MASTER
MATHEMATICAL FINANCE

MASTER'S FINAL WORK

PROJECT

Analysis of the Cryptocurrency market:
A Network Science Approach

MARTA MELO

OCTOBER – 2024

MASTER

MATHEMATICAL FINANCE

MASTER'S FINAL WORK

PROJECT

Analysis of the Cryptocurrency market:

A Network Science Approach

MARTA MELO

SUPERVISION:

PROF. CARLOS J. COSTA

JOÃO APARÍCIO

OCTOBER – 2024

Acknowledgments

I would like to express my gratefulness to all those who have supported me throughout my academic journey.

Firstly, I would like to thank my supervisors, Professor Carlos Costa and João Aparício, for their impeccable guidance, and without whom it would not have been possible to complete this challenging project.

I am also very grateful to my best friends, Rita, António and André, for their unconditional support and endless motivation, love you guys.

Last but not least, words cannot express my gratitude to my amazing family, especially to my mom, my sister Laura and my aunt Rosarinho, for all the love, encouragement and for supporting me in all my decisions. Your belief in me was one of the things that inspired me to continue this demanding journey.

Abstract

The rapid growth and increasing prominence of cryptocurrencies in the global financial market have brought new challenges in risk management and asset allocation. The high volatility and interconnectedness of digital assets make understanding risk contagion crucial for investors, portfolio managers, and regulators. Network Science provides a powerful framework for studying these interdependencies by modeling relationships as networks where assets are connected based on various metrics, such as correlations or causality measures. The primary objective of this research is to identify price contagion among cryptocurrencies, using Network Science methodologies to analyze these transmission effects and offering practical insights for risk management and portfolio optimization.

The methodology starts with a Network Science approach to model the relationships between cryptocurrencies. Correlation networks are created to visualize the connections between digital assets, indicating where strong relationships and potential contagion effects may occur. To enhance this analysis, Granger causality tests are applied to assess the directionality of these relationships, identifying predictive connections where the performance of one cryptocurrency may impact another. Finally, the Louvain algorithm, a community detection technique within Network Science, is used to cluster cryptocurrencies into groups based on the strength of their interconnections, providing insights into the structural composition of the cryptocurrency market.

The network-based approach reveals significant interconnections among cryptocurrencies, with correlation networks indicating clusters of assets that share strong relationships. Granger causality analysis provides evidence of directional risk transmission, suggesting specific paths through which risk may propagate. The Louvain algorithm identifies groups of highly interconnected cryptocurrencies, offering insights into potential diversification strategies and highlighting areas where risk mitigation may be necessary.

The results inform investors and portfolio managers on managing risk by identifying groups of cryptocurrencies with strong interdependencies, which may impact diversification strategies. Additionally, the findings provide valuable insights for regulators aiming to monitor systemic risk in the cryptocurrency market.

This study advances the understanding of risk contagion in the cryptocurrency market by integrating Network Science methodologies, including correlation networks, causality analysis, and community detection. It offers a comprehensive view of interdependencies and risk transmission, providing practical guidance for constructing more resilient cryptocurrency portfolios.

Keywords: Cryptocurrencies, Network Science, Community Finding, Granger Causality Tests, Correlation Network

Resumo

O rápido crescimento e a crescente proeminência das criptomoedas no mercado financeiro global trouxeram novos desafios em gestão de risco e alocação de ativos. Devido à elevada volatilidade e interconexão dos ativos digitais, compreender o contágio de risco é fundamental para investidores, gestores de portfólios e reguladores. A Ciência das Redes fornece uma forte estrutura para estudar estas interdependências, modelando relações como redes onde os ativos estão conectados com base em vários critérios, como correlações ou medidas de causalidade. O objetivo principal deste trabalho é identificar o contágio de preços entre criptomoedas, utilizando metodologias de Ciência das Redes para analisar estes efeitos de transmissão e oferecendo perspetivas práticas para a gestão de risco e otimização de portfólios.

A metodologia deste projeto começa com uma abordagem de Ciência das Redes para modelar as relações entre criptomoedas. São criadas redes de correlação para visualizar as conexões entre ativos digitais, indicando onde relacionamentos fortes e potenciais efeitos de contágio podem ocorrer. Para aprimorar esta análise, são aplicados testes de causalidade de Granger para avaliar a direção destas relações, identificando conexões preditivas onde o desempenho de uma criptomoeda pode impactar outra. Por fim, é utilizado o algoritmo de Louvain, uma técnica de deteção de comunidades dentro da Ciências das Redes, para agrupar criptomoedas em grupos com base na força das suas interligações, proporcionando perspetivas sobre a composição estrutural do mercado de criptomoedas.

A abordagem baseada em redes revela interligações significativas entre criptomoedas, com as redes de correlação a mostrar agrupamentos de ativos que partilham relações fortes. A análise de causalidade de Granger fornece evidências de transmissão direcional de risco, sugerindo caminhos específicos através dos quais o risco pode propagar-se. O algoritmo de Louvain identifica grupos de criptomoedas fortemente interligadas, dando perspetivas sobre potenciais estratégias de diversificação e destacando áreas onde pode ser necessária a mitigação de risco.

Os resultados informam investidores e gestores de carteiras sobre a gestão do risco, ao identificar grupos de criptomoedas com fortes interdependências, que podem impactar as estratégias de diversificação. Adicionalmente, as conclusões fornecem perspetivas importantes para os reguladores que visam monitorizar o risco sistémico no mercado de criptomoedas.

Esta investigação tenta contribuir para o aprofundamento do conhecimento e compreensão sobre a propagação de risco no mercado de ativos digitais, procurando incorporar metodologias provenientes da Ciência das Redes, tais como redes de correlação, análise de causalidade e deteção de comunidades. Por outro lado, procura proporcionar uma visão abrangente das diferentes interdependências e da transmissão de risco, oferecendo orientações práticas para a construção de carteiras de criptomoedas mais resilientes.

Palavras-chave: Criptomoedas, Ciência das Redes, Comunidades, Testes de causalidade de Granger, Rede de Correlação

Contents

List of Tables

List of Figures

List of Acronyms

ARMA: Autoregressive Moving-Average

CRISP-DM: Cross-industry Standard Process for Data Mining

DJIA: Dow Jones Industrial Index

GN: Girvan-Newman algorithm

LPA: Label Propagation Algorithm

PMFG: Planar Maximally Filtered Graph

S&P 500: Standard & Poor’s 500

SCAN: Structural Clustering Algorithm for Networks

TSE: Tokyo Stock Exchange

Chapter 1. Introduction

1.1 Motivation

Cryptocurrencies are a modern topic in the financial world, and we are talking about a market that has been constantly evolving. Cryptocurrencies are digital currencies that function as decentralized mediums of exchange through computer networks (Aparicio et al, 2020). To provide an idea of how long these digital assets have been around, a brief introduction to the first cryptocurrency, Bitcoin (BTC), is going to be presented. Launched in January 2009, Bitcoin reached a value of $1,000 by November 2013, establishing itself as the leading cryptocurrency. In the following figure (Figure 1), it is illustrated the price of BTC from its release (2009) to the present years.

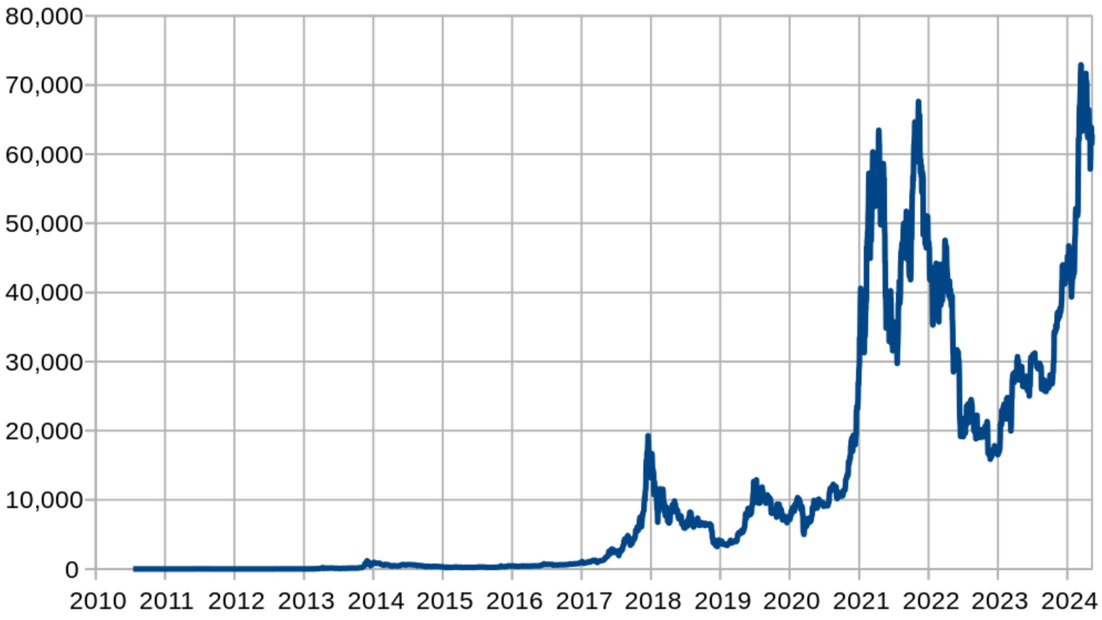

Figure 1: Bitcoin Price in US dollars

This work examines the price contagion based on Network Science, more specifically, on complex networks. These complex networks are graphs that show important topological features and that study individual elements, called nodes, and their interactions, called edges. In this case, these links will represent the correlations between several digital assets.

In order to study a way to construct portfolios from the cross-correlation networks of cryptocurrencies, some known methods will be adopted, such as Granger Causality, Community Detection, thresholds, and a few network metrics.

First of all, it is important to apply a threshold value after computing all Spearman correlations between the assets in order to preserve only strong and significant relationships. This also helps to prevent having a network that has too many weak connections and, consequently, that is too dense and difficult to analyze.

For the purpose of studying the temporal dependence between the data, the Granger Causality test was applied between any two cryptocurrencies that have significant correlations based on the current time window (window size of 6 months (183 days) and window step of 1 month (30 days)). If the results for each one of these tests show a p-value below 0.05, this means that there is statistical significance, and an edge is added to the graph, making a directed network.

Regarding community detection, methods are applied to identify clusters of strongly connected cryptocurrencies within the network, showing groups of digital assets that share strong connections. These clusters make the network structure easier to understand and provide insights into how the different nodes are connected.

Finally, some network metrics are computed in order to evaluate the performance of the Network, providing a meaningful understanding of several aspects of the behavior of the graph. This work measures the out-degree centrality, that is, the proportion of nodes that a certain node's outgoing edges are connected to. It also determines the network density, which measures the fraction of the number of actual edges in the Network by the maximum possible number of edges. Betweenness centrality is another metric that is studied and used to find nodes that serve as a bridge from one part of the Network to another, detecting the amount of a node's influence over the flow of information within the Network.

1.2 Objectives

The main objective of the present work is to study the price contagion between several digital assets, analyzing where a change in prices in one cryptocurrency can spread to other cryptocurrencies, influencing the stability of this market. This is achieved by analyzing a dynamic network of correlations and, from it, explore the diverse communities that those assets form along a temporal window and examine different key measures to evaluate these graphs.

1.3 Document Structure

The following describes the organization of the project and the content of each chapter:

- Chapter 2 briefly describes some works that already exist in this area, highlighting key findings and their connection to the current work.
- Chapter 3 describes the methodology employed in the research, outlining the techniques used to gather and analyze the data.
- Chapter 4 presents the results obtained from this study, as well as some discussion of the main findings and their implications.
- Chapter 5 concludes by synthesizing key findings and proposing potential avenues for future research.

Chapter 2. Literature Review

The main focus of this project is to study the relations between several financial assets based on Network Science and various other methods in order to construct solid and trustworthy complex networks. There are numerous related works on this subject, with many different important outcomes. In this section, I am going to mention some of these works, as well as their key findings and how they are related to my study.

2.1. Background

As it was already said before, cryptocurrencies are a relatively new type of assets that call into question financial stability and risk management. To study this special type of financial asset, it will explore in this sub-section some foundational theories related to risk theory and quantitative finance, namely fundamental concepts about risk contagion and correlation in the financial field.

Risk is a wide concept in finance that explains the uncertainty associated with the future performance of assets, which is defined as the likelihood an outcome or investment's actual gain will differ from the expected outcome or return. There are two different types of risk: systemic risk, which affects the entire market, and idiosyncratic risk, which is specific to individual assets. Regarding the cryptocurrency market, due to the speculative nature of these assets, they exhibit high levels of market risk, as well as a high level of contagion risk. Risk contagion refers to the transmission of financial shocks from one market or financial asset to others, which can generate a financial crisis.

To study this risk contagion, it is very useful to use Network Science, as it allows to understand the dynamics of contagion in financial systems. In this case, the cryptocurrency market can be represented by a network, where all the different digital assets are nodes, and the connections between them are the edges, which represent the correlations among these variables. This correlation is also an important matter in the financial sector, as it is a known statistical measure used to quantify the relationship between any two variables. (Xu & Gao, 2019), (Andersen, Bollerslev, Christoffersen, & Diebold, 2005), (Kenett, Huang, Vodenska, Havlin, & Stanley, 2015).

In cryptocurrency markets, it is fundamental to understand how correlations evolve, especially in periods of market stress. Unlike traditional assets, cryptocurrencies exhibit extreme volatility and speculative behavior, which can cause rapid shifts in correlation structures. For this reason, understanding correlation networks can be a powerful tool for investors to identify critical cryptocurrencies that may act as risk propagation channels, also detecting periods of more vulnerability to systemic events.

Community detection refers to the identification of groups of nodes within a network that are more densely connected to each other than the rest of the network. There are numerous algorithms for this process, namely the Louvain Algorithm, that which will be the one used in this study, which is a method based on the optimization of modularity, a measure of the relative density of edges inside the community in relation to the edges outside. Other known methods for Community Detection are the Girvan-Newman algorithm, that applies a divisive approach, deleting edges with maximum betweenness to find community structures; the INFOMAP method, which focuses on information compression and uses random walks to identify communities. The LPA algorithm a method that assigns labels to nodes based on their neighbors; and the SCAN algorithm, that focuses on the structural properties of Networks to identify clusters of interconnected nodes. (Huang, Chen, Ren, & Wang, 2021)

By incorporating these notions into the study, this background aims to provide a comprehensive understanding of the risks associated with cryptocurrencies and the approaches to evaluate and manage these risks effectively.

2.2. Related Works

In this section, it is presented some of the relevant research conducted in this field, structured in a Literature Review table, as it is shown next. This table points the main characteristics of each study, such as the central objective, what data was used, as well as the methods used. With the intention of enhancing the key points of the table and clarifying the insights provided by the literature, it will also explore some of the subjects discussed, as well as the main conclusions of each study.

Table 1: Literature Review Table

References	Objective	Data	Methods	Conclusions
(Ioannidis, Sarikeisoglou, & Angelidis, 2023)	Portfolio construction based on Network Science.	Daily returns of 26 stocks from the DJIA index from January 1998 to December 2022.	Pearson correlation; Transfer Entropy; Threshold technique: weak correlations between the interval [-0,1;0,1]; Rolling Window of 2 years.	Low centrality stock portfolios have the highest long-term risk-adjusted returns; High centrality portfolios are the riskiest; Transfer Entropy portfolios are the best in crises.
(Li, Jiang, Tian, Li, & Zheng, 2018)	Filter the noise interference and understand the driving mechanism of different network interactions.	Logarithmic returns of 200 stocks from the CSI 300 index (Chinese stock market) and 400 from the S&P 500 from 2009 to 2016.	125-day moving window; Global motion (eigenmode of the largest eigenvalue); PMFG filtered network.	Optimal assets can be found in peripheral positions of the global motion matrix.
(Wu, Tuo, & Xiong, 2015)	Stock correlation network built to investigate community structure.	SSE 180-index component stocks from 2009 to 2011.	Community Structure Detection: GN algorithm.	Stocks in the same industry are assigned to the same community; Correlations among different industries.
(Zheng & Song, 2018)	Networks of stock markets are constructed by using the Grager causality tests.	Index data of 34 major stock markets in Asia, America, Europe and Oceania from 2004 to 2017.	Sliding Window; Granger Causality tests.	Network topology shifts significantly during crises; A causal relationship between any two stock markets can usually be established with one stock market.
(Výrost, Lyócsa, &	Granger causality networks to model the	Daily closing prices from 20 stock market indices from 4	ARMA model; Rolling samples of 3 months; Granger	Impact of the US stock market has declined; Temporal proximity of market closing times

References	Objective	Data	Methods	Conclusions
Baumöhl, 2015)	complex relationships of return spillovers.	continents (20 developed countries), from 2006 to 2013.	Causality tests; Spatial probit.	influence return spillovers.
(Isogai, 2017)	Analysis of a dynamic correlation network of highly volatile financial asset returns.	Japanese Stock returns are listed on the TSE.	Hierarchical recursive Network clustering groups of stocks into a dynamic network using model-based correlation estimation.	Three sub-period correlation networks show stability over time, with higher correlations during stressed periods (T3) compared to normal periods (T1 and T2).

The work by Ioannidis, Sarikeisoglou, & Angelidis (2023), studies the construction of a portfolio of daily stocks from the Dow Jones Industrial Average (DJIA) based on Network Science. These networks are estimated from the Pearson correlation coefficient and Transfer Entropy. The main key findings conclude that peripheral portfolios of low centrality stocks are the best in the long term and that the Markowitz portfolio is the most reliable in the long term. In contrast, central portfolios of high centrality stocks are more uncertain and in times of crisis, portfolios based on Transfer Entropy perform better.

In the study by Li, Jiang, Tian, Li, & Zheng (2018), portfolios with varied performances are built from network filtering and peripherality measures. These networks are constructed using the full cross-correlation matrix and the global-motion one, respectively, and it is demonstrated that the peripherality in a global-motion network can work as an indicator for identifying optimal assets. This study also concludes that the global-motion cross-correlation matrix is useful in portfolio optimization and that peripheral nodes are more profitable and well-diversified than central nodes.

The paper "Network Structure Detection and Analysis of Shanghai Stock Market" (Wu, Tuo, & Xiong, 2015), aims to investigate the community structure of the component stocks of SSE (Shanghai Stock Exchange) 180-index, a stock correlation network is constructed to find the intra-community and inter-community. The methodology used for the community structure is based on the GN algorithm on an un-weighted network with different thresholds that are set according to coefficient distribution. The result of the network community structure analysis shows that the stock market has obvious industrial characteristics, as most of the stocks in the same industry or in the same supply chain are assigned to the same community. It also shows that there exist correlations among different industries.

In “Dynamic Contagion of Systemic Risks on Global Main Equity Markets Based on Granger Causality Networks” (Zheng & Song, 2018), 156 causal networks of stock markets are built using the Granger Causality tests and time series sliding window based on stock index data of 34 major stock markets in the world from 2004 to 2017. The results show that contagions between any two stock markets is established with one stock market on average and that the markets that have a significant impact in systemic risk contagion are the ones with a higher intermediate contagion ability. It is also concluded that despite having weak network correlations, markets with strong media ability perform a crucial role in risk contagion, such as the markets of Australia, Korea, and Japan.

The study by Výrost, Lyócsa, & Baumöhl (2015) examines the structure of return spillovers by creating Granger causality networks using daily closing prices of 20 developed markets from 2nd January 2006 to 31st December 2013. The main findings conclude that the most influential returns originate from European stock markets, while the influence of the US stock market was stronger before and during the financial crisis than afterward.

In the analysis made by Isogai (2017), it is used a network clustering algorithm to study a dynamic correlation network of 1324 Japanese stock returns over 8 years. This work contributes to two methods of dimensionality reduction that extract important information from complex correlation networks. The first method is the reduction of a large correlation network into a smaller factor correlation network, and the second is based in the reduction of a time series of a correlation network into a number of representative correlation networks.

In all of these previously presented studies, the authors apply several network metrics, that is, quantitative measures utilized to assess and track the performance and reliability of a network while also providing valuable insights into diverse aspects of network behavior. From the analysis of these metrics, it is easy to understand that the predominantly used are centrality measures, namely the Degree Centrality and the Betweenness Centrality, which represent the number of edges connected to a node and the number of shortest paths that pass through a given node, respectively.

One of the main divergences of this work relative to most of the papers analyzed it is the type of correlation coefficient used to analyze the interdependencies between the data. It is usual to compute the correlation coefficient using the Pearson correlation, which measures linear relationships between two variables, assuming that a change in one variable is proportional to a change in another, based on their raw data. However, in my analysis, I opted to use the Spearman correlation coefficient, as it captures monotonic relationships, whether these are linear or not. Considering that this project studies the cryptocurrency market and that this market is extremely volatile and dynamic, the use of the Spearman correlation coefficient is more appropriate for the construction of the networks.

Chapter 3. Methodology

This chapter explains the data collection process and methodologies used for the construction and analysis of the correlation networks, utilizing the CRISP-DM framework to guide the methodology (Chapman, et al., 2000, Costa & Aparicio, 2020, 2021) This process consisted of computing the daily logarithm returns of the prices of cryptocurrencies, determining the correlations matrices based on these returns using a moving window approach, and applying a threshold value to filter weak correlations. Following that, correlation networks were created, and the Granger causality tests were applied to determine the direction of the contagion between the nodes. Furthermore, using the Louvain algorithm, the detection of communities in the same networks was performed, with the aim of finding relevant groups of cryptocurrencies that show related correlation patterns.

3.1. Data

This study analyzes daily closing prices of 58 cryptocurrencies form January 1, 2020, to May 11, 2024. It is important to note that as cryptocurrencies are a relatively new topic, the lack of long-term historical data restricts the capability to capture wider market trends and extended dependencies between cryptocurrencies.

Daily logarithm returns were calculated from closing prices, $r_i(t) = \log\big(P_i(t)\big) - \log\big(P_i(t-1)\big)$. This data was collected from https://coinmarketcap.com/, a price-tracking website for cryptoassets in the rapidly growing cryptocurrency space. In the following table (Table 2) it is possible to see all cryptoassets used for this work, as well as their main descriptive statistics.

Table 2: Descriptive statistics of the cryptocurrencies

crypto	ticker	mean	std	min	max
Bitcoin	BTC	7.60E-05	3.19E-02	-1.74E-01	1.36E-01
Cardano	ADA	-1.19E-03	4.67E-02	-3.01E-01	2.19E-01
AGIX	AGIX	8.56E-04	1.11E-01	-1.93E+00	1.84E+00
AIOZ	AIOZ	-4.21E-05	8.57E-02	-4.67E-01	7.48E-01
Akash Network	AKT	1.33E-04	6.29E-02	-2.74E-01	3.30E-01
Algorand	ALGO	-1.86E-03	5.33E-02	-3.70E-01	4.18E-01

crypto	ticker	mean	std	min	max
Arweave	AR	4.67E-04	7.11E-02	-3.03E-01	4.21E-01
Cosmos	ATOM	-9.82E-04	5.74E-02	-4.93E-01	2.74E-01
Avalanche	AVAX	-3.18E-05	6.01E-02	-4.54E-01	2.46E-01
Axie Infinity	AXS	-2.35E-05	6.68E-02	-4.99E-01	5.30E-01
Bitcoin Cash	BCH	-1.02E-03	4.96E-02	-4.35E-01	4.60E-01
Binance	BNB	-6.16E-05	3.98E-02	-4.04E-01	2.72E-01
Bitcoin BEP2	BTCB	7.78E-05	3.17E-02	-1.77E-01	1.35E-01
Chiliz	CHZ	-1.16E-03	5.94E-02	-4.57E-01	3.77E-01
Cronos	CRO	-2.06E-04	4.81E-02	-2.34E-01	2.79E-01
Dai	DAI	-3.55E-07	1.60E-03	-2.54E-02	1.90E-02
Dogecoin	DOGE	-1.04E-03	5.40E-02	-4.52E-01	3.71E-01
Polkadot	DOT	-1.54E-03	5.17E-02	-4.77E-01	2.52E-01
eGold	EGLD	-1.35E-03	5.19E-02	-3.96E-01	3.23E-01
Ethereum Classic	ETC	-1.25E-03	5.30E-02	-3.89E-01	3.46E-01
Ethereum	ETH	-2.80E-04	4.07E-02	-3.17E-01	2.26E-01
Fetch.ai	FET	1.36E-03	6.90E-02	-4.38E-01	3.33E-01
Filecoin	FIL	-2.90E-03	5.81E-02	-4.26E-01	3.52E-01
Flow	FLOW	-3.13E-03	5.74E-02	-3.63E-01	3.38E-01
Fantom	FTM	-7.66E-05	7.20E-02	-5.63E-01	3.04E-01
Gala	GALA	1.02E-03	8.50E-02	-5.20E-01	8.10E-01
The Graph USD	GRT6719	-1.44E-03	6.37E-02	-4.87E-01	4.67E-01
Hedera Hashgraph	HBAR	-8.28E-04	5.45E-02	-4.21E-01	5.49E-01
Internet Computer	ICP	-3.26E-03	6.16E-02	-3.64E-01	3.48E-01
Injective	INJ	2.95E-04	6.69E-02	-4.20E-01	4.01E-01
JasmyCoin	JASMY	-4.23E-03	1.07E-01	-8.50E-01	1.29E+00
KuCoin	KCS	-1.33E-04	4.82E-02	-4.96E-01	4.16E-01
Lido DAO	LDO	-8.93E-04	7.99E-02	-5.10E-01	3.93E-01
UNUS SED LEO	LEO	4.79E-04	3.30E-02	-2.00E-01	4.41E-01
Chainlink	LINK	-1.12E-03	5.32E-02	-4.66E-01	2.76E-01
Litecoin	LTC	-1.35E-03	4.62E-02	-4.41E-01	2.48E-01
Polygon	MATIC	-2.26E-04	6.02E-02	-3.91E-01	4.58E-01
Maker	MKR	-5.61E-04	5.05E-02	-2.83E-01	4.23E-01
NEAR	NEAR	3.72E-04	6.64E-02	-4.44E-01	3.61E-01
Neo	NEO	-1.76E-03	5.52E-02	-4.54E-01	3.46E-01
OKB	OKB	3.88E-04	4.70E-02	-4.01E-01	2.90E-01
Quant	QNT	7.64E-04	5.45E-02	-3.19E-01	3.55E-01
Render	RNDR	2.42E-03	7.84E-02	-4.30E-01	3.61E-01
THORChain	RUNE	-9.47E-04	7.14E-02	-5.53E-01	3.20E-01
Sandbox	SAND	-8.76E-05	6.75E-02	-4.59E-01	4.56E-01
Solana	SOL	1.11E-03	6.24E-02	-5.50E-01	2.82E-01
Lido Staked ETH	STETH	-2.75E-04	4.05E-02	-3.03E-01	2.13E-01
Stacks USD	STX4847	1.19E-04	6.31E-02	-4.09E-01	5.23E-01
Theta Network	THETA	-1.51E-03	5.74E-02	-4.94E-01	2.48E-01

crypto	ticker	mean	std	min	max
Tron	TRX	-4.74E-06	3.70E-02	-3.83E-01	1.94E-01
Uniswap USD	UNI7083	-1.49E-03	5.49E-02	-4.03E-01	4.34E-01
USD Coin	USDC	-5.56E-08	1.14E-03	-2.84E-02	2.10E-02
Tether	USDT	-4.76E-07	4.23E-04	-3.93E-03	4.62E-03
VeChain	VET	-1.62E-03	5.20E-02	-4.09E-01	2.53E-01
Wrapped Bitcoin	WBTC	7.29E-05	3.19E-02	-1.75E-01	1.35E-01
Stellar	XLM	-1.67E-03	4.43E-02	-3.62E-01	4.76E-01
Monero	XMR	-1.10E-03	4.62E-02	-5.34E-01	3.45E-01
Ripple	XRP	-9.23E-04	4.76E-02	-3.96E-01	5.49E-01

3.2. Network Construction

The present section aims to describe in detail all the processes that were made to reach the final version of the dynamic correlation networks and their corresponding communities.

3.2.1. Threshold Value

As was already mentioned before, one of the key points in the construction of networks is the application of a threshold value to ensure that weak connections are put aside and do not interfere with the rest of the work.

For that, graphs were built where the various values of some network metrics, such as betweenness centrality, clustering coefficient, and density, for different threshold values were seen (Figures 2 to 6).

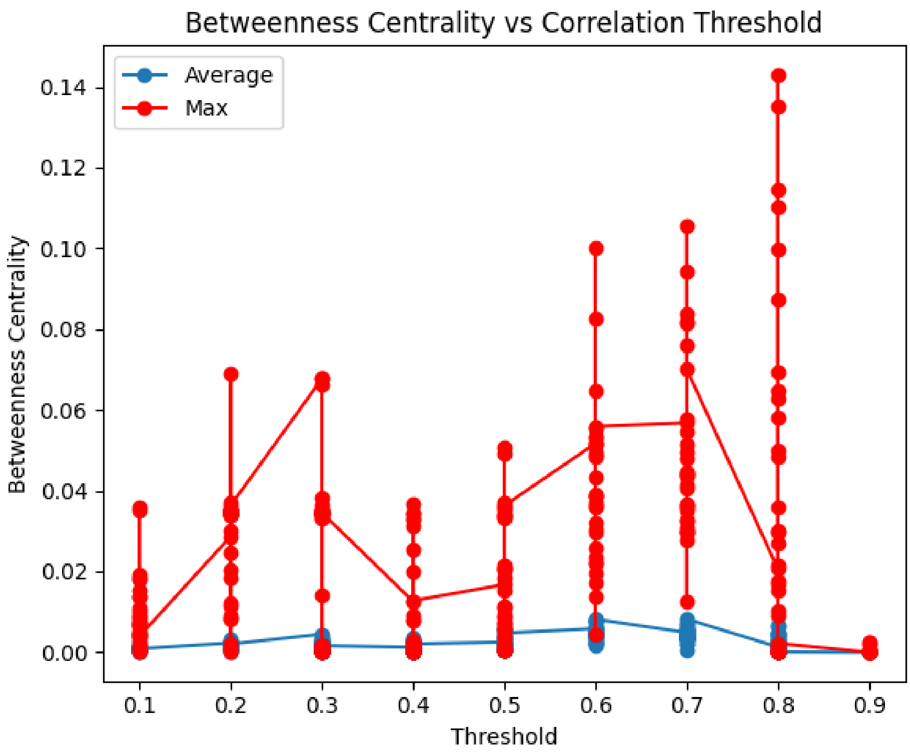

Figure 2: Betweenness Centrality for different Threshold values

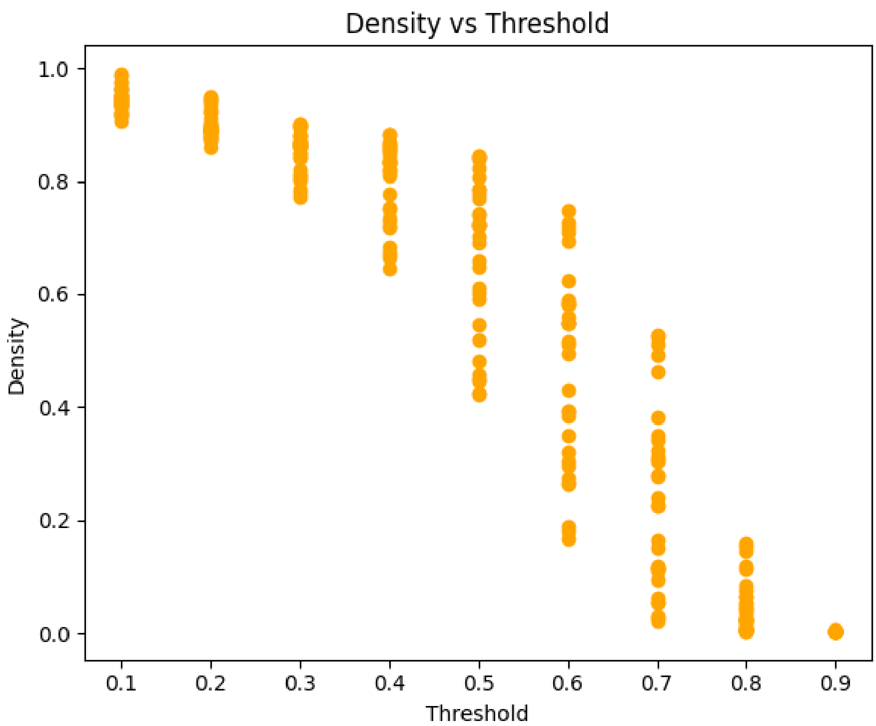

Figure 3: Density Values for different Threshold Values

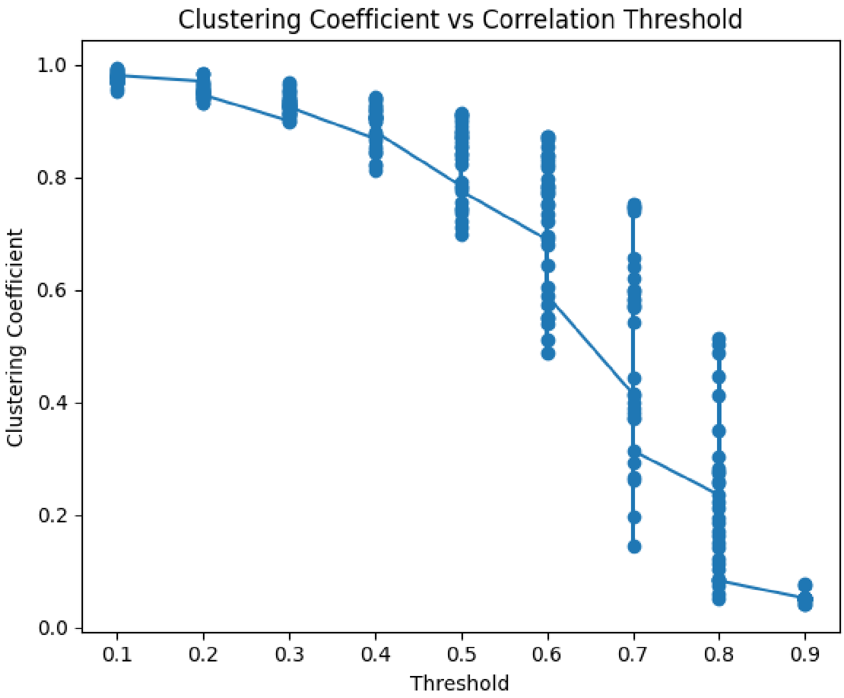

Figure 4: Clustering Coefficient for different values of threshold

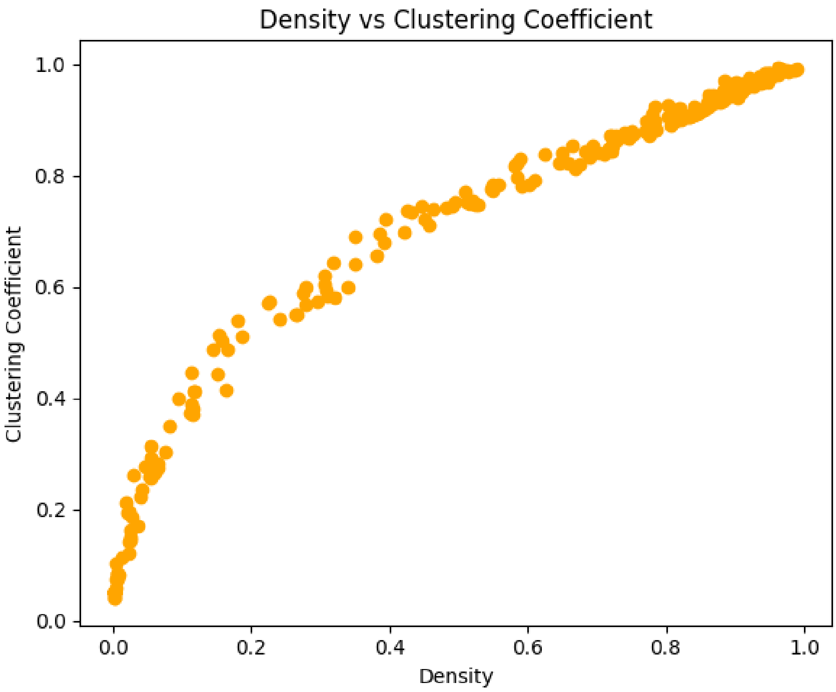

Figure 5: Clustering coefficient for different density values

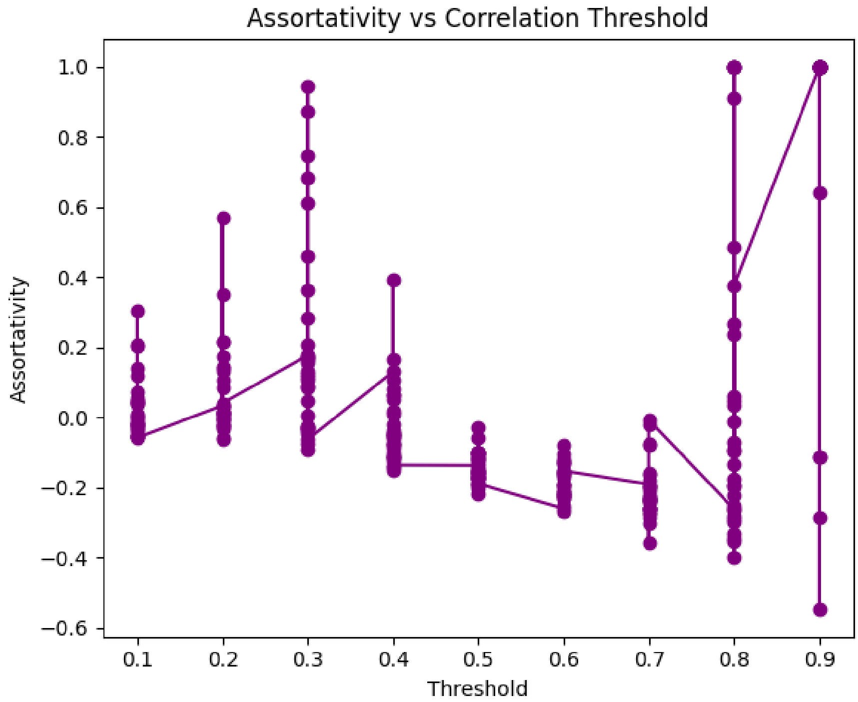

Figure 6: Assortativity for different threshold values

From the analysis of these figures, it is possible to conclude that betweenness centrality shows significant fluctuations at lower threshold values, making it easier to identify the role that certain cryptocurrencies play in connecting different communities.

It is also possible to analyze that regarding the density and the clustering coefficient, both values decrease as the threshold increases, translating into a less connected network and a reduction of dense subgraphs as weak correlations are filtered out.

In Figure 5, it is clear that as the Network becomes denser, that is, has more connections, and the clustering coefficient also increases, reflecting a higher local interconnectedness.

Regarding assortativity, this value floats around zero, indicating no clear tendency for nodes of similar degrees to connect, which goes along with the speculative and volatile nature of the cryptocurrency market.

From these figures, it was concluded that the threshold value to use in the filtration of correlations was 0.5, as it is the value that corresponds to a network that preserves the most important connections, allowing key nodes to be identified as central connectors, while also filtering out weak and irrelevant correlations that may make the Network's structure difficult to understand. This is appropriate since the project deals with cryptocurrencies that have higher values of correlations between them, and so, it is possible to have a clearer network structure that is focused only in strong connections, also simplifying the detection of communities.

3.2.2. Correlation matrix

The first step, after gathering and transforming all the data necessary to model the initial graphs, is to compute all cross-correlations using a sliding window approach. For this, it is defined a window size of 6 months (approximately 183 days), so that it is long enough to capture meaningful correlation patterns, and a step size of 1 month (approximately 30 days). Subsequently, a correlation matrix is computed, where each element represents the Spearman correlation coefficient (formula below) between two assets for the current time window's data. In this same matrix, it is applied a value of threshold, which has been explained in the previous sub-section, with the purpose of eliminating weak relationships, highlighting only relevant and impactful connections. This procedure translates into the following formula:

$$C_\delta = \begin{cases} C_{ij}, & if\ |C_{ij}| \geq \delta \\ 0, & if\ |C_{ij}| < \delta \end{cases}$$

Where C_δ corresponds to the thresholded value of the correlation between variable I and variable j, and δ represents the threshold value.

This whole process results in a total number of 31 matrices of thresholded correlations. As the Spearman correlation coefficient is defined as the Pearson correlation coefficient between the rank variables, we have that the formula is given by:

$$r_s = \rho[R[X], R[Y]] = \frac{cov[R[X], R[Y]]}{\sigma_{R[X]}\sigma_{R[Y]}}$$

Where ρ is the Pearson correlation coefficient applied to the rank variables, $cov[R[X], R[Y]]$ is the covariance of the rank variables and $\sigma_{R[X]}$ and $\sigma_{R[Y]}$ are the standard deviations of the rank variables.

3.2.3. Granger Causality

The next step is to implement the Granger Causality Tests so that it is possible to determine if one variable helps forecast the future values of another variable. It is important to acknowledge that Granger Causality Tests have certain limitations, such as the assumption of linearity and the stationary requirement. In the case of cryptocurrencies, these can exhibit non-linear patterns due to market sentiment or speculative behavior. However, despite these limitations, this method is a helpful resource for understanding predictive relations in time series data, particularly in the volatile cryptocurrency market.

In this case, a graph for each matrix, where the nodes represent all cryptocurrencies, was created from the correlation matrices, which were described in the previous subsection. Afterward, it is iterated over pairs of nodes with non-zero correlations in order to test the causality between each pair. For each test, the p-values that indicate whether the lagged values of one variable (cryptocurrency A) are statistically significant on the other (cryptocurrency B) are examined. If these p-values of the causality tests are below the significance level of 0.05, it adds an edge to the Network from cryptocurrency A to cryptocurrency B. At the end of this procedure, the result is 31 different directed networks that are already set to be analyzed in the matter of communities.

3.3. Community Detection

Lastly, the final step of this project is to find the different communities for each Network. Community detection plays an important role in understanding the natural divisions that exist in the Network, as well as showing how nodes are organized. The fact that the networks are divided into sub-groups also helps to analyze and interpret the information that each Network is giving, allowing to identify clusters that share similar characteristics or present stronger connections than with the rest of the Network.

For this, it is implemented the Louvain algorithm, a method used in large networks with the aim of optimizing the modularity of the Network. This modularity refers to a measure that quantifies the strength of the division of a network into clusters, analyzing the density of edges within a community compared to the density of edges between different communities. The modularity formula is defined as the following:

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

where A_{ij} represents the adjacency matrix (1 if there is an edge between nodes i and j, and 0 otherwise), k_i is the number of edges attached to node i, m is the total number of edges in the graph, N is the total number of nodes in the graph, c_i is the community to which the node i belongs, and δ is the Kronecker delta function:

$$\delta(c_i, c_j) = \begin{cases} 1 \ if \ c_i \ and \ c_j \ are \ the \ same \ cluster \\ 0 \ otherwise \end{cases}$$

The Louvain algorithm is split into two phases that are repeated iteratively, the first one involving assigning each node to its own community, then iteratively merging nodes into communities until the local modularity can no longer be increased. The second phase is where the detected communities are aggregated into a single node, and the steps in phase one are repeated to further improve modularity at a higher level.

Chapter 4. Results and discussion

This chapter will present the networks and their respective communities, as well as the main conclusions from its analysis. It will also be presented some comparisons relative to the studies presented in the Literature Review.

4.1. Results

In this sub-section, the results will be divided into two parts, the first one referring to the construction of the networks with the Granger Causality tests, and the second is where it is presented the communities that were found, derived from the same networks.

4.1.1. Networks

As it was already said before, one of the first steps after loading the data is to compute the cross-correlations for all the cryptocurrencies and then apply the threshold value, resulting in 31 correlation matrixes along the time frame of the variables. In the following figure, we can see the matrix correlation for the first time window, where all the cryptocurrencies are on and the x and y axes. It is possible to observe that meaningless correlations in the interval]-0.5;0.5[were despised, as we have only values in the interval [-1;-0.5] ∪ [0.5;1] or 0.

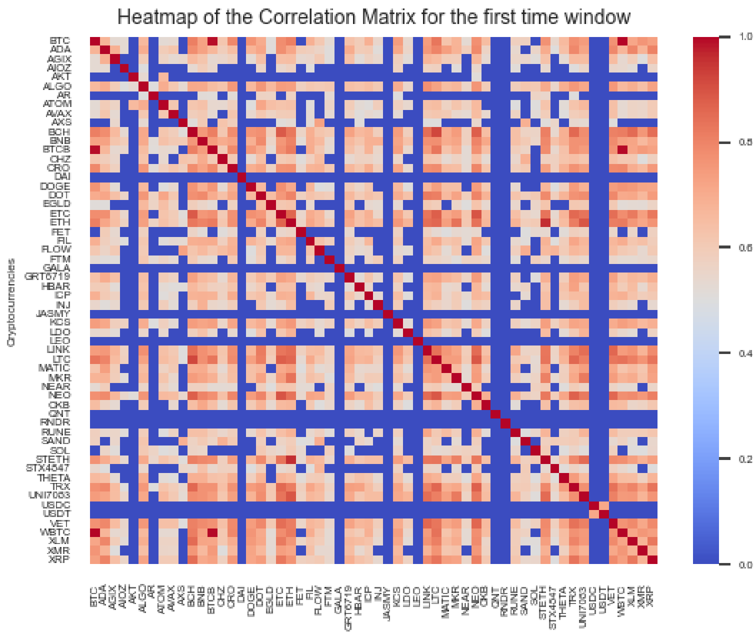

Figure 7: Correlation matrix for the first time window

Applying the Granger Causality Tests to the pairs of variables that show significant correlations is the next step, resulting in 31 directed networks that will be explored right away.

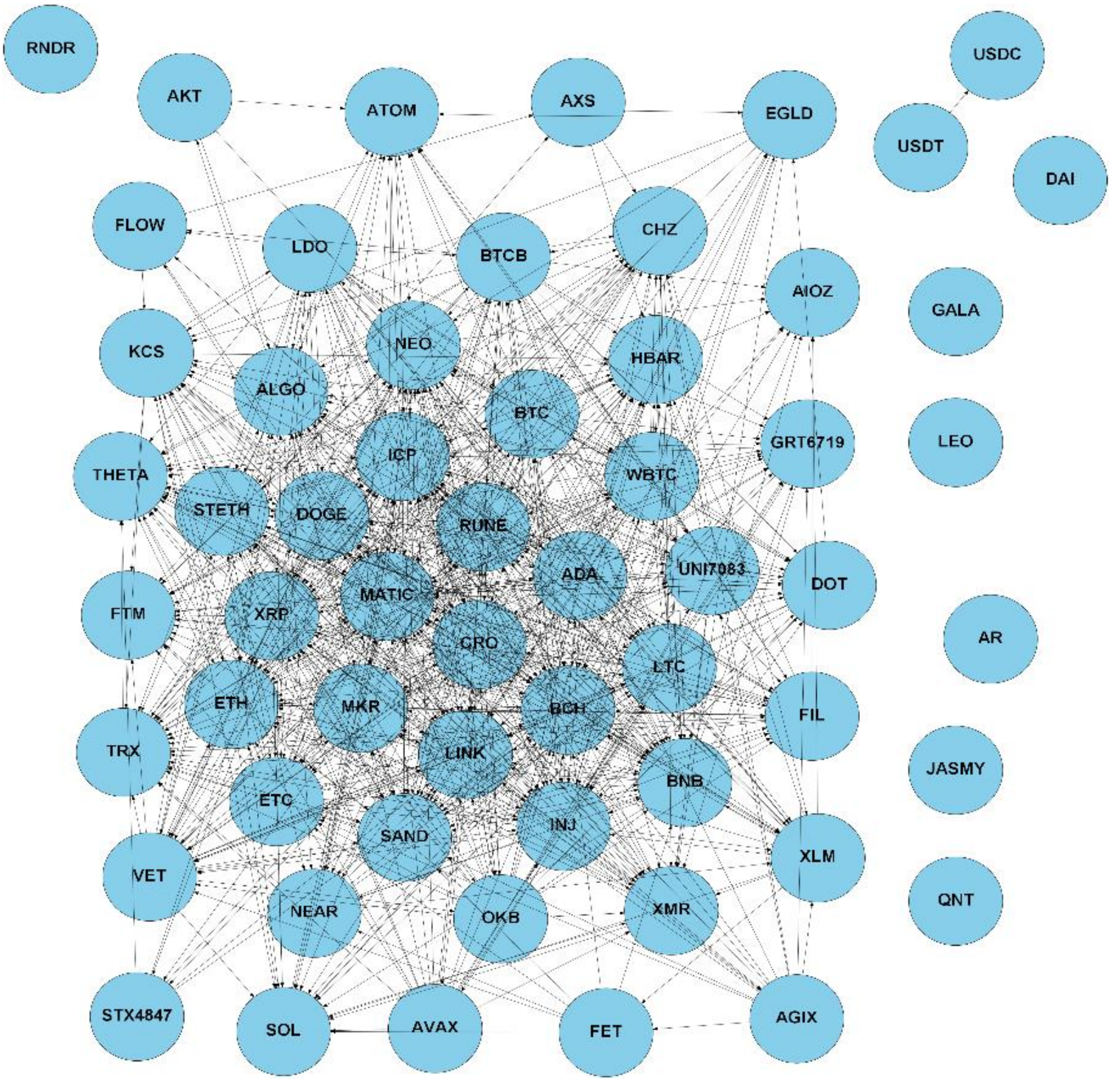

Figure 8: Granger Causality Network for the first moving window

In this figure, it is possible to see the network for the first time-window, characterized by a total of 789 edges and density of approximately 0.2387, meaning that about 23.87% of all possible connections between nodes are presented.

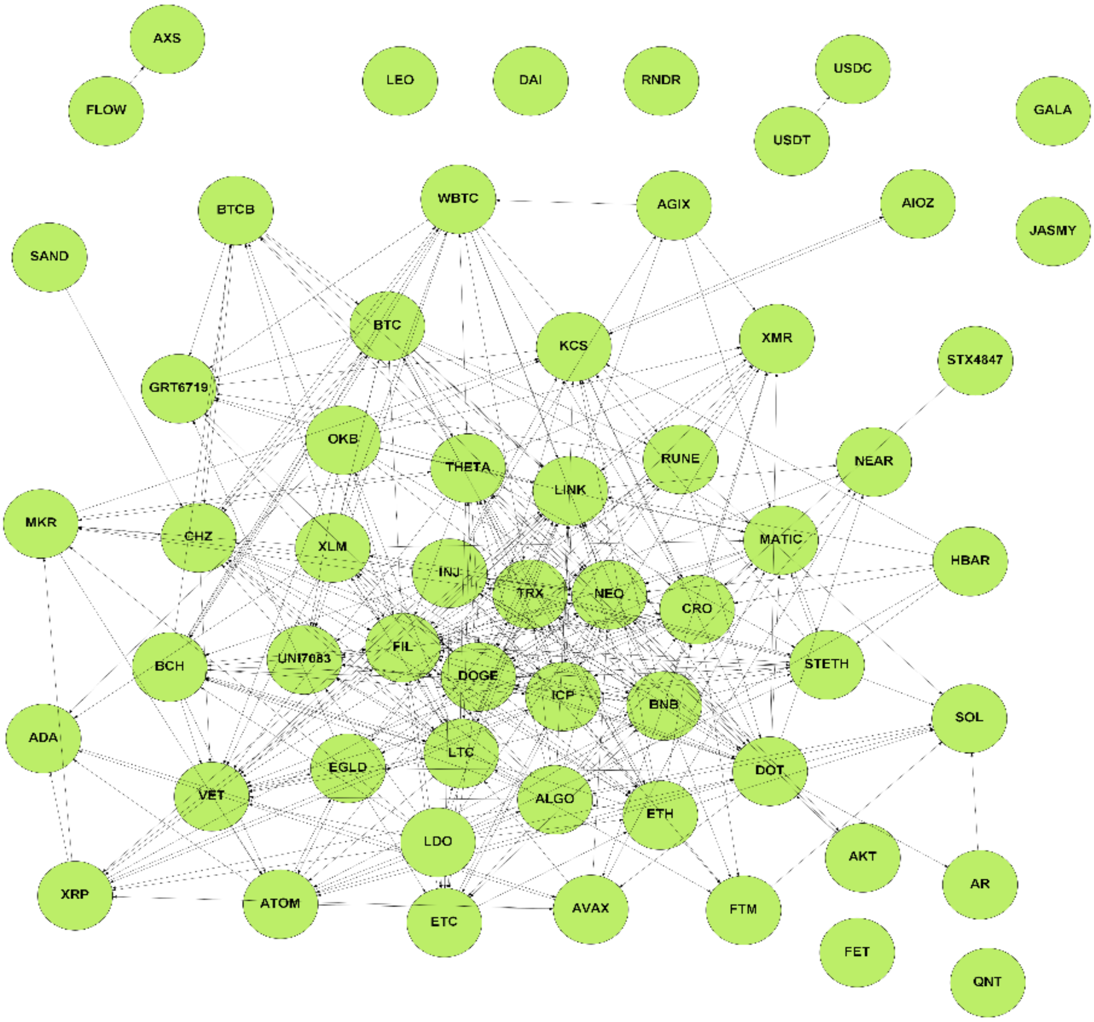

Figure 9: Granger Causality Network for the second moving window

In Figure 9, the number of edges decreases significantly from the first to the second moving-window, containing only a total of 342 edges. The density of this network also decreases to 0.1034, indicating a reduction in the connectivity.

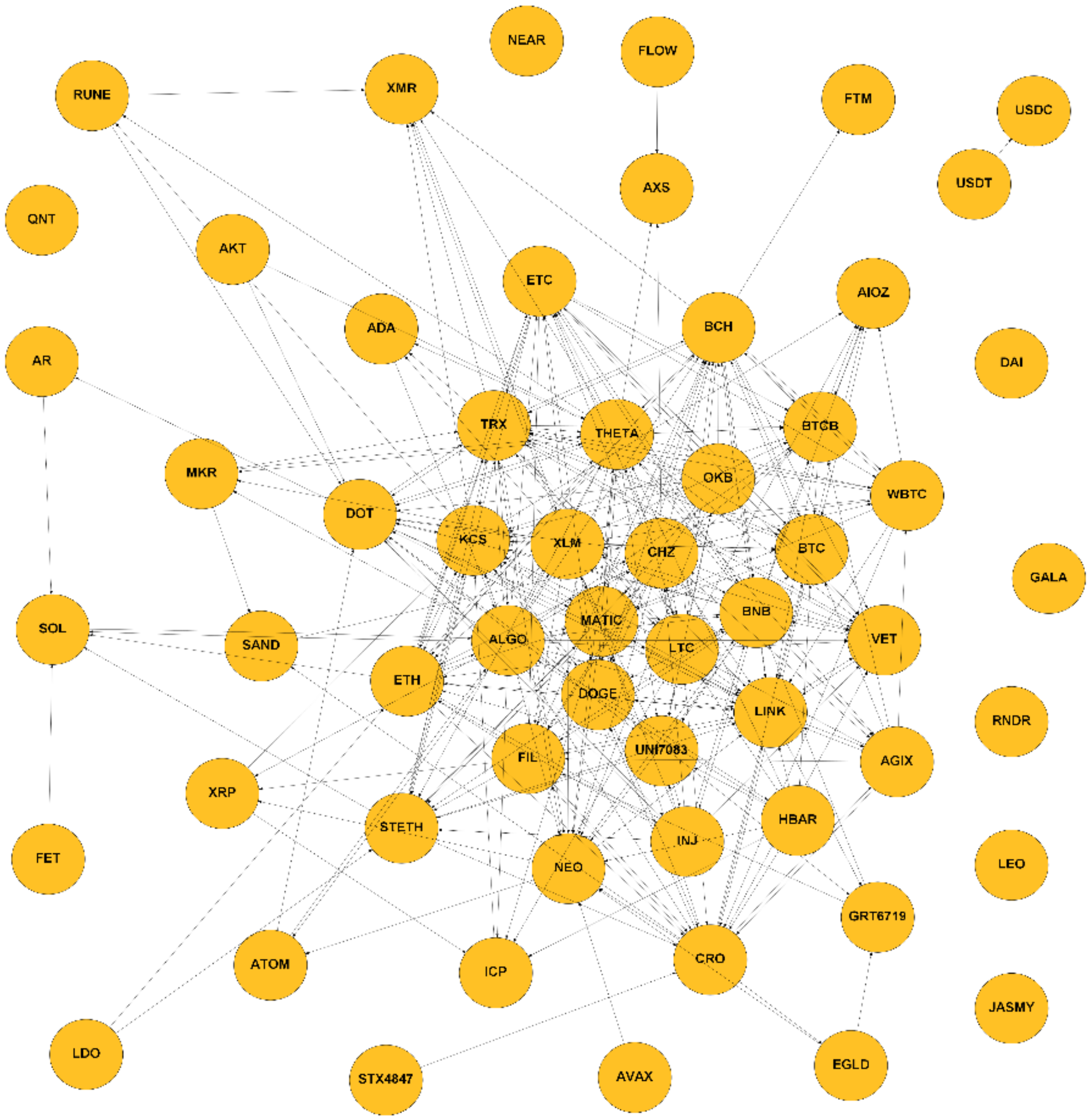

Figure 10: Granger Causality Network for the third moving window

Regarding the third moving window network (Figure 10), this graph reflects a continued reduction in the number of edges and in the density, as its values are, respectively, 311 and 0.0941, translating into an even less connected network.

Furthermore, the results for the in/out-degree and betweenness centrality measures and the density for each one of these networks were printed (Table 3).

Table 3: Metrics for the first moving window network

Ticker	In-Degree	Out-Degree	BC
BTC	0.2105	0.4912	0.0051
ADA	0.1579	0.4035	0.0031
AGIX	0.0351	0.2281	0.0034

Ticker	In-Degree	Out-Degree	BC
AIOZ	0.1754	0.0000	0.0000
AKT	0.0175	0.0526	0.0000
ALGO	0.3860	0.3860	0.0283
AR	0.0000	0.0000	0.0000
ATOM	0.2456	0.0877	0.0027
AVAX	0.0351	0.2982	0.0002
AXS	0.0351	0.0351	0.0004
BCH	0.5789	0.3158	0.0265
BNB	0.4561	0.4035	0.0151
BTCB	0.1579	0.4737	0.0056
CHZ	0.2982	0.2281	0.0157
CRO	0.2982	0.7018	0.0292
DAI	0.0000	0.0000	0.0000
DOGE	0.2807	0.5263	0.0277
DOT	0.2281	0.1754	0.0017
EGLD	0.0877	0.1579	0.0014
ETC	0.2982	0.3684	0.0071
ETH	0.3684	0.3509	0.0085
FET	0.0351	0.1228	0.0005
FIL	0.2632	0.2281	0.0025
FLOW	0.0702	0.0702	0.0019
FTM	0.3509	0.0175	0.0001
GALA	0.0000	0.0000	0.0000
GRT6719	0.2632	0.0877	0.0022
HBAR	0.4035	0.0877	0.0042
ICP	0.5088	0.5965	0.0682
INJ	0.4912	0.4211	0.0404
JASMY	0.0000	0.0000	0.0000
KCS	0.2982	0.2456	0.0063
LDO	0.1228	0.4912	0.0052
LEO	0.0000	0.0000	0.0000
LINK	0.5439	0.2632	0.0261
LTC	0.3684	0.4386	0.0104
MATIC	0.6491	0.5088	0.0529
MKR	0.2807	0.6140	0.0169
NEAR	0.1754	0.1404	0.0015
NEO	0.5263	0.3860	0.0281
OKB	0.0877	0.4211	0.0020
QNT	0.0000	0.0000	0.0000
RNDR	0.0000	0.0000	0.0000
RUNE	0.4386	0.0702	0.0032
SAND	0.4561	0.2632	0.0114
SOL	0.3158	0.0526	0.0011

Ticker	In-Degree	Out-Degree	BC
STETH	0.3333	0.3860	0.0218
STX4847	0.0526	0.0877	0.0000
THETA	0.3860	0.1228	0.0031
TRX	0.3860	0.1754	0.0024
UNI7083	0.2105	0.2632	0.0018
USDC	0.0175	0.0000	0.0000
USDT	0.0000	0.0175	0.0000
VET	0.2982	0.3158	0.0060
WBTC	0.1930	0.4912	0.0046
XLM	0.1930	0.2281	0.0136
XMR	0.3333	0.1579	0.0041
XRP	0.4386	0.3860	0.0186

From this analysis, it is shown that certain cryptocurrencies, like 'MATIC' and 'LINK', hold significant in-degree centrality, suggesting that they receive a high volume of influence from other assets, making them pivotal within the Network. This finding implies that these assets may serve as market leaders or key indicators of future trends. In contrast, cryptocurrencies with high out-degree centrality, such as 'CRO' and 'MKR', influence many other assets and can serve as tools for diversification in portfolio construction. These insights suggest practitioners should closely monitor assets with strong network centrality for potential leadership in market movements while isolating peripheral assets for safer, less volatile investment strategies.

On the contrary, looking only for the tickers that have value 0 for the three metrics, it is possible to conclude that 'AR', 'DAI', 'GALA', 'JASMY', 'LEO', 'QNT' and 'RNDR' are assets that do not communicate with any other variables, also visible in the network figure (Figure 10) as these are the nodes that are isolated from the Network.

In order to have an idea of the number of connections in each Network along the time frame, a graph was printed with the total number of edges (y-axis) for each of the time-window Network (x-axis) (Figure 11).

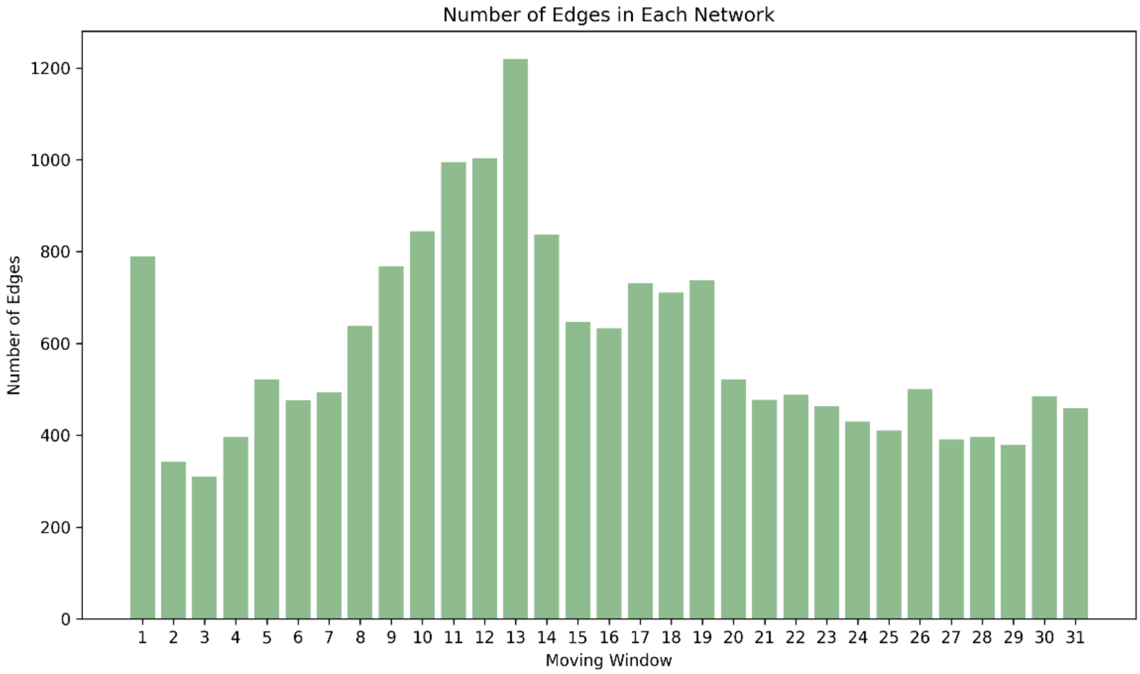

Figure 11: Number of edges for each moving-window Network

Analyzing the graph, it is evident that the moving windows 11, 12, and 13 have the higher values for the total of edges in the Network, corresponding to the dates between 25th February 2022 and 26th August 2022, 26th March 2022 and 24th September 2022, and 24th April 2022 and 23rd October 2022, respectively. These spikes in network density between February and October 2022 indicate a period of intensified correlation among cryptocurrencies, potentially driven by external events such as regulatory announcements or macroeconomic changes (e.g., inflation concerns and interest rate hikes). An example that could be related to these spikes is the geopolitical tension between Russia and Ukraine, which escalated in February 2022 and caused significant fluctuations in global markets (Izzeldin, Muradoğlu, Pappas, Petropoulou, & Sivaprasad, 2023). In periods of uncertainty, investors can either turn to cryptocurrencies as a safe haven or sell off their cryptocurrency assets to cover losses in traditional markets.

Practitioners can use these network insights to understand how external factors influence correlation spikes within cryptocurrency markets, offering an opportunity to hedge against broader market risks by diversifying across less correlated assets during market turbulence.

4.1.2. Community Detection

The last step to reach the main target of this project is to compute the different communities for each of the previously mentioned networks, applying the Louvain algorithm for community detection. The results are the following:

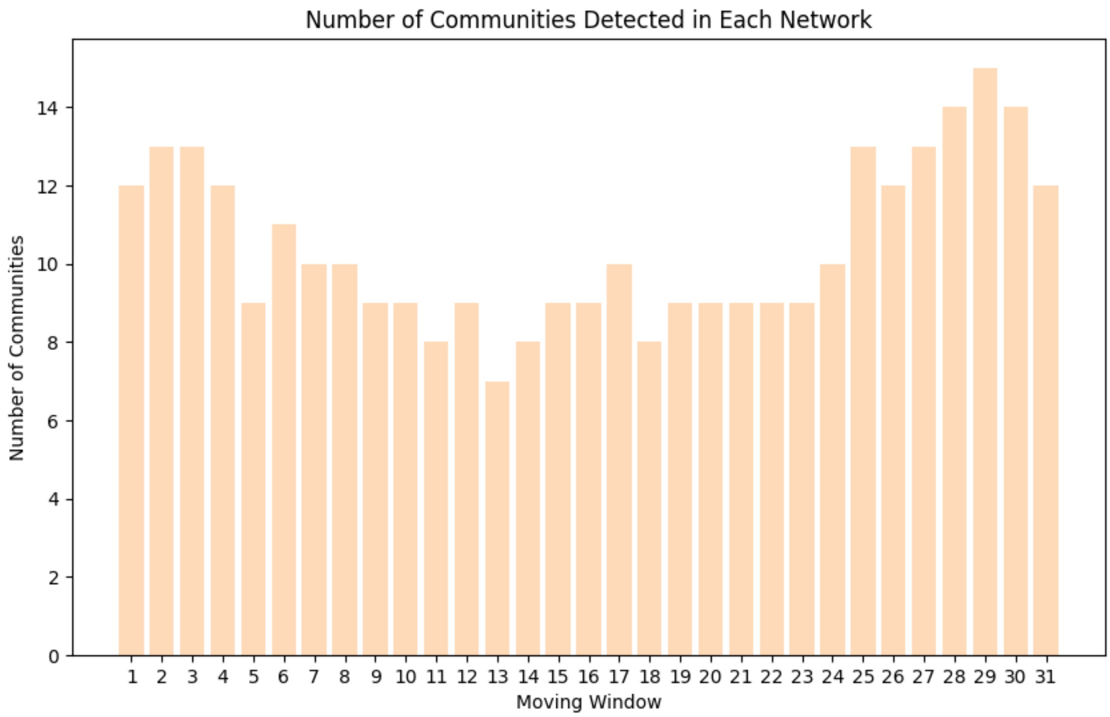

Figure 12: Number of communities for each moving-window Network

As demonstrated by the figure, in this case, the moving-window Network with a higher number of communities is the 29th, corresponding to the time span from 14th September 2023 to 14th March 2024 and having a total number of 15 communities.

Next, it will present some networks with distinct communities illustrated in different colors to simplify their interpretation (Figures 13, 14, and 15).

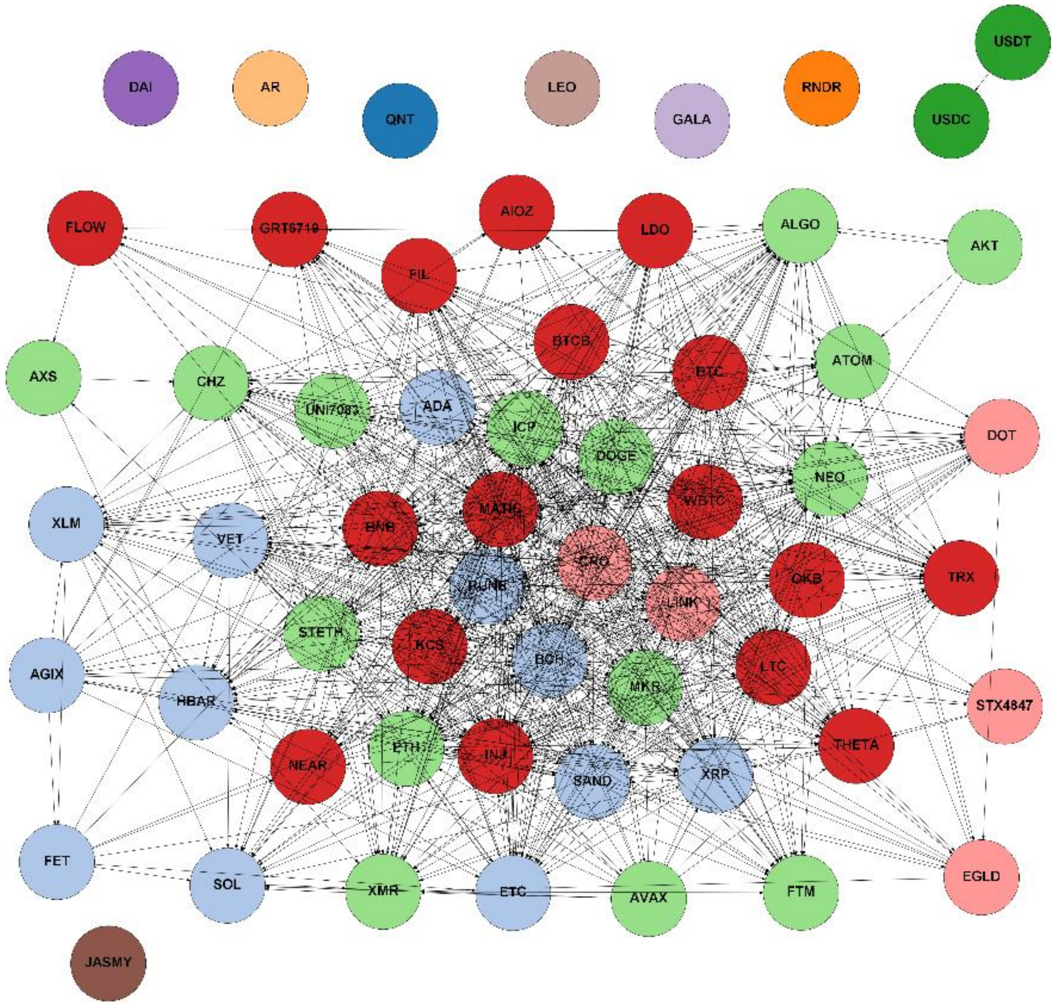

Figure 13: Communities for the first moving window network

The figure above presents the network of cryptocurrencies for the first moving window, with the different cryptocurrencies represented by the nodes and the relationships between them being the edges that link the different nodes. It is also possible to see the different communities in this network, each one of them being of a different color.

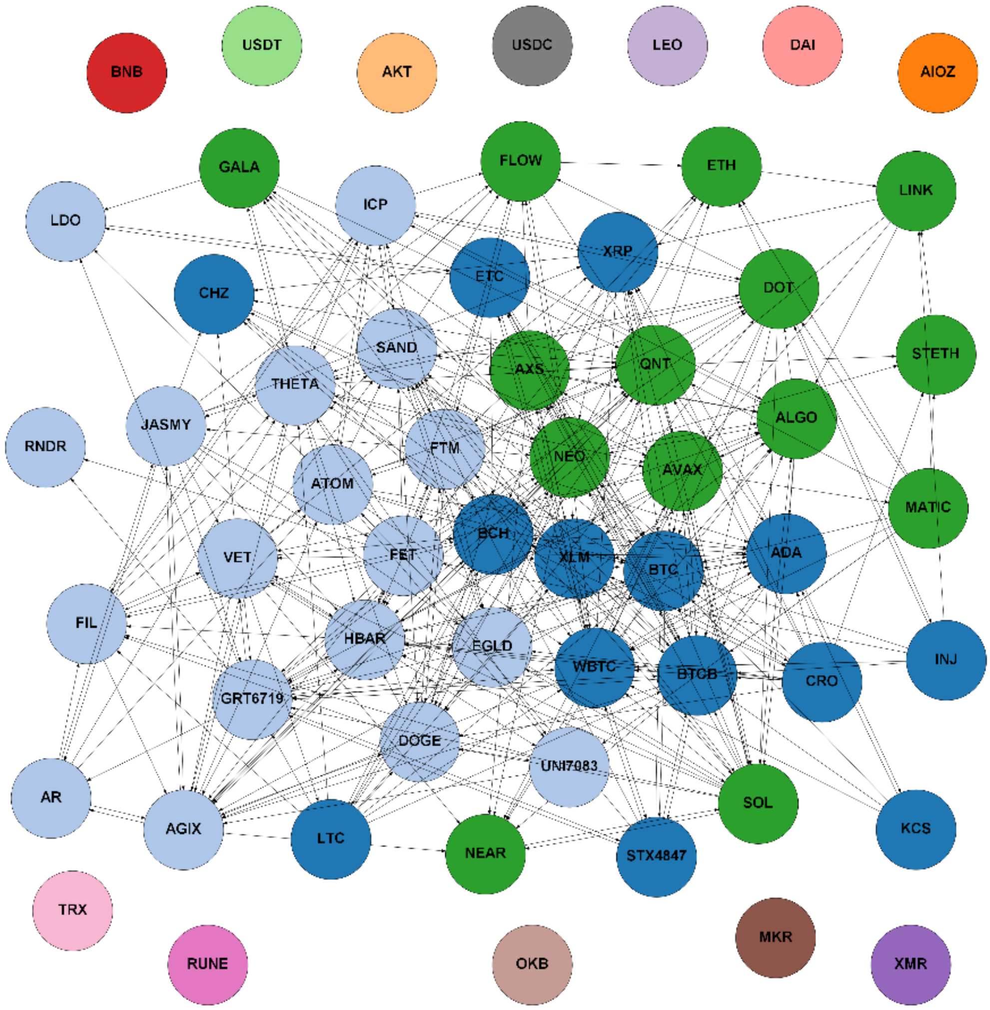

Figure 14: Communities for the 29th moving window network

In Figure 14, it is represented the same structure as Figure 13, and it is possible to conclude that cryptocurrencies like 'TRX', 'RUNE', 'OKB', 'MKR', 'XMR', 'BNB', 'USDT', 'AKT', 'USDC', 'LEO', 'DAI', and 'AIOZ' are isolated nodes, as they do not have any edges linked to them and form a single-node community.

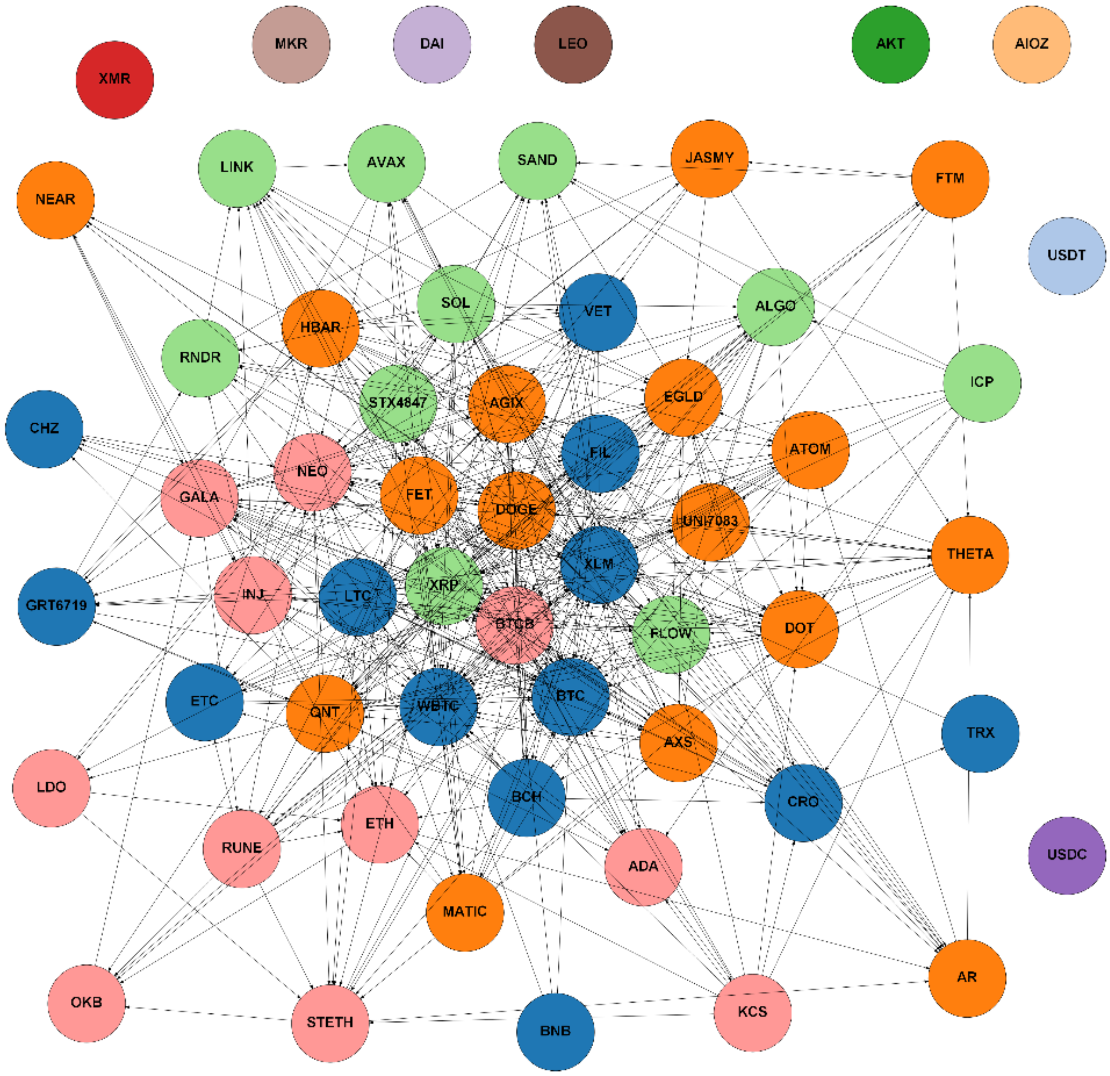

Figure 15: Communities for the last moving window network

The interpretation of Figure 15 is consistent with the previous two figures, where the community structures and relationships among cryptoassets are illustrated equally.

Analysis of the network visualizations reveals three major communities consistently appearing across all time periods, namely, the red, blue, and green ones in Network of time-window 1 (Figure 13), the green, light blue, and dark blue in the Network of time-window 29 (Figure 14), being this the one that has the highest number of communities. In the last moving-window network, four notable communities are present, in the colors pink, light green, orange, and dark blue (Figure 15). Regarding the first Network (Figure 13) and comparing with the results from Table 3, it can be noted that the nodes with higher betweenness centrality, such as 'ICP', 'MATIC' and 'INJ', are the ones in the larger communities mentioned above, implying that these variables act as connectors for nodes that might be distant from each other in terms of shortest paths.

With respect to the isolated nodes, it can be said that the ones that do not have any connection coming in or from it are the ones that correspondingly have zero or very low values in the network metrics that were already presented. This could translate into cryptocurrencies that are more independent from the rest of the dataset but also lack influence on the dynamics of the networks. Examples of these nodes are 'LEO' and 'DAI', which are disconnected nodes in all three networks introduced, but also 'USDC' and 'USDT' that in the first moving window, form a community with only each other and in the last networks are separated and act alone in both graphs.

4.2. Discussion

This project follows the primary procedures as most of the works referenced, in the sense that it also employs network theory to analyze financial markets. However, while the generality of the other studies applies different methodologies, such as Transfer Entropy, Pearson Correlation, and Granger causality tests, to define connections between financial assets, mostly stocks, and market indexes, the present work diverges slightly as it focuses on the investigation of the relationships in the cryptocurrency market. The fact that the dataset used in this analysis differs in various characteristics from datasets used in other works implies that the results have a different interpretation, although using similar methods.

In the research made by Wu, Tuo, & Xiong (2015), it is concluded that stocks in the same industry often tend to group together, forming the basis of communities, whereas in this case, community detection may reflect some market sentiment patterns, giving potential insights into how interconnected different segments of the cryptocurrency market are.

It is also important to mention that, in addition to similar approaches, some of the same network metrics were used in these studies, namely Degree Centrality and Betweenness Centrality. In agreement with the different interpretations of the outcomes, in this case, a high degree of centrality for a cryptocurrency variable could possibly mean that this asset is a leader in the involved market. In contrast, in the cases of traditional assets, this significant value could translate into key sectors.

Chapter 5. Conclusion

This study advances our understanding of cryptocurrency market dynamics through network analysis, yielding several key findings and implications while acknowledging certain limitations.

Summarizing what has been made, the correlation between 58 cryptocurrencies was investigated using the Granger Causality tests and Community Finding methods. This analysis was performed using a moving window approach, allowing to grasp the dynamics of the cryptocurrency market over the years, more specifically, over 3 years.

The key findings of this study reveal that these networks show meaningful variations in their structure and in the connections among the various nodes that represent the digital assets. Analysis of network evolution reveals a significant increase in cryproasset interdependencies between February and October 2022. Furthermore, along with the Louvain algorithm to detect different communities, the fundamental result is that there is a larger number of communities in the later time windows, particularly in the third to last, which exhibits a total of 15 distinct communities.

One of the most practical applications of this study is in portfolio optimization. High centrality assets like 'MATIC' may be included as risk-heavy components in an aggressive portfolio, while isolated assets like 'DAI' and 'LEO' can provide a hedge against market-wide movements, serving as risk-averse assets. By utilizing community detection insights, practitioners can segment their portfolios into core holdings, risk-on assets, and isolated hedges, thus creating a multi-tiered investment strategy that adapts to both short-term and long-term market fluctuations.

The main findings of this project provide important insights to investors who deal with this type of assets, as they can use them to build diversified portfolios that balance risk by grouping high-risk assets with lower-risk hedges and strategically managing exposure to potential contagion within the cryptocurrency market.

Future research directions include:

- Investigating the impact of external market factors;
- Extending the analysis to include more or other cryptocurrencies;
- Developing predictive models based on network metrics; and

- Exploring alternative community detection algorithms.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2005*). Practical volatility and correlation modeling for financial market risk management*. Cambridge, Mass., USA: National Bureau of Economic Research.

Aparicio, J, Romao. M & Costa, C (2022) Predicting Bitcoin prices : The effect of interest rate, search on the internet, and energy prices, *2022 17th Iberian Conference on Information Systems and Technologies (CISTI*), Madrid, Spain, 2022, pp. 1-5, d https://doi.org/10.23919/CISTI54924.2022.9820085.

Basu, S., Shojaie, A., & Michailidis, G. (14 de July de 2021). Network Granger Causality with Inherent Grouping Structure. *The Journal of Machine Learning Research*, *16*(1), 417-453.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.

Chunaev, P. (21 de July de 2020). Community Detection in node-attributed social networks: A survey. *Computer Science Review* 37, 100286

Costa, C. J. (2024). DeFi: Concepts and Ecosystem. arXiv preprint arXiv:2412.01357. https://doi.org/10.48550/arXiv.2412.01357

Costa, C. J., & Aparicio, J. T. (2020). POST-DS: A methodology to boost data science. *2020 15th Iberian Conference on Information Systems and* Technologies (CISTI) (pp. 1-6). IEEE.

Costa, C.J., Aparicio, J.T. (2021). A Methodology to Boost Data Science in the Context of COVID-19. In: Arabnia, H.R., et al. Advances in Parallel & Distributed Processing, and Applications. *Transactions on Computational Science and Computational Intelligence.* Springer, Cham.

Giudici, P., & Polinesi, G. (29 de May de 2019). Crypto price discovery through correlation networks. *Annals of Operations Research*, Springer, vol. 299(1), pages 443-457, April.

Huang, X., Chen, D., Ren, T., & Wang, D. (2021). A survey of community detection methods in multilayer networks. *Data Mining and Knowledge Discovery*, 1-45.

Ioannidis, E., Sarikeisoglou, I., & Angelidis, G. (16 de November de 2023). Portfolio Construction: A Network Approach. *Mathematics* 2023, 11(22), 4670; https://doi.org/10.3390/math11224670

Isogai, T. (23 de May de 2017). Dynamic correlation network analysis of financial asset returns with network clustering . *Applied Network Scien*ce 2, 8 (2017). https://doi.org/10.1007/s41109-017-0031-6

Izzeldin, M., Muradoğlu, Y. G., Pappas, V., Petropoulou, A., & Sivaprasad, S. (2023). The impact of the Russian-Ukrainian war on global financial markets. *International Review of Financial Analysis*, 102598.

Kenett, D. Y., Huang, X., Vodenska, I., Havlin, S., & Stanley, H. E. (2015). Partial correlation analysis: Applications for financial markets. *Quantitative Finance*, 569-578.

Li, Y., Jiang, X.-F., Tian, Y., Li, S.-P., & Zheng, B. (9 de October de 2018). Portfolio Optimization based on Network Topology. *Physica A: Statistical Mechanics and its Applications* V. 515, 1 February 2019, 671-681

Ma, M., Bao, T., & Wen, Y. (2023). Enhancing Portfolio Performance with Crypto Tokens: a Correlation Network Analysis. 2023 IEEE 43rd International Conference on Distributed Computing Systems Workshops (ICDCSW), Hong Kong, Hong Kong, 2023, pp. 31-36, https://doi.org/10.1109/ICDCSW60045.2023.00013.

Ricca, F., & Scozzari, A. (11 de July de 2023). Portfolio optimzation through a network approach: Network assortative mixing and portfolio diversification. *European Journal of Operational Research*, 312, 2, 16 January 2024, 700-717

Výrost, T., Lyócsa, Š., & Baumöhl, E. (11 de February de 2015). Granger Causality stock market networks: Temporal proximity and preferential attachment. *Physica A: Statistical Mechanics and its Applications* 427, 1 June 2015, 262-276

Wang, Y., Briola, A., & Aste, T. (November de 2023). Topological Portfolio Selection and Optimization. In *Proceedings of the Fourth ACM International Conference on AI in Finance (ICAIF '23*). Association for Computing Machinery, New York, NY, USA, 681–688. https://doi.org/10.1145/3604237.3626875

Wu, S., Tuo, M., & Xiong, D. (Janurary de 2015). Network Structure Detection and Analysis of Shanghai Stock Market *Journal of Industrial Engineering and Management (JIEM)*, *8*(2), 383-398

Xu, G., & Gao, W. (2019). Financial risk contagion in stock markets: causality and measurement aspects. *Sustainability*, 1402.

Yan, X., Yang, H., Yu, Z., & Zhang, S. (12 de August de 2021). A Network View of Portfolio Optimization Using Fundamental Information. *Frontiers in Physics*, *9*, 721007.

Zheng, Q., & Song, L. (7 de August de 2018). Dynamic Contagion of Systemic Risks on Global Main Equity Markets Based on Granger Causality Networks *Discrete Dynamics in Nature and Society*, *2018*(1), 9461870

Appendix

Code

```
# %%
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import spearmanr
from statsmodels.tsa.stattools import grangercausalitytests
import warnings
import sys
import io
import contextlib
import networkx as nx
import matplotlib.pyplot as plt
import community as community_louvain
from networkx.drawing.nx_agraph import graphviz_layout, to_agraph
import networkx.drawing.nx_agraph as nx_agraph
import os
import pygraphviz as pgv
import matplotlib.colors as mcolors
from collections import defaultdict


# %%
class SuppressOutput:
    def __enter__(self):
        self._original_stdout = sys.stdout
        sys.stdout = io.StringIO()

    def __exit__(self, exc_type, exc_val, exc_tb):
        sys.stdout = self._original_stdout


# %%
df = pd.read_excel("Crypto_Close_Prices.xlsx", sheet_name="Sheet2", usecols="B:BG")
returns = pd.DataFrame(df)

returns.replace([np.inf, -np.inf], np.nan, inplace=True)
returns.dropna(inplace=True)


# window_size represents a y-days window
# and window_step advances the window x trading days
window_size = 183  #  ~ 6 months
window_step = 30  # ~ 1 month
threshold = 0.5
```

```
# %%
threshold_correlation = []
for i in range(0, len(returns) - window_size + 1, window_step):
    j = i + window_size
    window_data = returns[i:j]

    corr_matrix, _ = spearmanr(window_data)
    # corr_matrix = window_data.corr(method="pearson")
    corr_matrix[np.abs(corr_matrix) < threshold] = 0

    corr_df = pd.DataFrame(
        corr_matrix, index=window_data.columns, columns=window_data.columns
    )

    threshold_correlation.append(corr_df)

# %%
# Plot the correlation matrix for the 1st time window

plt.figure(figsize=(8, 6))
plt.imshow(threshold_correlation[0], cmap="viridis", interpolation="nearest")
plt.colorbar()

plt.title("Heatmap of the Correlation Matrix for the first time window")

plt.savefig(r"C:\ISEG\Mathematical Finance\Tese\matrix_heatmap_05_28092024.png")

for i, corr_df in enumerate(threshold_correlation[:5]):
    print(f"window {i+1} thresholded correlation matrix:\n", corr_df)
```

```
# %%
networks = []

for i, corr_df in enumerate(threshold_correlation):
    G = nx.DiGraph()
    stocks = returns.columns
    G.add_nodes_from(stocks)
    granger_results = []
    for col in corr_df.columns:
        for row in corr_df.index:
            if col != row and corr_df.at[row, col] != 0:
                test_data = returns[[row, col]].iloc[
                    i * window_step : i * window_step + window_size
                ]
                test_data = test_data.dropna()
                if test_data.shape[0] > 1:
                    max_lag = min(5, test_data.shape[0] // 2 - 1)
                    if max_lag > 0:
                        with SuppressOutput():
                            result = grangercausalitytests(test_data, maxlag=max_lag)
                        granger_results.append((row, col, result))
    for row, col, result in granger_results:
        for key, res in result.items():
            f_test_pvalue = res[0]["ssr_ftest"][1]
            chi2_test_pvalue = res[0]["ssr_chi2test"][1]
            lr_test_pvalue = res[0]["lrtest"][1]
            params_test_pvalue = res[0]["params_ftest"][1]
            if (
                f_test_pvalue < 0.05
                or chi2_test_pvalue < 0.05
                or lr_test_pvalue < 0.05
                or params_test_pvalue < 0.05
            ):
                G.add_edge(row, col)
    networks.append(G)
```

```
# %%
G = networks[-1]
AG = nx.nx_agraph.to_agraph(G)

AG.node_attr["shape"] = "circle"
AG.edge_attr["color"] = "black"
AG.edge_attr["penwidth"] = 0.7
AG.node_attr["style"] = "filled"
AG.node_attr["fillcolor"] = "plum3"
AG.node_attr["fontsize"] = 40
AG.node_attr["fontname"] = "Helvetica-Bold"
AG.node_attr["shape"] = "circle"
AG.node_attr["width"] = 3

AG.layout(prog="fdp")
AG.draw("ntwk_graph_ultima_spearman_threshold05_28092024.png")
from PIL import Image

Image.MAX_IMAGE_PIXELS = None
img = Image.open("ntwk_graph_ultima_spearman_threshold05_28092024.png")
img.save(
    r"C:\ISEG\Mathematical Finance\Tese\Network Images 28Set\network_last_window_spearman_threshold05_28092024.png"
)
img.show()

# %%
# Example of computing and printing network metrics for each window
with open(
    r"C:\ISEG\Mathematical Finance\Tese\Network 28Set\metrics\first_3_network_indegree_spearman_threshold05_
28092024.txt",
    "w",
) as f:
    for i in range(0, 3):
        G = networks[0]
        density = nx.density(G)
        out_degree_centrality = nx.out_degree_centrality(G)
        in_degree_centrality = nx.in_degree_centrality(G)
        betweenness_centrality = nx.betweenness_centrality(G)
        # df = pd.DataFrame(list(out_degree_centrality.items()), columns = ['Ticker', 'Out-Degree'])
        # df.to_excel(r'C:\ISEG\Mathematical Finance\Tese\Network 28Set\metrics\outdegree_1.xlsx')
        f.write(f"Network Metrics for Window {i+1}:\n")
        f.write(f"Density: {density}\n")
        f.write(f"In-Degree Centrality: {in_degree_centrality}\n")
        f.write(f"In-Degree Centrality: {out_degree_centrality}\n")
        f.write(f"Betweenness Centrality: {betweenness_centrality}\n")
        f.write("\n")
```

```
# %%
# Community detection using Louvain method for each network

communities_list = []
number_communities = []
for G in networks:
    partition = community_louvain.best_partition(G.to_undirected())
    communities_list.append(partition)

    num_comun = len(set(partition.values()))
    number_communities.append(num_comun)
number_communities
plt.figure(figsize=(10, 6))
plt.bar(range(1, len(number_communities) + 1), number_communities, color="peachpuff")
plt.xlabel("Moving Window")
plt.ylabel("Number of Communities")
plt.title("Number of Communities Detected in Each Network")
plt.xticks(range(1, len(number_communities) + 1))

plt.savefig(
    r"C:\ISEG\Mathematical Finance\Tese\Network 28Set\number_of_communities.png"
)

plt.show()


# %%

# Save the communities to a text file
with open(
    r"C:\ISEG\Mathematical Finance\Tese\Network 28Set\communities\communities_antepenultimate_window_threshold05
_spearman_28092024.txt",
    "w",
) as file:
    # for i in range(3):  # Loop over the first 3 networks
    partition = communities_list[-3]  # Get the partition for the i-th network
    file.write(f"Communities for the antepenultimate Network:\n")
    community_dict = defaultdict(list)

    for node, community in partition.items():
        community_dict[community].append(node)

    # Write each community and its corresponding nodes to the file
    for community, nodes in community_dict.items():
        file.write(f"Community {community}: {', '.join(nodes)}\n")

    file.write("\n")
```

```
# %%

output_dir = r"C:\ISEG\Mathematical Finance\Tese\Network 28Set\communities images"
os.makedirs(output_dir, exist_ok=True)

G = networks[2]
AG = nx.nx_agraph.to_agraph(G)
partition = communities_list[2]
communities = list(set(partition.values()))

# Assign a color to each community
color_map = plt.get_cmap("tab20")
community_colors = {
    community: mcolors.to_hex(color_map(i), len(communities))
    for i, community in enumerate(communities)
}

# Map each node to its community color and set size in pygraphviz
for node in G.nodes():
    community = partition[node]
    color = community_colors[community]

    # Set the node attributes in pygraphviz for color, size, etc.
    AG.get_node(node).attr["style"] = "filled"
    AG.get_node(node).attr["fillcolor"] = color
    AG.get_node(node).attr["fontsize"] = 40
    AG.get_node(node).attr["fontname"] = "Helvetica-Bold"
    AG.get_node(node).attr["shape"] = "circle"

    # Set width and height for node size
    AG.get_node(node).attr["width"] = 3  # Adjust node width here
    AG.get_node(node).attr["height"] = 3  # Adjust node height here

# Set edge color
for edge in G.edges():
    AG.get_edge(edge[0], edge[1]).attr["color"] = "black"

# Apply layout and save the graph to a file
try:
    AG.layout(prog="fdp")  # Use 'dot', 'neato', 'circo', or other layouts
    AG.draw("network_with_communities.png")
except Exception as e:
    print(f"Error during layout or drawing: {e}")

# Display the saved image
from PIL import Image

Image.MAX_IMAGE_PIXELS = None
img = Image.open("network_with_communities.png")
img.save(f"{output_dir}/network_communities_third_window.png")
img.show()

print("Networks with colored communities saved as images.")
```

```
# %%
network_edges = []
# with open(
#     r"C:\ISEG\Mathematical Finance\Tese\Network 28Set\resumo_networks_threshold05_spearman_28092024.txt", "w"
# ) as f:
for i in range(len(networks)):
    # num_nodes = networks[i].number_of_nodes()
    num_edges = networks[i].number_of_edges()
    network_edges.append(num_edges)

ind = list(range(1, len(network_edges) + 1))

plt.figure(figsize=(10, 6))
plt.bar(ind, network_edges, color="darkseagreen")

plt.xlabel("Moving Window")
plt.ylabel("Number of Edges")
plt.title("Number of Edges in Each Network")
plt.xticks(ind)

# Show the plot
plt.tight_layout()
plt.savefig(
    r"C:\ISEG\Mathematical Finance\Tese\Network 28Set\number_edges.png",
    format="png",
    dpi=300,
)
plt.show()
```