos considerados, uma MST foi desenvolvida e analisou-se a existência de um link direto entre o par de ativos na estrutura da MST. Caso contrário, o resultado do backtesting proveniente desta cointegração foi desconsiderado. Fazendo-se a comparação do conjunto total de resultados com o subconjunto de resultados que levam em consideração a restrição da MST, avaliou-se o impacto da utilização da MST, como seletora de ativos, no resultado da estratégia de *pairs trading*.

Palavras-chave: Cointegração, correlação, *pairs trading*, reversão à média, *minimum Spanning Tree (MST)*, seleção de ativos, grafo.

ABSTRACT

This work aimed to evaluate the possibility of using a minimum spanning tree (MST) of correlation between assets as a selection tool in a pairs trading strategy based on cointegration. From a selection of Bovespa exchange index assets, cointegration tests were performed between pairs of assets and, subsequently, starting from positive results, backtestings were executed. After that, the backtestings were filtered by information included in an MST.

For a given period, when a cointegration was found between a pair of assets, a MST was developed and it was analyzed if there was a direct link between those assets in the MST's structure. Otherwise, the backtesting result from the cointegration would be disregarded. By comparing the total set of results with the subset of results that took into account the MST constraint, it was evaluated the impact of the use of the MST, as an asset selector, on the result of the pairs trading strategy.

Keywords: Cointegration, correlation, pairs trading strategy, mean reversion, minimum Spanning Tree (MST), stock selection, graph.

LISTA DE TABELAS

Tabela 1 – 30 maiores volumes de negócios da Bovespa em 01/11/2016	19
Tabela 2 – Mapa de Correlações	25
Tabela 3 – Resultados de todos os pares sem filtro por setor de atuação	33
Tabela 4 – Resultados de todos os pares com filtro por setor de atuação	33
Tabela 5 – Resultados filtrados para o setor de Materials	34
Tabela 6 – Resultados filtrados para o setor de Energy	34
Tabela 7 – Resultados filtrados para o setor de Financials	35
Tabela 8 – Resultados filtrados para o setor de Consumer Staples	35
Tabela 9 – Resultados filtrados para o setor de Consumer Discretionary	36
Tabela 10 – Resultados filtrados para o setor de Industrials	36
Tabela 11 – Resultados filtrados para o setor de Telecommunication Services	36
Tabela 12 – Resumo dos resultados	37

LISTA DE FIGURAS

Figura 1 – MST	27
Figura 2 – MST – Exemplo de <i>Trade</i> : ITSA4 com BBAS3	30
Figura 3 – MST – Exemplo de <i>Trade</i> : CSNA3 com USIM5	31

SUMÁRIO

1	INTRODUÇÃO	11
2	REVISÃO BIBLIOGRÁFICA.....	14
2.1	Correlação versus Cointegração.....	14
2.2	Grafo	15
2.3	MST	16
3	METODOLOGIA	18
3.1	Dados.....	18
3.2	Teste de cointegração.....	20
3.3	Backtesting	22
3.4	Criação da MST	24
3.5	Seleção de ativos através da MST.....	27
3.6	Exemplos de <i>Trades</i>	28
4	RESULTADOS.....	32
5	CONCLUSÕES	38
	REFERÊNCIA BIBLIOGRÁFICA.....	39

1 INTRODUÇÃO

Este trabalho teve como objetivo analisar a possibilidade de utilização de uma *Minimum Spanning Tree (MST)* como ferramenta de seleção de ativos em uma estratégia de *pairs trading*, a qual foi implementada segundo conceitos de cointegração de séries de preços. No escopo deste trabalho, uma MST é uma árvore de correlações entre ativos, otimizada de forma a possuir as mais importantes interações entre eles.

Os links entre ativos da MST foram utilizados como filtro seletor dos resultados das cointegrações. Isso foi feito com o intuito de estudar a existência de ganhos potenciais nos resultados, caso fossem selecionados, apenas, pares de ativos que fossem diretamente interligados na MST relacionada. Ou seja, analisou-se a possibilidade de se aumentar a chance de resultados positivos de uma estratégia de *pairs trading* por cointegração, com a utilização de uma MST.

Agentes do mercado financeiro (analistas, *traders* e gestores) buscam ineficiências no mercado de forma a se posicionar e poder, assim, alcançar retornos financeiros. Com a globalização e avanço tecnológico, a complexidade dos mercados vem se tornando cada vez maior, pois há uma grande interação entre os mercados e um constante desenvolvimento de novos produtos e instrumentos financeiros que são acessíveis tanto para grandes instituições quanto para pequenos investidores. Esta variedade de produtos em diversos mercados, junto a altas velocidades na transmissão de informações, faz com que um ativo possa ser influenciado por uma quantidade cada vez maior de variáveis. Isso dificulta a análise de relevância de quais fatores e de quais outros ativos têm influência na formação de preços de um ativo em análise.

Na busca de entender melhor estas interações, a aplicação de redes complexas para o estudo do mercado de ações tem gerado bastante interesse no meio acadêmico. Mantegna (1999) foi o primeiro a estudar estruturas hierárquicas em mercados financeiros, analisando o mercado de ações dos Estados Unidos por meio de MSTs. Ele descobriu um arranjo topológico de ações operadas em um mercado financeiro que possuía uma significativa taxonomia econômica. Empresas de um

mesmo setor se encontravam diretamente conectadas na MST. Isso demonstrou que através de uma MST se poderia extrair informação relevante antes dispersa. Estudos para outras classes de ativos como para moedas estrangeiras foram feitos por Mizuno et al. (2006) e Naylor et al. (2007) como também para *Credit Default Swaps* (CDS) por Marsh e Stevens (2003) e León et al. (2004). Estes estudos utilizaram a correlação entre ativos como métrica básica de análise.

Segundo Alexander (1999), apesar de ser muito disseminada, a correlação possui algumas ressalvas que a limitam como ferramenta exclusiva em uma estratégia de *pairs trading*. Como exemplo, a grande instabilidade ao longo do tempo, por ser uma métrica intrinsicamente de curto prazo. Por outro lado, a cointegração mede o movimento conjunto de preços no longo prazo, mesmo quando correlações estáticas aparentam ser baixas.

Com a ideia de que ações podem ser cointegradas (Bossaerts, 1988), o conceito por trás do *pairs trading* é o de primeiro identificar um par de ações com movimentos de preços históricos similares. Posteriormente, quando houver divergência suficiente entre os preços do par, definir simultaneamente uma posição *long-short*. Esta posição é definida com a aposta de que a divergência de preços é temporária e que haverá convergência no decorrer do tempo (Engelberg et al., 2008).

Para que a estratégia de *pairs trading* seja mais eficiente é importante que haja um processo de seleção dos pares de ativos. Este processo pode ser baseado nos fundamentos da empresa ou em uma abordagem puramente estatística, baseada em dados históricos (Vidyamurthy, 2004). Neste contexto, avaliamos a utilização de uma MST como processo decisório na seleção dos pares de ativos.

Neste trabalho, estudamos a união entre a vantagem de curto prazo da correlação (através do uso de uma MST de correlações) com a de longo prazo da cointegração. A partir de um conjunto de ações do mercado brasileiro, analisamos as cointegrações entre os pares de ativos para determinada janela de tempo. Ao constatar a presença de cointegração, fizemos um *backtesting* simulando o *trading* deste par de ativos segundo regras definidas na metodologia deste trabalho. Com os resultados de ganho/perda obtidos, analisamos se houve maximização no ganho

com a seleção dos pares segundo informações obtidas a partir de uma MST criada a partir de uma matriz de correlações de todos os ativos considerados na análise.

Os próximos tópicos seguem a seguinte sequência. No tópico 2 será feita uma revisão da literatura, passando por um comparativo entre correlação e cointegração, por conceitos básicos da teoria de grafos e posteriormente por detalhes de uma MST. No tópico 3 discutiremos acerca dos dados utilizados neste trabalho, detalhando em seguida a metodologia utilizada, que partirá dos testes de cointegração, passando pelo *backtesting* da estratégia de *pairs trading*, pela criação da MST e, finalmente, pela seleção dos pares de ativos através da MST. O tópico 4 contempla os resultados obtidos. O tópico final 5 contém as conclusões deste trabalho.

2 REVISÃO BIBLIOGRÁFICA

2.1 Correlação versus Cointegração

Correlação é uma medida de movimentos conjuntos entre duas séries de retornos. Uma forte correlação positiva indica que movimentos de alta em uma série tendem a ser acompanhados por movimentos de alta na outra série e, da mesma forma, movimentos de baixa nas séries tendem a ocorrer em conjunto. Se existe uma forte correlação negativa, então movimentos de alta em uma série estão associados a movimentos de baixa na outra série (ALEXANDER, 2001, p. 5).

Segundo Alexander e Dimitriu (2002), a correlação é válida apenas para variáveis estacionárias, desta forma é necessário neutralizar a tendência da série analisada que normalmente é integrada de ordem um ou superior. O cálculo da primeira diferença do log dos preços é o padrão utilizado para conseguir a estacionaridade e tem como consequência a mudança de análise do nível de preços para a de retornos. O processo de neutralização de tendência, acaba por limitar o escopo de análise já que exclui a possibilidade de estudos que apontem tendências em comum nos preços de dois ativos.

O termo cointegração foi definido por Granger (1983) como uma formulação do fenômeno de que processos não estacionários podem ter combinações lineares que são estacionárias (Johansen, 2014). Foi a investigação da relação entre cointegração e correção de erro que trouxe a modelagem de vetores autoregressivos com raiz unitária e a cointegração para o centro de atenção da econometria teórica e aplicada (Engle e Granger, 1987).

Segundo Alexander (1999), correlação e cointegração estão relacionadas, mas são conceitos distintos. Altas correlações de retornos não implicam necessariamente em alta cointegração de preços. Estratégias de investimento que são baseadas apenas em volatilidade ou correlação de retornos não conseguem garantir performance no longo prazo. Não há mecanismo que assegure a reversão do spread entre dois ativos em uma estratégia de *pairs trading* e nada que previna o desvio do spread (*tracking error*) de se comportar de maneira imprevisível, como um movimento

aleatório (*random walk*). Como altas correlações, sozinhas, não são suficientes para assegurar performances de longo prazo de estratégias *hedge*, há a necessidade de metodologias que levem em consideração tendências em comum de longo prazo nos preços. Isso é exatamente o que a cointegração leva em conta.

Para Alexander (2001) a análise de cointegração objetiva detectar tendências estocásticas nas séries de preços, usando estas tendências em comum em uma análise dinâmica de correlação de retornos. Segundo Stock e Watson (1991), um sistema de preços de ativos não estacionários em nível é capaz de compartilhar tendências estocásticas em comum.

2.2 Grafo

Diversas situações do mundo real podem ser descritas convenientemente com a utilização de diagramas contendo um conjunto de nós e linhas ligando estes nós. Poderíamos pensar em um conjunto de países como sendo nós e suas relações diplomáticas como sendo os links entre eles, estes podendo possuir um valor maior ou menor conforme a relevância das relações entre os países em questão. A busca por uma abstração matemática para configurações como esta fez com que o conceito de grafo fosse desenvolvido (BONDY, 1982).

Um grafo G pode ser definido como uma tripla ordenada:

$$(V(G), E(G), \psi_G) \quad (1)$$

onde:

$V(G)$ = um conjunto não-vazio de vértices (nós);

$E(G)$ = um conjunto de arestas, disjunto de $V(G)$;

ψ_G = uma função incidente que associa a cada aresta de G um par não ordenado de vértices (nós) de G .

Se e é uma aresta e u e v são vértices tais que $\psi_G(e) = uv$, então e é dito junção de u com v . Os vértices u e v são chamados de terminações de e .

Segundo Bondy (1982), uma árvore é um grafo acíclico conectado. Em uma árvore, quaisquer dois vértices são conectados por um único caminho (aresta).

Neste trabalho, foi criado um grafo a partir de uma matriz de correlações. Partindo de n ativos, pode-se gerar uma matriz $n \times n$ de dados com as interações dos ativos dois a dois. O grafo criado a partir desta matriz conterá então os n ativos, cada um deles se ligando a cada um dos outros $n-1$ ativos. Cada aresta será representada pela correlação entre os ativos.

Para um grande grupo de ativos, a extensa quantidade de informações deste grafo dificulta o estudo e a visualização das interações que de fato são relevantes. A criação de uma árvore de espalhamento mínimo, ou MST, a partir deste grafo, revela as correlações mais importantes, descartando as interações (arestas) irrelevantes entre ativos (Mantegna, 1999).

2.3 MST

Há anos, o problema da árvore de espalhamento mínimo é um dos mais típicos e bem conhecidos problemas de otimização combinatória. A sua importância e popularidade deriva do fato de ele possuir uma solução eficiente, o que o torna prático nos casos de utilização para grandes grafos. Há diversas aplicações para a MST como *design* de redes de computadores e de comunicações, redes de energia, conexões em uma rede de transportes e tubulações em uma rede de saneamento (Graham, 1985).

Singhal (2014) utilizou conceitos de MST para analisar o mercado de ações indiano, *National Stock Exchange* (NSE), para encontrar o impacto da crise financeira gerada pelo *Subprime* no mercado de ações da Índia. Ele analisou como a crise afetou ações de forma individual, como foi afetada a importância relativa entre setores para diferentes fases da crise e como as ações se agrupavam no decorrer do tempo.

Mizuno e Takaysu (2006) analisaram o mercado de moedas, derivando de MSTs uma taxonomia hierárquica de moedas. Encontraram uma estrutura agrupada de

moedas e a principal em cada um dos agrupamentos. Cada agrupamento de moedas casava perfeitamente com a região geográfica dos respectivos países das moedas do grupo, como por exemplo Asia ou Europa. As principais moedas de cada grupo de fato eram as de países com economias mais relevantes, como era de se esperar.

Segundo Mantegna (2008), uma *Minimum Spanning Tree* é um tipo particular de um grafo, contendo todos os n nós do grafo e possuindo $n-1$ arestas. Neste trabalho, o grafo criado possui a $C_{n,2}$ (combinação de n dois a dois) de arestas, ou seja, $n(n-1)/2$ arestas. Em uma MST, todos os vértices (nós) estão conectados sem que se forme um *loop*. Por definição, a soma total dos valores das arestas em uma MST possui valor mínimo se comparada às possíveis árvores formadas a partir do mesmo grafo, que contenham também $n-1$ arestas e não formem um loop.

Partindo de um grafo com arestas inversamente proporcionais às correlações de um mapa de correlações, a MST encontrada a partir deste grafo, conterá os links entre os ativos de tal forma que a soma total das correlações entre eles seja a máxima possível. Ou seja, a criação da MST acarreta uma seleção implícita das maiores e, assim, mais relevantes correlações entre ativos.

3 METODOLOGIA

3.1 Dados

Este trabalho se inicia com a definição do universo de ativos a ser estudado, o mercado brasileiro de ações (mercado à vista). A existência de trabalhos que obtiveram resultados positivos com a aplicação de estratégias de *pairs trading* (utilizando cointegração) para o mercado de ações brasileiro foi um motivador para esta escolha (Caldeira e Moura, 2013).

A amostra de dados definida para este estudo foi de 02/01/2009 até 01/11/2016. Foram então selecionadas as 30 ações com maior volume de negócios diários da bolsa de São Paulo (Bovespa) na data de 01/11/2016 e que já eram negociadas na bolsa de valores desde o início da amostra de dados considerada.

Não houve rebalanceamento da lista de ativos. Isso é justificado pelo intuito deste trabalho, que focou na comparação dos resultados com e sem a seleção de ativos por uma MST e não na maximização dos resultados de cointegração, independentemente dos critérios de seleção dos ativos.

Para cada um dos ativos, foram obtidas as cotações de preço de fechamento diário, já ajustadas por proventos, a partir da plataforma *Bloomberg*. Considerando que a amostra de dados pertence ao mesmo mercado, os feriados da bolsa de valores coincidem para todos os ativos, evitando tratamento nos dados para considerar apenas os dias úteis. Esse tratamento seria necessário caso a amostra de dados contivesse cotações de ativos de bolsas de países distintos. Neste caso, seria necessário considerar apenas os dias em que as duas bolsas de valores estivessem em funcionamento. Um tipo de estratégia de *trading* muito comum, entre ações locais e *American Depositary Receipts (ADRs)*, ou seja, entre ações negociadas na bolsa de São Paulo e suas respectivas *ADRs* negociadas nas bolsas dos Estados Unidos, possui este tipo de problema a ser considerado.

Foram calculados os retornos diários dos ativos para utilização nos cálculos de correlação, segundo a expressão abaixo:

$$retorno_{D-0} = \frac{P_{D-0}}{P_{D-1}} - 1 \quad (2)$$

onde:

$retorno_{D-0}$ = retorno do ativo em D-0;

P_{D-0} = preço de fechamento do ativo em D-0;

P_{D-1} = preço de fechamento do ativo em D-1.

Não houve necessidade de tratamento dos dados para os testes de cointegração, já que foram utilizadas diretamente as séries de preços de fechamento diário.

Os ativos foram classificados por setores segundo a *Global Industry Classification Standard* (GICS) que é uma taxonomia da indústria desenvolvida em 1999 por *Morgan Stanley Capital International* (MSCI) e *Standard & Poor's* (S&P) para uso da comunidade financeira global.

Tabela 1 – 30 maiores volumes de negócios da Bovespa em 01/11/2016

Nome do Ativo	Código	Setor	Nome do Ativo	Código	Setor
Petrobrás	PETR4	<i>Energy</i>	Sid. Nacional	CSNA3	<i>Materials</i>
Usiminas	USIM5	<i>Materials</i>	Vale	VALE3	<i>Materials</i>
Gerdau Met.	GOAU4	<i>Materials</i>	Tim Part. S/A	TIMP3	<i>Telecommunication Services</i>
Vale	VALE5	<i>Materials</i>	Cemig	CMIG4	<i>Utilities</i>
Itaúsa	ITSA4	<i>Financials</i>	Suzano Papel	SUZB5	<i>Materials</i>
Ambev S/A	ABEV3	<i>Consumer Staples</i>	BR Malls Part.	BRML3	<i>Real Estate</i>
Itaú Unibanco	ITUB4	<i>Financials</i>	Hypermarcas	HYPE3	<i>Health Care</i>
JBS	JBSS3	<i>Consumer Staples</i>	CCR S/A	CCRO3	<i>Industrials</i>
Kroton	KROT3	<i>Consumer Discretionary</i>	Bradespar	BRAP4	<i>Materials</i>
BM&F Bovespa	BVMF3	<i>Financials</i>	Estácio Part.	ESTC3	<i>Consumer Discretionary</i>
Gerdau	GGBR4	<i>Materials</i>	MRV	MRVE3	<i>Consumer Discretionary</i>
Petrobrás	PETR3	<i>Energy</i>	Lojas Renner	LREN3	<i>Consumer Discretionary</i>
Bradesco	BBDC4	<i>Financials</i>	Telef Brasil	VIVT4	<i>Telecommunication Services</i>
Embraer	EMBR3	<i>Industrials</i>	Cyrela Realty	CYRE3	<i>Consumer Discretionary</i>
Banco do Brasil	BBAS3	<i>Financials</i>	Weg S/A	WEGE3	<i>Industrials</i>

Fonte: Bloomberg

Nota 1: Considerados apenas ativos com cotações a partir de 02/01/2009

Nota 2: Ordenados por volume de negócios

Nota 3: Ativos classificados por setores segundo a *Global Industry Classification Standard* (GICS)

3.2 Teste de cointegração

O procedimento de dois passos de *Engle* e *Granger* (1987) tem como finalidade a obtenção dos parâmetros α , β e ρ mais adequados para o seguinte modelo:

$$Y[i] = \alpha + \beta * X[i] + R[i] \quad (3)$$

$$R[i] = \rho * R[i - 1] + \epsilon[i] \quad (4)$$

$$\epsilon[i] \sim N(0, \sigma^2) \quad (5)$$

onde:

$X[i]$ = preço do ativo X para o instante i ;

$Y[i]$ = preço do ativo Y para o instante i ;

$R[i]$ = erro da regressão linear de X sobre Y ;

$\epsilon[i]$ = erro da auto-regressão de R ;

Em um primeiro passo, os coeficientes α e β são encontrados usando a regressão linear de $X[i]$ com respeito a $Y[i]$. A série de resíduos $R[i]$ é então determinada.

Em um segundo passo, ρ é encontrado, novamente com o artifício de uma regressão linear.

Engle e *Granger* (1987) mostraram que se X e Y são cointegrados, então este procedimento resultará em estimativas consistentes para os parâmetros acima. Apesar disso, existem algumas possibilidades a serem testadas para evitar falhas no resultado deste procedimento:

1. X , Y (ou ambos) não são $I(0)$, ou seja, não são séries estacionárias;
2. A série de resíduos $R[i]$ é estacionária, ou seja, $I(0)$.

Para avaliar o primeiro item, foi utilizado o teste de Phillips-Perron, verificando então se as séries são $I(1)$. Para o segundo item, ou seja, para os testes que analisam se a série de resíduos contém raiz unitária, o teste de Johansen foi utilizado.

A ferramenta de análise estatística utilizada neste trabalho foi o *R script*. Foi utilizado o pacote *egcm* para a aferição dos testes descritos acima. Foi definido o *pValue* padrão de 5% para análise dos resultados dos testes.

Os dados foram divididos em janelas móveis de 252 dias úteis. Como exemplo, a primeira janela de tempo considerada se estende do dia 02/01/2009 até o dia 06/01/2010. Foram feitos testes para cada par de ativos e para cada deslocamento de um dia útil de cada janela, partindo do último dia da primeira janela (06/01/2010) até o final da amostra de dados disponível (01/11/2016), o que deu um total de 1.717 dias úteis.

Considerando todas as combinações entre os pares de ativos, existem 870 possibilidades. Foram considerados tanto testes de cointegração de um ativo A com B quanto de um B com A. Ao todo foram realizados em torno de 1,5 milhões de testes para os dados em questão.

Como, para serem testadas por cointegração, as duas séries de dados não podem ser consideradas estacionárias (Engle e Granger, 1987), devem na verdade ser integradas de primeira ordem, foram utilizados os níveis de preços, ao invés dos retornos calculados. Já para os cálculos de correlação, foram utilizados os retornos calculados, pois a correlação tem como premissa a estacionaridade dos dados.

Para cada teste realizado, obtemos a afirmação ou a negação da existência de cointegração da série de cotações do par de ativos. Esta informação é então utilizada para a realização de um backtesting de uma estratégia de *pairs trading* para este par de ativos, segundo as regras definidas no próximo tópico.

Além do resultado positivo (ou negativo) dos testes de cointegração, guardamos as seguintes variáveis para posterior utilização pelo backtesting: α , β , $Y_e[t]$ (Y estimado no tempo t = último dia da janela), $R[t]$ (erro calculado para t = último dia da janela), σ_R (desvio padrão dos resíduos R).

3.3 Backtesting

Foram obtidos no item anterior os dados necessários para o início de um backtesting da estratégia de *pairs trading*. Esta percorre os dias úteis de 06/10/2010 até 01/11/2016. Ao encontrar um resultado positivo de cointegração para determinado par de ativos, a estratégia inicia a busca por um sinal de entrada em um *trade*, com posição de compra em um primeiro ativo e de venda a descoberto em um segundo ativo.

A busca por sinal de entrada pode ser entendida da seguinte forma: a partir do dia em que foi encontrada a cointegração, a estratégia percorre até um total de 63 dias úteis, caso não encontre o sinal de entrada para o *trade* até o fim deste período, prossegue para o próximo resultado positivo de cointegração, repetindo então o processo descrito acima. Os critérios que definem o sinal de entrada serão descritos mais adiante.

Caso, dentro do período de 63 dias úteis, a estratégia encontre o sinal de entrada, as operações de compra e venda a descoberto nos respectivos ativos são efetuadas, e o algoritmo da estratégia passa agora a buscar por um sinal de fechamento da posição. Com isso, desfazendo então as operações do *trade* inicial. Como resultado, obtém-se um *PnL* (*profits and losses*) da operação que será utilizado como métrica de análise.

Após o fechamento da posição do *trade*, o algoritmo prossegue para o dia útil seguinte, analisando o resultado dos testes de cointegração para o dia em questão. Caso o resultado para este dia seja negativo, ele prossegue adiante por mais um dia útil, até que encontre novamente um resultado positivo, seguindo então os mesmos passos já descritos anteriormente.

Após percorrer toda a amostra de dados, os *PnLs* para todos os *trades* realizados para todos os pares de ativos da amostra são obtidos. Desta forma, temos os dados necessários para, finalmente, analisar a influência da seleção de pares de ativos através de uma MST. Analisamos o impacto da utilização da MST através do *PnL/trade* médio, caso ele seja maior com a utilização da MST, consideramos

vantajosa a seleção de ativos pela MST. No tópico de resultados, mais adiante, há mais detalhes sobre esta análise.

Os critérios para as regras dos sinais de entrada dos *trades* foram definidos da seguinte forma:

- Caso $Z[t] \geq 2\sigma_Z$, vende-se Y e compra-se X .
- Caso $Z[t] \leq -2\sigma_Z$, compra-se Y e vende-se X .

onde:

$Z[t]$ = spread entre o $Y[t]$ e o $Y_e[t]$ para o instante t ;

σ_Z = desvio padrão do spread Z . Calculado para cada dia analisado.

Os critérios para as regras dos sinais de saída dos *trades* foram definidos da seguinte forma:

- Caso $Z[t]$ mude de sinal, ou se torne 0 (aferimos lucro, neste caso);
- Caso $Z[t] \geq 4\sigma_Z$ ou $Z[t] \leq -4\sigma_Z$ (respectivamente, ao primeiro e ao segundo casos de entrada acima) (aferimos prejuízo, neste caso).
- Caso o *trade* demore mais que 42 dias úteis para ser finalizado (resultado indeterminado).

A relação entre as posições dos ativos X e Y é definida através do cálculo de:

$$H = \frac{\beta X[t]}{Y[t]} \quad (6)$$

onde:

H = *hedge ratio* entre X e Y ;

$X[t]$ = preço do ativo X para o último dia t da janela;

$Y[t]$ = preço do ativo Y para o último dia t da janela;

β = coeficiente da equação (3).

H é utilizado para relacionar o financeiro operado do ativo X com o do ativo Y no momento em que a posição do par é iniciada. São operados $-H$ reais do ativo X para cada um real de Y . Neste trabalho, para cada *trade* realizado, foi definido o financeiro de 10.000 reais para o ativo Y , conseqüentemente, o financeiro do ativo X ficou determinado por $-10.000H$ reais.

O PnL de cada ativo e o PnL total do *trade* é definido segundo as expressões abaixo:

$$PnL_X = c * \frac{-H * 10,000}{Px_e} * (Px_s - Px_e) \quad (7)$$

$$PnL_Y = c * \frac{10,000}{Py_e} * (Py_s - Py_e) \quad (8)$$

$$PnL = PnL_X + PnL_Y \quad (9)$$

onde:

PnL_X = resultado da operação do ativo X.

PnL_Y = resultado da operação do ativo Y.

PnL = resultado total do *trade*.

Px_e = preço de entrada na operação do ativo X;

Px_s = preço de saída na operação do ativo X;

Py_e = preço de entrada na operação do ativo Y;

Py_s = preço de saída na operação do ativo Y;

c = define lado do *trade* dos ativos X e Y dependendo do valor de $Z[t]$. Ou seja, $c=1$ se $Z[t] \geq 2\sigma_Z$ ou $c=-1$ se $Z[t] \leq -2\sigma_Z$.

Só foram considerados os resultados obtidos com $\beta > 0$, para que os ativos X e Y sempre sejam operados um contra o outro, ou seja, caso haja compra de X, haverá venda a descoberto de Y e vice-versa. Isso foi feito para que o net financeiro do par de ativos negociados seja minimizado e da própria natureza *long and short* da estratégia de *pairs trading*.

A janela de 252 dias para o teste de cointegração, o período limite de 63 dias para se entrar em um *trade*, os valores limites de $Z[t]$ ($-2\sigma_Z, -4\sigma_Z, 2\sigma_Z, 4\sigma_Z$), a duração máxima de um *trade* (42 dias) e o financeiro de 10.000 reais para o ativo Y foram selecionados de forma empírica.

3.4 Criação da MST

Os testes de cointegração buscam uma relação de equilíbrio de longo prazo entre os ativos, diferentemente da correlação que tende a perder relevância conforme a

amostra de dados aumenta. Desta forma, definimos uma janela menor de dados para as correlações, 63 dias úteis (3 meses).

Para cada teste de cointegração que confirme a cointegração de um par de ativos é gerada uma matrix simétrica de correlações 30 x 30, contendo as correlações entre todos os pares de ativos. O último dia da janela de tempo considerada para o cálculo da correlação é definido como o último dia da janela considerada no teste de cointegração. A correlação é calculada considerando os retornos diários de cada ativo.

Na tabela 2 é possível ver um exemplo de matriz de correlações criada para o dia 01/09/2016. A tonalidade de cada célula da tabela varia conforme o valor da correlação calculada, desde vermelho (para correlações com valor -1) até verde (para correlações com valor +1), passando por amarelo (para correlações com valor 0). Neste mapa de correlações, para a janela considerada, é fácil notar a predominância de correlações positivas entre os ativos.

Tabela 2 – Mapa de Correlações

	PET4	USIM5	GOAU4	VALE5	ITSA4	ABEV3	ITUB4	JBSS3	KROT3	BVMF3	GGBR4	PETR3	BBD4	EMBR3	BBAS3	CSNA3	VALE3	TIMP3	CMIG4	SUZB5	BRML3	HYPE3	CCRO3	BRAP4	ESTC3	MRVE3	LREN3	VIVT4	CYRE3	WEGE3
PET4	1,0	0,5	0,5	0,6	0,6	0,1	0,7	0,3	0,3	0,4	0,4	0,9	0,6	0,1	0,5	0,7	0,6	0,5	0,5	0,1	0,3	0,0	0,4	0,6	0,4	0,4	0,3	0,3	0,5	0,3
USIM5	0,5	1,0	0,7	0,5	0,3	0,0	0,3	0,1	0,1	0,3	0,5	0,5	0,4	-0,1	0,2	0,6	0,5	0,4	0,3	0,0	0,3	0,2	0,3	0,6	0,1	0,3	0,2	0,2	0,3	0,2
GOAU4	0,5	0,7	1,0	0,6	0,4	0,2	0,4	0,2	0,2	0,4	0,9	0,5	0,5	0,1	0,4	0,7	0,6	0,3	0,3	0,1	0,2	0,1	0,2	0,7	0,2	0,1	0,2	0,3	0,3	0,3
VALE5	0,6	0,5	0,6	1,0	0,5	0,2	0,6	0,2	0,1	0,3	0,6	0,6	0,4	0,2	0,4	0,7	0,9	0,5	0,5	0,2	0,3	0,1	0,3	0,9	0,1	0,3	0,3	0,4	0,4	0,4
ITSA4	0,6	0,3	0,4	0,5	1,0	0,2	0,9	0,3	0,3	0,5	0,4	0,6	0,9	0,2	0,7	0,5	0,4	0,5	0,4	0,1	0,5	0,1	0,5	0,5	0,2	0,4	0,3	0,5	0,5	0,5
ABEV3	0,1	0,0	0,2	0,2	0,2	1,0	0,2	0,1	0,4	0,1	0,3	0,1	0,1	0,5	0,2	0,3	0,3	0,3	0,3	0,3	0,2	0,0	0,2	0,2	0,2	0,0	0,2	0,3	0,1	0,4
ITUB4	0,7	0,3	0,4	0,6	0,9	0,2	1,0	0,3	0,3	0,4	0,5	0,7	0,8	0,2	0,6	0,6	0,5	0,6	0,4	0,2	0,5	0,1	0,5	0,5	0,2	0,4	0,4	0,5	0,5	0,4
JBSS3	0,3	0,1	0,2	0,2	0,3	0,1	0,3	1,0	0,2	0,2	0,2	0,3	0,3	0,0	0,3	0,3	0,2	0,2	0,1	0,1	0,4	-0,1	0,3	0,3	0,2	0,2	0,2	0,3	0,5	0,2
KROT3	0,3	0,1	0,2	0,1	0,3	0,4	0,3	0,2	1,0	0,4	0,3	0,3	0,3	0,2	0,1	0,4	0,2	0,3	0,3	0,3	0,3	0,1	0,2	0,2	0,6	0,3	0,3	0,3	0,3	0,3
BVMF3	0,4	0,3	0,4	0,3	0,5	0,1	0,4	0,2	0,4	1,0	0,4	0,4	0,5	0,2	0,5	0,6	0,4	0,4	0,3	-0,1	0,4	0,1	0,3	0,4	0,1	0,4	0,4	0,2	0,5	0,2
GGBR4	0,4	0,5	0,9	0,6	0,4	0,3	0,5	0,2	0,3	0,4	1,0	0,4	0,5	0,3	0,4	0,7	0,6	0,4	0,3	0,2	0,2	0,1	0,3	0,6	0,2	0,2	0,3	0,2	0,4	0,4
PETR3	0,9	0,5	0,5	0,6	0,6	0,1	0,7	0,3	0,3	0,4	0,4	1,0	0,6	0,1	0,5	0,7	0,6	0,5	0,5	0,1	0,2	0,1	0,4	0,6	0,3	0,3	0,3	0,4	0,5	0,3
BBD4	0,6	0,4	0,5	0,4	0,9	0,1	0,8	0,3	0,3	0,5	0,5	0,6	1,0	0,0	0,7	0,6	0,3	0,5	0,4	0,0	0,5	0,0	0,4	0,5	0,2	0,4	0,4	0,4	0,6	0,5
EMBR3	0,1	-0,1	0,1	0,2	0,2	0,5	0,2	0,0	0,2	0,2	0,3	0,1	0,0	1,0	0,1	0,3	0,3	0,2	0,2	0,3	0,0	0,1	0,0	0,3	0,0	0,1	0,1	0,1	0,0	0,3
BBAS3	0,5	0,2	0,4	0,4	0,7	0,2	0,6	0,3	0,1	0,5	0,4	0,5	0,7	0,1	1,0	0,5	0,3	0,4	0,4	0,0	0,4	0,1	0,4	0,4	0,0	0,3	0,2	0,3	0,5	0,4
CSNA3	0,7	0,6	0,7	0,7	0,5	0,3	0,6	0,3	0,4	0,6	0,7	0,7	0,6	0,3	0,5	1,0	0,7	0,4	0,6	0,2	0,4	0,2	0,4	0,8	0,2	0,4	0,4	0,3	0,4	0,2
VALE3	0,6	0,5	0,6	0,9	0,4	0,3	0,5	0,2	0,2	0,4	0,6	0,6	0,3	0,3	0,3	0,7	1,0	0,5	0,5	0,2	0,2	0,1	0,3	0,9	0,1	0,3	0,2	0,4	0,4	0,3
TIMP3	0,5	0,4	0,3	0,5	0,5	0,3	0,6	0,2	0,3	0,4	0,4	0,5	0,5	0,2	0,4	0,4	0,5	1,0	0,4	0,3	0,3	0,0	0,3	0,5	0,2	0,5	0,4	0,5	0,4	0,4
CMIG4	0,5	0,3	0,3	0,5	0,4	0,3	0,4	0,1	0,3	0,3	0,3	0,5	0,4	0,2	0,4	0,6	0,5	0,4	1,0	0,1	0,3	0,1	0,2	0,5	0,3	0,2	0,4	0,3	0,4	0,2
SUZB5	0,1	0,0	0,1	0,2	0,1	0,3	0,2	0,1	0,3	-0,1	0,2	0,1	0,0	0,3	0,0	0,2	0,2	0,3	0,1	1,0	0,0	0,1	0,0	0,2	0,0	0,0	-0,1	0,2	-0,2	0,1
BRML3	0,3	0,3	0,2	0,3	0,5	0,2	0,5	0,4	0,3	0,4	0,2	0,2	0,5	0,0	0,4	0,4	0,2	0,3	0,3	0,0	1,0	-0,1	0,6	0,3	0,3	0,4	0,4	0,3	0,7	0,4
HYPE3	0,0	0,2	0,1	0,1	0,1	0,0	0,1	-0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,2	0,1	0,0	0,1	0,1	-0,1	1,0	0,0	0,1	0,0	-0,1	0,1	-0,2	0,1	0,2
CCRO3	0,4	0,3	0,2	0,3	0,5	0,2	0,5	0,3	0,2	0,3	0,3	0,4	0,4	0,0	0,4	0,4	0,3	0,3	0,2	0,0	0,6	0,0	1,0	0,4	0,2	0,6	0,3	0,4	0,6	0,3
BRAP4	0,6	0,6	0,7	0,9	0,5	0,2	0,5	0,3	0,2	0,4	0,6	0,6	0,5	0,3	0,4	0,8	0,9	0,5	0,5	0,2	0,3	0,1	0,4	1,0	0,1	0,3	0,3	0,4	0,4	0,3
ESTC3	0,4	0,1	0,2	0,1	0,2	0,2	0,2	0,6	0,1	0,2	0,3	0,2	0,0	0,0	0,2	0,1	0,2	0,3	0,0	0,3	0,0	0,2	0,1	1,0	0,1	0,3	0,2	0,2	0,2	0,2
MRVE3	0,4	0,3	0,1	0,3	0,4	0,0	0,4	0,2	0,3	0,4	0,2	0,3	0,4	0,1	0,3	0,4	0,3	0,5	0,2	0,0	0,4	-0,1	0,6	0,3	0,1	1,0	0,4	0,4	0,5	0,1
LREN3	0,3	0,2	0,2	0,3	0,3	0,2	0,4	0,2	0,3	0,4	0,3	0,3	0,4	0,1	0,2	0,4	0,2	0,4	0,4	-0,1	0,4	0,1	0,3	0,3	0,3	0,4	1,0	0,4	0,4	0,3
VIVT4	0,3	0,2	0,3	0,4	0,5	0,3	0,5	0,3	0,3	0,2	0,3	0,4	0,4	0,1	0,3	0,3	0,4	0,5	0,3	0,2	0,3	-0,2	0,4	0,4	0,2	0,4	0,4	1,0	0,4	0,3
CYRE3	0,5	0,3	0,3	0,4	0,5	0,1	0,5	0,5	0,3	0,5	0,2	0,5	0,6	0,0	0,5	0,5	0,4	0,4	0,4	-0,2	0,7	0,1	0,6	0,4	0,2	0,5	0,4	0,4	1,0	0,4
WEGE3	0,3	0,2	0,3	0,4	0,5	0,4	0,4	0,2	0,3	0,2	0,4	0,3	0,5	0,3	0,4	0,4	0,3	0,4	0,2	0,1	0,4	0,2	0,3	0,3	0,2	0,1	0,3	0,3	0,4	1,0

Fonte: Elaboração própria

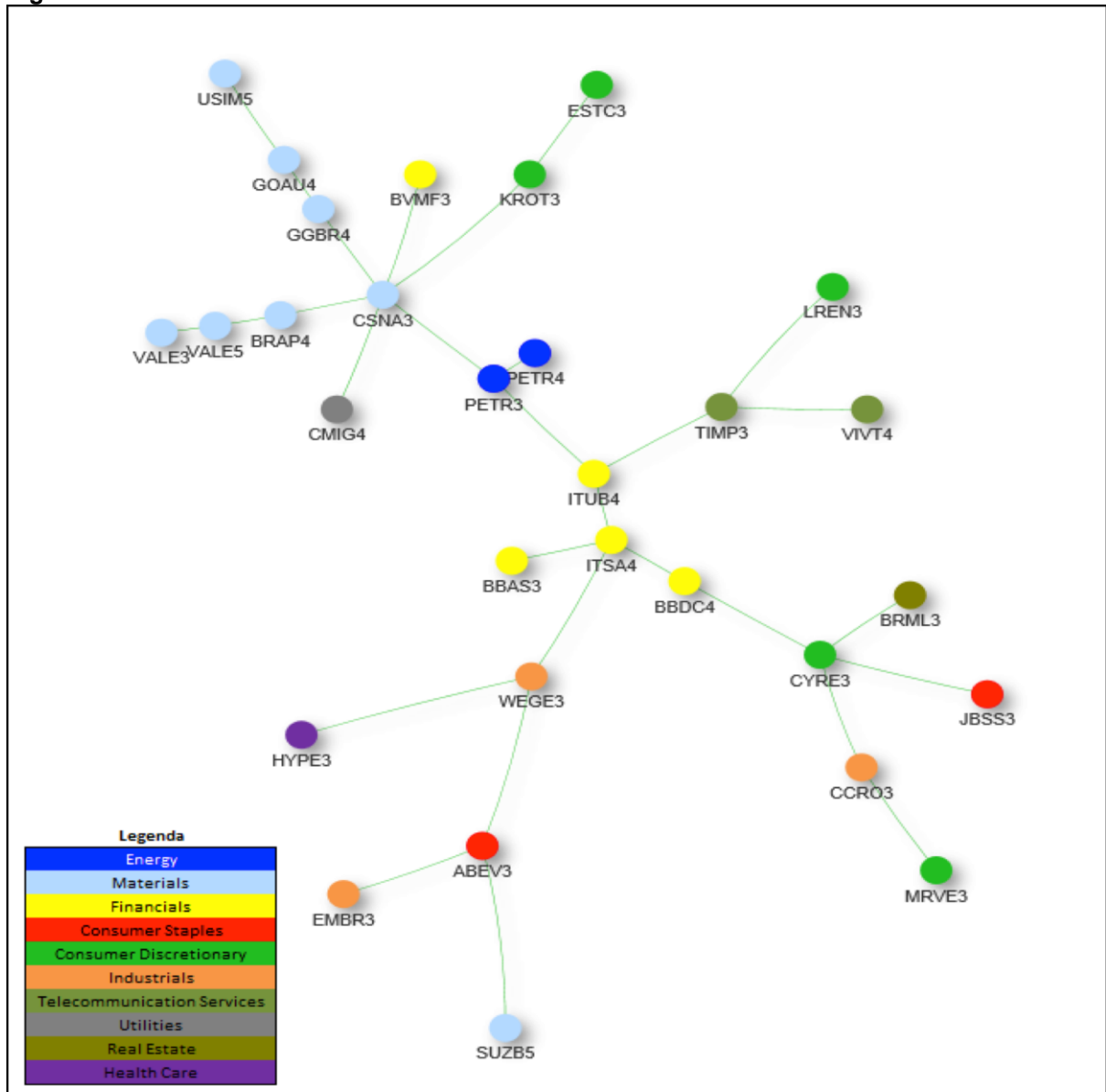
Nota 1: Correlações de 63 dias úteis antecedentes ao dia 01/09/2016

Há dois algoritmos bastantes utilizados na literatura para a criação de uma MST: algoritmo de Kruskal e algoritmo de Prim (1957). Foi utilizado neste trabalho o algoritmo de Prim.

Como a geração de uma MST envolve a minimização da soma das arestas da árvore, precisamos fazer uma conversão dos dados da matriz de correlações de forma que a minimização da soma das arestas nos dê a maximização da soma de correlações.

Primeiramente, foi calculado o módulo das correlações, pois há interesse no nível de correlação, independentemente do seu sinal. Foi feita então uma conversão linear dos valores resultantes de forma a transformar uma sequência crescente de correlações em uma sequência decrescente de distâncias (arestas da MST) que serão utilizadas para construir efetivamente a MST, conforme exemplo da Figura 1 a seguir.

Figura 1 – MST



Fonte: Elaboração própria

Nota 1: MST gerada a partir da Matriz de Correlações da tabela 2

Em seguida os dados das distâncias foram utilizados como input para o algoritmo de Prim que, percorrendo ativo por ativo, constrói a MST, gerando como saída os links entre ativos e seus respectivos pesos (distâncias).

3.5 Seleção de ativos através da MST

Partindo da árvore que contém todos os ativos analisados e que possui somente os links entre eles de forma que a soma de suas correlações seja máxima, foi possível filtrar as correlações mais relevantes e visualizar mais claramente como os ativos se inter-relacionam através dos links da MST.

Este trabalho utilizou estes links da MST como filtro seletor para os resultados das cointegrações, no intuito de estudar se há ganhos potenciais nos resultados caso fossem selecionados, apenas, pares que contivessem os ativos diretamente interligados em uma MST relacionada.

Foram analisados os resultados absolutos dos *backtestings* para cada setor dos ativos, comparando os resultados com e sem a seleção pelas MSTs.

3.6 Exemplos de *Trades*

Neste subtópico foram detalhados 2 *trades* com o intuito de exemplificar os passos mencionados nos itens anteriores desta metodologia.

No primeiro *trade* analisado, ITSA4 com BBAS3, o teste de cointegração, feito para o dia 04/01/2016, retornou resultado positivo para a existência de cointegração entre o par de ativos em questão. Foram utilizados 252 dias úteis para os dados do teste, conforme mencionado anteriormente nos critérios adotados.

Com a confirmação da cointegração, iniciou-se a análise de entrada no *trade* que ocorreu 35 dias úteis depois, quando $\frac{Z[t]}{\sigma_Z} = -2,0$. Então, o ativo ITSA4 foi vendido a descoberto ao preço de 5,95 reais e o ativo BBAS3 foi comprado ao preço de 12,53 reais. O *trade* teve duração de 12 dias úteis, quando $\frac{Z[t]}{\sigma_Z} = 0,5$. O ativo ITSA4 foi, então, recomprado ao preço de 7,65 reais (perda de 4.737 reais) e o ativo BBAS3 foi vendido ao preço de 22,04 reais (ganho de 7.591 reais). Assim, o resultado obtido pelo *pairs trading* de ITSA4 com BBAS3 foi de um ganho de 2.854 reais. As quantidades de ações negociadas foram calculadas considerando o financeiro de

10.000 reais para o ativo BBAS3 e o financeiro de $10.000H$ ($H = 1,66$) para o ativo ITSA4, tomando como referência o preço de entrada em cada ativo para o cálculo.

Para a análise de seleção do par de ativos, a partir da informação da existência de cointegração entre o par ITSA4 e BBAS3, foi construída uma matrix de correlações entre todos os pares da amostra e, em seguida, foi calculada uma MST, figura 2. A MST criada é derivada do grafo associado à matriz de correlações. Na figura 2, há uma aresta ligando os ativos ITSA4 e BBAS3, ou seja, o resultado do *trade* acima é então considerado como pertencente ao grupo de *trades* com pares selecionados por uma MST.

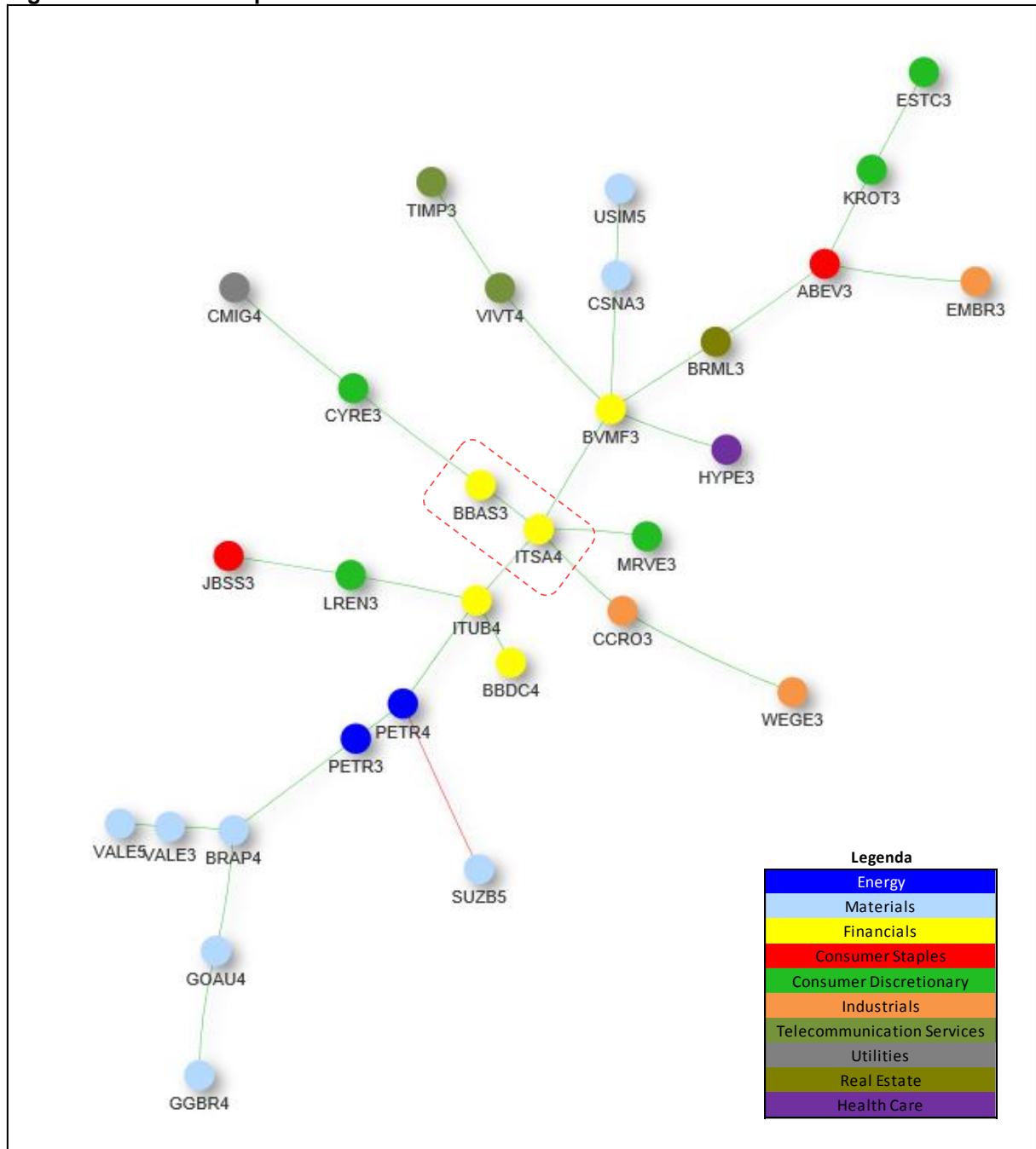
Em um segundo *trade* analisado, CSNA3 com USIM5, o teste de cointegração, feito para o dia 25/09/2015, também retornou resultado positivo para a existência de cointegração entre o par de ativos em questão.

Após a confirmação da cointegração, iniciou-se a análise de entrada no *trade* que ocorreu 14 dias úteis depois, quando $\frac{Z[t]}{\sigma_Z} = -2,4$. Então, o ativo CSNA3 foi vendido a descoberto ao preço de 5,02 reais e o ativo USIM5 foi comprado ao preço de 3,27 reais. O *trade* teve duração de 14 dias úteis, quando $\frac{Z[t]}{\sigma_Z} = -4,1$. O ativo CSNA3 foi, então, recomprado ao preço de 5,16 reais (perda de 236 reais) e o ativo USIM5 foi vendido ao preço de 2,66 reais (perda de 1.891 reais). Assim, o resultado obtido pelo *pairs trading* de CSNA3 com USIM5 foi de uma perda de 2.127 reais. As quantidades de ações negociadas foram calculadas considerando o financeiro de 10.000 reais para o ativo USIM5 e o financeiro de $10.000H$ ($H = 0,85$) para o ativo CSNA3, tomando como referência o preço de entrada em cada ativo para o cálculo.

Na figura 3, pode-se ver a MST criada para a análise de seleção de ativos derivada do grafo associado à matriz de correlações. Verifica-se a existência de uma aresta ligando os ativos CSNA3 e USIM5, ou seja, o resultado do *trade* acima também é considerado como pertencente ao grupo de *trades* com pares selecionados por uma MST.

Interessante notar que, tanto na figura 2 quanto na figura 3, os ativos de mesmo setor tendem a se agrupar na MST, como, por exemplo, para os setores de *Materials* e *Financials*.

Figura 2 – MST – Exemplo de Trade: ITSA4 com BBAS3

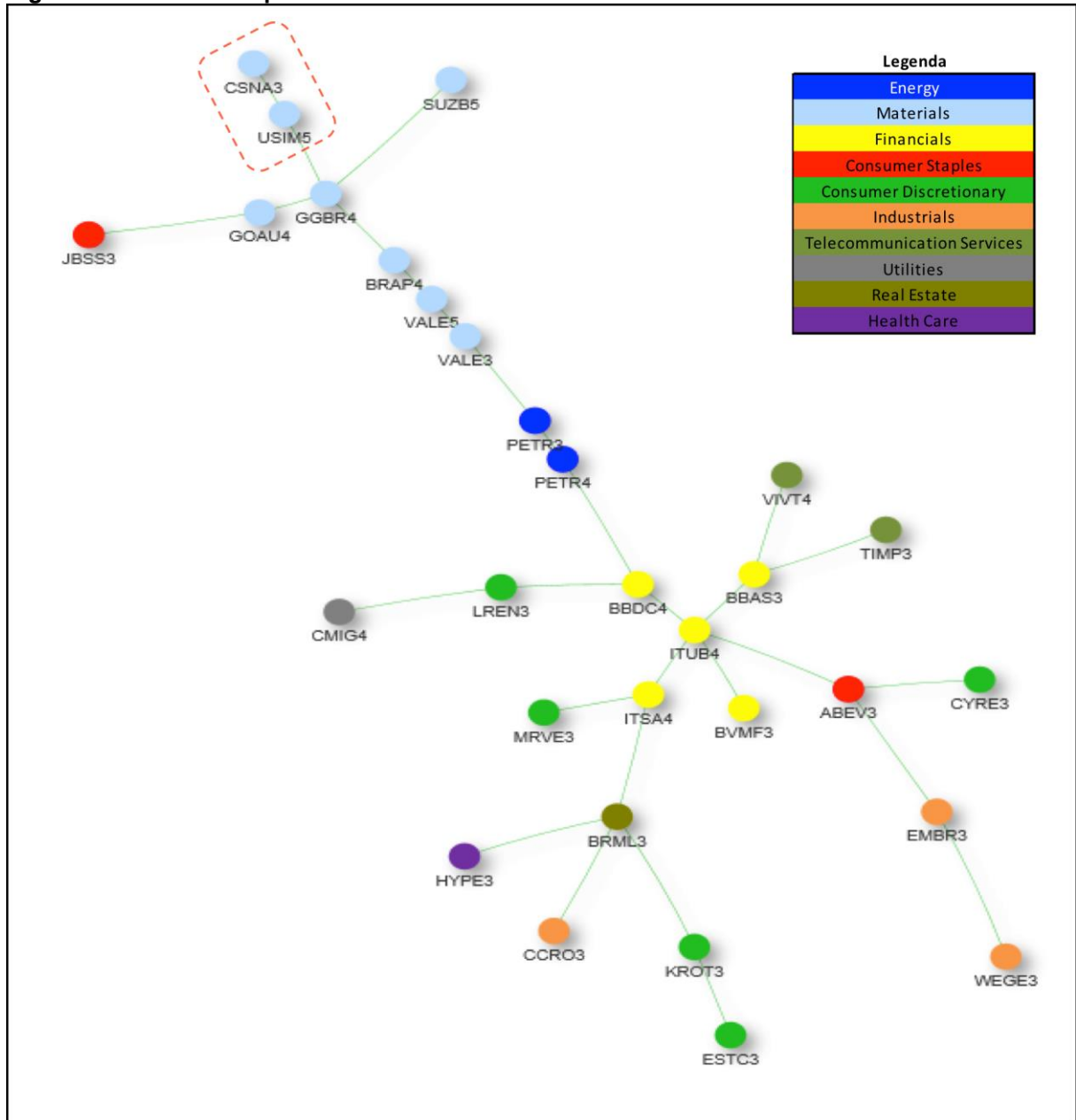


Fonte: Elaboração própria

Nota 1: MST gerada para o dia 04/01/2016. Data referente ao teste de cointegração.

Nota 2: Utilizados 63 dias úteis para os cálculos das correlações.

Figura 3 – MST – Exemplo de *Trade*: CSNA3 com USIM5



Fonte: Elaboração própria

Nota 1: MST gerada para o dia 25/09/2015. Data referente ao teste de cointegração.

Nota 2: Utilizados 63 dias úteis para os cálculos das correlações.

4 RESULTADOS

Após efetuar todos os testes de cointegração para todos os pares, foi feito o backtesting para os casos em que o teste de cointegração deu positivo para cointegração segundo as regras descritas na metodologia e foram guardados os resultados dos *PnLs* para posterior análise.

Estes resultados foram analisados segundo os critérios de agregação abaixo:

1. Agregado geral dos resultados, independentemente do setor dos ativos.
2. Para cada setor, agregado dos resultados em que os dois ativos pertençam ao mesmo setor analisado.
3. Agregado de todos os resultados de pares de ativos de mesmo setor. Ou seja, agregado dos valores do item 2.

Para cada um dos casos acima, os resultados foram separados em valores positivos e negativos e, além disso, foi feito um filtro (seleção) pela MST relacionada.

Este filtro foi feito de forma a manter apenas os *trades* de pares de ativos com links entre si em sua MST correspondente, ou seja, foi feita a seleção pela MST e foram guardados os resultados para comparação com o caso sem este filtro. Assim, foi analisado o impacto da seleção pela MST através do *PnL/trade* médio pré e pós seleção pela MST. Caso o *PnL/trade* médio seja superior de forma consistente para cada um dos casos acima, consideramos que a utilização de uma MST como seletor de pares de ativos otimizaria os resultados.

Através das tabelas seguintes é possível verificar os resultados agregados. Se considerarmos o caso mais geral (caso 1 acima), no qual foram comparados os resultados agregados de todos os pares de ativos, apenas diferenciando pela seleção pela MST, obtemos um *PnL/trade* inferior com a utilização da seleção (63 reais por *trade* médio versus 98 reais por *trade* médio sem a seleção), o que pode ser visto na tabela 3 abaixo.

Tabela 3 – Resultados de todos os pares sem filtro por setor de atuação

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	1.316	1.272	2.588	117	134	251
% trades	51%	49%	100%	47%	53%	100%
pnl/trade*	-1.077	1.313	98	-952	950	63
pnL total**	-1.417.305	1.670.733	253.427	-111.407	127.236	15.830
até início do trade ***	23	25	24	21	24	23
duração do trade ****	28	31	29	27	27	27

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Pulando o caso 2 que será analisado em seguida, para o caso 3 (tabela 4), agregando todos os ativos que pertencem ao mesmo setor de atuação e considerando todos os setores da amostra na análise, também houve desvantagem com a utilização da seleção pela MST, temos 52 reais por *trade* médio versus 294 reais por *trade* médio sem a seleção.

Tabela 4 – Resultados de todos os pares com filtro por setor de atuação

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	178	235	413	63	83	146
% trades	43%	57%	100%	43%	57%	100%
pnl/trade*	-978	1.257	294	-924	793	52
pnL total**	-174.079	295.437	121.357	-58.217	65.802	7.585
até início do trade ***	23	25	24	21	25	23
duração do trade ****	27	30	29	25	25	25

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Interessante notar que nos dois casos acima, independentemente do uso das MSTs para seleção dos pares de ativos, a utilização das cointegrações nos trouxe resultados positivos quando analisamos os *PnLs* totais de cada caso. Ou seja, desconsiderando os custos de transação, no agregado, os *trades* foram rentáveis. Para fazer uma análise mais fiel, teríamos que considerar todos os custos envolvidos como corretagens da corretora e emolumentos da bolsa de valores, além dos custos de venda a descoberto.

Considerando agora o caso 2, temos os agregados dos resultados para cada setor de atuação dos ativos, de forma isolada. Partindo da nossa lista de ativos, o setor mais relevante em termos de *PnL* total foi o de *Materials*. Mas ao adicionar a seleção por MST, além de uma queda elevada no *PnL* total (de 135.661 reais para 91 reais), o *PnL/trade* médio foi reduzido a 1 real apenas, conforme tabela 5. A seleção dos *trades* com a utilização da MST zerou o resultado em relação ao caso sem a seleção pela MST.

Tabela 5 – Resultados filtrados para o setor de *Materials*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	88	142	230	39	42	81
% trades	38%	62%	100%	48%	52%	100%
pnl/trade*	-1.085	1.628	590	-1.108	1.031	1
pnl total**	-95.450	231.111	135.661	-43.195	43.286	91
até início do trade ***	23	27	25	19	25	22
duração do trade ****	27	31	30	23	24	24

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Para os demais setores, tivemos resultados maiores de *PnL/trade* ao utilizar a seleção para o setor de *Financials* (tabela 7). Já para o setor de *Energy* (tabela 6), os resultados com a seleção pela MST foram equivalentes aos sem a seleção, de 272 reais por *trade* médio. Isso ocorreu pois todos *trades* existentes possuíam pares de ativos com links nas MSTs respectivas, não havendo influência da seleção.

Tabela 6 – Resultados filtrados para o setor de *Energy*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Com ou Sem seleção pela MST		
	pnl < 0	pnl > 0	total
trades	3	8	11
% trades	27%	73%	100%
pnl/trade*	-142	428	272
pnl total**	-425	3.420	2.996
até início do trade ***	42	19	26
duração do trade ****	42	37	38

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Nota1: Para este setor não houve diferença entre com ou sem seleção pela MST

Tabela 7 – Resultados filtrados para o setor de *Financials*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	31	60	91	9	28	37
% trades	34%	66%	100%	24%	76%	100%
pnl/trade*	-550	680	261	-370	527	309
pnL total**	-17.047	40.812	23.765	-3.328	14.757	11.429
até início do trade ***	21	24	23	14	27	24
duração do trade ****	23	25	24	25	22	23

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Já para os setores de *Consumer Staples*, *Consumer Discretionary*, *Industrials* e *Telecommunication Services* os resultados do *PnL* total foram negativos. De qualquer forma, dentre esses setores, verificamos que apenas o setor de *Consumer Discretionary* poderia ser considerado relevante na nossa análise, ao considerar quantidade total de *trades* efetuada (66 *trades* para o caso sem seleção). Já para *Industrials* houve 7 *trades* e para *Consumer Staples* e *Telecommunication Services* houve 4 *trades* no total para cada setor, no caso sem seleção por MST. Assim, esses setores podem ser desconsiderados da análise.

Tabela 8 – Resultados filtrados para o setor de *Consumer Staples*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	4	0	4	2	0	2
% trades	100%	0%	100%	100%	0%	100%
pnl/trade*	-685	0	-685	-818	0	-818
pnL total**	-2.739	0	-2.739	-1.636	0	-1.636
até início do trade ***	8	0	8	12	0	12
duração do trade ****	32	0	32	34	0	34

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Nota1: Para este setor, independentemente da seleção pela MST, houve perda.

Tabela 9 – Resultados filtrados para o setor de *Consumer Discretionary*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	45	21	66	8	5	13
% trades	68%	32%	100%	62%	38%	100%
pnl/trade*	-1.150	779	-537	-943	868	-246
pnL total**	-51.763	16.353	-35.410	-7.540	4.338	-3.202
até início do trade ***	24	22	23	23	12	19
duração do trade ****	29	36	31	22	21	21

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Nota1: Para este setor, independentemente da seleção pela MST, houve perda.

Tabela 10 – Resultados filtrados para o setor de *Industrials*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	3	4	7	0	0	0
% trades	43%	57%	100%	0%	0%	0%
pnl/trade*	-1.094	935	65	0	0	0
pnL total**	-3.283	3.740	457	0	0	0
até início do trade ***	19	13	15	0	0	0
duração do trade ****	16	22	19	0	0	0

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Nota1: Para este setor, independentemente da seleção pela MST, houve perda.

Tabela 11 – Resultados filtrados para o setor de *Telecommunication Services*

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

	Sem seleção pela MST			Com seleção pela MST		
	pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
trades	4	0	4	2	0	2
% trades	100%	0%	100%	100%	0%	100%
pnl/trade*	-843	0	-843	-1.047	0	-1.047
pnL total**	-3.373	0	-3.373	-2.093	0	-2.093
até início do trade ***	26	0	26	43	0	43
duração do trade ****	30	0	30	30	0	30

Fonte: Elaboração própria

* valores médios e em reais

** valores em reais

*** quantidade média de dias úteis decorridos desde o dia em que foi feito o teste de cointegração até o início do trade.

**** quantidade média de dias úteis decorridos desde o início do trade até o fechamento da posição

Nota1: Para este setor, independentemente da seleção pela MST, houve perda.

Tabela 12 – Resumo dos resultados

Comparação entre os resultados das cointegrações para os casos com e sem seleção pela MST

		Sem seleção pela MST			Com seleção pela MST		
		pnl < 0	pnl > 0	total	pnl < 0	pnl > 0	total
Caso 1 (sem filtro por setor)	trades	1.316	1.272	2.588	117	134	251
	% trades	51%	49%	100%	47%	53%	100%
	pnl/trade*	-1.077	1.313	98	-952	950	63
Caso 3 (com filtro por setor)	trades	178	235	413	63	83	146
	% trades	43%	57%	100%	43%	57%	100%
	pnl/trade*	-978	1.257	294	-924	793	52
Materials	trades	88	142	230	39	42	81
	% trades	38%	62%	100%	48%	52%	100%
	pnl/trade*	-1.085	1.628	590	-1.108	1.031	1
Financial	trades	31	60	91	9	28	37
	% trades	34%	66%	100%	24%	76%	100%
	pnl/trade*	-550	680	261	-370	527	309
Consumer Discretionary	trades	45	21	66	8	5	13
	% trades	68%	32%	100%	62%	38%	100%
	pnl/trade*	-1.150	779	-537	-943	868	-246

Fonte: Elaboração própria

* valores médios e em reais

Na tabela 12 acima, foi feito um resumo dos resultados das tabelas anteriores.

5 CONCLUSÕES

Conforme visto nos resultados, ao analisar a amostra total de *trades*, sem nenhum filtro por setor, apesar de o *PnL/trade* médio ser positivo para os casos com e sem seleção pela MST, houve queda em seu valor ao aplicar a seleção dos ativos pela MST.

Ao agregar os ativos por setores de atuação, seria esperado que houvesse uma melhora no resultado já que os ativos de um mesmo setor têm características similares. De fato, é possível notar aumento no *PnL/trade* médio para o conjunto de todos os setores (tabela 4), e para alguns setores analisados individualmente como para *Materials*, *Energy* e *Financials*, mas apenas sem a seleção dos *trades* pela MST. Ao aplicar a seleção, só não houve queda no *PnL/trade* médio para o setor de *Financials*.

Concluimos, então, que não houve evidência de que o uso de uma MST como seletora de pares de ativos em uma estratégia de *pairs trading*, que utilize cointegração, gera otimização na escolha dos pares a serem operados. Pelo contrário, houve piora dos resultados com sua utilização.

Possíveis melhorias para esse trabalho seriam no intuito de aumentar a amostra analisada tanto na quantidade de ativos quanto no intervalo de tempo utilizado. Foi verificado que para uma lista de 30 ativos, houve certa concentração em alguns setores, como no de *Materials*. Elevando a quantidade de ativos, potencialmente, obtém-se uma maior distribuição deles entre setores de forma a se conseguir comparar melhor os setores entre si. Outras métricas de aferição de resultados poderiam ser utilizadas, como os retornos percentuais, índices de *Sharpe* e retornos relativos a algum *benchmark* como o índice *bovespa*.

Uma outra métrica de seleção de pares através da utilização de uma MST a ser analisada, seria a de centralidade da MST. Verificando se um par de ativos mais central traria melhores resultados do que um par de ativos periférico da MST, já que o par mais central tem maior relevância na MST.

REFERÊNCIA BIBLIOGRÁFICA

ALEXANDER, C., Dimitriu, A., *The Cointegration Alpha: Enhanced Index Tracking and Long – Short Equity Market Neutral Strategies*, ISMA Finance Discussion Paper, n. 2002-08, 2002.

ALEXANDER, C., *Optimal hedging using cointegration*, *Optimal hedging using cointegration*, Philosophical Transactions of the Royal Society, v.357, n. 1758, p. 2039-2058, 1999.

ALEXANDER, Carol, *Market Models: A Guide to Financial Data Analysis*, New York: John Wiley & sons, 2001. 514 p.

BANERJEE, A., HENDRY, D., *Testing Integration and Cointegration: An Overview*, Oxford Bulletin of Economics and Statistics, v. 54, p. 225–255, 1992.

BONANNO, G., CALDARELLI, G., LILLO, F., MICCICHÈ, S., VANDEWALLE, N., MANTEGNA R. N., *Network of Equities in Financial Markets*, The European Physical Journal B, v. 38, n. 2, p. 363-371, 2004.

BONANNO, G., CALDARELLI, G., LILLO, F., MANTEGNA R. N., *Topology of Correlation Based Minimal Spanning Trees in Real and Model Markets*, PHYSICAL REVIEW E, v. 68, n. 4, p. 046130, 2003.

BONDY, J., A., MURTY, U., S., R., *Graph Theory with Applications*, North-Holland: Elsevier Science Ltd, 1976. 264 p.

BOSSAERTS, P., *Common nonstationary components of asset prices*, Journal of Economic Dynamics and Control, v. 12, n. 2-3, p. 347-364, 1988.

BUCKLE, M., CHEN, J., TONG, C., *A Network Visualization Approach and Global Stock Market Integration*, SSRN, 2559628, 2015.

CALDEIRA, J., F., MOURA, G., V., *Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage*, Brazilian Review of Finance, v.11, n. 1, 2013.

CHANG, E. C., LEWELLEN, W. G., *An Arbitrage Pricing Approach to Evaluating Mutual Fund Performance*, Journal of Financial Research, v. 8, p. 15-30, 1985.

ENGELBERG, J., GAO, P., JAGANNATHAN, R., *An Anatomy of Pairs Trading: the role of idiosyncratic news, common information and liquidity*, Third Singapore International Conference on Finance 2009, 2008.

ENGLE, R., F., GRANGER, C., W., J., *Co-integration and error correction: Representation, estimation and testing*, Econometrica, v. 55, p. 251—276, 1987.

GARAS, A., ARGYRAKIS, P., *Correlation study of Athens Stock Exchange*, Physica: Statistical Mechanics and its Applications, v. 380, n. 5, p. 399-410, 2007.

GRAHAM, R., L., HELL, P., *On the History of the Minimum Spanning Tree Problem*, IEEE Annals of the History of Computing, V. 7, N. 1, P. 43-57, 1985.

HJALMARSSON, E., ÖSTERHOLM, P., *Testing for Cointegration Using the Johansen Methodology when Variables are Near-Integrated*, International Finance Discussion Papers, BiblioGov, 2013.

JOHANSEN, S., *Cointegration: An Overview*, Symposium on Econometric Theory and Application (SETA), 2006.

JOHANSEN, S., *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*, Econometrica, v. 59, n. 6, p. 1551-1580, 1991.

JOHANSEN, S., *Times Series: Cointegration*, CREATES Research Paper, 38, 2014.

KHEYRKHAH, A., ROODPOSHTI, F. R., KAZEMI, M. A. A., *Using the Theory of Network in Finance*, International Journal of Finance and Managerial Accounting, v. 1, n. 2, 2016.

KHULLER, S., RAGHAVACHARI, B., YOUNG, N., *Balancing Minimum Spanning Trees and Shortest Path Trees*, Algorithmica, v. 14, n.4, p. 305-322, 1995.

LÉON C., LEITON, K., PERÉZ, J., *Extracting the sovereign's CDS market hierarchy: a correlation-filtering approach*. Physica A, v. 415, p. 407-420, 2014.

MAENG, S. E., CHOI, H. W., LEE, J. W., *Complex Networks and Minimal Spanning Trees in International Trade Network*, International Journal of Modern Physics: Conference Series, v.16, p. 51-60, 2012.

MANTEGNA, R. N., *Hierarchical Structure in Financial Markets*, The European Physical Journal B, v. 11, n. 1, p. 193-197, 1999.

MANTEGNA, Rosario. N., STANLEY, H. Eugene, *An Introduction to Econophysics: Correlations and Complexity Finance*, Cambridge University in Press, 2007. 167 p.

MARSH, I. W., STEVENS, I., *Large complex financial institutions: common influences on asset price behaviour?*, Financial Stability Review, Bank of England, p. 91-101, 2003.

MIAO, G. J., *High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach*, International Journal of Economics & Finance, v. 6, n. 3, p. 96, 2014.

MIZUNO, T., TAKAYASU, H., TAKAYASU, M., *Correlation networks among currencies*, Physica A: Statistical Mechanics and its Applications, v. 364, p. 336-342, 2006.

NAYLOR, M. J., LAWRENCE, C. R., MOYLE, B. J., *Topology of foreign exchange markets using hierarchical structure methods*, Physica A: Statistical Mechanics and its Applications, v. 382, p. 199-208, 2007.

REŠOVSKÝ, M., HORVÁTH, D., GAZDA, V., SINIČÁKOVÁ, M., *Minimum Spanning Tree Application in the Currency Market*, Biatec, v. 21, n. 7, p. 21-23, 2013.

RINCON, C. E. L., KIM, G-Y., MARTÍNEZ, C., LEE, D., *Equity Markets' Clustering and the Global Financial Crisis*, CentER Discussion Paper, v. 2016-016, 2016.

SINGHAL, P., SINHA, S., *Network Analysis of An Indian Stock Market Using The Minimum Spanning Tree Algorithm*, Journal of Quantitative Economics, v. 12, n. 2, jul/2014.

STOCK, J., H., WATSON, M., W., *Testing for common trends*, Journal of American Statistical Association, v.83, n. 404, p. 1097-1107, 1988.

TABAK, B., SERRA, T. R., CAJUEIRO D. O., *Topological Properties of Stock Market Networks: the case of Brazil*, Physica A: Statistical Mechanics and its Applications, v. 389, n. 16, p. 3240-3249, 2009.

TUMMINELLO, M. et al. *A Tool for Filtering Information in Complex Systems*, v. 102, n. 30, 2005.

VIDYAMURTHY, Ganapathy, *Pairs Trading, Quantitative Methods and Analysis*, Canada: John Wiley & Sons, 2014. 224 p.

YINGHUA, Z., *Stock Market Network Topology Analysis Based on a Minimum Spanning Tree Approach*, 2009.

Disponível em: <<https://etd.ohiolink.edu/>>. Acesso em: 10 ago. 2016.