

Inspere
Ciências Econômicas

Guilherme Augusto do Prado Giannasi

**Análise de interações *clusterizadas* do mercado acionário brasileiro antes e durante a
crise da Covid-19**

São Paulo

2022

GUILHERME AUGUSTO DO PRADO GIANNASI

**ANÁLISE DE INTERAÇÕES *CLUSTERIZADAS* DO MERCADO ACIONÁRIO
BRASILEIRO ANTES E DURANTE A CRISE DA COVID-19**

Trabalho de Conclusão de Curso
apresentado ao curso de Graduação em
Ciências Econômicas, como requisito
parcial para obtenção do grau de Bacharel
em Economia do Insper Instituto de
Ensino e Pesquisa

Orientador: Leonidas Sandoval Junior

SÃO PAULO

2022

Ficha Catalográfica

Giannasi, Guilherme Augusto do Prado

Análise de interações *clusterizadas* do mercado acionário brasileiro antes e durante a crise da Covid-19.

Guilherme Augusto do Prado Giannasi – São Paulo, 2022.

22 f.

Trabalho de Conclusão de Curso (Graduação) – Insper, 2022

Orientador: Leonidas Sandoval Junio

1. Crise do coronavírus 2. Bolsa de Valores 3. Brasil 4. Correlação no mercado de ações 5. *Clusterização* I. Guilherme Augusto do Prado Giannasi II. Análise de interações *clusterizadas* do mercado acionário brasileiro antes e durante a crise da Covid-19.

GUILHERME AUGUSTO DO PRADO GIANNASI

Análise de interações *clusterizadas* do mercado acionário brasileiro antes e durante a crise da Covid-19

Trabalho de Conclusão de Curso
apresentado ao curso de Graduação em
Ciências Econômicas, como requisito
parcial para obtenção do grau de Bacharel
em Economia do Insper Instituto de
Ensino e Pesquisa

Orientador: Leonidas Sandoval Junior

Banca Examinadora

Leonidas Sandoval Junior

Insper

Eduardo Correia de Souza

Insper

Resumo

A ocorrência da pandemia do novo coronavírus no mundo globalizado trouxe uma onda de incertezas ao mercado, fazendo com que bolsas caíssem livremente e com que a volatilidade mundial saltasse. Como resultado, deixou os investidores desprotegidos, uma vez que não se sabia como se portar diante de um fenômeno biológico cuja última aparição foi há aproximadamente 100 anos. Sendo assim, este artigo busca estudar as principais correlações que existiram e persistiram durante a crise provocada pelo Sars-CoV-2 no mercado de ações brasileiro. Para a realização, manuseiam-se dados financeiros diários e mensais da série histórica das ações. Ainda, propõe-se a criação de distâncias euclidianas a partir da correlação entre os preços das ações, usando, principalmente, as *Minimum Spanning Trees (MST)*, árvores que conectam os principais vértices (ações) e criam *clusters* que podem ser separados em áreas econômicas. Como resultados, espera-se que os investidores consigam sobreviver melhor a crises semelhantes que atinjam o Brasil.

Palavras-chave: Crise do Coronavírus. Bolsa de Valores. Brasil. Correlação no mercado de ações. *Clusterização*.

Abstract

The occurrence of the pandemics of the new coronavirus in the globalized world brought a wave of uncertainties to the market, causing stock markets to fall freely and world volatility to increase. As a result, it left investors unprotected, since it was not known how to behave in the face of a biological phenomenon whose last appearance was approximately 100 years ago. Thus, this article seeks to study the main correlations that existed and persisted during the crisis caused by Sars-CoV-2 in the Brazilian stock market. In order to analyze the crisis, daily and monthly financial data of the historical series of stocks were used. Furthermore, it is proposed to create Euclidean distances from the correlation between stock prices, mainly using the Minimum Spanning Trees (MST), trees that connect the main vertices (stocks) and create clusters that can be separated into economic areas. As a result, investors are expected to be able to survive similar crises that hit Brazil.

Keywords: Coronavirus crisis. Stock Market. Brazil. Correlation in the stock market. Clustering.

Sumário

1. Introdução	8
2. Revisão de Literatura	11
3. Metodologia	13
4. Análise de Dados	15
4.1. Dados	15
4.2. Preparação dos dados	16
4.3. Análise descritiva.....	16
5. <i>Minimum Spanning Trees</i>	25
5.1. Antes da crise	25
5.2. Durante a crise da Covid-19	32
6. Plano de ações	39
7. Conclusões	41
Referências	43
Anexos	45
Anexo 01	45
Anexo 02	45
Anexo 03	46

1. Introdução

O impacto de uma das maiores crises sanitárias do mundo (OMS, 2020) está começando a ser mensurado após certa estabilização do novo coronavírus. A repercussão da pandemia não ocorreu apenas no setor da saúde, mas trouxe inúmeras modificações a praticamente todos os aspectos da vida no mundo (NATIONS, 2020). Segundo o Relatório de Desenvolvimento Humano do Programa das Nações Unidas para o Desenvolvimento (2020), a Covid-19 intensificou, entre outros aspectos, a fome, o desemprego, e a desigualdade social no mundo. A economia, portanto, passou por um choque fortíssimo, e ainda o experencia por meio das diferenças de crescimento, inflação alta, perda de força do momentum da retomada do crescimento e ruptura na cadeia de suprimentos (OECD, 2021).

Boa parte dessa incerteza criada pela doença também foi repassada aos mercados financeiros. A partir do início da alta exponencial das infecções, quanto mais as bolsas mundiais entravam em queda livre, mais se via a volatilidade crescer (ALI; ALAM; RIZVI, 2020). Conforme Albulescu (2020) conclui em seu *paper*, quanto mais casos e mais mortes eram reportados, maior era a volatilidade no mercado financeiro, ao tomar S&P 500 como uma *proxy* mundial. No Brasil, essa situação se deteriora: a B3 ativou o botão do pânico – *circuit breaker* – 6 vezes em apenas 8 pregões durante março de 2020. Em 2021, por outro lado, a bolsa se manteve mais estável, porém ainda assim com quedas bruscas em certos momentos, devido principalmente a incertezas econômicas e políticas internas ao País (G1, 2021).

Não obstante, o cenário externo – do qual o Brasil apresenta uma certa dependência – tem se demonstrado rigoroso. O desaceleramento do crescimento mundial (OECD, 2021) pode ter efeitos no País, especialmente sobre as *commodities* exportadas. Além disso, líderes mundiais, como o presidente francês e o governo chinês, sinalizaram uma redução na demanda por exportações brasileiras. Isso tudo traz consequências a um dos *clusters* mais importantes da bolsa de valores brasileira: o de *commodities*, uma vez que empresas desse setor desempenham um papel importante na B3, como mostra Sandoval (2012) em um mapeamento do mercado brasileiro de ações no período da crise de 2008.

É importante ressaltar que o colapso dos *subprimes*, por um lado, foi marcado por uma queda acentuada no setor financeiro e nas *commodities* no Brasil, uma vez que as empresas brasileiras abertas desses setores apresentaram uma correlação significativa e positiva ao longo dessa crise (SANDOVAL; FRANCA, 2012). A crise da Covid-19, por outro lado, não se iniciou

no setor financeiro da economia, indicando possíveis efeitos díspares dos que aconteceram em 2008 na bolsa de valores brasileira. Apesar dessas diferenças entre as crises, o que ambas têm em comum são que afetaram negativamente e rapidamente o mercado financeiro.

Dessa forma, busca-se, neste artigo, examinar quais são as correlações mais fortes que perduraram ao longo da crise provocada pela pandemia nas empresas e nos setores do mercado financeiro do Brasil. Ainda, trar-se-ão variáveis macroeconômicas de controle para que se definam melhor as interações e correlações existentes entre cada parte da B3. Definido o escopo da proposta deste artigo, trazem-se dois trabalhos recentes com objetivos similares à análise deste estudo a fim de introduzir a tendência de pesquisa acadêmica na segmentação aqui proposta.

Zhang, Hu e Ji (2020), em *Financial Markets under the global pandemic of COVID-19*, estudam o comportamento do risco sistêmico entre os mercados financeiros de diversos países durante o início da crise da COVID-19. Para tal, os autores levantam os dados do risco de mercado por meio do desvio-padrão das bolsas de valores dos países considerados focos do novo vírus à época. Seguidamente, é feita uma análise de correlações para identificar a reação sistêmica à pandemia. Os autores alegam que houve um aumento do risco sistêmico por meio da diferença apresentada entre as correlações do início da pandemia e durante ela. Além do mais, algo que se assemelha às análises em Sandoval (2012, 2013) e em Sandoval e Franca (2012), Zhang, Hu e Ji trazem o método de *minimum spanning tree* a fim de mapear os *clusters* entre os mercados mundiais antes e depois do início da pandemia, sinalizando diferenças visíveis sobre as mudanças no risco. Por fim, os autores concluem que o vírus deixou as bolsas mais voláteis e incertas, sendo o risco conectado à gravidade da epidemia da COVID-19 no país em específico.

Em conformidade com as pesquisas recentes, este artigo propõe os elementos da análise por meio de correlações para criar *clusters* mapeando o comportamento do risco dos ativos financeiros. Há, também, a inclusão de variáveis macroeconômicas em determinadas análises para que se tenha uma visão mais geral do contexto em que se está fazendo o estudo das correlações. Ainda, levar-se-á em conta o tempo pelo qual as correlações perduram, a fim de construir uma análise de sobrevivência dos *clusters*, que, por conseguinte, identifica os setores mais robustos do mercado a fim de complementar a estratégia de sobrevivência do investidor em crises similares.

Para realizar as análises de *clusters*, primeiramente, extraem-se dados financeiros. Após tratá-los, parte-se para a metodologia, que se baseará em Mantegna (1999). Em seu artigo, demonstra-se uma técnica para tratar os retornos das ações por meio da correlação existente entre as empresas. Secundariamente, depois da engenharia de dados com as técnicas listadas pelo autor, é possível criar a *MST (Minimum Spanning Tree)*: uma rede de conexões entre os retornos das ações de modo que todos estejam conectados a pelo menos um outro, fazendo com que a soma das distâncias seja minimizada.

Tal visualização permite identificar os principais *clusters* no mercado brasileiro e ver o comportamento presente nos setores da economia a fim de trazer mais *insights* à análise da crise da Covid-19. Ao trabalhar com correlações, porém, nota-se que algumas relações exibidas podem ser originárias de ruídos aleatórios, o que requer uma atenção na interpretação do modelo. Investigar-se-á também quais setores da economia foram mais ou menos impactados pelo efeito do vírus, agregando esses resultados à estratégia traçada para sobreviver à crise da COVID-19.

Por fim, espera-se que os resultados indiquem uma correlação mais forte no setor exportador brasileiro, uma vez que foi um dos principais para segurar a economia durante a crise de 2020 (SALATI, 2021). Ainda, o setor financeiro, em comparação com 2008, pode não ser tão impactado devido às origens diferentes das crises. Adicionalmente, é importante destacar a onda de IPOs (*Initial Public Offering*) que a B3 presenciou durante esse período, indicando que novos setores possam emergir como *clusters* quando comparados ao estudo de Sandoval sobre o mapeamento da bolsa de valores brasileira (2012).

De maneira geral, após a pesquisa, espera-se que este artigo contribua às análises financeiras receitas com melhores estratégias de diversificação a fim de os investidores ao menos sobreviverem a crises semelhantes à causada pela Covid-19.

Definido o propósito deste artigo na primeira seção, tem-se a Revisão da Literatura na segunda seção. A terceira seção é a Metodologia, seguida da Análise de Dados em quarto. Nas seções cinco e seis, aplicam-se as metodologias discutidas aos dados e se discutem os planos de ações para os investimentos, respectivamente. A última e sétima é a conclusão do artigo.

2. Revisão da Literatura

São 4 os principais trabalhos em que este artigo se inspira. O primeiro é o *Correlation of financial markets in times of crisis*, em que Sandoval e Franca (2012) estudaram a correlação entre índices de bolsas de valores ao redor do mundo. Nele, os autores olham para as principais e mais recentes crises mundiais: a *Black Monday* em 1987, a Crise Russa em 1998, a bolha ponto-com em 2001, o choque depois do atentado terrorista em 11/09/2001 e a crise do *subprime* em 2008. Como resultados, as crises indicaram um aumento na correlação entre os índices antes do aumento da volatilidade de mercado e, ainda, a covariância se apresentou positiva para a correlação média e a volatilidade. Ademais, os autores destacam o peso que cada bolsa de valores tem na análise de correlações, mostrando que boa parte das contribuições para os resultados advém de mercados mais desenvolvidos. Ao fim, os pesquisadores indicam que os mercados tendem a andar juntos em momentos de crise e que as economias mais desenvolvidas se envolvem mais no cenário do que as outras.

No segundo estudo, Sandoval (2013), em *Cluster formation and evolution in networks of financial market indices*, traz uma perspectiva sobre o comportamento ao longo do tempo de *clusters* das bolsas mundiais tanto em períodos de estabilidade econômica quanto em períodos de maior volatilidade. Para a análise dos dados, há a utilização de técnicas que criam uma distância euclidiana a partir da correlação dos índices usados. Dessa forma, a visualização dos resultados é facilmente interpretável para os leitores. Ainda, como forma de mostrar robustez, Sandoval também executa outras técnicas que possam ser mais adequadas aos dados – como a correlação de Spearman – e encontra resultados parecidos. Depois da análise de sobrevivência dos *clusters*, a principal conclusão é que, muito embora haja mudanças na sobrevivência em períodos de crises, a maior formação das interações se dá por meios geográficos e entre mercados economicamente desenvolvidos, sendo os *clusters* norte-americano e o europeu mais persistentes.

O terceiro estudo em que este artigo se inspira é o *A map of the Brazilian stock market*, de Sandoval (2012), o qual visa mapear o comportamento do mercado de ações no Brasil durante a crise de 2008. A metodologia usada é parecida com o *paper* de Sandoval (2013), no qual há uso de correlações e medidas de centralidade a fim de se obter uma melhor visualização. Para a conclusão, então, o autor traz que, durante a crise do *subprime*, os setores financeiro e o de *commodities* foram os que mais obtiveram correlação e perduraram ao longo desse período conturbado.

Finalmente, o 4º artigo mostra como se aplicam as técnicas citadas de correlações e distâncias a dados financeiros. Em *Hierarchical structure in financial markets*, Mantegna (1999) define uma distância métrica a partir da correlação entre o preço de duas ações chegando a uma matriz de distância entre cada empresa listada. Essa matriz, por sua vez, indica que, quanto maior a correlação entre dois ativos, menor será a distância entre eles. O autor, então, parte para a construção de uma *MST (Minimal Spanning Tree)* associada à matriz de distância euclidiana feita. A árvore criada organiza as ações ao selecionar as conexões mais importantes da base de dados, formando os *clusters* e, ainda, permitindo a separação dos setores econômicos a que cada empresa pertence para trazer mais informações à análise.

Com a motivação dos 3 primeiros estudos e o aperfeiçoamento técnico deste último artigo, propor-se-á estudar as correlações por meio da *MST* das empresas na B3, a bolsa de valores brasileira, durante o período da crise provada pelo novo coronavírus. Assim, o mercado brasileiro pode ser melhor analisado mediante crises semelhantes.

3. Metodologia

O método utilizado neste artigo é aquele exposto por Mantegna (1999) em seu artigo, no qual se faz uso das *Minimum Spanning Trees* para indicar as relações de distanciamento entre ativos de um portfólio de ações. As *MSTs* são uma rede de nós conectados entre si por pelo menos uma aresta cuja soma das distâncias entre os vértices é minimizada. Para o problema proposto neste artigo, pretende-se criar uma rede de relações entre ativos do Índice Ibovespa em que os papéis são os pontos – os vértices – e são conectados entre cada um por meio da proximidade do comportamento entre cada ativo – as arestas.

A criação de tal árvore se dá por meio do estabelecimento de distâncias euclidianas entre cada ativo para que o algoritmo das *MSTs* escolha as melhores combinações entre os papéis. O primeiro passo, portanto, seria a definição de retornos dos preços de fechamentos para que os ativos tenham uma mesma base entre si:

$$Y_i \approx \ln P_i(t) - \ln P_i(t - 1) \quad (1)$$

em que Y_i é o retorno da ação i por meio da aproximação logarítmica da diferença do preço da ação no tempo t , $P_i(t)$, e seu preço um dia anterior, $P_i(t - 1)$. O próximo passo é fazer a matriz de correlação com todas as combinações possíveis de Y_i da base.

Secundariamente, a partir da matriz de correlação, as distâncias euclidianas são criadas mediante a equação (2):

$$d(i, j) = \sqrt{2(1 - \rho_{ij})} \quad (2)$$

em que $d(i, j)$ é a distância euclidiana entre as ações i e j , e ρ_{ij} é a correlação entre os retornos das ações i e j . $d(i, j)$, portanto, consegue tratar o quão próximos (ou distantes) estão os ativos, podendo gerar clusters. É importante ressaltar que $d(i, j)$ é uma distância euclidiana por satisfazer as três propriedades de uma distância no espaço métrico: simetria, não-negatividade, e obedecer à desigualdade triangular. Ainda, quando $d(i, j) = 0$, implica-se correlação perfeita entre os ativos i e j ($\rho_{ij} = 1$); e, quando $d(i, j) = 2$, há uma correlação totalmente imperfeita entre os ativos ($\rho_{ij} = -1$).

Transformada a matriz de correlação em uma de distâncias euclidianas, o método para construir uma *MST* é uma criação progressiva ligando todos os ativos em um gráfico caracterizado por uma distância mínima entre ações, realizado pelo algoritmo *emstreeR*

(QUADROS; GARMONSWAY, 2022) de forma direta. Como resultado, tem-se uma matriz das distâncias mínimas entre os ativos, de fácil visualização. No entanto, devido à presença de muitos ativos, faz-se necessário selecionar bandas para as distâncias, caso contrário, a visualização dos dados seria de difícil interpretação. Cria-se, portanto, uma análise baseada em *thresholds* a fim de estabelecer as conexões mais fortes e as mais fracas ao filtrar os valores das distâncias entre os ativos. Os filtros são mais bem explicados na seção 5, das *Minimum Spanning Trees*.

Paralelamente à construção das árvores de menores distâncias entre os ativos, usa-se a matriz de distâncias para definir os pontos que os ativos ocupam dentro do plano cartesiano. Isso é feito para que se consiga imprimir de forma mais fácil as árvores a partir da posição dos pontos e das distâncias mínimas definidas. Essa definição é feita pela técnica do *Multidimensional Scaling*, criada por Kruskal e Wish (1978), em que os pontos são encontrados por meio das distâncias.

Após a seleção e o tratamento dos dados do modelo, a interpretação dos resultados pode ser feita por meio da visualização dos gráficos gerados. Isso auxilia na detecção de quais *clusters* são possivelmente formados. Definida a metodologia deste artigo, parte-se para o aprofundamento nos dados e na execução do que fora aqui exposto.

4. Análise de dados

4.1. Dados

Definida a metodologia, parte-se para a extração e manipulação de dados, sendo esta feita em R. Trabalhando em um primeiro momento apenas com a parte de *equity*, obtiveram-se 1.476 observações de fechamentos diários ajustados para proventos para 90 empresas listadas na bolsa brasileira e presentes no Índice Bovespa de 31/12/2016 a 26/08/2022 da plataforma *Economática* (“Economática”, 2022). Isso permite uma análise extensiva do mercado antes e durante a crise da Covid-19. Em adição aos principais dados de análise, extraiu-se a carteira teórica do índice Ibovespa do dia 29/08/2022 do site da B3¹ (“B3”, 2017); e os pontos de fechamento do Ibovespa do site *Yahoo Finance* (“Yahoo Finance”, 2022).

No entanto, algumas empresas da base de dados abriram seu capital durante o intervalo de tempo, fazendo com que haja muitas observações faltantes. Além disso, o período desde o início de 2017 se mostra longo para o proposto neste artigo, uma vez que há fechamentos para pouco mais de 2 anos após o choque da pandemia. Faz-se necessário, então, uma filtragem para 1) as empresas com capital fechado durante o período selecionado; e 2) para o íterim selecionado a fim de manter a comparabilidade entre os períodos pré e durante a pandemia.

Para fazer *clusters* supervisionados, ainda, foi extraída uma base de dados da B3 que lista as empresas e seus respectivos setores econômicos, subsetores e segmentos. Ao decorrer da análise, serão usados apenas os setores econômicos. Os segmentos e subsetores serão deixados de lado, uma vez que são classificações muito específicas de cada empresa, o que extrapola o intuito macroeconômico da análise.

Finalmente, separa-se o cerne dessa análise em dois grandes intervalos de tempo. O primeiro se dá pelo período antes da pandemia, de 21/05/2018 a 10/03/2020. O segundo, pelo íterim entre a caracterização da Covid-19 como pandemia pela OMS (2020) (11/03/2020) até a estabilização da doença no Brasil ao final de 2021 (31/12/2021). A escolha do filtro de tempo foi baseada primeiramente no período durante a pandemia, no qual há menos dados. Definido tal intervalo, escolheu-se o dia anterior ao choque exposto pela OMS e subtraiu-se dessa data a mesma quantidade de dias existente no decorrer da Covid-19, o que resultou no intervalo apontado.

¹ Apesar de a carteira ser de 29/08/2022, as perdas de informações não são altas uma vez que não se usará tal base para o cerne deste artigo.

4.2. Preparação dos dados

Para realizar as próximas etapas com os dados, é preciso tratá-los. Filtra-se para os fechamentos obtidos depois ou iguais a 21/05/2018. Dentre esse intervalo, descobrem-se quais foram as empresas que abriram seu capital a fim de retirá-las do portfólio, pois elas trariam incomparabilidade aos dados. Da amostra de 90 papéis, tiram-se 10 empresas que abriram seu capital no decorrer de tempo estabelecido para a análise.

A fim de transformar os preços de fechamento em uma série temporal comparável com os outros ativos e os colocar no modelo, é preciso transformar os preços de fechamento em retornos. Para tal, os fechamentos são transformados em escala logarítmica e, então, subtrai-se a observação do log do preço de fechamento em t pela em $t - 1$, como supracitado na equação (1): $Y_i \approx \ln P_i(t) - \ln P_i(t - 1)$. Os log-retornos são quase iguais aos retornos da ação, podendo aproximá-los. A matriz de correlação, então, é feita com os log-retornos para ser usada no modelo das *Minimum Spanning Trees*.

4.3. Análise descritiva

A análise descritiva primeiramente traz dados mais relacionados aos retornos das ações para depois aprofundar nos setores e então nos períodos selecionados da amostra. Ainda, as análises focam mais na média dos retornos obtidos de todas as ações por se tratar de uma base com 80 ativos, fazendo-se inviável o olhar individual, e há mais foco nos setores econômicos, pois já representam uma análise *clusterizada* dos ativos.

Começando pela análise descritiva dos ativos e dos setores, tem-se a tabela 1. Nela, é possível ver todas as medidas descritivas mais importantes para entender o comportamento de cada variável. Como a inclusão de cada ativo na tabela seria inviável para a análise, decidiu-se por fazer a média por dia dos retornos diários de cada ativo, chegando à variável *Retorno médio diário*. Da tabela, assim, destaca-se o que já se imaginava sobre a Bolsa: uma média de retornos médios diários menor durante a crise e maior antes dela, sendo cerca de 8 vezes maior.

Ainda, alguns setores se destacam, como o de Materiais Básicos e o de Tecnologia da Informação. O primeiro apresentou uma média de queda antes do choque, mas, durante a pandemia, foi o setor com a maior média de retornos. O segundo, por sua vez, se mostrou robusto, tendo uma média não tão díspar tanto no período de maior austeridade econômica quanto no de menor, diferentemente do que aconteceu com os outros setores. Além disso, o de

TI foi o que obteve a maior queda (retorno mais negativo) e a maior subida (retorno mais positivo) durante a pandemia, corroborando para mostrar o quão volátil esse setor é.

No mais, é possível identificar que a amplitude (percentil 100% - percentil 0%) aumentou durante a crise, o que ajuda a visualizar a maior volatilidade nos mercados causada pelo vírus. A mediana (percentil 50%), por sua vez, ajuda a mostrar a crise: enquanto no período antes da pandemia há apenas um setor com mediana negativa, observam-se três setores com medianas de retornos negativas durante a crise. Finalmente, com os histogramas, é possível ter uma melhor visualização de onde se encontram os retornos ocorridos ao longo do tempo.

Não obstante as medidas descritivas, analisa-se também especificamente o retorno médio diário dos ativos. O gráfico 01 mostra a densidade dessa variável no conjunto da amostra inteira. A ponta da distribuição é bem alta, sendo a moda um pouco maior que zero, ao passo que a média, segundo a tabela 01, é próxima de zero, e a mediana parece estar mais próxima à moda. Ainda, nota-se uma cauda mais pesada à esquerda do que à direita, indicando que possa haver mais retornos negativos do que positivos longe da média. A maior quantidade de quedas grandes, por sua vez, pode representar uma maior correlação nos mercados, como mencionado na Revisão de Literatura, indicando que as árvores de mínima distância possam ter distâncias mais próximas de 1.

Como muitos dos ativos estão presentes no índice Bovespa, é interessante analisar a variável do retorno médio diário com o índice a fim de identificar o quão parecidos são. Dessa forma, fez-se o gráfico 02, o que mostra a semelhança entre os retornos médio diários com os retornos do Ibovespa, reforçada pela inclinação da reta presente, que se aproxima de 45° . Investigando as causas, tem-se que 96% do índice Bovespa em 29/08/2022 era formado pelos ativos presentes na base de dados², indicando, então, que os dados para a análise da *MST* representam bem o Ibovespa.

No entanto, para se ter uma ideia melhor de como os ativos se comportam diante dos retornos do Ibovespa, traz-se a divisão por setores em um gráfico similar ao 02. No gráfico 03, o retorno médio diário é feito de acordo com o setor de cada ativo, o que ajuda a ressaltar o comportamento de cada setor frente ao índice. Destacam-se, assim, o setor de Bens Industriais, Consumo Cíclico, Petróleo, Gás e Biocombustíveis e o de Tecnologia da Informações, cujas retas possuem uma inclinação maior. Isso significa que as movimentações desses setores são mais intensas que as movimentações do mercado, ou seja, em situações de crescimento do

² De acordo com os pesos da carteira teórica, retirada do site da B3

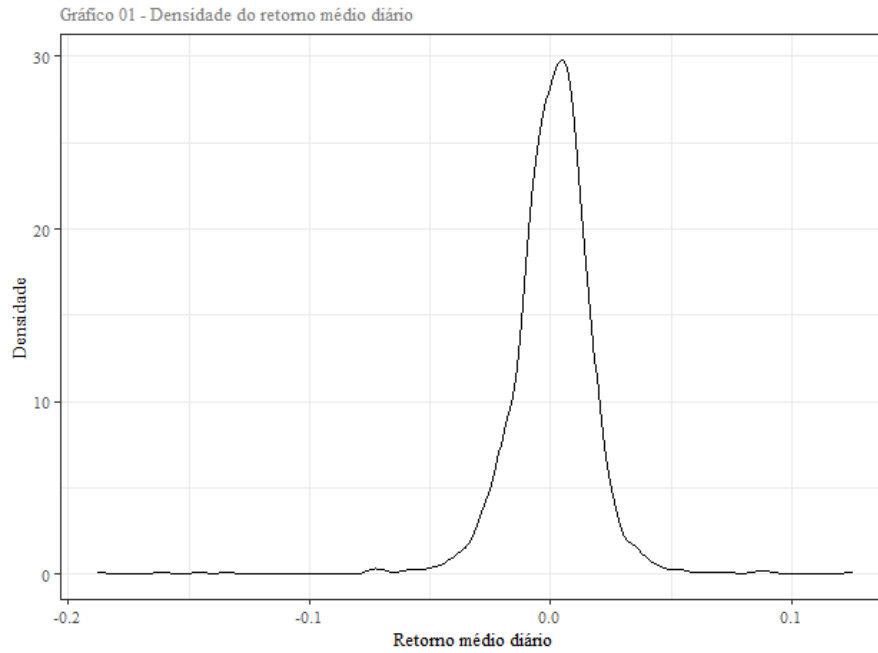
Ibovespa, há um crescimento maior que o da média em ativos desses setores. Em outras palavras, os betas de mercado desses setores tendem a ser maiores que os dos outros setores.

Tabela 01 - Medidas descritivas

	Média	Desvio Padrão	Percentil 0%	Percentil 25%	Percentil 50%	Percentil 75%	Percentil 100%	19 Histograma
1. Período todo - de 22/05/2018 a 30/12/2021								
Retorno médio diário	0,047%	1,971%	-18,774%	-0,775%	0,168%	0,999%	12,482%	
1.1 Setores								
Bens Industriais	0,028%	2,713%	-26,971%	-1,134%	0,101%	1,267%	15,172%	
Comunicações	0,031%	1,816%	-11,547%	-0,917%	-0,005%	0,950%	9,194%	
Consumo Cíclico	0,020%	2,702%	-21,284%	-1,153%	0,024%	1,447%	11,856%	
Consumo não Cíclico	0,062%	1,826%	-16,655%	-0,764%	0,139%	0,944%	12,951%	
Financeiro	0,013%	2,148%	-17,980%	-0,999%	0,083%	1,126%	13,315%	
Materiais Básicos	0,067%	2,224%	-18,651%	-0,988%	0,108%	1,255%	14,159%	
Petróleo, Gás e Biocombustíveis	0,076%	2,676%	-24,508%	-0,937%	0,128%	1,288%	16,579%	
Saúde	0,028%	1,890%	-15,238%	-0,911%	0,011%	1,074%	7,560%	
Tecnologia da Informação	0,135%	2,959%	-29,539%	-1,209%	0,136%	1,436%	20,671%	
Utilidade Pública	0,085%	1,841%	-14,925%	-0,842%	0,174%	1,059%	11,094%	
2. Pré Covid - de 22/05/2018 a 10/03/2020								
Retorno médio diário	0,084%	1,517%	-13,434%	-0,627%	0,224%	0,945%	7,299%	
2.1 Setores								
Bens Industriais	0,070%	1,917%	-13,614%	-0,888%	0,190%	1,161%	8,627%	
Comunicações	0,069%	1,536%	-7,366%	-0,811%	0,078%	1,054%	8,138%	
Consumo Cíclico	0,137%	2,052%	-14,642%	-0,957%	0,275%	1,448%	11,163%	
Consumo não Cíclico	0,059%	1,576%	-12,907%	-0,732%	0,146%	0,940%	6,403%	
Financeiro	0,077%	1,535%	-11,568%	-0,810%	0,155%	1,024%	4,508%	
Materiais Básicos	0,019%	1,903%	-15,935%	-0,951%	-0,003%	0,992%	9,195%	
Petróleo, Gás e Biocombustíveis	0,043%	2,255%	-24,486%	-0,849%	0,126%	1,175%	10,105%	
Saúde	0,101%	1,542%	-9,971%	-0,743%	0,041%	1,050%	6,656%	
Tecnologia da Informação	0,142%	2,238%	-14,300%	-0,943%	0,137%	1,164%	10,547%	
Utilidade Pública	0,150%	1,515%	-9,405%	-0,629%	0,332%	1,052%	5,784%	
3. Durante Covid - de 11/03/2020 a 30/12/2021								
Retorno médio diário	0,011%	2,337%	-18,774%	-0,899%	0,110%	1,089%	12,482%	
3.1 Setores								
Bens Industriais	0,012%	3,320%	-26,971%	-1,365%	0,004%	1,379%	15,172%	
Comunicações	0,006%	2,058%	-11,547%	-1,004%	-0,087%	0,864%	9,194%	
Consumo Cíclico	0,096%	3,217%	-21,284%	-1,352%	-0,156%	1,411%	11,856%	
Consumo não Cíclico	0,064%	2,045%	-16,655%	-0,770%	0,125%	0,947%	12,951%	
Financeiro	0,051%	2,619%	-17,980%	-1,305%	-0,036%	1,241%	13,315%	
Materiais Básicos	0,153%	2,501%	-18,651%	-1,065%	0,220%	1,494%	14,159%	

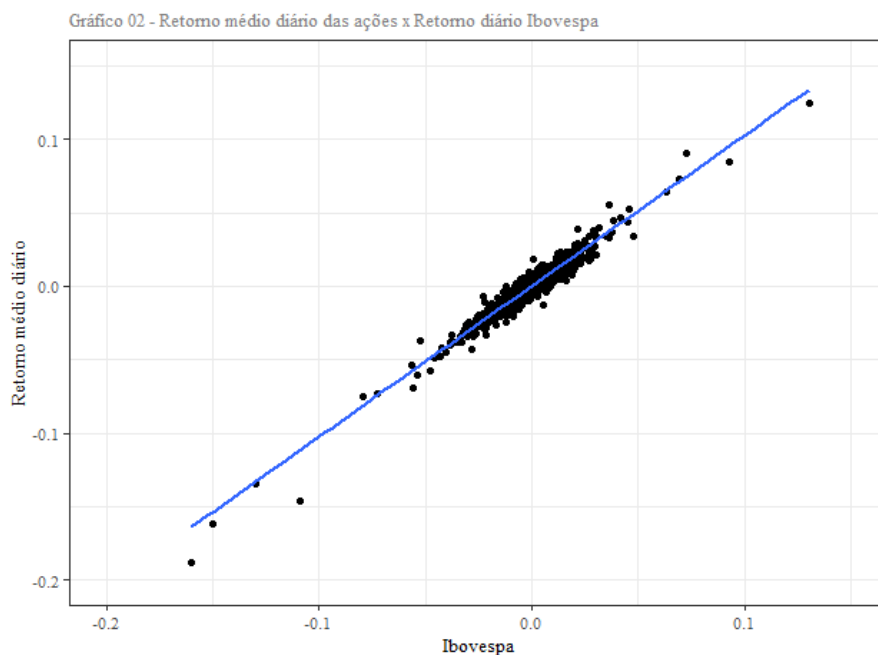
Petróleo, Gás e Biocombustíveis	0,109%	3,039%	-24,508%	-1,173%	0,142%	1,395%	16,579%	---■---
Saúde	0,043%	2,180%	-15,238%	-1,075%	-0,011%	1,097%	7,560%	---■---
Tecnologia da Informação	0,128%	3,534%	-29,539%	-1,632%	0,128%	1,877%	20,671%	---■---
Utilidade Pública	0,021%	2,115%	-14,925%	-0,966%	0,077%	1,080%	11,094%	---■---

Fonte: elaboração própria, dados do Economática e da B3



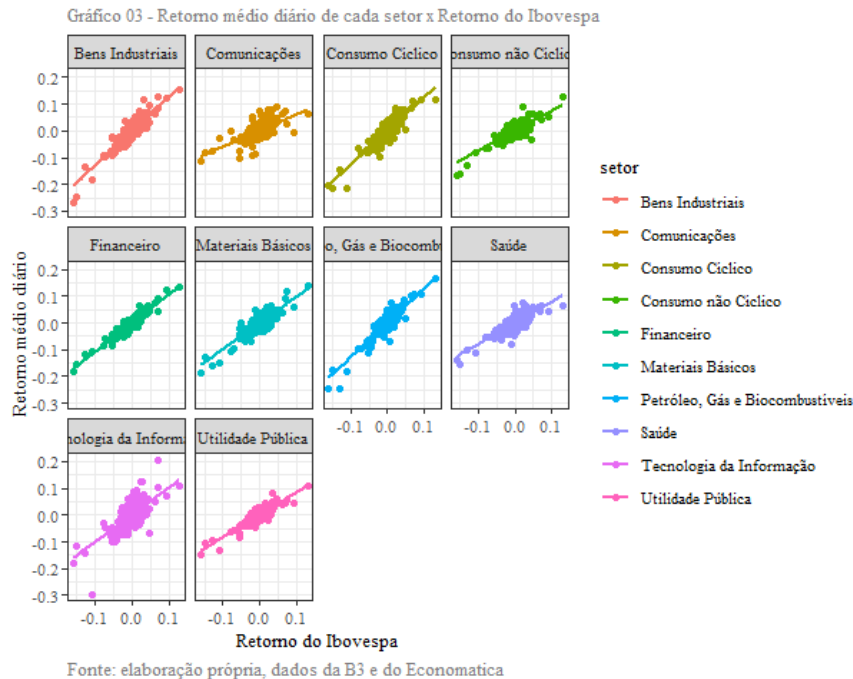
Fonte: elaboração própria, dados do Economática

O gráfico 01 mostra os o histograma dos retornos médios diários por meio da densidade a cada valor mostrado, assim se tem uma maior concentração da variável x próximo do ponto 0, com caudas para ambos os lados.



Fonte: elaboração própria, dados do Economática e do Yahoo

O gráfico 02 mostra uma dispersão entre os retornos médios diários contra os retornos diários do Ibovespa. Com isso, é possível ver o quanto o retorno médio diário das ações varia em relação ao do Ibovespa. Neste caso, percebe-se uma variação quase idêntica ao do Ibovespa.

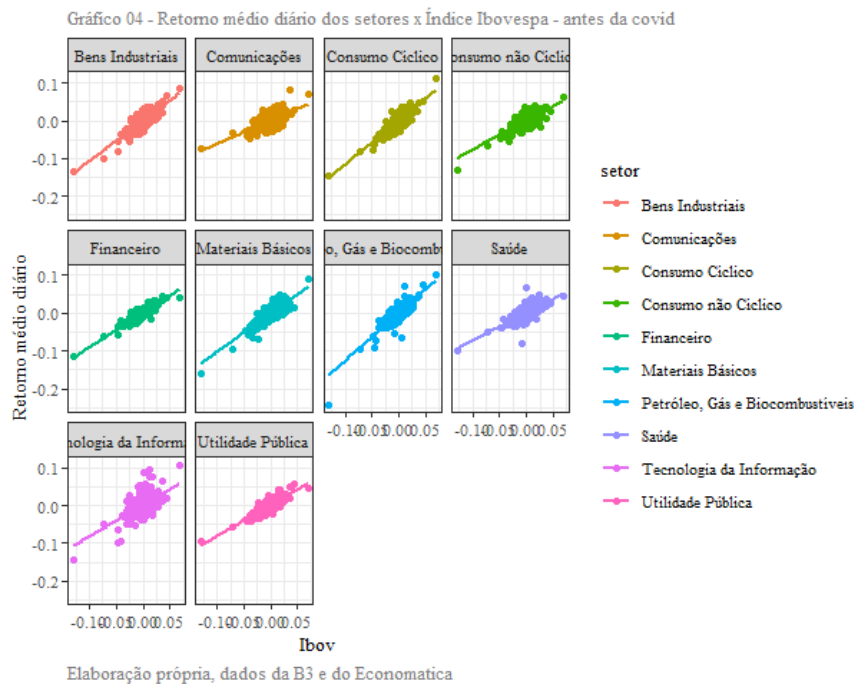


O gráfico 03 segue a mesma lógica do 02, porém separando os retornos médios por setores, também contra o retorno do Ibovespa. É possível ver, pois, o quanto o retorno médio diário de cada setor varia em relação ao índice do mercado e se determina os que costumam variar mais ou menos.

Ainda, conjuntamente com a tabela 01, identifica-se tal comportamento por meio dos desvios padrão. Os 4 setores destacados apresentam as maiores volatilidades dos setores: todos têm desvio padrão maior que 2,6%, pouco mais de 0,6pp a mais do que o desvio do retorno médio diário dos ativos. Esse comportamento mais volátil pode indicar distâncias menores entre ativos dentro desses setores na *MST*, devido ao fato de que mercados mais voláteis também apresentam maiores correlações, como demonstrado na Revisão de Literatura.

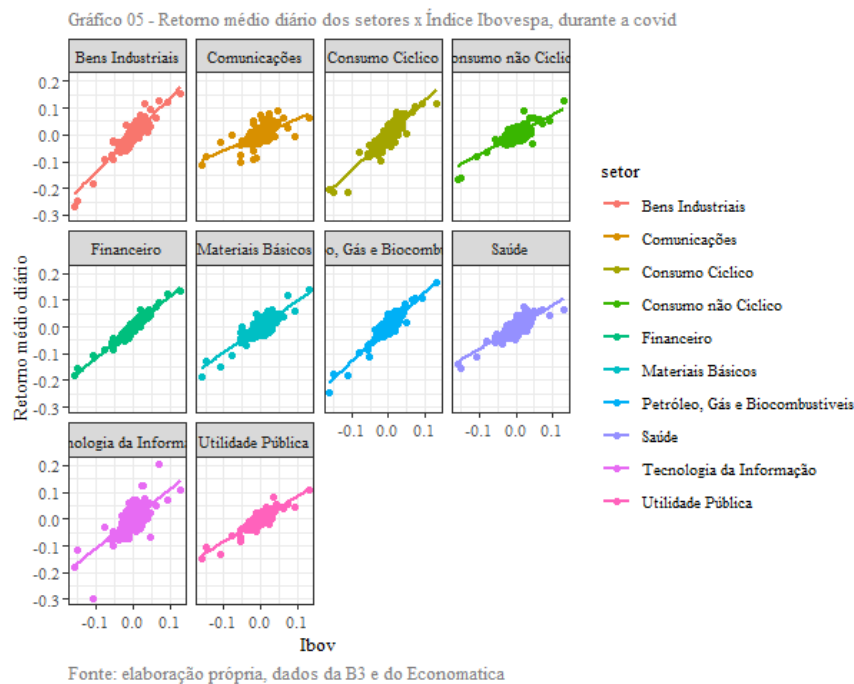
Finalmente, separa-se o gráfico 03 nos períodos que serão analisados com o intuito de identificar quaisquer possíveis mudanças de comportamento antes e após o choque do vírus. Os gráficos 04 e 05 mostram, principalmente, que houve mais ocorrências de retornos fora da nuvem de pontos, significando possivelmente mais casos de *outliers* para a amostra. Assim, fica visível como se deu o aumento da volatilidade durante a pandemia: mais retornos díspares entre si.

Por fim, faz-se uma breve análise de correlação entre os retornos médios diários dos setores



O gráfico 04 é o gráfico 03 filtrado para o período anterior à crise da covid, mostrando o que seria um comportamento típico dos setores e do mercado antes da crise.

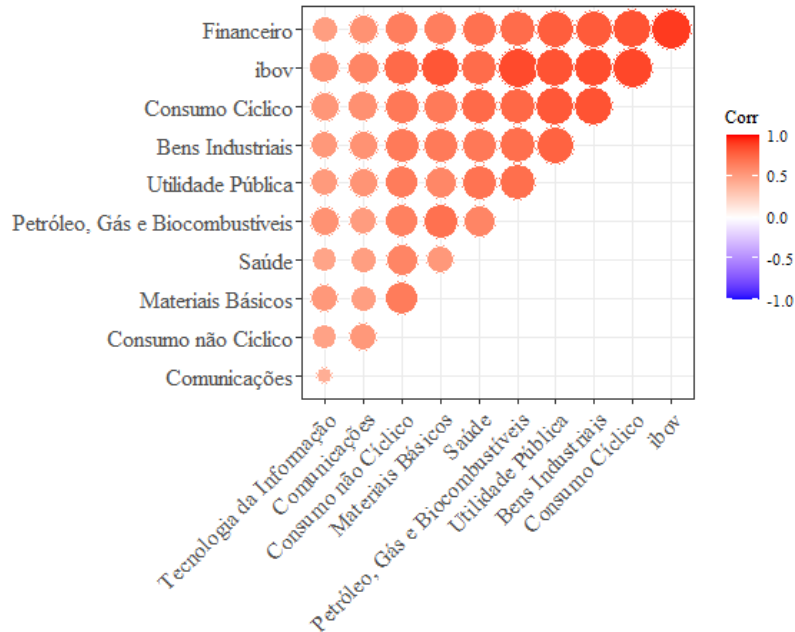
dos ativos a fim de determinar o que se espera de resultados da *MST*. Os gráficos 06 e 07 – feitos com os retornos médios diários de cada setor – mostram que de fato houve um aumento de correlação entre os setores durante a crise, como fora supracitado ao longo deste artigo.



O gráfico 05 é o gráfico 03 filtrado para o período durante a crise da covid, mostrando o que seria um comportamento atípico dos setores e do mercado por se encontrar dentro do choque.

Ainda, há alguns setores que possuem maiores correlações com os outros, como o de Finanças, que se interliga significativamente e todos, e especialmente ao Ibovespa. O setor de Tecnologia da Informação, por sua vez, apesar de ser o que menos demonstrou correlação com os outros

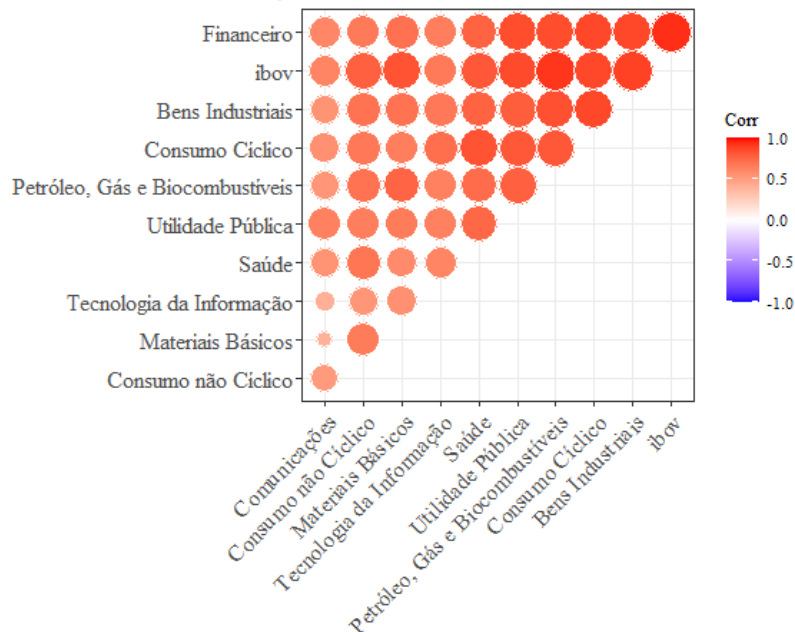
Gráfico 06 - Matrix de correlação antes da covid



Fonte: elaboração própria, dados da B3, do Economatica e do Yahoo Finance

O gráfico 06 é um mapa de correlação para todos os setores da B3 com o que fora observado antes da crise da covid, mostrando o que seria uma correlação comum neste período. Quanto maiores as bolas, mais fortes são as correlações entre os ativos que se cruzam nos eixos. Ainda, a cor, além de ser mais intensa com correlações mais fortes, representa o sinal da correlação, sendo vermelho positiva e azul negativa. No caso, não há correlações negativas.

Gráfico 07 - Matrix de correlação durante a covid



Fonte: elaboração própria, dados da B3, do Economatica e do Yahoo Finance

O gráfico 07 é também um mapa de correlação para todos os setores da B3 com o que fora observado durante a crise da covid, mostrando o que seria um comportamento incomum para os setores em tempos de crise. Novamente, observa-se a falta de uma correlação negativa entre os setores da B3 durante a crise da covid-19.

antes da pandemia, subiu algumas posições após o choque. Ainda, há evidências de que o setor de Comunicações se manteve com correlações baixas com os setores mesmo após o choque, chegando até a reduzir sua interação com o setor de Materiais Básicos, por exemplo.

Após as análises descritivas feitas, esperam-se as distâncias nas *MSTs* sejam menores durante a pandemia devido aos aumentos de correlação entre os ativos e à maior volatilidade no período. Ainda, alguns setores serão mais presentes em toda a análise, como o financeiro; enquanto outros podem se apresentar mais tímidos antes da pandemia, mas serem mais significativos na análise depois do choque. Isso traz importantes implicações de estratégias de investimento de acordo com o perfil do investidor, por exemplo estar comprado no setor Financeiro e no de Tecnologia da Informação pode significar maiores volatilidades e correlações positivas, intensificando os ainda mais os riscos e os retornos; ao passo que o setor de Comunicações parece ser para alguém mais avesso ao risco no período antes da pandemia.

5. *Minimum Spanning Trees*

Como supramencionado na parte de Metodologia, para o algoritmo calcular as *MSTs*, é preciso que primeiramente se tenha uma matriz de correlação entre todos os ativos. Este, então, é transformado em uma matriz de distâncias entre os ativos por meio da fórmula de Mantegna (1999). Assim, o algoritmo calcula qual é a árvore que interliga todos os pontos de forma a se ter a menor soma das distâncias entre todos a partir da matriz quadrada de distância.

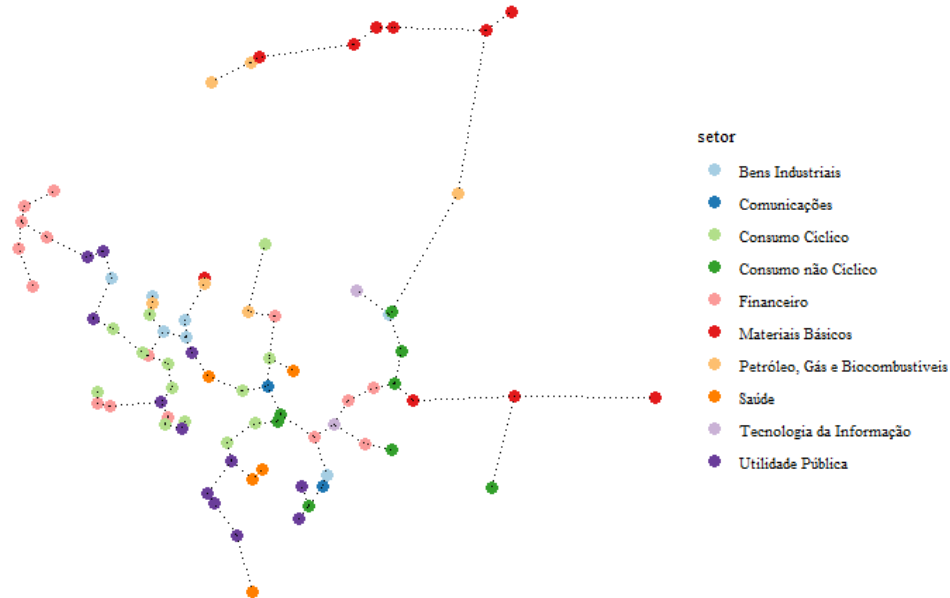
Após os cálculos do algoritmo, inserem-se os dados em gráficos para que se consiga visualizar melhor os resultados obtidos. Começa-se primeiramente pela análise do período todo já com os *clusters* a fim de identificar possíveis interações agrupadas. Depois, repete-se tal análise para os períodos estabelecidos e fazendo uso de *thresholds* com o objetivo de reconhecer as interações mais relevantes para ambos os períodos e diferenciá-las para uma melhor sobrevivência dos investimentos ao longo de crises parecidas. Os níveis de *thresholds* foram definidos de acordo com uma análise das distâncias entre os dois períodos listados e o total, indicando a filtragem pelos níveis: menores que 0,025; 0,05; 0,075; 0,125; e para todos os valores. A amplitude de 0,025 nos três primeiros ocorre devido às maiores concentrações de distâncias, o que ajuda a identificar mais precisamente os surgimentos dos *clusters* e suas respectivas interações. O aumento da amplitude, por outro lado, é justificado pela baixa concentração de distâncias nos intervalos, o que carrega menos informações para o proposto neste artigo. Os gráficos com os nomes dos ativos, por sua vez, encontram-se na parte dos anexos, sendo o anexo 01 o gráfico para o período antes da crise, e o 02, para o durante a crise.

5.1. Antes da crise

A análise de antes da pandemia começa com o gráfico 08, o qual representa a *Minimum Spanning Tree* para o período antes da crise, quando se pode ter uma ideia de como o mercado comumente se comportava antes da pandemia. Fora da nuvem de pontos, observa-se uma forte e isolada relação entre os ativos do setor Financeiro e de Materiais Básicos, de forma separada, com este ainda comportando dois ativos da área petrolífera: PETR4 e PETR3, como mostra o anexo 01. Tal interação provavelmente se dá devido à exposição em comum que as empresas de Materiais Básicos e a Petrobrás têm aos preços de matérias-primas básicas no mundo, como as *commodities* para o primeiro e o preço internacional do petróleo para o segundo. O de Utilidade Pública, apesar de ser um tanto quanto disperso, possui seus grupos, com CMIG4, ELET3 e ELET6 mais próximos do setor Financeiro; ao passo que ENGI11, ENGIE3, ENBR3 e EQTL3 formam uma interação ao fundo do gráfico, sendo muito possível derivada da alta

correlação que o setor energético possui no Brasil. A diferença entre os ativos de energia pode ser originária nas diferenças de tamanhos entre as empresas, sendo as mais correlacionadas com

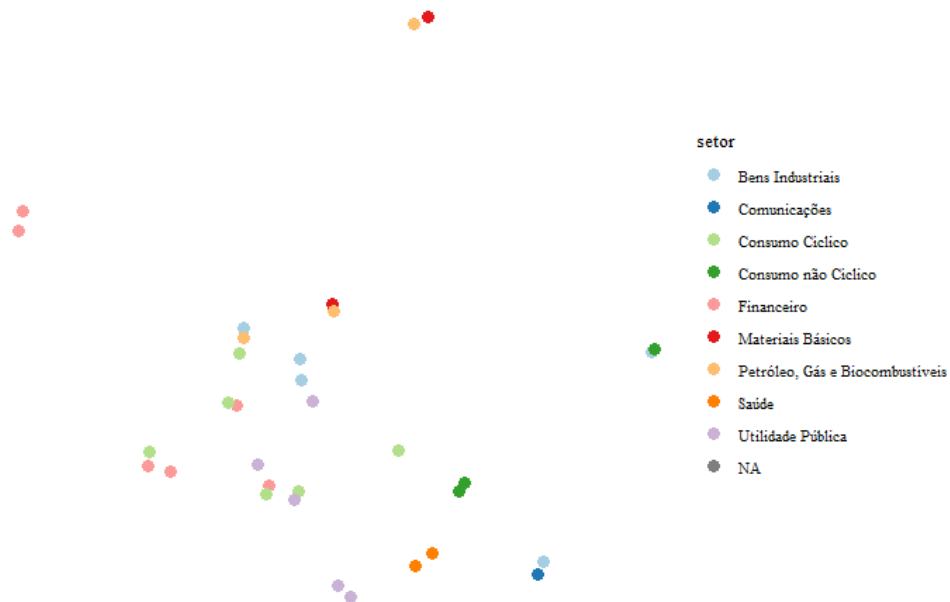
Gráfico 08 - Minimum Spanning Tree antes da crise da covid



Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 08 mostra a árvore de distâncias mínimas para todos os ativos antes da pandemia e colorida de acordo com seus setores. É possível identificar relações peculiares entre os ativos e seus setores.

Gráfico 09 - Minimum Spanning Tree antes da crise da covid - threshold de 0,025



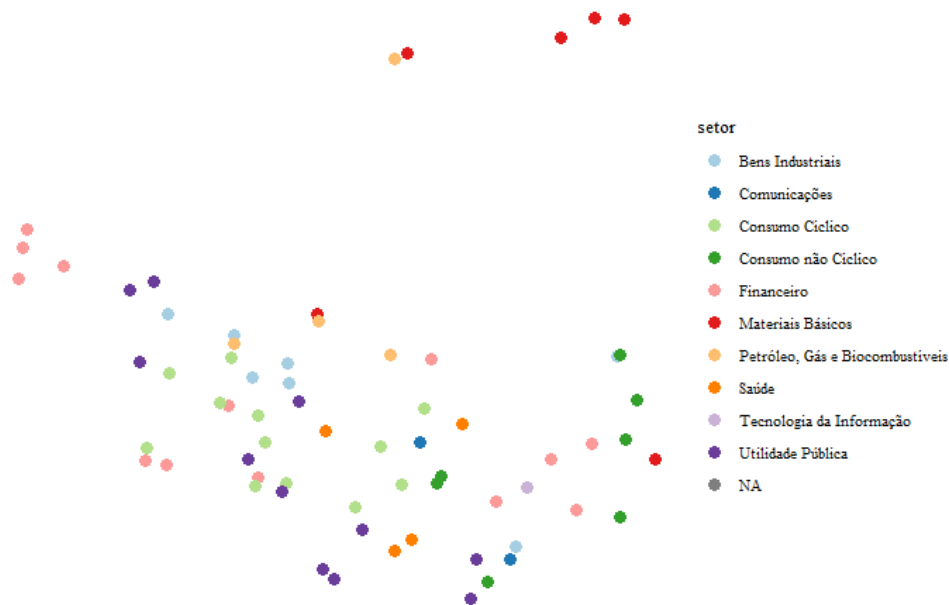
Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 09 mostra a MST filtrada para distâncias menores que 0,025 no período anterior à pandemia. É possível identificar os ativos com maiores correlações.

o Financeiro maiores e mais alavancadas pelos bancos. No mais, os setores de Consumo Cíclico e Consumo não Cíclico aparecem misturados a outros ativos de outros setores, indicando que ambos, antes da crise, possuíam um amplo potencial de mistura com ativos de outros setores.

Olhando agora para a formação dos *clusters* de acordo com os *thresholds*, tem-se os gráficos 09, 10, 11, 12. Tal sequência mostra a evolução do surgimento dos ativos na *MST* criada no início da análise (video gráfico 08). Em um primeiro momento, com o gráfico 09, é possível identificar a formação de interações mais fortes entre alguns ativos de diferentes setores, como o de Utilidade Pública, Consumo Cíclico e Petróleo, Gás e Biocombustíveis. Ainda, nota-se também o isolamento do setor Financeiro, e o da PETR3 com USIM5, sendo ambos os inícios de seus *clusters* como visto no gráfico 08. O *threshold* de 0,025 traz também a formação de 4 possíveis *clusters* além dos de setores, sendo um formado pelos ativos do setor Financeiro à esquerda com os bancos BBSA3 e BBDC4, outro pela interação entre PETR3 e USIM5, o terceiro sendo a nuvem de pontos formada ao meio do gráfico, e o último, a nuvem ao sudeste. Estes dois últimos, por sua vez, indicam que há interações próximas com os setores presentes, porém de forma mais segregada no ao sudete, com possíveis trocas pontuais.

Gráfico 10 - Minimum Spanning Tree antes da crise da covid - threshold de 0,05



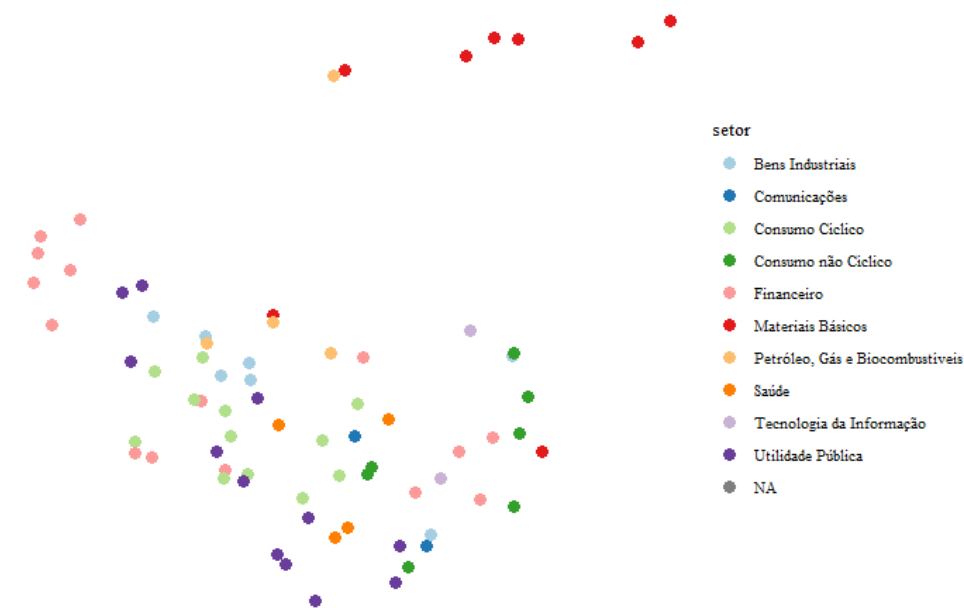
Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 10 mostra a MST filtrada para distâncias menores que 0,05 no período anterior à pandemia. É possível identificar as mudanças ocorridas para o gráfico 09, de um *threshold* menor.

Observadas essas tendências, parte-se para a análise do próximo *threshold*: o de 0,05, como mostrado no gráfico 10. Os novos pontos corroboram para marcar os *clusters* dos setores Financeiros e de Materiais Básicos junto à PETR3. Por outro lado, a nuvem de pontos no centro e sudeste do gráfico se tornou mais intensa, indicando que os ativos ali são bem correlacionados, provavelmente ditando parte do andamento do mercado. Ainda assim, é possível destacar a relação próxima do setor de Consumo Cíclico, Bens Industriais, Petróleo, Gás e Biocombustíveis e, em partes o de Utilidade Pública. Os três primeiros setores mencionados,

ainda, foram destacados na seção da Análise Descritiva como sendo mais voláteis que o mercado (vide gráfico 04), o que ajuda a explicar as interações fortes identificadas pela *Minimum Spanning Tree*. Com os novos pontos, observa-se a inserção do setor de Consumo Cíclico à nuvem de pontos ao sudeste, e se identifica a formação de outro possível *cluster* à direita do gráfico, com interações entre os setores de Consumo não Cíclico e Financeiro mais um ativo de Tecnologia da Informação e um de Materiais Básicos. IRBR3, BPAN4, SULA11, IRBR3, MRFG3, BEEF3, BRFS3, JBSS3, BRKM5, TOTS3 e CIEL3 fazem parte desse grupo, o que leva destaque a financeiros com foco em crédito e pagamentos com o setor de comidas.

Gráfico 11 - Minimum Spanning Tree antes da crise da covid - threshold de 0,075



Fonte: elaboração própria, dados da B3 e do Economática

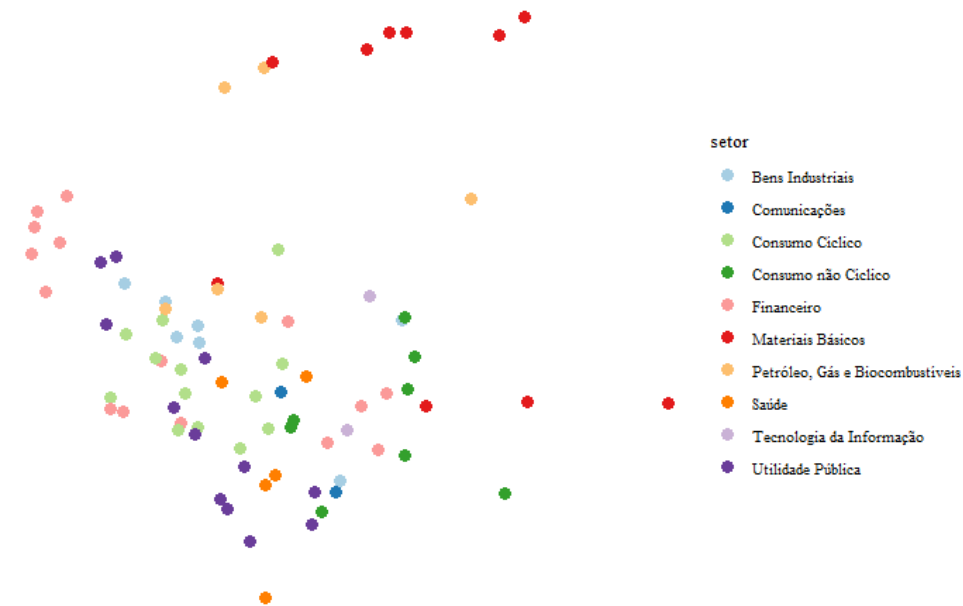
O gráfico 11 mostra a MST filtrada para distâncias menores que 0,075 no período anterior à pandemia. É possível identificar as mudanças ocorridas para o gráfico 10, de um *threshold* menor.

Adiante, com o *threshold* de 0,075, poucas mudanças são identificadas de acordo com o gráfico 11, dessa forma destacam-se algumas interações pontuais ocorridas. Há o aparecimento de mais pontos no *cluster* do setor Financeiro (sendo todos bancos) e no de Materiais Básicos (grande foco na parte de mineração), reforçando a proximidade entre esses ativos não só pelo núcleo do negócio, mas também em movimentos de mercado. No mais, surgem mais pontos de Utilidade Pública, que passam a integrar a nuvem de pontos cada vez maior e mais heterogênea.

Por último, o gráfico 12 já não discerne tanto do gráfico 11 devido à presença de quase todos os ativos da base. Há a adição de apenas alguns ativos isolados, o que possivelmente indica

interações isoladas e não tão fortes com o restante dos outros ativos, com um comportamento característico da própria empresa.

Gráfico 12 - Minimum Spanning Tree antes da crise da covid - threshold de 0,3



Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 12 mostra a MST filtrada para distâncias menores que 0,125 no período anterior à pandemia. É possível identificar as mudanças ocorridas para o gráfico 11, de um *threshold* menor.

Para o próximo nível, contudo, englobam-se todas as variáveis da base, o que implica a volta ao gráfico 08. Antes de partir para as conclusões desta seção, vale identificar quais são os ativos com mais ramificações, que interagem de uma forma mais abrangente com a média do mercado antes da pandemia.

Tabela 02 - Quantidade de conexões entre cada ativo

Ativo	Conexões
ABEV3	3
B3SA3	3
BBDC4	3
BRAP4	3
BRFS3	3
CCRO3	3
CIEL3	3
CPLE6	3
EMBR3	3
EQTL3	3
NTCO3	3
KLBN11	3
RAIL3	3
TIMS3	3
TOTS3	3
VIIA3	3

ALPA4	2
AMER3	2
BPAN4	2
BBSE3	2
BRML3	2
BBDC3	2
BBAS3	2
BRKM5	2
BPAC11	2
CRFB3	2
CMIG4	2
COGN3	2
CSAN3	2
CVCB3	2
CYRE3	2
ECOR3	2
ELET3	2
ELET6	2
ENBR3	2
ENGI11	2
EGIE3	2
EZTC3	2
FLRY3	2
GGBR4	2
GOAU4	2
GOLL4	2
IRBR3	2
ITUB4	2
RENT3	2
MRFG3	2
BEEF3	2
MRVE3	2
MULT3	2
PETR3	2
PRIO3	2
QUAL3	2
SBSP3	2
CSNA3	2
SULA11	2
TAAE11	2
VIVT3	2
UGPA3	2
USIM5	2
VBBR3	2
WEGE3	2
YDUQ3	2
AZUL4	1
CPFE3	1

DXCO3	1
ENEV3	1
HAPV3	1
HYPE3	1
ITSA4	1
JBSS3	1
JHSF3	1
LREN3	1
MGLU3	1
PETR4	1
POSI3	1
RADL3	1
SANB11	1
SLCE3	1
SUZB3	1
VALE3	1

Fonte: elaboração própria, dados do Economatica e da B3

De acordo com a tabela 02, identificam-se os ativos ABEV3, B3SA, BBDC4, BRAP4, BRFS3, CCRO3, CIEL3, CPLE6, EMBR3, EQTL3, NTCO3, KLBN11, RAIL3, TIMS3, TOTS3 e VIIA3 como os que possuem maiores conexões na *Minimum Spanning Tree*. Isso possivelmente indica que suas movimentações no mercado são mais influenciadas por outros ativos ou mais influentes neles do que a média do mercado, devido à maior proximidade (e correlação) existente em volta de cada ativo.

Com a análise de formação de *clusters* por meio de filtros limites para as distâncias geradas pela *Minimum Spanning Tree* e da descrição dos eventos demonstrados, é possível concluir alguns pontos. Antes da pandemia, por mais que os setores definidos pela B3 sejam um ponto de encontro entre seus ativos, nem todos formaram grupos exclusivos com interações fortes, sendo a atividade da empresa um ponto crucial para determinar sua posição no mapa. Há, por outro lado, dois setores que se destacam por uma certa exclusividade: o Financeiro por meio dos bancos e os Materiais Básicos por meio das empresas focadas em recursos naturais, como os minerais.

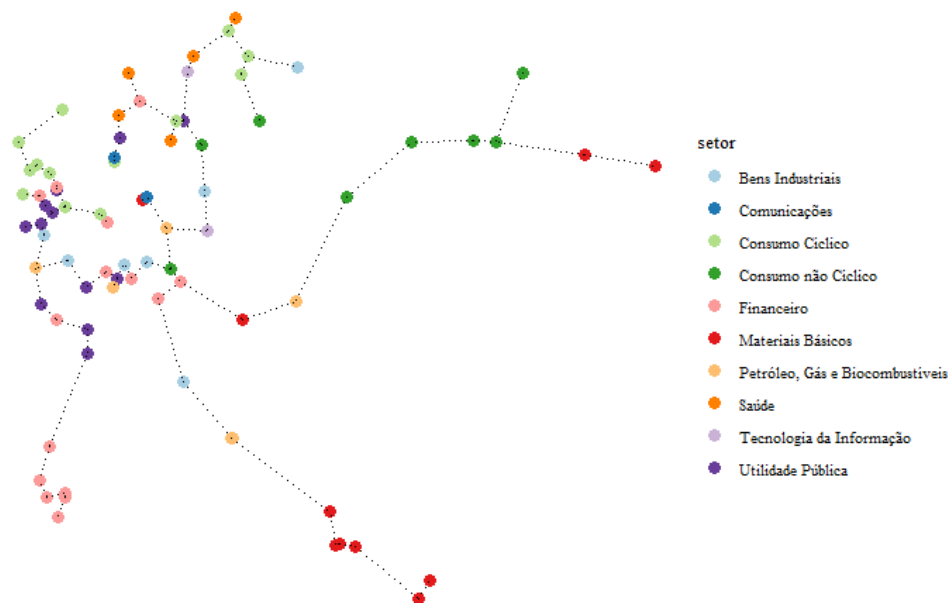
As interações, por sua vez, são mais fortes nos gráficos 09 e 10, sendo os setores de Consumo Cíclico, Utilidade Pública, Bens Industriais e Petróleo, Gás e Biocombustíveis fortemente correlacionados; da mesma forma que os de Consumo não Cíclico (agropecuarista), Financeiro (de pagamentos e créditos) e um ativo da Tecnologia da Informação e um de Materiais básicos também apresentam fortes interações. Tais relações indicam parâmetros para as mudanças ocorridas no próximo período, que se encontra na próxima seção.

5.2. Durante a crise da Covid-19

Ressalta-se que o gráfico com os *tickers* dos ativos se encontra no anexo 02.

Para começar a análise da *Minimum Spanning Trees* ao longo da pandemia, faz-se o gráfico 13, que mostra as mínimas distâncias para todos os ativos da base, coloridos de acordo com seus respectivos setores. Como se esperava, há uma proximidade maior entre os ativos por se tratar de uma época de crise, na qual a maior volatilidade gera maior correlação entre os retornos observados na Bolsa. Em adição, os ativos tenderam a se isolar mais nos seus setores, com

Gráfico 13 - Minimum Spanning Tree durante a crise da covid



Fonte: elaboração própria, dados da B3 e do Economática

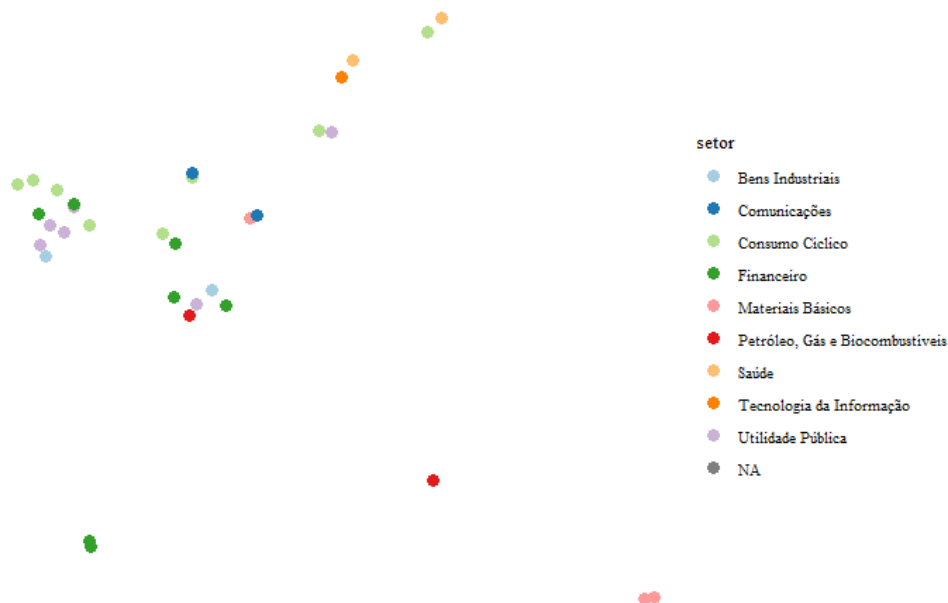
O gráfico 13 mostra a árvore de distâncias mínimas para todos os ativos durante a pandemia e coloridos de acordo com seus setores. Identificam-se padrões similares e diferentes aos observados antes da crise da covid.

destaques para os de Consumo Cíclico, Consumo não Cíclico, Materiais Básicos, Financeiro (bancos), Utilidade Pública e Saúde. A nuvem de pontos observada no gráfico 08 se repetiu em menor intensidade no período durante a crise, o que corrobora para o maior isolamento dos ativos em setores parecidos, contudo ainda há interações entre setores diversos, como o de Bens Industriais, Petróleo, Gás e Biocombustíveis, Financeiro (não bancos) e uma parte do de Utilidade Pública. Uma nova interação neste gráfico pode ser vista entre o setor de Saúde com o de Consumo Cíclico, indicando que a parte de saúde passou a ter características mais parecidas com esse setor, uma vez que, antes da pandemia, as ações de Saúde eram dispersas no gráfico sem apresentar um entrosamento com algum outro em específico. Tal modificação de comportamento provavelmente se desdobra das origens da crise mundial, que mobilizou a

área da saúde para combater a pandemia do novo coronavírus. Assim, pode-se afirmar que houve mudanças no comportamento do setor de saúde com o choque sanitário.

Investigando a formação desses clusters, parte-se para os gráficos com os *thresholds* definidos no começo da seção da mesma forma como foram construídos os de antes da pandemia. Tem-se, portanto, os gráficos 14, 15, 16 e 17, filtrando para distâncias de até 0,025; 0,05; 0,075 e 0,125, respectivamente.

Gráfico 14 - Minimum Spanning Tree durante a crise da covid - threshold de 0,025



Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 14 mostra a MST filtrada para distâncias menores que 0,025 no intervalo da pandemia. É possível identificar os ativos com maiores correlações e as principais mudanças para o período anterior à covid.

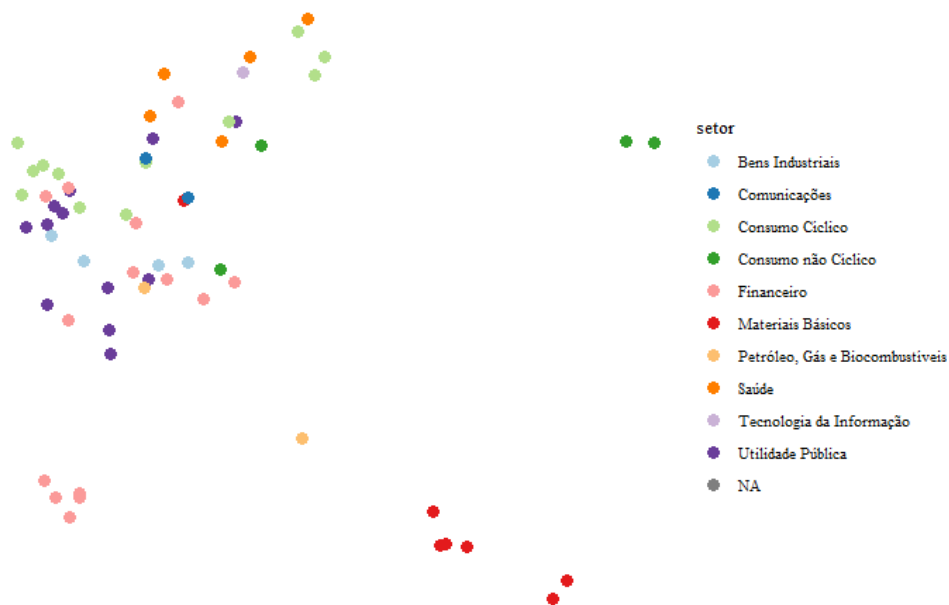
Como definido na análise anterior à pandemia, o *threshold* de 0,025 mostra os ativos com as menores distâncias entre si e permite identificar o início das formações dos *clusters* independentemente do setor da ação. Com o gráfico 14, identificam-se também possíveis tendências ou mudanças de comportamento nas situações de estresse caudas pela pandemia. Ao olhar para a imagem, nota-se de fato um maior isolamento entre os ativos, com alguns apontamentos para seus setores. Assim como antes da pandemia, já há um início dos *clusters* dos bancos, com SANB11 e BBDC4 ao sudoeste do gráfico; e do de Materiais Básicos, com GOAU4 e GGBR4.

Característico do período durante a pandemia, tem-se a elaboração de um grupo de ativos com interações fortíssimas à esquerda do gráfico. Na presença dos setores de Consumo Cíclico, Utilidade Pública, Financeiro e Bens Industriais, há a mistura das áreas de energia, educação, construção civil, investimento em mercado imobiliário e uma empresa de vestimentas e outras

de rodovias. Tal heterogeneidade pode indicar uma certa relação com o mercado imobiliário, sendo este afetado pela crise por meio da maior permanência em casa durante a crise, além do esvaziamento dos escritórios. Esse *cluster*, por se convergir ao *real estate*, não seria uma opção de investimentos durante a pandemia, uma vez que a característica que une essas empresas foi duramente afetada.

Nota-se também a presença de um grupo formado por uma mistura de ações de diferentes setores à centro-esquerda do gráfico, porém que não aparenta ter uma área convergente nos negócios de suas empresas. Isso implica que a associação entre eles pode ser apenas existente devido à maior volatilidade dos ativos em questão dada a maior incerteza no mercado no período.

Gráfico 15 - Minimum Spanning Tree durante a crise da covid - threshold de 0,05



Fonte: elaboração própria, dados da B3 e do Economática

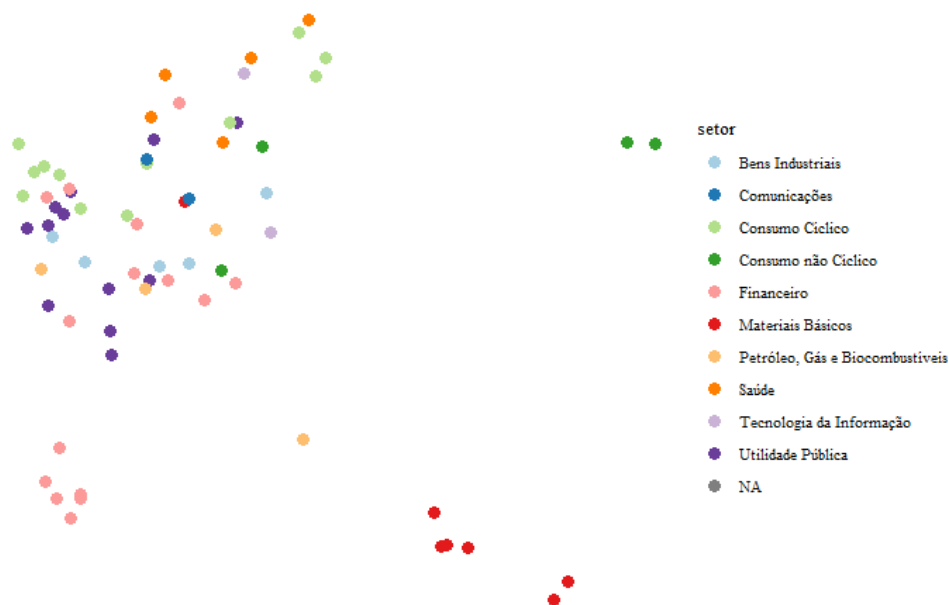
O gráfico 15 mostra a MST filtrada para distâncias menores que 0,05 no intervalo da pandemia. É possível identificar as formações dos *clusters* mais fortes e os diferenciar com o período antes da crise.

Aumentando o valor do limite, tem-se o gráfico 15 acima. Nele se presencia uma formação mais desenvolvida dos *clusters* mencionados com o gráfico 14: mais ativos de Materiais Básicos, Financeiro e Utilidade Pública; além do surgimento mais expressivo do setor de Saúde e do de Consumo não Cíclico. No mais, como ocorrido antes da crise, o setor Financeiro de bancos e o de Materiais Básicos de matéria bruta se mantiveram isolados em seus cantos com fortes correlações intragrupos.

No mais, surgem mais ativos da área de energia, pertencentes ao setor de Utilidade Pública, e integram o grupo de ações à esquerda. Os ativos do grupo à centro-esquerda também recebem mais vizinhos de Utilidade Pública, além de surgirem mais ações de Financeiro (não bancos).

O destaque de variação entre o *threshold* de 0,025 e o de 0,05 é o surgimento de mais ativos do setor de Saúde, formando um *cluster* relativamente grande na parte superior. Esse conjunto, por mais que tenha sido destacado pela presença dos setores de Saúde e de Consumo Cíclico, percebe-se uma maior força de interação com os ativos de Saúde, estando presentes em grande parte do grupo e cercando os outros pontos. Isso indica que as empresas das áreas sanitárias podem ter puxado consigo outros ativos em suas movimentações devido à proximidade indicada pela *MST*.

Gráfico 16 - Minimum Spanning Tree durante a crise da covid - threshold de 0,075



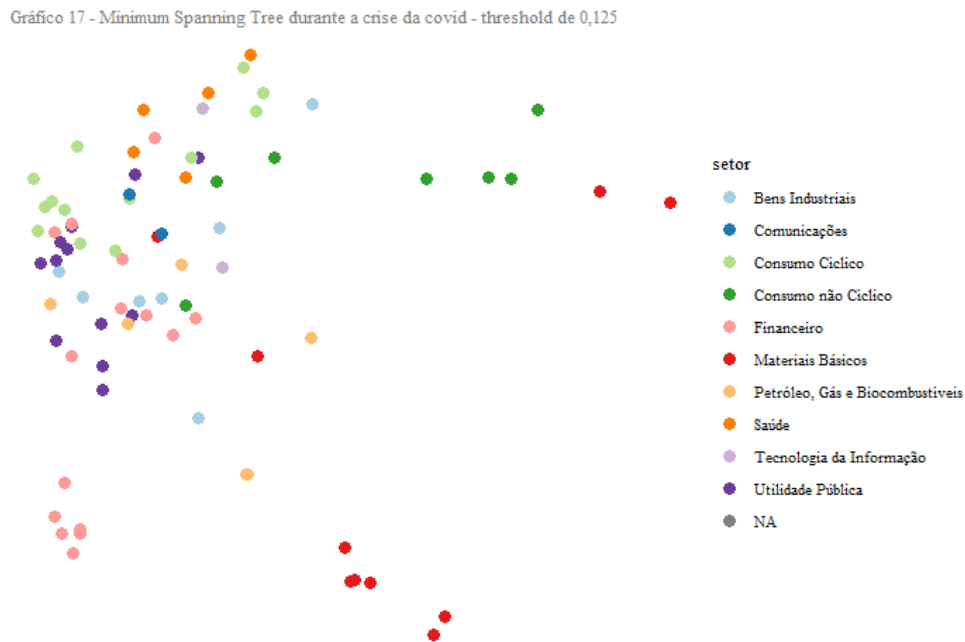
Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 16 mostra a MST filtrada para distâncias menores que 0,075 durante a pandemia. É possível identificar as mudanças ocorridas para o gráfico 15, de um limite menor.

O próximo limite alocado para as distâncias é de 0,075 e há a adição de apenas 5 ações como se observa no gráfico 16, mostrando que as interações mais importantes entre os ativos foram capturadas pelos gráficos 14 e 15, o que também ocorreu no outro intervalo em análise. Desta vez, completa-se a formação do grupo dos bancos, e se vê mais ativos de Petróleo, Gás e Biocombustíveis, indicando que esses ativos não possuem interações tão fortes com o restante do grupo das variáveis por estarem relativamente longes. No entanto, os pontos adicionados podem ser uma ligação entre os grupos montados à esquerda, à centro-esquerda e na parte superior. Isso implica que possivelmente há um equilíbrio em cada ativo com distâncias entre

0,05 e 0,075 de suas ligações, sendo uma escolha para estratégias que tendem a não correr muitos riscos em situações de crise parecidas devido à falta de correlação com os outros.

Com o *threshold* seguinte, de 0,125, apesar de não trazer informações relevantes para os *clusters* até aqui discutidos, pode-se dar mais atenção ao grupo de ativos formado a nordeste do gráfico, com os setores de Consumo não Cíclico e Materiais Básicos, das áreas de agropecuária e de papel e celulose. Assim, o gráfico 17 mostra que as ações desse grupo não são próximas entre si, mas apresentam certos elementos em comum: os subsetores das empresas de cada setor e o quão longe estão do restante dos pontos. Isso pode indicar um bom plano de investimento devido à maior distância tanto entre si quanto entre os outros ativos.



Fonte: elaboração própria, dados da B3 e do Economática

O gráfico 17 mostra a MST filtrada para distâncias menores que 0,125 durante a pandemia. É possível identificar as mudanças ocorridas para o gráfico 16, de um *threshold* menor.

O próximo *threshold*, por sua vez, englobaria todas as distâncias possíveis, o que faz retornar ao gráfico 13. No entanto, antes de se concluir a análise de *clusters* e a intensidade das interações entre os ativos, traz-se a tabela de conexões de cada ativo para que se tenha uma noção de quais possuem mais ligações e influência ou que mais é influenciado.

Tabela 03 - Quantidade de conexões entre cada ativo

Ativo	Conexões
ALPA4	3
ABEV3	3
COGN3	3
CPLE6	3

CSAN3	3
ENBR3	3
ENEV3	3
EQTL3	3
IRBR3	3
MGLU3	3
MRFG3	3
SULA11	3
VIIA3	3
VBBR3	3
AMER3	2
AZUL4	2
BPAN4	2
BBSE3	2
BRML3	2
BBDC3	2
BBDC4	2
BBAS3	2
BRKM5	2
BRFS3	2
BPAC11	2
CCRO3	2
CMIG4	2
CIEL3	2
CPFE3	2
CVCB3	2
ECOR3	2
ELET3	2
ELET6	2
EMBR3	2
ENGI11	2
EZTC3	2
GGBR4	2
GOAU4	2
GOLL4	2
NTCO3	2
HAPV3	2
ITSA4	2
JBSS3	2
KLBN11	2
LREN3	2
MRVE3	2
MULT3	2
PETR3	2
PETR4	2
PRIO3	2
POS3	2
QUAL3	2

RAIL3	2
SBSP3	2
SANB11	2
CSNA3	2
SLCE3	2
TAEE11	2
VIVT3	2
TIMS3	2
TOTS3	2
USIM5	2
VALE3	2
YDUQ3	2
B3SA3	1
BRAP4	1
CRFB3	1
CYRE3	1
DXCO3	1
EGIE3	1
FLRY3	1
HYPE3	1
ITUB4	1
JHSF3	1
RENT3	1
BEEF3	1
RADL3	1
SUZB3	1
UGPA3	1
WEGE3	1

Fonte: elaboração própria, dados do Economatica e da B3

Assim, a tabela 03 mostra que os ativos ALPA4, ABEV3, COGN3, CPLE6, CSAN3, ENBR3, ENEV3, EQTL3, IRBR3, MGLU3, MRFG3, SULA11, VIIA3 e VBBR3 são os que possuem a maior quantidade conexões na *Minimum Spanning Tree* feitas nesta seção. Em comparação com o exposto na tabela 02, vê-se a presença dos ativos ABEV3, EQTL3 e VIIA3, mostrando que possivelmente tais ativos são mais influenciadores de movimentos do que influenciados pois aparecem em ambos os períodos, ao passo que todos os outros são novos.

Assim, por meio da análise da *MST* dividida em *thresholds* e das interações entre os ativos, é possível construir algumas estratégias de investimentos para que ajude os investidores em próximas crises parecidas. No mais, do que fora discutido na seção 5.2, tira-se que, apesar de os ativos da Bolsa de Valores de São Paulo tenham sofrido algumas quebras estruturais, algumas interações são tão fortes que perduram por qualquer que seja o contexto local e global.

6. Plano de ações

Concluída e discutida a 5ª seção, estabelece-se nesta seção um conjunto de planos e ações para que os investidores consigam sobreviver a crises parecidas com a ocorrida. Ainda, é possível que tais planos sejam usados em outras crises, desde que fortemente adaptados pelo agente econômico.

Começa-se, então, o plano por meio de uma inferência do que foi analisado em toda a seção 5: os setores divulgados pela B3 não são determinantes de relações fortes entre os ativos integrantes. Apesar de haver um agrupamento por atividade econômica, é possível que certas crises impactem os ativos de formas diferentes, como se observou com o grupo que tinha como característica em comum o mercado imobiliário forte. Ainda, como Sandoval (2012) também mostra, os ativos dos mercados impactados pela crise financeira sofreram um certo tipo de mudança estrutura, o que ocasionou na formação de *clusters* baseados nessa quebra. Outro ponto a se ponderar é a incerteza do investimento no setor em que a crise se originou, dado que tanto na pandemia quanto em 2008 houve quebras estruturais no setor de saúde e financeiro, respectivamente.

Além de considerar os setores próprios dos ativos, deve-se considerar também a proposta de valor das empresas e todos os mercados significativos dos quais a empresa faz parte e todos os mercados e áreas mais impactados pela crise. Com ambos os grupos de informações feitos, é possível cruzá-los para determinar quais ações serão possivelmente mais surpreendidas em tempos adversos.

Por mais que, no geral, os setores indicados pela B3 não estabelecem *clusters* em todos os momentos, há grupos de ativos de um mesmo setor que podem ajudar na sobrevivência do investidor em crises. O gráfico 17 mostra que os ativos do setor de Consumo não Cíclico se isolaram dos outros e mantiveram distâncias relativamente grandes entre si. Isso indica que ter investido nesse setor seria um bom plano de ação. Há duas características para incluí-lo ao conjunto: é um setor que não depende fortemente do ciclo econômico por ter uma demanda mais inelástica; e, principalmente, foi um setor que se isolou dos outros ativos e houve um distanciamento razoável dentro do grupo, corroborando ainda mais para sua solidez.

Por fim, é importante que o investidor estude sobre o peso da empresa na Bolsa para que evite ações que sofram muita influência de outras, uma vez que uma queda generalizada pode ser intensificada por tal elemento. Ações que influenciam mais as outras, por outro lado, tendem

a ser mais seguras devido à sua maior solidez (como tamanho de mercado) e independência de outras empresas dado seu poder.

7. Conclusões

Este buscou identificar padrões de comportamento do mercado de ações do Brasil antes de uma crise e durante ela para desenhar estratégias de como se comportar a fim de não tomar muitos danos durante os tempos turbulentos. Começou-se por meio de um estudo bibliográfico para determinar quais seriam as expectativas para o desenvolvimento do artigo e as metodologias usadas. Baseando-se extensivamente em Sandoval (2012, 2013), Sandoval e Franca (2012) e Mantegna (1999), foi possível montar uma metodologia que se encaixasse na crise mais recente vivenciada: a da Covid-19.

A estrutura da análise, então, consistiu em duas bases: a descritiva e a metodológica. A primeira definiu uma parte dos comportamentos esperados, como maiores volatilidades em certos setores e a similaridade da base de dados com o índice Ibovespa. A segunda, por sua vez, baseou-se na aplicação do método das *Minimum Spanning Trees*, em que, a partir de uma matriz de distâncias entre os ativos, calcula qual seria a árvore de menor distância que interligasse todos os ativos pelo menos uma vez. Com ela, então, foi possível estabelecer de forma mais específica os comportamentos *clusterizados* de cada ativo antes da pandemia a fim de estabelecer parâmetros para a análise do período ao longo da pandemia.

Apesar de alguns padrões claros, como as ações referentes a bancos e a materiais brutos, identificou-se muitas misturas de setores dos ativos, sendo essas com uma leve tendência aos mercados mais importantes para cada empresa. A análise no decorrer da pandemia mostrou que tal tendência se consolidou, especialmente no mercado imobiliário e no setor de saúde, dado que ambos foram intensamente impactados. Ainda, como uma surpresa, o subsetor agropecuário do setor de Consumo não Cíclico se isolou do restante dos outros ativos e suas ações também mantiveram uma distância relativamente grande entre si, mostrando que esse setor possivelmente seria uma boa escolha de investimentos para sobreviver à crise.

Ao final, com todas as informações obtidas, foi possível desenhar um plano de ações para serem tomadas a fim de sobreviver em crises similares, ou então para adaptá-las a próximas crises que tenham características próximas. Do conjunto de ações, tem-se que considerar os setores definidos pela B3 pode ser fatal, pois os ativos não se comportam de acordo com isso. Para tal, então, deve-se considerar um estudo dos principais mercados das empresas e os mais impactados pela crise para que se cruze informações e se defina se as companhias podem ser afetadas. Ainda, buscar setores que possuam inelasticidade em sua demanda e se isolam dos

outros é interessante para manter uma solidez nos investimentos; entretanto deve-se manter atento às ações que sofrem mais influências das outras.

Apesar das conclusões aqui mencionadas, é importante levar em consideração as limitações deste artigo. A primeira diz respeito à suposição de que as relações entre os ativos são lineares, porém podem existir outros tipos não-lineares, estes que não foram consideradas. A segunda se projeta na análise generalizada dos 80 ativos na base, que poderia ser realizada de forma individual a fim de estabelecer critérios mais rígidos para a formação dos *clusters* e para a análise das interações. De forma mais específica também, poder-se-ia utilizar os pesos da Carteira Teórica do Ibovespa a fim de ponderar as ações na análise, todavia não havia disponibilidade de descarregamento para dados passados, apenas para os usados na análise. A terceira e última, por sua vez, refere-se ao uso de métodos qualitativos para a definição de *clusters*: tanto a definição da B3 quanto a observação de interações mais fortes entre os ativos. Métodos quantitativos, como *k-means*, trariam uma visão diferente e mais exata sobre a formação dos *clusters*, além de gerarem interpretações mais fáceis para o leitor. Tais estatísticas não foram empregadas para que a análise não se estendesse além do esperado, podendo trazer resultados confusos no meio de tanta informação.

Não obstante as limitações, os resultados obtidos neste artigo apresentam uma forma de se analisar as crises a fim não de evitá-las, mas de contorná-las. Ainda, fomentam mais análises de interações em períodos similares, trazendo mais informações relevantes para o investidor.

Referências

ALBULESCU, C. T. COVID-19 and the United States financial markets' volatility. **Finance Research Letters**, v. 38, p. 101699, jul. 2020.

ALI, M.; ALAM, N.; RIZVI, S. A. R. Coronavirus (COVID-19) – An epidemic or pandemic for financial markets. **Journal of Behavioral and Experimental Finance**, v. 27, p. 100341, maio 2020.

B3. Disponível em: <https://www.b3.com.br/pt_br/>. Acesso em: 15 nov. 2022.

Economática. Disponível em: <<https://economica.com/>>. Acesso em: 15 nov. 2022.

G1. Bovespa fecha queda pressionada por Vale e incerteza fiscal. Disponível em: <<https://g1.globo.com/economia/noticia/2021/09/16/bovespa.ghtml>>. Acesso em: 6 nov. 2021.

KRUSKAL, J. B.; WISH, M. **Multidimensional Scaling**. Beverly Hills, Calif.: Sage Publications, 1978.

MANTEGNA, R. N. Hierarchical structure in financial markets. **The European Physical Journal B**, v. 11, n. 1, p. 193–197, set. 1999.

MCKINSEY. The coronavirus effect on global economic sentiment | McKinsey.

Disponível em: <<https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/the-coronavirus-effect-on-global-economic-sentiment>>. Acesso em: 19 set. 2022.

NATIONS, U. How COVID-19 is changing the world: a statistical perspective. Disponível em: <https://hdr.undp.org/content/how-covid-19-changing-world-statistical-perspective?utm_source=EN&utm_medium=GSR&utm_content=US_UNDP_PaidSearch_Brand_English&utm_campaign=CENTRAL&c_src=CENTRAL&c_src2=GSR&gclid=CjwKC-AjwgaeYBhBAEiwAvMgp2hcUkG7sF75DnaqIbxE5fnLCn6QtY2v5aSEkr46LAv0udTvxwXDrkBoCy9wQAvD_BwE>. Acesso em: 19 set. 2022.

OECD. OECD Economic Outlook, Interim Report September 2021. [s.l.] OECD, 2021.

OECD. **First lessons from government evaluations of COVID-19 responses: A synthesis.** Disponível em: <<https://www.oecd.org/coronavirus/policy-responses/first-lessons-from-government-evaluations-of-covid-19-responses-a-synthesis-483507d6/>>.

OMS. **Histórico da pandemia de COVID-19 - OPAS/OMS | Organização Pan-Americana da Saúde.** Disponível em: <<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>>.

QUADROS, A.; GARMONSWAY, D. **emstreeR: Tools for Fast Computing and Plotting Euclidean Minimum Spanning Trees.** Disponível em: <<https://cran.r-project.org/web/packages/emstreeR/index.html>>. Acesso em: 19 set. 2022.

SALATI, P. **Agropecuária foi o único setor que cresceu no PIB de 2020; entenda.**

Disponível em:

<<https://g1.globo.com/economia/agronegocios/noticia/2021/03/03/agropecuaria-foi-o-unico-setor-que-cresceu-no-pib-de-2020-entenda.ghtml>>. Acesso em: 12 out. 2021.

SANDOVAL, L. A MAP OF THE BRAZILIAN STOCK MARKET. **Advances in Complex Systems**, v. 15, n. 05, p. 1250042, jul. 2012.

SANDOVAL, L. Cluster formation and evolution in networks of financial market indices. **Algorithmic Finance**, v. 2, n. 1, p. 3–43, 2013.

SANDOVAL, L.; FRANCA, I. D. P. Correlation of financial markets in times of crisis. **Physica A: Statistical Mechanics and its Applications**, v. 391, n. 1, p. 187–208, 1 jan. 2012.

Yahoo Finance. Disponível em: <<https://finance.yahoo.com/>>. Acesso em: 15 nov. 2022.

ZHANG, D.; HU, M.; JI, Q. Financial markets under the global pandemic of COVID-19. **Finance Research Letters**, v. 36, p. 101528, abr. 2020.

Anexo 03 – Código em R do que foi analisado

```
# Chamando as bibliotecas
```

```
library("tidyverse")
```

```
library("emstreeR")
```

```
library("openxlsx")
```

```
library("lubridate")
```

```
library("extrafont")
```

```
library("corr")
```

```
library("skimr")
```

```
library("ggpubr")
```

```
font_import()
```

```
loadfonts(device = "win")
```

```
theme_abnt <- theme(text = element_text(size = 10, family = "serif"),
```

```
  plot.caption = element_text(hjust = 0, colour = "#808080", size = 10),
```

```
  plot.title = element_text(colour = "#808080", size = 10))
```

```
theme_set(theme_bw() + theme_abnt)
```

```
intervalo_pre <- interval(ymd("2018-05-21"), ymd("2020-03-10"))
```

```

intervalo_dur <- interval(ymd("2020-03-11"), ymd("2021-12-31"))

intervalo_tot <- interval(ymd("2018-05-21"), ymd("2021-12-31"))

# Usar todas as variáveis no singular

dados <- read.xlsx("economica.xlsx", detectDates = TRUE,

                  na.strings = "-") %>%

as_tibble() %>%

mutate(Data = ymd(Data)) %>%

select(-`Fechamento.ajust.p/.prov.Em.moeda.orig`) %>%

pivot_longer(-Data, names_to = "Ativo", values_to = "Fechamento") %>%

mutate(Ativo = str_remove_all(

  Ativo, "Fechamento\\.ajust\\.p\\/\\.prov\\.Em\\.moeda\\.orig\\.")

))

# Fazer join com dados do setor

setor <- read.xlsx("tickers.xlsx") %>%

as_tibble() %>%

select(ticker = Ticker, setor = Setor.econômico)

```

```
# 909 observações é normal em conjunto com 46 NAs, a partir de 2019
```

```
ativo_na <- dados %>%
```

```
  filter(Data %within% intervalo_tot) %>%
```

```
  group_by(Ativo) %>%
```

```
  count(is.na(Fechamento)) %>%
```

```
  filter(`is.na(Fechamento)` == T & n > 50) %>% .[[1]]
```

```
# são 10 ativos na lista ativo_na, indicando que são 80 ativos que existem entre
```

```
# 2019 e agora
```

```
# As correlações só podem ser feitas por meio do retorno de cada ação
```

```
retorno <- dados %>%
```

```
  filter(!(Ativo %in% ativo_na), Data %within% intervalo_tot) %>%
```

```
  drop_na() %>%
```

```
  pivot_wider(names_from = Ativo, values_from = Fechamento) %>%
```

```
  mutate(across(c(-1), ~log(.x)),
```

```
    across(c(-1), ~.x - lag(.x))) %>%
```

```
  slice(-1)
```

```
# df do ibovespa para comparar com os retornos do índice
```

```
ibov <- read.csv("ibov.csv") %>%
```

```
as_tibble() %>%
```

```
select(data = Date, fechamento = Adj.Close) %>%
```

```
mutate(fechamento = as.double(fechamento),
```

```
  data = ymd(data)) %>%
```

```
drop_na() %>%
```

```
mutate(ibov = log(fechamento) - lag(log(fechamento))) %>%
```

```
select(-fechamento)
```

```
# df com os setores dos ativos, long
```

```
df_setor <- retorno %>%
```

```
  pivot_longer(cols = -Data, names_to = "ativo", values_to = "retorno") %>%
```

```
  mutate(radical = str_sub(ativo, 1, 4)) %>%
```

```
  left_join(setor, by = c("radical" = "ticker")) %>%
```

```
  select(-radical)
```

```
# carteira teorica
```

```
cart_teorica <- read.xlsx("cart teorica.xlsx") %>%
```

```
as_tibble() %>%
```

```
select(ativo = Código, empresa = Ação, qtde = Qtde..Teórica, porcen = `Part..(%)`)
```

```
### Descritivas
```

```
skim(retorno %>%
```

```
  pivot_longer(cols = -Data, names_to = "ativo", values_to = "retorno") %>%
```

```
  group_by(Data) %>%
```

```
  summarise(media = mean(retorno)))
```

```
# Jogar os dados pro excel
```

```
df_skim <- retorno %>%
```

```
  pivot_longer(cols = -Data, names_to = "ativo", values_to = "retorno") %>%
```

```
  group_by(Data) %>%
```

```
  summarise(media = mean(retorno)) %>%
```

```
  ungroup()
```

```
df_skim <- df_setor %>%
```

```
  group_by(Data, setor) %>%
```

```
  summarise(med_ret = mean(retorno)) %>%
```

```
  ungroup() %>%
```

```
  pivot_wider(names_from = setor, values_from = med_ret)
```

```
skim_tudo <- skim(df_skim) %>% as_tibble()

skim_pre <- skim(df_skim %>% filter(Data %within% intervalo_pre)) %>% as_tibble()

skim_dur <- skim(df_skim %>% filter(Data %within% intervalo_dur)) %>% as_tibble()

descritivas_ret <- rbind(skim_tudo, skim_pre, skim_dur)

write.xlsx(descritivas_ret, "descritivas_ret.xlsx")
```

```
### Dispersões com o índice Bovespa
```

```
# retorno médio de cada setor com o ibov ---- Gráfico 3
```

```
df_setor %>%

  group_by(setor, Data) %>%

  summarise(ret_medio = mean(retorno)) %>%

  left_join(ibov, by = c("Data" = "data")) %>%

  ggplot(aes(x = ibov, y = ret_medio, colour = setor)) +

  geom_point() +

  geom_smooth(se = F, method = "lm") +
```

```

facet_wrap(. ~ setor) +

labs(y = "Retorno médio diário", x = "Retorno do Ibovespa",

      title = "Gráfico 03 - Retorno médio diário de cada setor x Retorno do Ibovespa",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica")

# Descrição de quantos % cada ação vale

retorno %>%

pivot_longer(cols = -Data, names_to = "ativo", values_to = "retorno") %>%

distinct(ativo) %>%

left_join(cart_teorica, by = "ativo") %>%

select(-empresa, -qtde) %>%

summarise(soma = sum(porcen))

# Histograma do ret médio diário ----- Gráfico 1

retorno %>%

pivot_longer(cols = -Data, names_to = "ativo", values_to = "retorno") %>%

group_by(Data) %>%

summarise(ret_medio = mean(retorno)) %>%

ggplot(aes(x = ret_medio)) +

```

```
geom_density() +

labs(x = "Retorno médio diário", y = "Densidade",

      title = "Gráfico 01 - Densidade do retorno médio diário",

      caption = "Fonte: elaboração própria, dados do Economatica")
```

```
# retorno médio diário das ações x ibov ---- Gráfico 2
```

```
retorno %>%

pivot_longer(cols = -Data, names_to = "ativo", values_to = "retorno") %>%

group_by(Data) %>%

summarise(ret_medio = mean(retorno)) %>%

left_join(ibov, by = c("Data" = "data")) %>%

ggplot(aes(x = ibov, y = ret_medio)) +

  geom_point() +

  geom_smooth(se = F, method = "lm") +

  lims(x = c(-0.2, 0.15), y = c(-0.2, 0.15)) +

  labs(x = "Ibovespa", y = "Retorno médio diário",

        title = "Gráfico 02 - Retorno médio diário das ações x Retorno diário Ibovespa",

        caption = "Fonte: elaboração própria, dados do Economatica e do Yahoo")
```

```
##### Períodos analisados #####

retorno_antes <- retorno %>%

  filter(Data % within% intervalo_pre)

retorno_dur <- retorno %>%

  filter(Data % within% intervalo_dur)

# Dispersão com ibov nos períodos analisados

# par(mfrow = c(2, 1)) não funcionou

# Antes da covid ---- Gráfico 4

df_setor %>%

  group_by(Data, setor) %>%

  summarise(ret_medio = mean(retorno)) %>%

  left_join(ibov, by = c("Data" = "data")) %>%

  filter(Data % within% intervalo_pre) %>%

  ggplot(aes(x = ibov, y = ret_medio, colour = setor)) +

  geom_point() +

  geom_smooth(se = F, method = "lm") +
```

```

facet_wrap(. ~ setor) +

labs(x = "Ibov", y = "Retorno médio diário",

      title = "Gráfico 04 - Retorno médio diário dos setores x Índice Ibovespa, antes da
covid",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica")

df_setor %>%

group_by(Data, setor) %>%

summarise(ret_medio = mean(retorno)) %>%

left_join(ibov, by = c("Data" = "data")) %>%

filter(Data % within% intervalo_dur) %>%

ggplot(aes(x = ibov, y = ret_medio, colour = setor)) +

geom_point() +

geom_smooth(se = F, method = "lm") +

facet_wrap(. ~ setor) +

labs(x = "Ibov", y = "Retorno médio diário",

      title = "Gráfico 05 - Retorno médio diário dos setores x Índice Ibovespa, durante a
covid",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica")

## Correlação antes ---- Gráfico 06

```

```

df_setor %>%

group_by(Data, setor) %>%

summarise(ret_medio = mean(retorno)) %>%

left_join(ibov, by = c("Data" = "data")) %>%

filter(Data % within% intervalo_pre) %>%

pivot_wider(names_from = setor, values_from = ret_medio) %>%

corrr::correlate() %>%

corrr::as_matrix() %>%

ggcorrplot::ggcorrplot(method = "circle", hc.order = T, type = "upper",

                        outline.color = "white", ggtheme = theme_bw()) +

labs(title = "Gráfico 06 - Matrix de correlação antes da covid",

      caption = "Fonte: elaboração própria, dados da B3, do Economatca e do Yahoo
Finance") +

theme_abnt +

theme(plot.caption = element_text(hjust = 1.5),

      plot.title = element_text(hjust = -10))

## Correlação antes ---- Gráfico 07

df_setor %>%

group_by(Data, setor) %>%

summarise(ret_medio = mean(retorno)) %>%

```

```

left_join(ibov, by = c("Data" = "data")) %>%

filter(Data % within% intervalo_dur) %>%

pivot_wider(names_from = setor, values_from = ret_medio) %>%

corrr::correlate() %>%

corrr::as_matrix() %>%

ggcorrplot::ggcorrplot(method = "circle", hc.order = T, type = "upper",

                        outline.color = "white", ggtheme = theme_bw()) +

labs(title = "Gráfico 07 - Matrix de correlação durante a covid",

      caption = "Fonte: elaboração própria, dados da B3, do Economatica e do Yahoo
Finance") +

theme_abnt +

theme(plot.caption = element_text(hjust = 1.5),

      plot.title = element_text(hjust = -10))

##### MST - geral #####

# Matriz de correlação com todos os pares de ativos

distancias <- retorno %>%

filter(Data % within% intervalo_tot) %>%

correlate(diagonal = 1) %>%

```

```
select(x = 1, everything()) %>%  
  
mutate(across(-1, ~sqrt(2*(1 - .x))))  
  
# pivot_longer(cols = -1, names_to = "y", values_to = "dist") %>%  
  
ordem_ativos <- distancias$x  
  
ativo_setor <- df_setor %>%  
  
select(ativo, setor) %>%  
  
distinct()  
  
pontos <- distancias %>%  
  
select(-1) %>%  
  
cmdscale() %>%  
  
as_tibble() %>%  
  
mutate(ativo = ordem_ativos) %>%  
  
select(x = V1, y = V2, ativo)  
  
# fazendo a MST  
  
set.seed(1234)  
  
mst <- ComputeMST(pontos[-3])
```

```
#df com todas as informações necessárias para plotar

df_plots <- mst %>%

  as_tibble() %>%

  mutate(ativo = ordem_ativos) %>%

  left_join(ativo_setor, by = "ativo")

# tabela de conexões entre os pontos (ver quais pontos aparecem mais)

df_plots %>%

  mutate(from = factor(from))

  group_by(from) %>%

  count() %>%

  arrange(desc(n))

my_pal <- RColorBrewer::brewer.pal(n = 10, name = "Paired")

##### Antes da crise #####

distancias <- retorno %>%
```

```
filter(Data %within% intervalo_pre) %>%
```

```
correlate(diagonal = 1) %>%
```

```
select(x = 1, everything()) %>%
```

```
mutate(across(-1, ~sqrt(2*(1 - .x))))
```

```
ordem_ativos <- distancias$x
```

```
ativo_setor <- df_setor %>%
```

```
select(ativo, setor) %>%
```

```
distinct()
```

```
set.seed(1234)
```

```
pontos <- distancias %>%
```

```
select(-1) %>%
```

```
cmdscale() %>%
```

```
as_tibble() %>%
```

```
mutate(ativo = ordem_ativos) %>%
```

```
select(x = V1, y = V2, ativo)
```

```
# fazendo a MST
```

```
set.seed(1234)

mst <- ComputeMST(pontos[-3])

#df com todas as informações necessárias para plotar

df_plots <- mst %>%

  as_tibble() %>%

  mutate(ativo = ordem_ativos) %>%

  left_join(ativo_setor, by = "ativo")

df_chave <- data.frame(ativo = ordem_ativos, numero = 1:80) %>%

  as_tibble()

tab_freq_ativo_antes <- df_plots %>%

  select(-x, -y, -distance, -setor) %>%

  slice(-nrow(.)) %>%

  pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%

  group_by(numero) %>%

  count() %>%

  arrange(desc(n)) %>%

  left_join(df_chave, by = "numero")
```

```
write.xlsx(tab_freq_ativo_antes, "tab_freq_antes.xlsx")
```

```
df_plots %>%
```

```
  ggplot(aes(x = distance)) +
```

```
  geom_histogram()
```

```
# Gráfico 08 - sem thresholds
```

```
df_plots %>%
```

```
  ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",
```

```
                    palette = my_pal, size = 3,
```

```
                    ellipse.type = "convex", legend = "right", theme = theme_bw()) +
```

```
  stat_MST(aes(x, y, from = from, to = to)) +
```

```
  labs(x = "", y = "",
```

```
        title = "Gráfico 08 - Minimum Spanning Tree antes da crise da covid",
```

```
        caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +
```

```
  theme_void() +
```

```
  theme_abnt +
```

```
  theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),
```

```
plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))
```

```
# histograma das distâncias para ver os thresholds
```

```
skim(df_plots %>% select(distance))
```

```
df_plots %>%
```

```
ggplot(aes(x = distance)) +
```

```
geom_histogram()
```

```
# Gráfico 09 - antes e com thresholds 0.025
```

```
mst %>%
```

```
as_tibble() %>%
```

```
select(3:5) %>%
```

```
filter(distance < 0.025) %>%
```

```
pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%
```

```
left_join(df_chave, by = "numero") %>%
```

```
pivot_wider(names_from = from_to, values_from = numero) %>%
```

```
distinct(ativo) -> df_lista_ativos
```

```

mst %>%

as_tibble() %>%

mutate(ativo = ordem_ativos,

       across(

         c("from", "to", "distance"), ~case_when(distance < 0.025 ~ .x,

                                                  TRUE ~ NA_real_)),

       across(

         c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,

                                  TRUE ~ NA_real_)),

       ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,

                         TRUE ~ "")) %>%

left_join(ativo_setor, by = "ativo") %>%

select(-distance) %>%

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",

                  palette = my_pal, size = 3,

                  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to)) +

labs(x = "", y = "",

      title = "Gráfico 09 - Minimum Spanning Tree antes da crise da covid - threshold de
0,025",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

```

```

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

# Gráfico 10 - antes e com thresholds 0.05

mst %>%

as_tibble() %>%

select(3:5) %>%

filter(distance < 0.05) %>%

pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%

left_join(df_chave, by = "numero") %>%

pivot_wider(names_from = from_to, values_from = numero) %>%

distinct(ativo) -> df_lista_ativos

mst %>%

as_tibble() %>%

mutate(ativo = ordem_ativos,

      across(

        c("from", "to", "distance"), ~case_when(distance < 0.05 ~ .x,

```

```

TRUE ~ NA_real_)),

across(

  c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,

    TRUE ~ NA_real_)),

ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,

  TRUE ~ "")) %>%

left_join(ativo_setor, by = "ativo") %>%

select(-distance) %>%

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",

  palette = my_pal, size = 3,

  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to), na.rm = T, linetype = 0) +

labs(x = "", y = "",

  title = "Gráfico 10 - Minimum Spanning Tree antes da crise da covid - threshold de 0,05",

  caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

  plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

```

```
# Gráfico 11 - antes e com thresholds 0.075
```

```
mst %>%
```

```
  as_tibble() %>%
```

```
  select(3:5) %>%
```

```
  filter(distance < 0.075) %>%
```

```
  pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%
```

```
  left_join(df_chave, by = "numero") %>%
```

```
  pivot_wider(names_from = from_to, values_from = numero) %>%
```

```
  distinct(ativo) -> df_lista_ativos
```

```
mst %>%
```

```
  as_tibble() %>%
```

```
  mutate(ativo = ordem_ativos,
```

```
    across(
```

```
      c("from", "to", "distance"), ~case_when(distance < 0.075 ~ .x,
```

```
        TRUE ~ NA_real_)),
```

```
    across(
```

```
      c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,
```

```
        TRUE ~ NA_real_)),
```

```
    ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,
```

```

      TRUE ~ "")) %>%

left_join(ativo_setor, by = "ativo") %>%

select(-distance) %>%

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",

                  palette = my_pal, size = 3,

                  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to), na.rm = T, linetype = 0) +

labs(x = "", y = "",

      title = "Gráfico 11 - Minimum Spanning Tree antes da crise da covid - threshold de
0,075",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

# Gráfico 12 - antes e com thresholds 0.125

mst %>%

as_tibble() %>%

select(3:5) %>%

filter(distance < 0.125) %>%

```

```
pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%
```

```
left_join(df_chave, by = "numero") %>%
```

```
pivot_wider(names_from = from_to, values_from = numero) %>%
```

```
distinct(ativo) -> df_lista_ativos
```

```
mst %>%
```

```
as_tibble() %>%
```

```
mutate(ativo = ordem_ativos,
```

```
  across(
```

```
    c("from", "to", "distance"), ~case_when(distance < 0.125 ~ .x,
```

```
      TRUE ~ NA_real_)),
```

```
  across(
```

```
    c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,
```

```
      TRUE ~ NA_real_)),
```

```
  ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,
```

```
    TRUE ~ "")) %>%
```

```
left_join(ativo_setor, by = "ativo") %>%
```

```
select(-distance) %>%
```

```
ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",
```

```
  palette = my_pal, size = 3,
```

```

    ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to), na.rm = T, linetype = 0) +

labs(x = "", y = "",

      title = "Gráfico 12 - Minimum Spanning Tree antes da crise da covid - threshold de
0,125",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

```

Gráfico de anexo 01 - antes da pandemia

```
df_plots %>%
```

```

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",

                  palette = my_pal, size = 3,

                  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to)) +

labs(x = "", y = "",

      title = "Anexo 01 - Minimum Spanning Tree antes da crise",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

```

```

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1)) +

ggrepel::geom_text_repel(aes(x, y, label = ativo))

```

```
##### Durante a crise #####
```

```
distancias <- retorno %>%
```

```
filter(Data % within% intervalo_dur) %>%
```

```
correlate(diagonal = 1) %>%
```

```
select(x = 1, everything()) %>%
```

```
mutate(across(-1, ~sqrt(2*(1 - .x))))
```

```
ordem_ativos <- distancias$x
```

```
ativo_setor <- df_setor %>%
```

```
select(ativo, setor) %>%
```

```
distinct()
```

```
set.seed(1234)

pontos <- distancias %>%

select(-1) %>%

cmdscale() %>%

as_tibble() %>%

mutate(ativo = ordem_ativos) %>%

select(x = V1, y = V2, ativo)

# fazendo a MST

set.seed(1234)

mst <- ComputeMST(pontos[-3])

#df com todas as informações necessárias para plotar

df_plots <- mst %>%

as_tibble() %>%

mutate(ativo = ordem_ativos) %>%

left_join(ativo_setor, by = "ativo")

df_chave <- data.frame(ativo = ordem_ativos, numero = 1:80) %>%

as_tibble()
```

```
tab_freq_ativo_durante <- df_plots %>%  
  
  select(-x, -y, -distance, -setor) %>%  
  
  slice(-nrow(.)) %>%  
  
  pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%  
  
  group_by(numero) %>%  
  
  count() %>%  
  
  arrange(desc(n)) %>%  
  
  left_join(df_chave, by = "numero")  
  
write.xlsx(tab_freq_ativo_durante, "tab_freq_durante.xlsx")  
  
df_plots %>%  
  
  ggplot(aes(x = distance)) +  
  
  geom_histogram()  
  
# Gráfico 13 - sem thresholds  
  
df_plots %>%  
  
  ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",
```

```

palette = my_pal, size = 3,

ellipse.type = "convex", legend = "right", theme = theme_bw() +

stat_MST(aes(x, y, from = from, to = to)) +

labs(x = "", y = "",

title = "Gráfico 13 - Minimum Spanning Tree durante a crise da covid",

caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

# histograma das distâncias para ver os thresholds

skim(df_plots %>% select(distance))

df_plots %>%

ggplot(aes(x = distance)) +

geom_histogram()

# Gráfico 14 - durante e com thresholds 0.025

mst %>%

```

```
as_tibble() %>%
```

```
select(3:5) %>%
```

```
filter(distance < 0.025) %>%
```

```
pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%
```

```
left_join(df_chave, by = "numero") %>%
```

```
pivot_wider(names_from = from_to, values_from = numero) %>%
```

```
distinct(ativo) -> df_lista_ativos
```

```
mst %>%
```

```
as_tibble() %>%
```

```
mutate(ativo = ordem_ativos,
```

```
  across(
```

```
    c("from", "to", "distance"), ~case_when(distance < 0.025 ~ .x,
```

```
      TRUE ~ NA_real_)),
```

```
  across(
```

```
    c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,
```

```
      TRUE ~ NA_real_)),
```

```
  ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,
```

```
    TRUE ~ "")) %>%
```

```
left_join(ativo_setor, by = "ativo") %>%
```

```

select(-distance) %>%

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",

                  palette = my_pal, size = 3,

                  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to)) +

labs(x = "", y = "",

      title = "Gráfico 14 - Minimum Spanning Tree durante a crise da covid - threshold de
0,025",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

# Gráfico 15 - durante e com thresholds 0.05

mst %>%

as_tibble() %>%

select(3:5) %>%

filter(distance < 0.05) %>%

pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%

left_join(df_chave, by = "numero") %>%

```

```
pivot_wider(names_from = from_to, values_from = numero) %>%
```

```
distinct(ativo) -> df_lista_ativos
```

```
mst %>%
```

```
as_tibble() %>%
```

```
mutate(ativo = ordem_ativos,
```

```
  across(
```

```
    c("from", "to", "distance"), ~case_when(distance < 0.05 ~ .x,
```

```
      TRUE ~ NA_real_)),
```

```
  across(
```

```
    c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,
```

```
      TRUE ~ NA_real_)),
```

```
  ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,
```

```
    TRUE ~ "")) %>%
```

```
left_join(ativo_setor, by = "ativo") %>%
```

```
select(-distance) %>%
```

```
ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",
```

```
  palette = my_pal, size = 3,
```

```
  ellipse.type = "convex", legend = "right", theme = theme_bw()) +
```

```
stat_MST(aes(x, y, from = from, to = to), na.rm = T, linetype = 0) +
```

```

labs(x = "", y = "",

      title = "Gráfico 15 - Minimum Spanning Tree durante a crise da covid - threshold de
0,05",

      caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),

      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

# Gráfico 16 - durante e com thresholds 0.075

mst %>%

as_tibble() %>%

select(3:5) %>%

filter(distance < 0.075) %>%

pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%

left_join(df_chave, by = "numero") %>%

pivot_wider(names_from = from_to, values_from = numero) %>%

distinct(ativo) -> df_lista_ativos

mst %>%

as_tibble() %>%

```

```

mutate(ativo = ordem_ativos,

  across(

    c("from", "to", "distance"), ~case_when(distance < 0.075 ~ .x,

      TRUE ~ NA_real_)),

  across(

    c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,

      TRUE ~ NA_real_)),

  ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,

    TRUE ~ "")) %>%

left_join(ativo_setor, by = "ativo") %>%

select(-distance) %>%

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",

  palette = my_pal, size = 3,

  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to), na.rm = T, linetype = 0) +

labs(x = "", y = "",

  title = "Gráfico 16 - Minimum Spanning Tree durante a crise da covid - threshold de
0,075",

  caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

```

```

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),
      plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

```

```
# Gráfico 17 - durante e com thresholds 0.125
```

```
mst %>%
```

```
  as_tibble() %>%
```

```
  select(3:5) %>%
```

```
  filter(distance < 0.125) %>%
```

```
  pivot_longer(cols = c("from", "to"), names_to = "from_to", values_to = "numero") %>%
```

```
  left_join(df_chave, by = "numero") %>%
```

```
  pivot_wider(names_from = from_to, values_from = numero) %>%
```

```
  distinct(ativo) -> df_lista_ativos
```

```
mst %>%
```

```
  as_tibble() %>%
```

```
  mutate(ativo = ordem_ativos,
```

```
         across(
```

```
           c("from", "to", "distance"), ~case_when(distance < 0.125 ~ .x,
```

```
           TRUE ~ NA_real_)),
```

```
         across(
```

```

c("x", "y"), ~case_when(ativo %in% df_lista_ativos$ativo ~ .x,
  TRUE ~ NA_real_)),

ativo = case_when(ativo %in% df_lista_ativos$ativo ~ ativo,
  TRUE ~ "")) %>%

left_join(ativo_setor, by = "ativo") %>%

select(-distance) %>%

ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",
  palette = my_pal, size = 3,
  ellipse.type = "convex", legend = "right", theme = theme_bw()) +

stat_MST(aes(x, y, from = from, to = to), na.rm = T, linetype = 0) +

labs(x = "", y = "",
  title = "Gráfico 17 - Minimum Spanning Tree durante a crise da covid - threshold de
0,125",
  caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +

theme_void() +

theme_abnt +

theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),
  plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1))

# Gráfico de anexo 02 - durante a pandemia

df_plots %>%

```

```
ggpubr::ggscatter(x = "x", y = "y", ellipse = F, color = "setor",  
  
  palette = my_pal, size = 3,  
  
  ellipse.type = "convex", legend = "right", theme = theme_bw()) +  
  
stat_MST(aes(x, y, from = from, to = to)) +  
  
labs(x = "", y = "",  
  
  title = "Gráfico de anexo 02 - Minimum Spanning Tree durante a crise da covid",  
  
  caption = "Fonte: elaboração própria, dados da B3 e do Economatica") +  
  
theme_void() +  
  
theme_abnt +  
  
theme(plot.caption = element_text(hjust = 0.05, colour = "#808080", size = 10),  
  
  plot.title = element_text(colour = "#808080", size = 10, hjust = 0.1)) +  
  
ggrepel::geom_text_repel(aes(x, y, label = ativo))
```