



Miguel Marques

**Gestão de portefólio baseado em correlações:
Detecção de anomalias e agrupamento de ações**

**Portfolio Management based on correlations:
Anomaly detection and stock grouping**

DOCUMENTO PROVISÓRIO



Miguel Marques

Gestão de portefólio baseado em correlações:
Deteção de anomalias e agrupamento de ações

Portfolio Management based on correlations:
Anomaly detection and stock grouping

DOCUMENTO PROVISÓRIO

“The greatest challenge to any thinker is stating the problem in a way that will allow a solution”

— Bertrand Russell



Miguel Marques

**Gestão de portefólio baseado em correlações:
Deteção de anomalias e agrupamento de ações**

**Portfolio Management based on correlations:
Anomaly detection and stock grouping**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Licenciatura em Engenharia Computacional, realizada sob a orientação científica da Professora Doutora Rita Bastião, Professora auxiliar do Departamento de Economia, Gestão, Engenharia Industrial e Turismo da Universidade de Aveiro.

o júri / the jury

presidente / president

Prof. Doutor Carlos Brites

professor auxiliar do Departamento de Física da Universidade de Aveiro

vogais / examiners committee

Prof. Doutora Rita Bastião

professora auxiliar convidada do Departamento de Economia, Gestão, Engenharia Industrial e Turismo da Universidade de Aveiro

Prof. Doutora Mara Madaleno

professora auxiliar do Departamento de Economia, Gestão, Engenharia Industrial e Turismo da Universidade de Aveiro

**agradecimentos /
acknowledgements**

Agradeço toda a ajuda a todos os meus colegas, aos meus pais, ao meu irmão e à Prof. Doutora Rita Bastião por participar no meu projeto como orientadora.

Palavras Chave

ações, mercados financeiros, correlação, Local Outlier Factor, Planal Maximally Filtered Graph

Resumo

Este projeto apresenta o desenvolvimento de uma aplicação com o objetivo de analisar as dinâmicas do mercado de ações através da detecção de anomalias e visualização de dados. O mercado de ações é um mercado complexo, este é influenciado por vários fatores que complicam a dinâmica do mesmo. A compreensão destas dinâmicas é algo que investidores e profissionais financeiros procuram perceber para poderem tomar melhores decisões na compra ou venda de ações. As ações são títulos que representam uma fração do capital social de uma sociedade anônima. O investimento em ações implica a ponderação de três variáveis de decisão designadamente a liquidez, o risco e a rentabilidade. A aplicação aborda estes desafios através de técnicas como detecção de anomalias e visualização de dados. A detecção de anomalias consiste em implementar a detecção de comportamentos anormais nas correlações de ações. Estas anomalias podem ser indicadores de potenciais oportunidades ou riscos no mercado. A aplicação em geral interpreta e analisa em tempo real as dinâmicas do mercado de ações, conduzindo a uma tomada de decisão informada e mais célere.

Keywords

stocks, financial markets, correlation, Local Outlier Factor, Planal Maximally Filtered Graph

Abstract

This project presents the development of an application aimed at analyzing the dynamics of the stock market by detecting anomalies and providing comprehensive data visualization. The stock market, as a complex financial market, is influenced by various factors that hamper its dynamics. Investors and financial professionals constantly seek to comprehend these dynamics in order to take informed decisions regarding buying, selling. Stocks, being a prevalent financial instrument, represent fractional ownership in a company's share capital. Investing in stocks involves assessing three key variables, namely liquidity, risk, and profitability. The application developed addresses these challenges by employing techniques for anomaly detection and data visualization. Anomaly detection consists in the detection of abnormal behavior within correlations movements between two stocks. These detected anomalies serve as crucial indicators of potential market opportunities or risks. Overall, the application interprets and analyzes the dynamics of the stock market in real time, leading to a informed and faster decision making.

Conteúdo

Conteúdo	i
Lista de Figuras	iii
Lista de Tabelas	v
Glossário	vii
1 Introdução	2
2 Metodologias	3
2.1 Bibliotecas	3
2.2 Webscraping	3
2.3 Pré processamento de dados	3
2.4 Análise do comportamento de ações	4
2.4.1 Pearson Correlation Coefficient	4
2.5 Detecção de anomalias	5
2.5.1 Detecção de anomalias nos valores de correlação	6
2.6 Detecção de anomalias em tempo real	12
2.7 Planar Maximally Filtered Graph	12
2.7.1 Louvain community detection	13
3 Aplicação	19
4 Conclusão	23
5 Trabalho Futuro	25
Referências	27

Lista de Figuras

2.1	Anomalias detetadas utilizando o MAD	8
2.2	Exemplo do algoritmo Local Outlier Factor	8
2.3	Exemplo de <i>reachability-distance</i> para $k=3$	9
2.4	Anomalias detetadas com um <i>threshold</i> 1 para o LOF no dia 08/06/2023 depois do mercado ter fechado	10
2.5	Anomalias detetadas com um <i>threshold</i> 1.2 para o LOF calculadas no dia 08/06/2023 depois do mercado ter fechado	10
2.6	Anomalias detetadas com um <i>threshold</i> 1.2 para o LOF calculadas no dia 09/06/2023 depois do mercado ter fechado	11
2.7	Exemplo de cliques para um grafo	12
2.8	Grafo PMFG com 63 ações	14
2.9	Grafo PMFG com 63 ações com grupos identificados	14
2.10	Grafo PMFG com 63 ações utilizando o <i>threshold</i>	15
2.11	Subgrafo isolado na rede com 63 ações com o <i>threshold</i> aplicado	15
2.12	Grafo PMFG com 63 ações utilizando o <i>threshold</i> com grupos identificados	16
2.13	Subgrafo isolado na rede com 63 ações com o <i>threshold</i> aplicado identificados como grupo	16
2.14	Análise de um grupo (amarelo) na rede com 63 ações com o <i>threshold</i>	17
2.15	Análise de um grupo (verde) na rede com 63 ações com o <i>threshold</i>	18
3.1	Interface inicial da aplicação	19
3.2	Informação sobre uma ação	20
3.3	Interface do heatmap	20
3.4	Cores atribuídas conforme valores do coeficiente de correlação	20
3.5	Interface para visualizar anomalias	21
3.6	Interface para visualizar o sistema	21
3.7	Diagrama do <i>update</i> de valores de ações e detecção de anomalias	22
3.8	Exemplo de <i>pop-up</i>	22

Lista de Tabelas

2.1	Tabela de verdade para os modelos	11
-----	---	----

Glossário

LOF	Local Outlier Factor	MAD	Median Absolute Deviation
PMFG	Planar Maximally Filtered Graph	MST	Minimum Spanning Tree

CAPÍTULO 1

Introdução

Os mercados financeiros distribuem recursos e providenciam liquidez para firmas e empresários, tornando a troca de participações financeiras simples para compradores e vendedores. Os mercados financeiros recompensam os investidores e credores que tenham dinheiro a mais, deixando esse dinheiro disponível para quem precisa.

O mercado de ações é discutivelmente o mercado financeiro que mais prevalece. Neste mercado os investidores e intermediários podem comprar e vender ações cotadas. As empresas utilizam os mercados de ações para obter dinheiro através da Oferta Pública Inicial, através da qual as ações são negociadas entre diferentes compradores e vendedores.

No mercado de ações o preço de cada ação não se altera, por regra de forma independente, pois na realidade alterações no preço individual poderão afetar a padrão de evolução do mercado, da mesma forma que dinâmicas de grupos de ações tem impacto nas ações que constituem o grupo. Diferentes ações podem ter comportamentos similares ou opostos, estando diretamente correlacionadas ou inversamente correlacionadas respetivamente. Esta relação entre ações pode ser utilizada para prever o comportamento de ações correlacionadas relativamente ao comportamento de uma ação.[1], [2]

O objetivo deste projeto é analisar a dinâmica do mercado de ações, mais especificamente o comportamento de diferentes ações e avaliá-las em tempo real. Neste sentido a aplicação desenvolvida deteta anomalias e emite alertas para o utilizador. Paralelamente, o utilizador também consegue visualizar diferentes grupos de ações que tenham comportamentos similares. A aplicação utiliza webscraping para fazer a recolha de dados relevantes das ações em tempo real e processa os dados obtidos para o utilizador. Desta forma a aplicação permite que o utilizador esteja informado sobre uma possível mudança nos valores dos seus ativos, que poderá resultar numa minimização de perdas ou maximização de lucros.

A avaliação da dinâmica dos mercados de ações tem sido um tópico de investigação recorrente na literatura, geralmente baseado numa base de dados empírica estática criando lacunas na continuidade futura [3], [4]. O presente trabalho apresenta um duplo contributo. Por um lado evolui favoravelmente por se basear numa base de dados com a componente de recolha de informação em tempo real, com técnicas de deteção de anomalias e sinalização na visualização gráfica. Por outro lado, a inovação por via de uma aplicação disponível para utilizadores potencia a tomada de decisão mais célere e informada aos investidores contribuindo para uma maior eficiência dos mercados.

Metodologias

Este capítulo tem como objetivo apresentar os métodos e bibliotecas utilizadas para o processamento de dados e desenvolvimento da aplicação.

2.1 BIBLIOTECAS

- Para desenvolvimento da aplicação utilizou-se *tkinter*, *customTkinter* [5];
- Na extração de dados são utilizadas as bibliotecas *Requests* e *Beautiful Soup* com o intuito de obter o código fonte da página e extrair dados do código fonte respetivamente;
- Para fazer os gráficos na aplicação utilizou-se *Matplotlib*;
- Para criar grafos utilizou-se *NetworkX* [6];
- Utiliza-se [7] para o algoritmo *Louvain community detection*;
- Utiliza-se *SciPy* [8] para o cálculo do *Z-Score* e para o teste de normalidade *Kolmogorov-Smirnov*;
- Utiliza-se *Scikit-learn* [9] para o algoritmo Local Outlier Factor e *MinMaxScaler*;

2.2 WEBSCRAPING

Webscraping é o processo utilizado para a recolha de dados estruturados, sendo os dados extraídos através do *website Yahoo Finance* [10]. Os *URLs* são criados tendo em conta o símbolo da ação, a biblioteca *Requests* obtém o código fonte do endereço de rede e através de *Beautiful Soup* extrai-se a informação relevante. Este processo será utilizado principalmente para obter os valores de fecho das ações.

2.3 PRÉ PROCESSAMENTO DE DADOS

Os dados recolhidos através de *Webscraping* são convertidos para uma escala logarítmica. Esta conversão irá reduzir a diferença das escalas dos valores de fecho de diferentes ações, facilitando e obtendo resultados mais precisos no cálculo da correlação.

2.4 ANÁLISE DO COMPORTAMENTO DE AÇÕES

Para analisar o comportamento de ações utiliza-se a correlação destas. Existem diferentes métodos de correlação, no entanto uma vez que se utilizam os valores do fecho diário das ações, as ações irão ter o mesmo número de valores de fecho, assim como estarão igualmente espaçadas ao longo do tempo (exceto em diferentes mercados), por estes motivos utiliza-se o método *Pearson Correlation Coefficient*. Para o cálculo de coeficiente de correlação, são utilizados os valores diários de fecho de 30 dias, que por sua vez implica o uso de 29 valores de rentabilidade e considera-se ações diretamente correlacionadas, aquelas que tem um valor para o coeficiente de correlação superior ou igual a 0.75 ou inferior ou igual a -0,75. Os valores do coeficiente de correlação são calculados nos últimos 30 dias para obter uma distribuição dos valores de correlação para as ações ao longo do tempo, sendo que esta distribuição será utilizada para detetar anomalias. Segundo o modelo de Markowitz [11], pode-se minimizar o risco de um portefólio de ações através da construção do mesmo utilizando ações que não estejam correlacionadas. No mercado de ações existem diferentes Bolsas, no entanto como é referido por *Agarwal et al.* [12] poderá haver ações numa Bolsa a influenciarem outras em Bolsas de Valores diferentes. Tendo isto presente, as ações serão correlacionadas independentemente da bolsa em que se encontram.

2.4.1 Pearson Correlation Coefficient

Este método mede a correlação linear entre dois conjuntos de dados, os valores de correlação estão entre -1 e 1, caso o valor perto de 1, têm-se uma correlação positiva (diretamente correlacionadas), se o valor estiver perto de -1 obtemos uma correlação negativa (inversamente correlacionadas) por fim se o valor estiver perto de 0 não se encontram correlacionadas. O coeficiente de correlação pode ser descrito como o coeficiente de covariância de duas variáveis, neste caso ações, dividido pelo produto do desvio padrão de cada variável.

Tendo em conta o nosso conjunto de dados inicial, valores de fecho das ações é necessário efetuar alguns cálculos intermédios para calcular a covariância entre duas ações.

Rentabilidade

Rentabilidade é uma métrica relativa aos ganhos potenciais de um investimento, quanto maior os ganhos potenciais será o investimento, o mesmo se verifica para o oposto. Esta pode ser calculada com os dados históricos para o valor de fecho através da seguinte expressão:

$$R_{x,t} = \frac{C_t - C_{t-1}}{C_{t-1}}, \quad (2.1)$$

Para escalas logarítmicas:

$$r_{x,t} = \ln\left(\frac{C_t}{C_{t-1}}\right) = \ln(C_t) - \ln(C_{t-1}) \quad (2.2)$$

onde C_t e C_{t-1} representam o valor de fecho no momento t e $t - 1$ respetivamente, $R_{x,t}$ e $r_{x,t}$ é a rentabilidade da ação no instante t .

A rentabilidade média ou retorno \bar{r}_x é dado por:

$$\bar{r}_x = \frac{1}{n} \sum_{t=1}^n R_{x,t} \quad ou \quad \bar{r}_x = \frac{1}{n} \sum_{t=1}^n r_{x,t} \quad (2.3)$$

onde n é a dimensão da amostra.

Risco

O risco é uma métrica que é dependente das condições do mercado, este está associado à variabilidade dos preços. A variância irá avaliar a dispersão da rentabilidade das ações, relativamente ao *retorno*. A variância pode ser calculada da seguinte forma:

$$\sigma_x^2 = \frac{1}{n-1} \sum_{t=1}^n (r_{x,t} - \bar{r}_x)^2 \quad (2.4)$$

onde n é a dimensão da amostra, neste caso o número de dias com o preço de fecho, $r_{x,t}$ é a rentabilidade no instante t e \bar{r}_x é o retorno.

O desvio-padrão nos mercados financeiros também denominado por volatilidade pode ser calculado através da variância:

$$\sigma_i = \sqrt{\sigma_i^2} \quad (2.5)$$

Covariância

A covariância é uma métrica estatística que compara a variabilidade conjunta entre duas variáveis aleatórias, neste caso as variáveis são ações. Esta pode ser calculada da seguinte forma:

$$\sigma_{xy} = \frac{1}{n-1} \sum_{t=1}^n (r_{x,t} - \bar{r}_x)(r_{y,t} - \bar{r}_y) \quad (2.6)$$

n é a dimensão da amostra.

Correlação

Utilizando as métricas calculadas previamente pode-se então calcular a correlação entre duas ações com a seguinte fórmula:

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \quad (2.7)$$

2.5 DETEÇÃO DE ANOMALIAS

Outliers

Um *outlier* é uma observação que difere das restantes observações, estes *outliers* poderão ter significados, como é referido em [13] onde os três maiores *crashes* são considerados *outliers*. Um *crash* no mercado ações é o declínio rápido e severo dos valores de ações. Este pode ser causado por vários fatores ou a combinação destes nomeadamente económicos, financeiros e psicológicos. Tendo em conta o efeito que os *outliers* têm na dinâmica dos mercados financeiros, estes serão utilizados para ajudar a perceber esta dinâmica.

2.5.1 Detecção de anomalias nos valores de correlação

Uma vez que não existe uma base de dados com anomalias previamente identificadas, é utilizado um método para detecção de anomalias não supervisionado. Neste sentido foram avaliadas várias possibilidades [14], [15], onde são mencionados diferentes métodos para detetar *outliers*.

A dimensão da amostra utilizada para os valores de correlação entre ações é composta por 100 valores, que com base na análise feita pelas autoras *S. Saleem et al.* [15], pode-se dizer que tem-se uma amostra de dimensão média, uma vez que eles consideram uma amostra com 66 observações pequena e uma amostra de dimensão 161 grande. Por seu turno, a revisão feita pelo autor *S. Seo* [14] que apresentou uma sistematização relevante de vários métodos para detetar anomalias, elegeu-se o método *Modified Z-Score* para distribuições normais, enquanto que para distribuições que sejam assimétricas, em outras palavras, não seguem uma distribuição normal, utiliza-se o método *MAD_e*, no presente trabalho foram seguidos os métodos *2MAD_e* e *3MAD_e*.

Para perceber qual dos métodos *2MAD_e* ou *3MAD_e* utilizar, tendo por base a discussão de resultados do estudo [15], de salientar que o desempenho do método *2MAD_e* para distribuições ligeiramente assimétricas e uma amostra pequena, originou a detecção de mais anomalias e que para distribuições com assimetria alta e dimensão grande o método é o mais sensível a anomalias. Neste sentido é possível concluir que o método *2MAD_e* é superior comparativamente ao *3MAD_e* que não se destacou em nenhum ponto do estudo.

Para verificar se a amostra segue uma distribuição normal para posteriormente aplicar o método adequado tendo em conta a distribuição *Modified Z-Score* ou *2MAD_e*, é feito um teste de normalidade.

Teste de normalidade

O teste de Kolmogorov-Smirnov [16] é utilizado para avaliar a distribuição da amostra, neste caso os valores históricos de correlação. Este testa a hipótese da amostra seguir uma distribuição normal. Este teste é feito com um nível de significância de 5%.

$$H_0 \sim N(\mu, \sigma) \quad vs \quad H_1 \not\sim N(\mu, \sigma)$$

Caso o *p-value* obtido for inferior ao nível de significância definido, *p-value* < 0.05, rejeita-se a hipótese nula, caso o nível de significância seja superior não se rejeita a hipótese nula *H₀*.

Modified Z-Score

O método *Z-Score* utiliza a média e o desvio padrão, no entanto este método tem uma limitação quando utilizado para detetar *outliers*, uma vez que o desvio padrão pode ser altamente influenciado por *outliers*. O método *modified Z-score* ultrapassa esta barreira, utilizando a mediana e o desvio mediano absoluto (MAD) em substituição da média e desvio padrão. A mediana e o desvio mediano absoluto são métricas mais robustas comparativamente à média e desvio padrão uma vez que não são influenciadas por *outliers*.

$$MAD = median\{|x_i - \tilde{x}|\} \quad (2.8)$$

onde:

\tilde{x} é a mediana da amostra;

x_i são observações da amostra, $i = 1, 2, 3, \dots, n$ para uma amostra de dimensão n

Com o valor obtido para o desvio mediano absoluto é calculado o *Modified Z-Score* para cada observação da amostra que é dado por M_i .

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD} \quad (2.9)$$

onde:

$E(MAD) = 0.6745\sigma$ para uma amostra grande

Para definir o critério que se utiliza para categorizar uma observação como anomalia, *Iglewicz and Hoaglin (1993)* [17] propõem que o valor de *Modified Z-Score* de uma observação deve ser superior a 3.5 para ser considerado um *outlier*.

Median Absolute Deviation

O desvio mediano absoluto é um método robusto que não surge tão enviesado pela presença de valores extremos na amostra.

Os autores dos estudos [14], [15] calculam MAD_e multiplicando o valor MAD que pode ser obtido através da expressão 2.8 por 1.483. No entanto como é referido em *Leys et al.* [18], esta constante apenas deve ser utilizada caso a amostra tenha uma distribuição normal. Uma vez que tal não é aplicável nesta amostra, para o caso da amostra seguir uma distribuição normal, já se definiu o método *Modified Z-Score*. No estudo mencionado anteriormente é referido como proceder para o cálculo desta constante para o caso de distribuições não normais, a constante c pode ser calculada da seguinte forma:

$$c = \frac{1}{Q(0.75)} \quad (2.10)$$

onde:

$Q(0.75)$ é o quantil 0.75, também denominado por terceiro quartil da amostra;

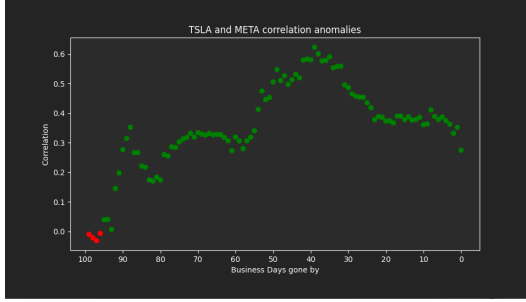
O método $2MAD$ pode ser calculado através da seguinte expressão:

$$2MAD_e \text{ Method} = Median \pm 2MAD_e \quad (2.11)$$

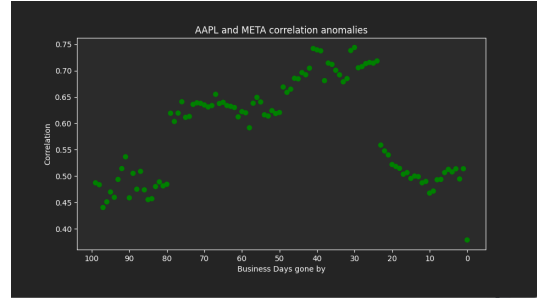
onde:

MAD_e é o valor calculado através da expressão 2.8 multiplicado por c ;

Através deste método é obtido um intervalo, se uma observação se encontrar fora deste, é considerado um *outlier*.



(a) Anomalias detetadas nos valores de correlação das ações TSLA e META



(b) Anomalias detetadas nos valores de correlação das ações AAPL e META

Figura 2.1: Anomalias detetadas utilizando o MAD

No entanto estes métodos não são o suficiente, uma vez que eles apenas permitem detetar anomalias se estes forem *outliers*, isto implica que os *outliers* sejam máximos ou mínimos da amostra como é visível na figura 2.1a. De forma a considerar eventuais anomalias que não sejam máximos nem mínimos utiliza-se o Local Outlier Factor.

Local Outlier Factor

Local Outlier Factor é um algoritmo proposto por *Breunig et al.* [19] para detetar anomalias em observações através do desvio local relativamente aos vizinhos destas. Este algoritmo é baseado na densidade local, onde localidade é definida pelo número de vizinhos. Comparando a densidade local com a dos seus vizinhos, é possível detetar regiões com densidade similar e observações que tenham densidade inferior que os vizinhos, neste último caso são considerados *outliers*. Como é referido por *Braei e Wagner* [20], apresentam-se alguns desafios, nomeadamente a escolha dos parâmetros mais adequados para LOF, quando a série temporal é convertida para espacial, ambas as dimensões são importantes, o que numa série temporal a sequência ordenada contém informação importante e a sequência não é considerada por este algoritmo.

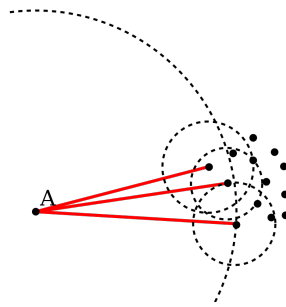


Figura 2.2: Exemplo do algoritmo Local Outlier Factor ¹

¹Wikipedia, Local Outlier Factor. URL: https://en.wikipedia.org/wiki/Local_outlier_factor, [Online; acedido em 19 de Julho, 2023]

Sendo $k\text{-distance}(A)$ a distância de uma observação A para o vizinho k , a distância entre estes, também denominada de *reachability-distance* pode ser calculada da seguinte forma:

$$\text{reachability-distance}_k(A, B) = \max\{k - \text{distance}(B), d(A, B)\} \quad (2.12)$$

Isto implica que a *reachability distance* é no mínimo a distância de um objeto A para o vizinho k .

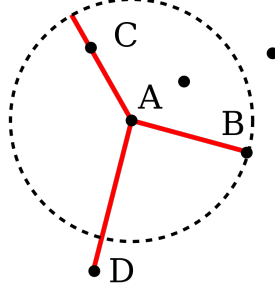


Figura 2.3: Exemplo de *reachability-distance* para $k=3$ ²

Na figura 2.3, pode-se dizer que B e C têm o mesmo valor para *reachability-distance*_k uma vez que o máximo para estas observações é $k - \text{distance}(A)$, onde $k = 3$. D, não é um 3 vizinho mais próximo.

A *local reachability density* de uma observação A é definida por:

$$\text{lrd}_k(A) := \frac{1}{\frac{\sum_{B \in N_k(A)} \text{reachability-distance}_k(A, B)}{|N_k(A)|}} \quad (2.13)$$

A *local reachability density* é o inverso da média da *reachability distance* de uma observação A do seu vizinho, ou seja a distância a que a observação A pode ser alcançada pelos seus vizinhos.

Por fim a *local reachability density* é comparada com a dos vizinhos através da seguinte expressão:

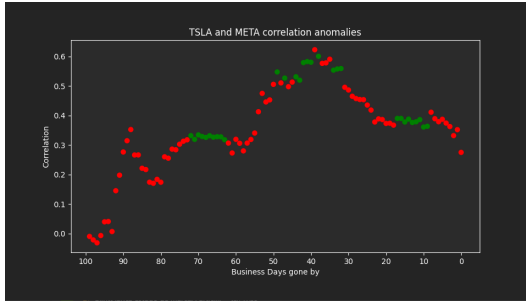
$$\text{LOF}_k(A) := \frac{\sum_{B \in N_k(A)} \frac{\text{lrd}_k(B)}{\text{lrd}_k(A)}}{|N_k(A)|} = \frac{\sum_{B \in N_k(A)} \text{lrd}_k(B)}{|N_k(A)| \text{lrd}_k(A)} \quad (2.14)$$

Os valores obtidos para $\text{LOF}(k)$ indicam como a densidade de uma amostra se compara aos seus vizinhos, $\text{LOF}(k) \sim 1$ significa que tem uma densidade similar à dos vizinhos, $\text{LOF}(k) < 1$ significa que tem uma densidade superior relativamente aos vizinhos e por fim $\text{LOF}(k) > 1$ demonstra que tem uma densidade inferior comparativamente aos vizinhos, neste caso é um *outlier*, no LOF definiu-se o número de vizinhos $k = 10$, o que o permite ser mais sensível a desvios locais da densidade média.

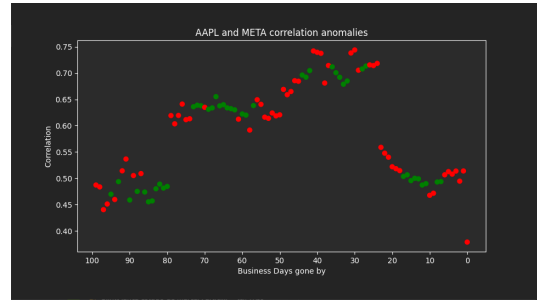
Antes de aplicar o algoritmo LOF, para as *features* deste algoritmo que são nomeadamente os valores de correlação e a escala temporal representada sequencialmente em numeração

²Wikipedia, Local Outlier Factor. URL: https://en.wikipedia.org/wiki/Local_outlier_factor, [Online; acedido em 19 de Julho, 2023]

numérica, foram transformados para a mesma escala $[-1, 1]$ utilizando *MinMaxScaler* [21], este método comparativamente a outros *Standardization*, *Normalization* têm a vantagem de ampliar os *outliers* o que é benéfico para a aplicação do LOF no âmbito de detecção de anomalias. A normalização é utilizada para colocar ambas as dimensões no mesmo intervalo, embora os valores de correlação estão entre -1 e 1 estes também são redimensionados, tendo uma distribuição mais ampla neste intervalo, os valores temporais também foram normalizados para a escala -1 e 1.



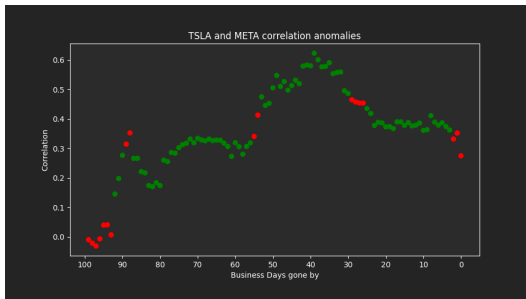
(a) Anomalias detetadas nos valores de correlação das ações TSLA e META



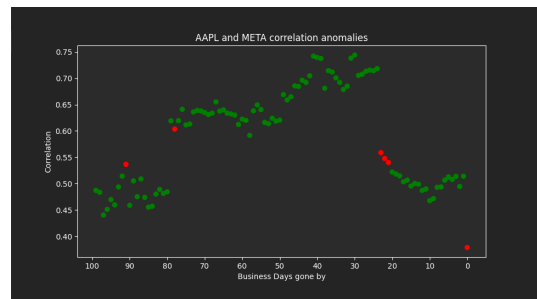
(b) Anomalias detetadas nos valores de correlação das ações AAPL e META

Figura 2.4: Anomalias detetadas com um *threshold* 1 para o LOF no dia 08/06/2023 depois do mercado ter fechado

Como se pode observar na figura 2.4 grande parte das observações são consideradas anomalias, para minimizar o número de anomalias detetadas aumenta-se o valor *threshold* para as quais se considera um *outlier*. Como foi referido anteriormente, o valor padrão é 1, mas utilizando um *threshold* com o valor 1.2 reduz significativamente o número de anomalias detetadas como se pode verificar na seguinte figura:



(a) Anomalias detetadas nos valores de correlação das ações TSLA e META

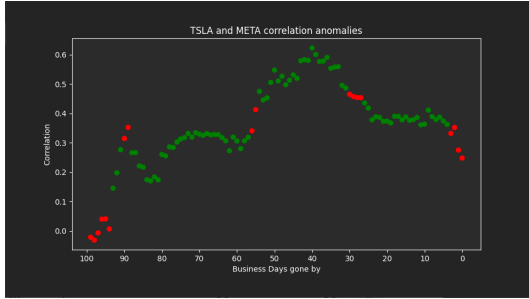


(b) Anomalias detetadas nos valores de correlação das ações AAPL e META

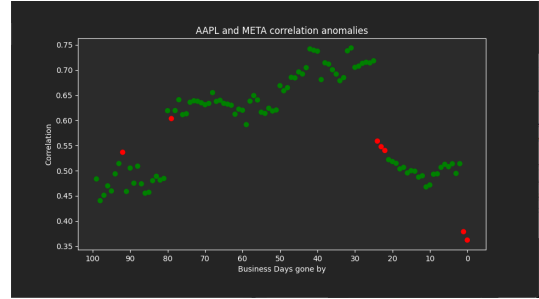
Figura 2.5: Anomalias detetadas com um *threshold* 1.2 para o LOF calculadas no dia 08/06/2023 depois do mercado ter fechado

O valor para o *threshold* foi baseado em alguns dos valores calculados pelo algoritmo LOF, uma vez que não se possui uma base de dados com anomalias detetadas. Desta forma não é possível fazer o *fine tuning* dos parâmetros e, para além disso este aumento ao *threshold* também reduz um possível problema que é a tendência de observações no início e fim da série temporal serem consideradas como anomalias, uma vez que apenas têm vizinhos à sua direita ou esquerda, respetivamente que resulta numa distância ao longo do eixo X (eixo

temporal) superior às observações que têm vizinhos à esquerda e direita. Esta diferença é possível observar na comparação entre as figuras 2.4 e 2.5.



(a) Anomalias detetadas nos valores de correlação das ações TSLA e META



(b) Anomalias detetadas nos valores de correlação das ações AAPL e META

Figura 2.6: Anomalias detetadas com um *threshold* 1.2 para o LOF calculadas no dia 09/06/2023 depois do mercado ter fechado

Comparando diferentes figuras, a primeira 2.5 observa-se que as anomalias detetadas no dia 08/06/2023 para ambos os casos (TSLA e META, AAPL e META) corresponderam efetivamente para um decréscimo nos valores de correlação no dia 09/06/2023 correspondente à figura 2.6.

Combinar ambos os modelos ou seja *Modified Z-Score* ou MAD dependendo da distribuição, com o algoritmo LOF não traz nenhuma vantagem, uma vez que a vantagem deste último algoritmo é que consegue determinar anomalias que os outros modelos não conseguem.

Para o caso do *Modified Z-Score* para valores M_i for inferior a 3.5, este método não deteta anomalias e recorro ao LOF para detetar anomalias. Outra situação possível de ocorrer advém de quando as observações que têm o valor M_i , calculado pelo *Modified Z-Score* é superior a 3.5, vejo se para esta observação o valor $LOF_k(A)$ é superior a 1.2 e, se for o caso, considero como uma anomalia, $(Modified\ Z-Score \wedge LOF)$. Isto pode ser representado pela seguinte tabela de verdade:

Modified Z-Score	LOF	Modified Z-Score \wedge LOF	$(Modified\ Z-Score \wedge LOF) \vee (LOF)$
1	1	1	1
1	0	0	0
0	1	0	1
0	0	0	0

Tabela 2.1: Tabela de verdade para os modelos

onde:

1 representa que o método detetou com anomalia e 0 que não detetou como anomalia;
 $(Modified\ Z-Score \wedge LOF) \vee LOF$ representa a *label* final para a observação;

Na tabela 2.1 anterior pode-se observar que a *label* final depende apenas da classificação do LOF pelo que basta apenas usar o LOF uma vez que se utiliza menos recursos computacionais, o mesmo se aplica no caso em que se combina MAD com LOF.

2.6 DETEÇÃO DE ANOMALIAS EM TEMPO REAL

Para detetar anomalias em tempo real, é necessário perceber quando os mercados fecham, para ser obtido o preço de fecho de uma ação e atualizam-se os valores de correlação para as ações para quais os mercados fecharam, se o módulo do valor mais recente de correlação $|p_{x,y}|$ for superior 0.5, utiliza-se o LOF para detetar anomalias na série temporal, após isto verifica-se se o valor mais recente de correlação foi considerado uma anomalia, e se for é emitido um alerta. Foi definido um *threshold* de 0.5 uma vez que só é relevante detetar anomalias em tempo real para ações correlacionadas, ações que não estavam correlacionadas e agora estão, ações que estiveram correlacionadas e que deixaram de estar ou ações que poderão eventualmente correlacionar num futuro próximo. Para cumprir estas condições o *threshold* tem que ser inferior ao limite em que se considera ações correlacionadas e têm que ter um intervalo significativo para poder capturar grandes alterações no mercado.

2.7 PLANAR MAXIMALLY FILTERED GRAPH

Tumminello et al. [22] propõe a utilização de *Planar Maximally Filtered Graph (PMFG)* para filtrar informações de um sistema complexo, este tipo de grafo é adequado para grafos baseados em correlação, também é feita a comparação entre o PMFG e o MST (*Minimum Spanning Tree*), onde a diferença é o número de conexões o MST possui $(n-1)$ conexões enquanto que o PMFG com genus k igual 0, contém $3(n-2 + 2k)$ conexões, sendo ou seja, $3(n-2)$, também é referido que o MST está contido no grafo PMFG. Os autores também referem que numa rede baseada na correlação, tendo três elementos A, B e C se A está correlacionado com B e B está bastante correlacionado com C, existe uma probabilidade alta de A estar correlacionado com C criando um *clique* triangular. Caso exista outro elemento D que esteja bastante correlacionado com qualquer um dos três elementos previamente mencionados, os restantes dois elementos também deverão estar correlacionados com o elemento D gerando um clique com quatro vértices, tetraedro.

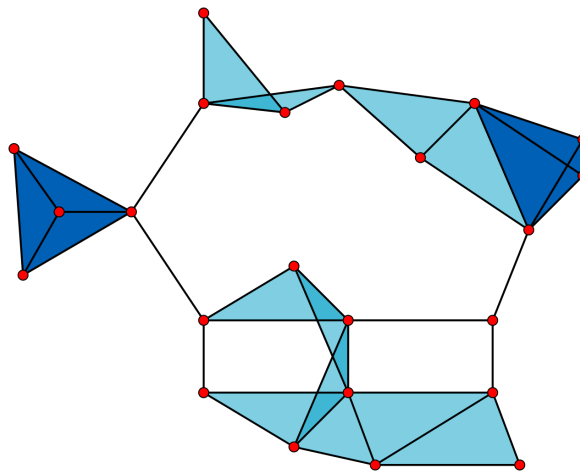


Figura 2.7: Exemplo de cliques para um grafo ³

No PMFG cada vértice ou nó representa uma ação e as arestas os valores de correlação entre elas. Um clique é um subconjunto de vértices de um grafo não direcionado tal que cada dois vértices distintos no clique são adjacentes.

Como é possível observar na figura 2.7, cliques com 1 vértice são os respetivos vértices, cliques com 2 são arestas, com 3 são as áreas azuis claras e escuras e com 4 são as áreas escuras.

Huang et al. [23] sugerem utilizar um *threshold* θ no intervalo $[0.55, 0.69]$ para ter redes altamente agrupadas. Tendo este intervalo como base definiu-se $\theta = 0.55$, para ter uma grande quantidade de arestas, uma vez que os valores de correlação não são tão elevados. Isto significa que só serão adicionadas arestas entre duas ações se o valor de correlação delas $\rho_{x,y} \geq 0.55$

Diferentes estudos [24]–[26] analisam diferentes métodos para detetar *clusters*, dentre estes utiliza-se o *Louvain community detection*[27] para detetar aglomerados no PMFG criado.

2.7.1 Louvain community detection

Este método foi inspirado pela otimização de modularidade conforme o algoritmo faz progressos. A modularidade é uma medida de uma estrutura de redes ou grafos onde se mede a intensidade da divisão de uma rede em grupos. A escala de modularidade está entre -0.5 e 1, que reflete a densidade relativa de arestas dentro de uma comunidade comparativamente a arestas fora da comunidade. Este método começa por detetar as comunidades pequenas através da otimização de modularidade local em todos os nós, após isto cada comunidade é agrupada em um nó ou vértice e repete-se o primeiro passo.

Uma vez que a complexidade de uma rede é elevada e calcular a melhor solução para o ajuste de ações em grupos é computacionalmente exigente, este método utiliza métodos heurísticos, ou seja, os grupos calculados por este método não são os mais otimizados, mas são uma aproximação da solução otimizada. Paralelamente a isto também implica que aplicar o mesmo algoritmo ao mesmo grafo pode resultar em grupos ligeiramente diferentes.

O valor para a modularidade pode ser calculado da seguinte forma:

$$Q = \frac{1}{2m} \sum_{i,j} [A_{i,j} - \frac{k_i k_j}{2m} \delta(c_i, c_j)] \quad (2.15)$$

³Wikipedia, Clique (graph theory). URL:[https://en.wikipedia.org/wiki/Clique_\(graph_theory\)](https://en.wikipedia.org/wiki/Clique_(graph_theory)), [Online; acedido em 19 de Julho, 2023]

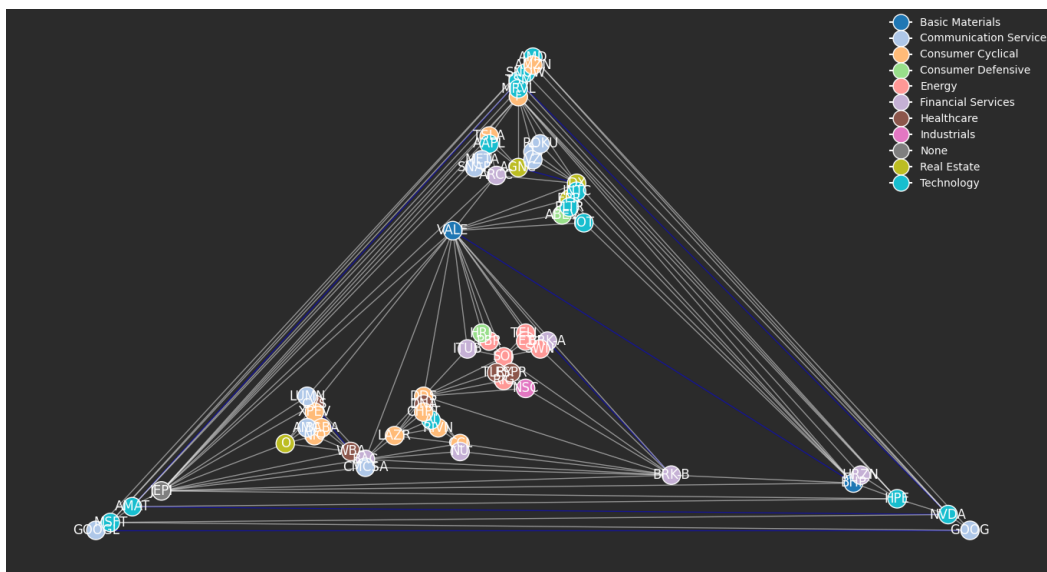


Figura 2.8: Grafo PMFG com 63 ações

Pode-se observar que leitura do grafo não é a mais intuitiva nem a mais legível, tendo *cliques* bastante distanciados.

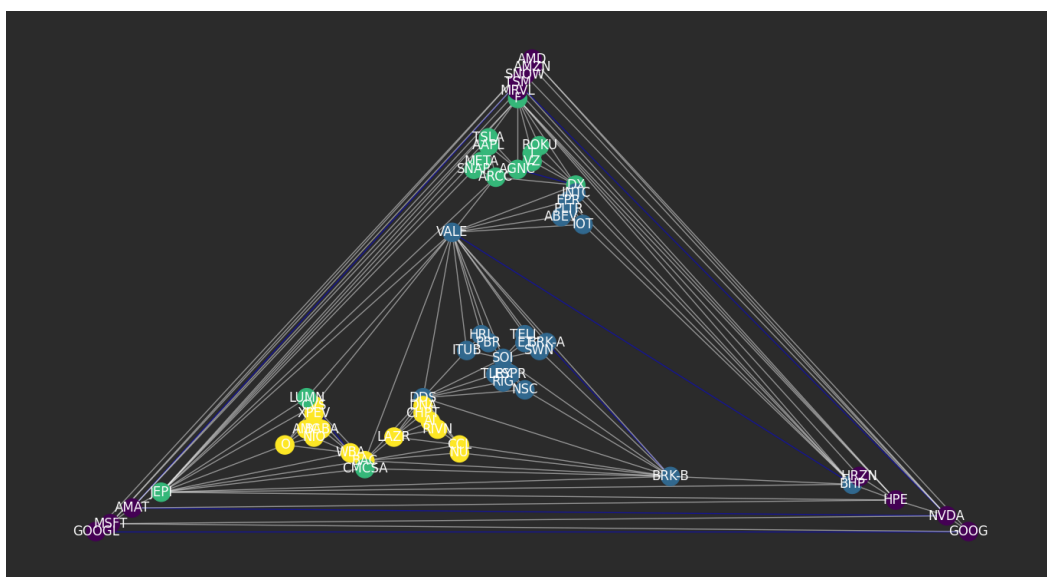


Figura 2.9: Grafo PMFG com 63 ações com grupos identificados

Na figura 2.9 pode-se verificar que o algoritmo *Louvain community detection* agrupou as ações em quatro grupos.

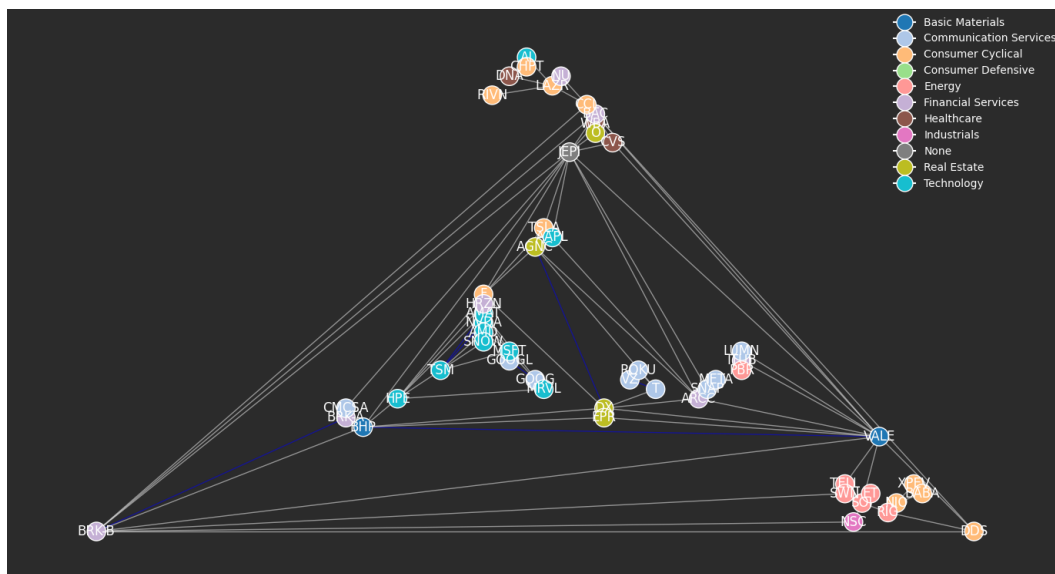


Figura 2.10: Grafo PMFG com 63 ações utilizando o *threshold*

Comparando a figura 2.10 com a figura 2.8, conclui-se que a dinâmica entre as ações alterou significativamente. No canto inferior direito é possível observar um subgrafo que está isolado da restante rede.

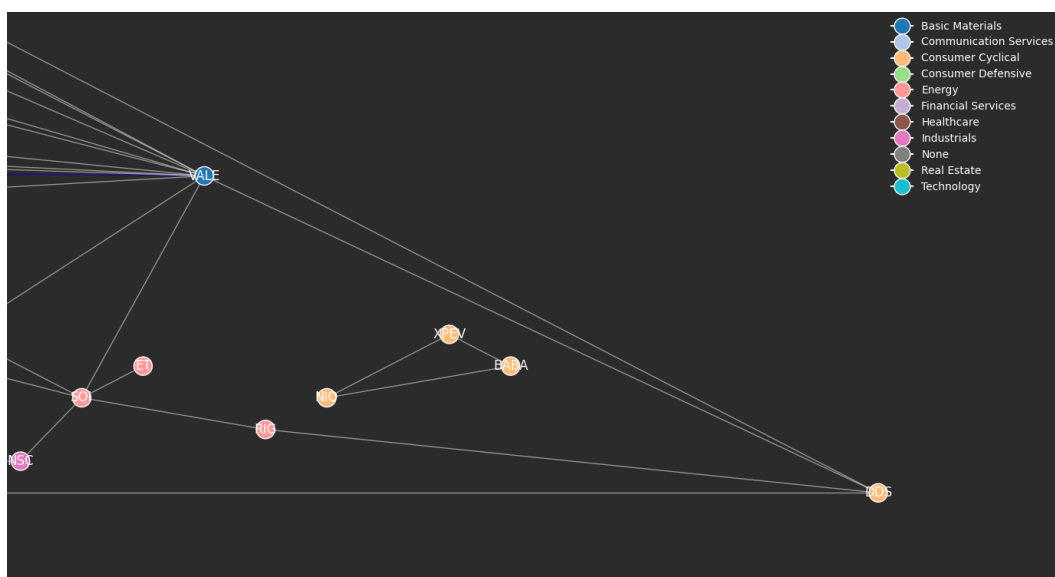


Figura 2.11: Subgrafo isolado na rede com 63 ações com o *threshold* aplicado

Na figura 2.11 observa-se que estas ações pertencem ao mesmo setor.

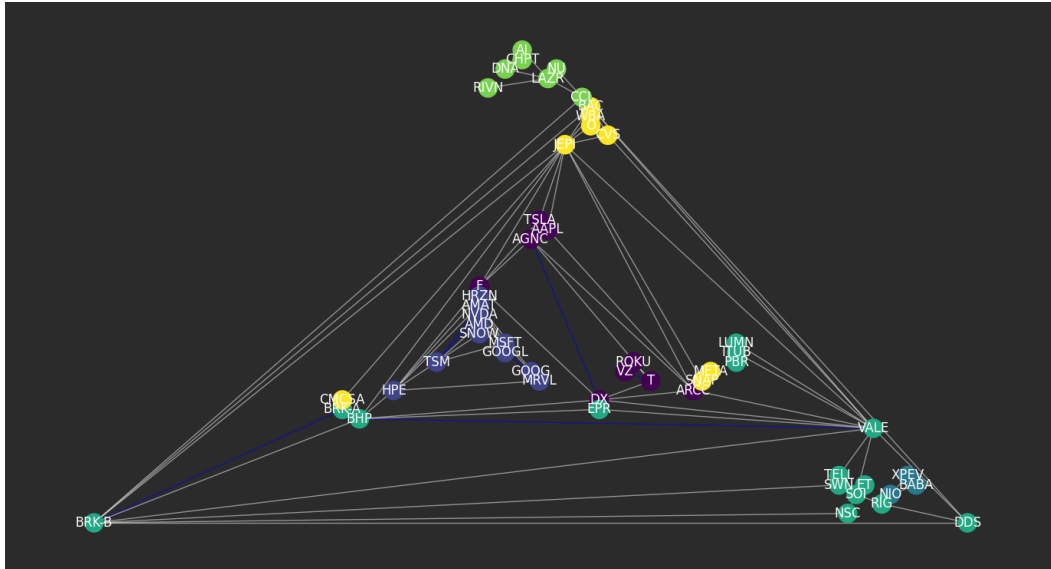


Figura 2.12: Grafo PMFG com 63 ações utilizando o *threshold* com grupos identificados

Com o *threshold* aplicado, verifica-se na figura 2.12 que efetivamente a rede está mais agrupada, com 6 grupos comparativamente aos 4 grupos observados na figura 2.9.

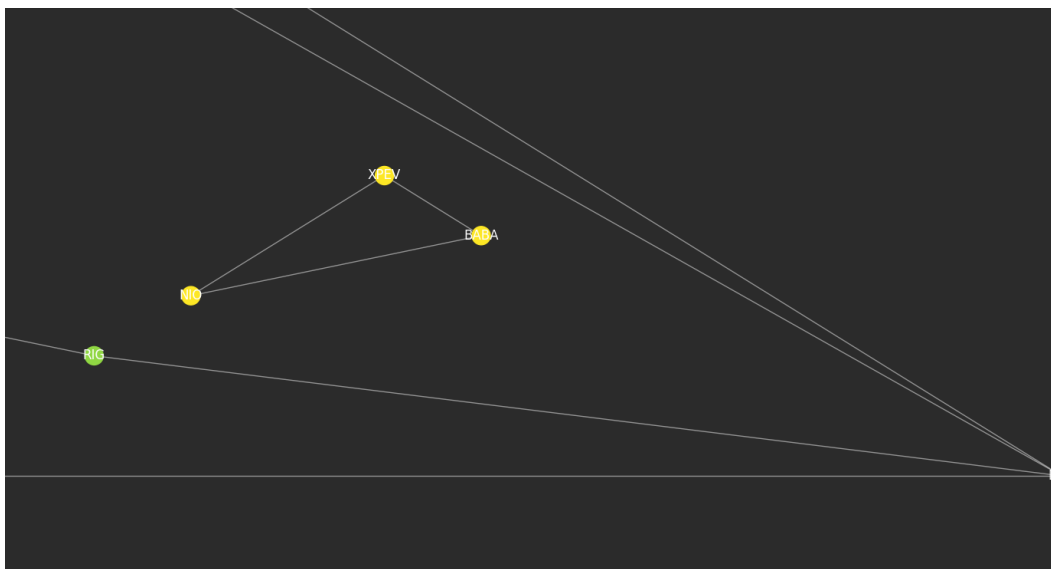


Figura 2.13: Subgrafo isolado na rede com 63 ações com o *threshold* aplicado identificados como grupo

Analisando a figura 2.13, pode-se concluir que o algoritmo *Louvain community detection* considerou este subgrafo como um grupo.

Pozzi *et al.* [28] investigaram quais ações teriam menor risco e maior retorno conforme a sua localização no grafo PMFG e MST, e concluíram que ações localizadas nas periferias dos grafos tinham menor risco e maior retorno comparativamente às ações mais centrais dos grafos. Uma das possíveis razões para isto, segundo os autores, é que as ações mais centrais estão mais sujeitas a perturbações repentinas durante períodos de *booms* e *crashes* em que o mercado fica bastante correlacionado e os investidores tendem a comprar ou vender ações que estejam centradas.

Apesar destes autores utilizarem diferentes algoritmos para categorizar ações centradas e ações nas periferias, o mesmo poderá ser aplicado ao utilizar o PMFG com o *threshold*, embora utilizar o algoritmo *Louvain community detection* torna mais fácil identificar uma vez que há mais grupos, os grupos terão menos ações para analisar dentro destes. A ação que tiver o maior número de arestas seria a ação central uma vez que esta é aquela que pode impactar mais rapidamente o grupo onde ela se encontra, as ações que se iriam encontrar nas periferias seriam aquelas com menos conexões sendo influenciada por um menor número de ações e mais distantes da ação central, o que permite que caso a ação central seja perturbada, esta perturbação demorar mais tempo a chegar a esta ação.

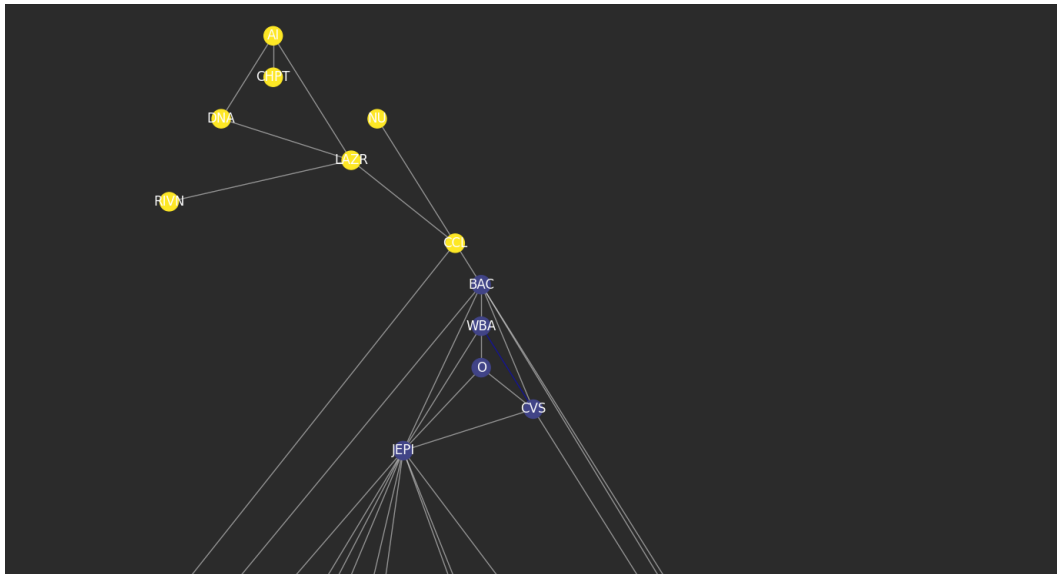


Figura 2.14: Análise de um grupo (amarelo) na rede com 63 ações com o *threshold*

Na figura 2.14 a ação central seria a ação *LAZR* e uma ação na periferia seria *NU* a ação *CHPT* também está a mesma distância da ação central, a duas arestas desta, no entanto têm mais caminhos possíveis pelo que a melhor escolha seria *NU*.

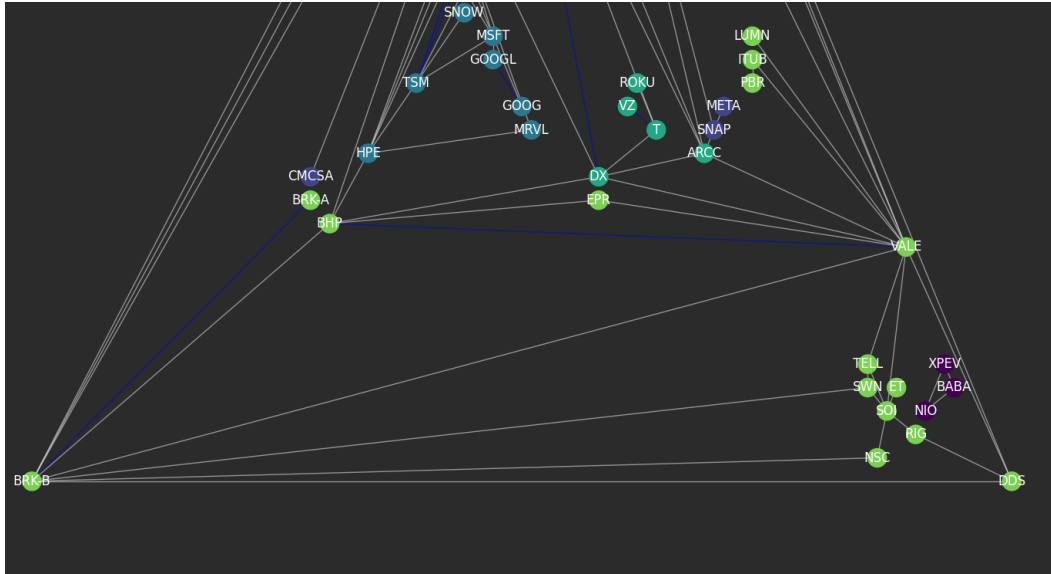


Figura 2.15: Análise de um grupo (verde) na rede com 63 ações com o *threshold*

Na figura 2.15 o grupo já tem uma dimensão maior, pelo que se torna mais difícil determinar qual ação está na periferia, a ação central é a *VALE*, as ações neste grupo têm no máximo 2 arestas de distância, no entanto a melhor opção seria a ação *PBR*, a *BRK-A* também poderia ser uma opção, mas têm mais caminhos para esta e para além disso têm uma correlação mais elevada com a ação *BRK-B* notado pela linha azul.

Aplicação

Quando se abre a aplicação obtém-se a seguinte interface:

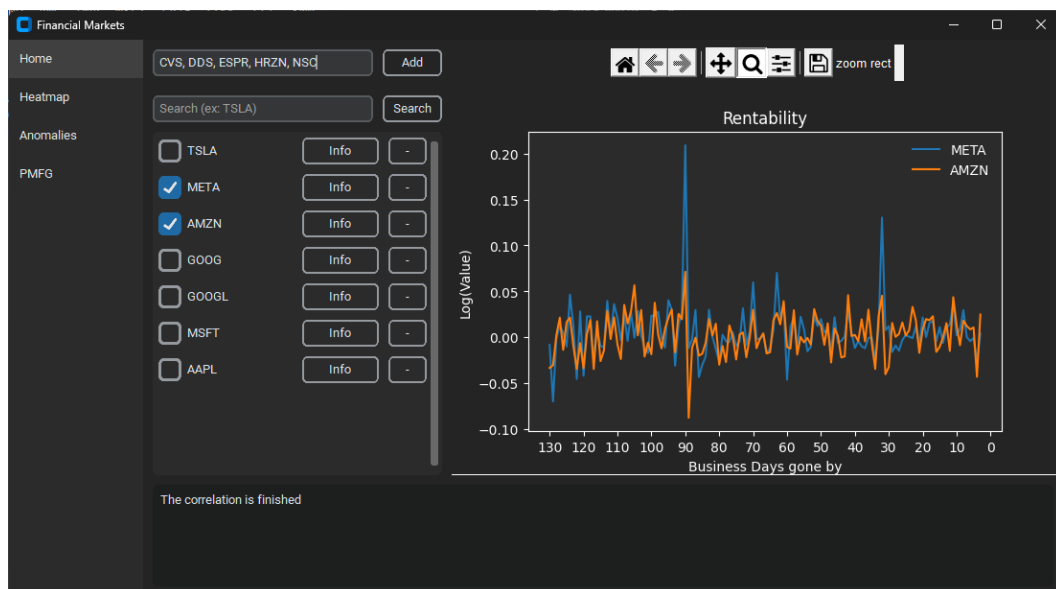


Figura 3.1: Interface inicial da aplicação

Na figura 3.1 pode-se introduzir as ações através dos seus símbolos, e se estiverem listadas no Yahoo Finance [10] são extraídos os dados. Apenas dar nota que, caso hajam demasiados *requests* ao *website*, este pode bloquear temporariamente o acesso, pelo que neste caso não se consegue extrair dados de ações e deve-se tentar mais tarde. Depois de adicionadas as ações também é possível removê-las. A aplicação permite visualizar e comparar a rentabilidade de diferentes ações, clicando na *checkbox* da lista de ações, caso se tenha uma lista extensa de ações pode-se procurar por ações para utilizando a opção *Search*, para além disso clicando em *Info* pode-se abrir a seguinte janela:

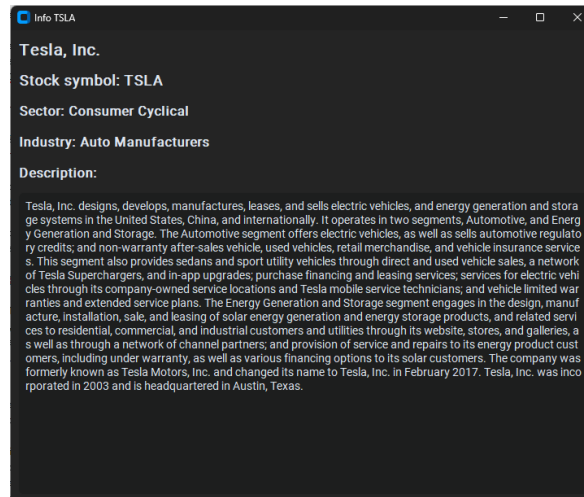


Figura 3.2: Informação sobre uma ação

Nesta janela é possível observar informações como o setor, a indústria e a descrição da mesma.

Para visualizar o *heatmap* das ações clica-se em *heatmap* na barra lateral e obtém-se a seguinte interface:

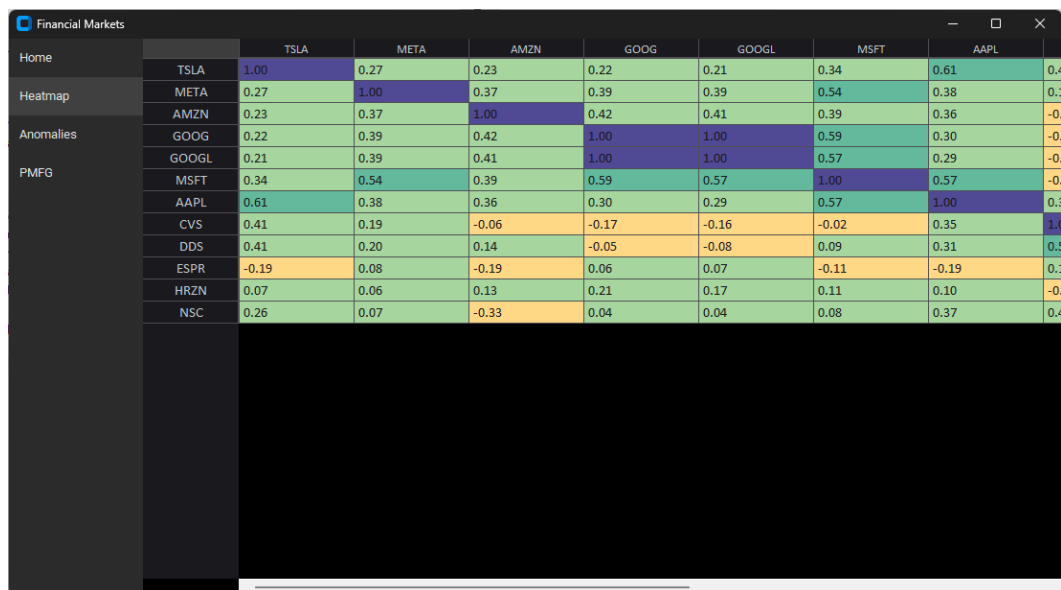


Figura 3.3: Interface do heatmap

Nesta figura observa-se uma tabela com os valores de correlação aos quais são atribuídas cores dependendo destes valores, caso a tabela seja extensa, pode-se fazer *scroll* para visualizar os restantes valores.

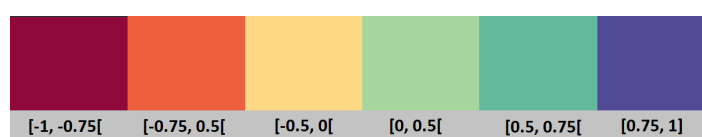


Figura 3.4: Cores atribuídas conforme valores do coeficiente de correlação

A figura 3.4 mostra os intervalos definidos para os valores do coeficiente de correlação com a respetiva cor associada.

Para visualizar os valores de correlação ao longo do tempo entre diferentes ações e se há anomalias neste valores, clica-se em *Anomalies* na barra lateral onde será direcionado para a seguinte interface:

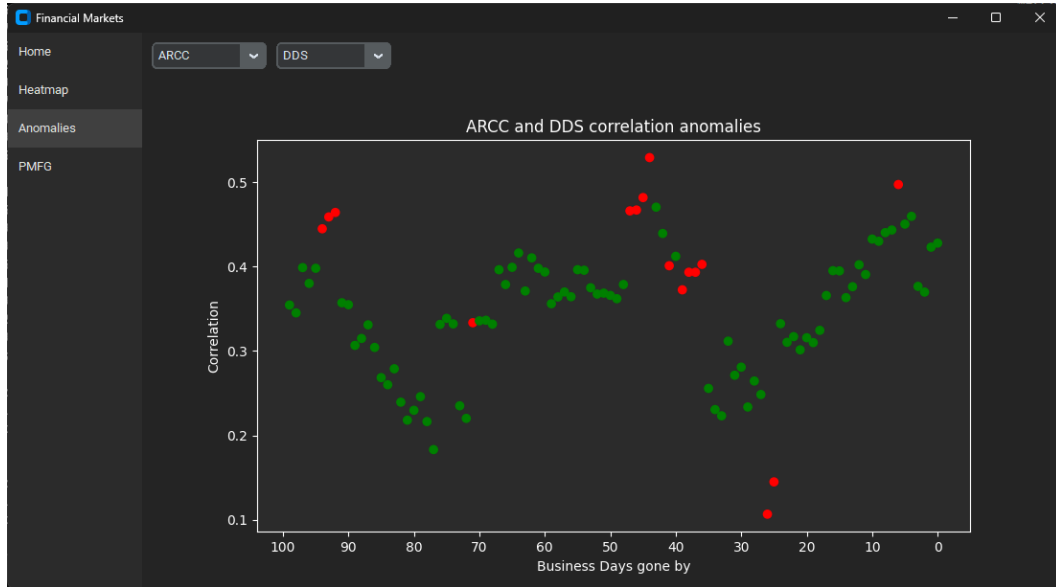


Figura 3.5: Interface para visualizar anomalias

Nesta interface utiliza-se as caixas de combinação para selecionar quais os pares de ações que se pretende observar os valores de correlação.

Para observar os grafos que representam o sistema de ações clica-se em PMFG na barra lateral, onde o utilizador irá encontrar a seguinte interface:

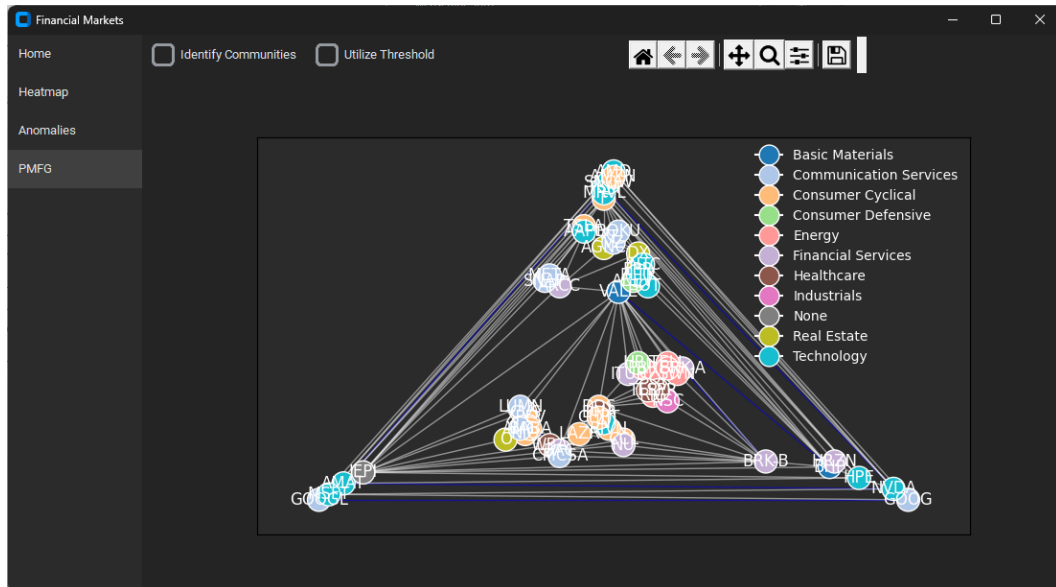


Figura 3.6: Interface para visualizar o sistema

Nesta interface o utilizador pode visualizar o PMFG, clicando na *checkbox Identify Communities* utilizar o algoritmo *Louvain community detection* para agrupar as ações em grupos e ao clicar na *checkbox Utilize Threshold* o utilizador ativa o *threshold* $\theta = 0.55$ para filtrar algumas das arestas dos grafos.

A figura seguinte demonstra o processo de detecção de anomalias.

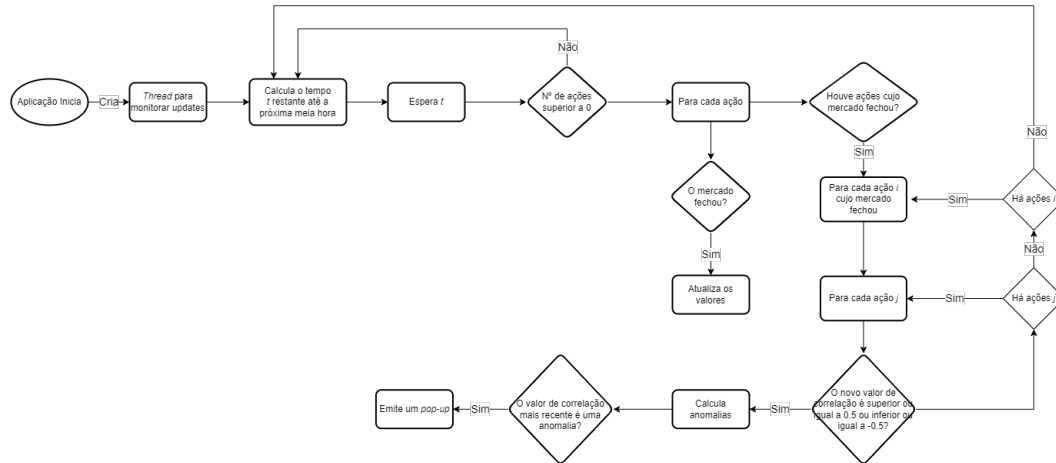


Figura 3.7: Diagrama do *update* de valores de ações e detecção de anomalias

O *pop-up* tem o seguinte aspeto:

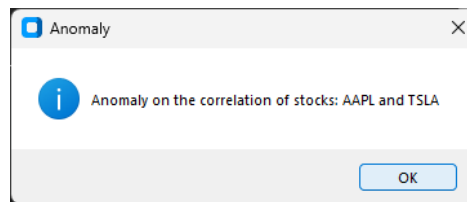


Figura 3.8: Exemplo de *pop-up*

O código da aplicação pode ser consultado em [29]

Conclusão

Através do *heatmap* o utilizador pode analisar os coeficientes de correlação de ações para a construção do seu portefólio, minimizando o risco do mesmo. Para além disso com o auxílio do PMFG, o algoritmo *Louvain community detection* e também o *threshold* que pode ser aplicado no PMFG, pode-se perceber algumas das dinâmicas do sistema complexo que é o mercado de ações assim como identificar grupos de ações.

Para minimizar o risco de um portefólio pode-se utilizar os grupos para a construção do portefólio escolhendo apenas uma ação num determinado grupo e como foi dito anteriormente pode-se escolher ações nas periferias dos grupos para minimizar o risco e maximizar o retorno, uma vez que as ações localizadas nas periferias estão menos sujeitas a perturbações do mercado. Em conjunto com a deteção de anomalias é expectável que caso haja uma perturbação numa ação central de um grupo, seja detetada uma anomalia nesta ação, e analisando o gráfico PMFG se possa avaliar esta perturbação antes desta alcançar a ação localizada na periferia.

Tendo em conta dois cenários onde a ação central tem uma subida repentina no valor, poderia ser uma boa decisão investir numa ação da periferia na expectativa que o preço desta seria influenciado, para capitalizar num possível aumento do valor desta ação na periferia, por outro lado caso a ação central de um grupo tenha uma descida repentina e um dos meus ativos seja uma ação localizada na periferia, talvez seja melhor vender para diminuir o risco e minimizar perdas ou maximizar lucros dependendo da situação.

Neste sentido a deteção de anomalias nos valores de correlação desempenha um papel fundamental uma vez que permite alertar para estes comportamentos, no entanto deve-se analisar a fonte da anomalia para perceber melhor se poderá afetar as outras ações.

A aplicação permite então analisar dinâmicas dos mercados de ações e o utilizador pode tirar proveito desta para tomar decisões informadas, permitindo uma melhor gestão do seu portefólio.

Trabalho Futuro

No âmbito da aplicação algumas operações da aplicação podem ser otimizadas, obtendo assim um melhor desempenho. Para além disto os dados são guardados em memória, pelo que uma base de dados iria melhorar a manipulação de dados, assim como permitir uma paralelização de certas tarefas melhorando o desempenho. Isto torna-se difícil com os dados na memória uma vez que é necessário que gerir o acesso e manipulação de dados, para garantir o funcionamento correto das operações e poder visualizar o gráfico PMFG com a legenda das indústrias para além dos setores.

Este trabalho poderia ser ainda mais relevante caso se tivesse acesso a um *dataset* com anomalias detetadas para os valores de correlação de forma a treinar modelos de *AI* assim como outros modelos matemáticos e avaliar o desempenho destes, uma vez que a deteção de anomalias no valor de correlação é um passo fundamental para obter uma perceção melhor sobre o mercado.

Para além disso pode-se implementar diferentes grafos para além do PMFG que podem transmitir outras informações sobre o mercado assim como diferentes algoritmos para agrupar ações e algoritmos que permitem calcular automaticamente quais as ações que se localizam nas periferias, também permitir a manipulação do valor de *threshold* no grafo PMFG por parte do utilizador que pode gerar grafos diferentes e dar diferentes pontos de vista sobre o sistema complexo que é o mercado de ações.

Referências

- [1] R. Levine e S. Zervos, «Stock market development and long-run growth,» *The world bank economic review*, vol. 10, n.º 2, pp. 323–339, 1996, <https://documents1.worldbank.org/curated/en/396111468766528277/pdf/multi0page.pdf>.
- [2] N. M. Masoud, «The Impact of Stock Market Performance upon Economic Growth,» *International Journal of Economics and Financial Issues*, vol. 3, pp. 788–798, 1996, <https://dergipark.org.tr/tr/download/article-file/362817>.
- [3] W. H. Chan e J. M. Maheu, «Conditional Jump Dynamics in Stock Market Returns,» *Journal of Business & Economic Statistics*, vol. 20, n.º 3, pp. 377–389, 2002, <https://doi.org/10.1198/073500102288618513>. DOI: 10.1198/073500102288618513.
- [4] L. Zhao, W. Li e X. Cai, «Structure and dynamics of stock market in times of crisis,» *Physics Letters A*, vol. 380, n.º 5, pp. 654–666, 2016, <https://www.sciencedirect.com/science/article/pii/S0375960115009937>, ISSN: 0375-9601. DOI: <https://doi.org/10.1016/j.physleta.2015.11.015>.
- [5] «Custom Tkinter,» <https://customtkinter.tomschimansky.com>.
- [6] «NetworkX,» <https://networkx.org>.
- [7] «Louvain Community Detection,» <https://github.com/taynaud/python-louvain>. (acedido em 09/06/2023).
- [8] «SciPy,» <https://scipy.org/install/>.
- [9] «Scikit-learn: Machine Learning in Python,» <https://scikit-learn.org/stable/index.html>.
- [10] *Yahoo Finance*. URL: <https://finance.yahoo.com>.
- [11] H. Markowitz, «Portfolio Selection,» *The Journal of Finance*, vol. 7, n.º 1, pp. 77–91, 1952, ISSN: 00221082, 15406261. URL: <http://www.jstor.org/stable/2975974>.
- [12] S. Agarwal, C. Liu e S. G. Rhee, «Where does price discovery occur for stocks traded in multiple markets? Evidence from Hong Kong and London,» *Journal of International Money and Finance*, vol. 26, n.º 1, pp. 46–63, 2013, <https://www.sciencedirect.com/science/article/pii/S0261560606001069>. DOI: <https://doi.org/10.1016/j.jimonfin.2006.10.011>.
- [13] A. Johansen e D. Sornette, «Stock market crashes are outliers,» *The European Physical Journal B*, vol. 3, pp. 141–143, 1998, <https://link.springer.com/article/10.1007/s100510050163>.
- [14] S. Seo, *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*, University of Pittsburgh, 2006. URL: <http://d-scholarship.pitt.edu/7948/1/Seo.pdf>.
- [15] S. Sehar, M. Aslam e M. R. Shaukat, «A review and empirical comparison of univariate outlier detection methods,» *Pakistan Journal of Statistics*, vol. 37, n.º 4, 2021. URL: <https://www.pakjs.com/wp-content/uploads/2021/08/37406.pdf>.
- [16] A. Kolmogorov-Smirnov, A. N. Kolmogorov e M. Kolmogorov, «Sulla determinazione empirica di una legge di distribuzione,» 1933.
- [17] B. I. e D. C. Hoaglin, «How to detect and handle outliers,» *ASQC Quality Press*, p. 8, 1993, <https://hwbdocuments.env.nm.gov/Los%20Alamos%20National%20Labs/TA%2054/11587.pdf>.

- [18] C. Leys, C. Ley, O. Klein, P. Bernard e L. Licata, «Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median,» *Journal of Experimental Social Psychology*, vol. 49, n.º 4, pp. 764–766, 2013, ISSN: 0022-1031. DOI: <https://doi.org/10.1016/j.jesp.2013.03.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0022103113000668>.
- [19] M. M. Breunig, H.-P. Kriege, R. T. Ng e J. Sander, «LOF: Identifying Density-Based Local Outliers,» 2000, <https://www.dbs.ifi.lmu.de/Publikationen/Papers/LOF.pdf>.
- [20] M. Braei e D.-I. S. Wagner, «Anomaly Detection in Univariate Time-Series: A Survey on the State-of-the-Art,» 2020, <https://arxiv.org/pdf/2004.00433.pdf>.
- [21] «Min-max normalization,» <https://www.linkedin.com/advice/0/what-benefits-drawbacks-different-normalization>.
- [22] M. Tumminello, T. Aste, T. D. Matteo e R. N. Mantegna, «A tool for filtering information in complex systems,» *Proceedings of the National Academy of Sciences*, vol. 102, n.º 30, pp. 10 421–10 426, 2005. DOI: 10.1073/pnas.0500298102. eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.0500298102>. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.0500298102>.
- [23] W.-Q. Huang, X.-T. Zhuang e S. Yao, «A network analysis of the Chinese stock market,» *Physica A: Statistical Mechanics and its Applications*, vol. 388, n.º 14, pp. 2956–2964, 2009, ISSN: 0378-4371. DOI: <https://doi.org/10.1016/j.physa.2009.03.028>. URL: <https://www.sciencedirect.com/science/article/pii/S0378437109002519>.
- [24] T. Di Matteo e T. Aste, «Extracting the correlation structure by means of planar embedding - art. no. 60390P,» *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6039, pp. 227–236, dez. de 2005. DOI: 10.1117/12.637543.
- [25] S. S. Hosseini, N. Wormald e T. Tian, «A Weight-based Information Filtration Algorithm for Stock-correlation Networks,» *Physica A: Statistical Mechanics and its Applications*, vol. 563, p. 125 489, 2021, ISSN: 0378-4371. DOI: <https://doi.org/10.1016/j.physa.2020.125489>. URL: <https://www.sciencedirect.com/science/article/pii/S0378437120307883>.
- [26] N. Musmeci, T. Aste e T. Di Matteo, «Relation between Financial Market Structure and the Real Economy: Comparison between Clustering Methods,» *SSRN Electronic Journal*, vol. 10, jan. de 2014. DOI: 10.2139/ssrn.2525291.
- [27] V. D. Blondel, J.-L. Guillaume, R. Lambiotte e E. Lefebvre, «Fast unfolding of communities in large networks,» *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, n.º 10, P10008, out. de 2008. DOI: 10.1088/1742-5468/2008/10/p10008. URL: <https://doi.org/10.1088/1742-5468/2008/10/p10008>.
- [28] T. D. M. F. Pozzi e T. Aste, «Spread of risk across financial markets: better to invest in the peripheries,» *Nature*, 2013, <https://www.nature.com/articles/srep01665>. DOI: <https://doi.org/10.1038/srep01665>.
- [29] «Código da aplicação,» <https://github.com/rosas-m/Projeto>.