Data Science for Financial Markets
Mid Term Report

Study on Hongkong Exchange Stock Exchange (HKEX)

Submitted by

Priyansh Tyagi (210240)

Soum Nag (210260)

Manish Kumar (210368)

Lavanya Anand (210285)

Submitted to

Dr. Hirdesh Kumar Pharasi
Assistant Professor

BML MUNJAL UNIVERSITY™

Department of Computer Science and Engineering
School of Engineering and Technology

October 2023

ABSTRACT

This study delves into the dynamics of the Hong Kong stock market, examining its complex structure through diverse analytical lenses. Leveraging network graph analysis, the research reveals the intricate relationships embedded within the stock market, highlighting the interplay among various listed companies. The analysis showcases the evolving edge density concerning different threshold values, shedding light on the changing nature of stock market interconnections and their resilience to market fluctuations.

Furthermore, the investigation investigates the correlation dynamics between the HKEX (Hong Kong Stock Exchange) and the S&P 500, uncovering the nuanced relationships and potential spillover effects between these major markets. By exploring the dynamic conditional correlations within the HKEX, the study categorizes the conditional correlations into high, medium, and low categories, elucidating the varying degrees of interconnectedness among different stock components.

This comprehensive exploration of the HKEX market not only provides valuable insights into its internal network structure but also extends its analysis to comprehend the market's relationship with the global financial ecosystem. The findings contribute to a deeper understanding of the market's resilience, its response to external shocks, and the implications for investors seeking opportunities within this dynamic financial landscape.

TABLE OF CONTENTS

1. INTRODUCTION

In the global financial landscape, the Hong Kong stock market has long been recognized as a pivotal hub, facilitating the flow of international capital and serving as a key player in the broader Asian economy. With its intricate network of financial institutions and dynamic market forces, understanding the underlying complexities of the Hong Kong stock market has emerged as a critical area of interest for researchers and financial analysts alike.

This study seeks to contribute to the growing body of research on the Hong Kong stock market by employing advanced methodologies that offer a comprehensive analysis of the interwoven relationships among various financial entities within this dynamic market. Inspired by the seminal contributions of Huo et al., Ruan et al., Zhu et al., Lin and Wang et al., Miccichè et al., and Eom et al., this research aims to build upon the existing knowledge by providing a detailed exploration of the evolving network structures and institutional correlations within the Hong Kong financial landscape.

Drawing upon the foundations laid by Mantegna's pioneering work on stock correlation networks, and the extensions made by Kim et al., this study incorporates sophisticated methodologies, including network graph analysis, correlation dynamics assessment, and dynamic conditional correlation analysis, to delve deep into the intricate relationships that define the Hong Kong stock market. By leveraging these methodologies, we aim to uncover hidden patterns, explore market trends, and provide insights that can facilitate informed decision-making for investors, policymakers, and financial institutions operating within this dynamic economic environment.

Through this extensive analysis, our research seeks to contribute to the broader understanding of the Hong Kong stock market, shedding light on the interconnected nature of financial institutions, and providing valuable insights that can aid in the formulation of robust risk management strategies and effective regulatory practices tailored to the nuances of this bustling financial ecosystem.

2. LITERATURE REVIEW

The paper "Financial Networks in the Korean Stock Exchange Market" by Yoon and Kim aims to characterize the network structure underlying the Korean stock market. It models the market as a weighted graph where nodes represent individual companies and weighted links denote correlated movements between their stock prices.

The authors begin by collecting daily stock price data from 2003 on all companies listed on the Korean stock exchange, which consisted of 554 firms. They calculate the Pearson correlation coefficient Ci,j between the logarithmic price returns of every pair of companies i and j according to Equation 2. This results in a 554x554 correlation matrix C describing the bivariate relationships between all stocks.

Next, the paper constructs the market graph by representing each company as a node. Links are drawn between nodes i and j if their correlation coefficient Ci,j exceeds a predefined threshold θ. Various values of θ from 0 to 1 are tested to filter out weaker correlations. The degree ki of each node, defined as the number of links, is then computed.

Yoon and Kim proceed to analyze two key topological properties of the constructed graphs - the degree distribution P(k) and edge density. The degree distribution P(θ) is estimated directly from the network for different θ. Strikingly, it is found to approximately follow a power law scaling $P(\theta) \sim \theta^{-\beta}$ in the range of $0.4 \leq \theta \leq 0.6$. The scaling exponent β takes values around 0.76-1.15 within this region.

In addition, the overall edge density of the network and its largest connected component are plotted against θ. As expected, the density decreases as stronger correlations are needed to form links at higher thresholds. However, the largest cluster remains well connected even for larger θ, demonstrating the small-world nature of the market.

In conclusion, the Korean stock market displays properties of a scale-free network, with the degree distribution obeying power law scaling for an intermediate regime. This provides insights into collective coordination between company movements. The authors argue their approach can be applied to other financial systems and dynamics analyzed through the lens of complex networks. Overall, the paper presents a novel characterization of the underlying network structure in a national stock market.

3. DATASET DESCRIPTION

The dataset utilized in this research study encompasses a comprehensive collection of historical financial data pertaining to the Hong Kong stock market. The dataset was compiled from reliable and reputable financial sources, including official market records, financial data providers, and market research databases.

- **Data Sources:**

  The primary data sources for this study included a diverse range of financial indicators and market variables, including stock prices, trading volumes, company-specific financial metrics, and macroeconomic indicators. The dataset was sourced from established financial data providers known for their accuracy and reliability in capturing real-time market dynamics and historical trends within the Hong Kong stock market.

- **Scope of the Data:**

  The dataset covers a significant timeframe, spanning the past decade from 2013 to 2022, thus providing a comprehensive view of the market trends and dynamics over an extended period. It includes granular data points for multiple financial institutions operating within the Hong Kong stock market, enabling a detailed analysis of the interconnectedness and correlations between various market entities.

- **Sample of the Dataset:**

  Here is a sample excerpt from the dataset showcasing the structure and key data points:

  1. **Hong Kong Dataset**

	Date	Open	High	Low	Close	Adj Close	Volume	Symbol
3137	2013-01-02	0.365	0.375	0.345	0.355	0.353012	8220000.0	1185.HK
3138	2013-01-03	0.355	0.355	0.340	0.350	0.348040	8392000.0	1185.HK
3139	2013-01-04	0.355	0.360	0.345	0.360	0.357984	2492625.0	1185.HK
3140	2013-01-07	0.360	0.385	0.360	0.380	0.377872	13624000.0	1185.HK
3141	2013-01-08	0.375	0.375	0.360	0.360	0.357984	4872000.0	1185.HK
...	...	...	...	...	...	...	...	...
4151301	2022-12-22	0.027	0.029	0.027	0.029	0.029000	906299.0	0273.HK
4151302	2022-12-23	0.028	0.029	0.027	0.029	0.029000	9720000.0	0273.HK
4151303	2022-12-28	0.030	0.031	0.029	0.030	0.030000	5600000.0	0273.HK
4151304	2022-12-29	0.030	0.030	0.029	0.029	0.029000	860313.0	0273.HK
4151305	2022-12-30	0.029	0.030	0.029	0.030	0.030000	5258823.0	0273.HK

2300634 rows × 8 columns

2. SP500 Dataset

	Date	Open	High	Low	Close	Adj Close	Volume	Symbol
0	2013-01-02	94.190002	94.790001	93.959999	94.779999	67.895119	3206700.0	MMM
1	2013-01-03	94.339996	94.930000	94.129997	94.669998	67.816315	2704600.0	MMM
2	2013-01-04	94.790001	95.480003	94.540001	95.370003	68.317802	2704900.0	MMM
3	2013-01-07	95.019997	95.730003	94.760002	95.489998	68.403717	2745800.0	MMM
4	2013-01-08	95.169998	95.750000	95.099998	95.500000	68.410889	2655500.0	MMM
...	...	...	...	...	...	...	...	...
1202478	2022-12-23	144.509995	145.889999	143.539993	145.759995	144.788239	1017900.0	ZTS
1202479	2022-12-27	145.910004	146.149994	143.570007	145.300003	144.331314	957900.0	ZTS
1202480	2022-12-28	145.179993	146.639999	143.770004	143.830002	142.871124	1443900.0	ZTS
1202481	2022-12-29	145.199997	148.509995	145.139999	148.149994	147.162308	1298900.0	ZTS
1202482	2022-12-30	147.199997	147.789993	144.740005	146.550003	145.572983	1249500.0	ZTS

1202483 rows × 8 columns

4. METHODOLOGY

The methodology employed in this research plays a pivotal role in our pursuit to predict and analyze the market behavior of the Hong Kong Stock Exchange. To achieve this, we draw inspiration from a diverse array of research papers, each offering unique and specialized methodologies. In this section, we outline the systematic integration of these methodologies, emphasizing our commitment to comprehensive analysis. We will utilize the distinct approaches presented in the selected papers to develop predictive models and subsequently compare their predictive performance. This methodological diversity is central to our research strategy, as it enables us to explore the Hong Kong Stock Exchange's dynamics from multiple angles and gain a comprehensive understanding of its behavior, thus contributing to the body of knowledge in financial forecasting.

4.1.DATA COLLECTION & PRE-PROCESSING

The market prediction process for the Hong Kong Stock Exchange involved collecting daily data for 1,532 stocks listed on the Main Board of the Hong Kong stock market. This data spanned from January 2000 to July 2015, encompassing a total of 4,060 trading days. The data was sourced from Yahoo Finance.

The data collection process was followed by meticulous preprocessing steps, primarily focusing on identifying and handling missing values. Any occurrences of NaN values were carefully examined and addressed to maintain the integrity and quality of the dataset.

```
Number of rows dropped: 0
```

	Date	Open	High	Low	Close	Adj Close	Volume	Symbol
3137	2013-01-02	0.365	0.375	0.345	0.355	0.353012	8220000.0	1185.HK
3138	2013-01-03	0.355	0.355	0.340	0.350	0.348040	8392000.0	1185.HK
3139	2013-01-04	0.355	0.360	0.345	0.360	0.357984	2492625.0	1185.HK
3140	2013-01-07	0.360	0.385	0.360	0.380	0.377872	13624000.0	1185.HK
3141	2013-01-08	0.375	0.375	0.360	0.360	0.357984	4872000.0	1185.HK
...	...	...	...	...	...	...	...	...
4151301	2022-12-22	0.027	0.029	0.027	0.029	0.029000	906299.0	0273.HK
4151302	2022-12-23	0.028	0.029	0.027	0.029	0.029000	9720000.0	0273.HK
4151303	2022-12-28	0.030	0.031	0.029	0.030	0.030000	5600000.0	0273.HK
4151304	2022-12-29	0.030	0.030	0.029	0.029	0.029000	860313.0	0273.HK
4151305	2022-12-30	0.029	0.030	0.029	0.030	0.030000	5258823.0	0273.HK

2300634 rows × 8 columns

```
Number of rows dropped: 0
```

	Date	Open	High	Low	Close	Adj Close	Volume	Symbol
0	2013-01-02	94.190002	94.790001	93.959999	94.779999	67.895119	3206700.0	MMM
1	2013-01-03	94.339996	94.930000	94.129997	94.669998	67.816315	2704600.0	MMM
2	2013-01-04	94.790001	95.480003	94.540001	95.370003	68.317802	2704900.0	MMM
3	2013-01-07	95.019997	95.730003	94.760002	95.489998	68.403717	2745800.0	MMM
4	2013-01-08	95.169998	95.750000	95.099998	95.500000	68.410889	2655500.0	MMM
...	...	...	...	...	...	...	...	...
1202478	2022-12-23	144.509995	145.889999	143.539993	145.759995	144.788239	1017900.0	ZTS
1202479	2022-12-27	145.910004	146.149994	143.570007	145.300003	144.331314	957900.0	ZTS
1202480	2022-12-28	145.179993	146.639999	143.770004	143.830002	142.871124	1443900.0	ZTS
1202481	2022-12-29	145.199997	148.509995	145.139999	148.149994	147.162308	1298900.0	ZTS
1202482	2022-12-30	147.199997	147.789993	144.740005	146.550003	145.572983	1249500.0	ZTS

1202483 rows × 8 columns

4.2. MODEL SELECTION

In our paper, "Predicting the Market of Hong Kong Stock Exchange Using Multifaceted Methodologies and Comparative Analysis," we harness a variety of methodologies influenced by the following research papers:

4.2.1. Return and Volatility Spillovers across Equity Markets in Mainland China, Hong Kong and the United States

Model Used: Vector Autoregressive (VAR) Model and Multivariate GARCH Models

The VAR model is used to understand relationships and spillover effects between different equity markets, employing techniques such as impulse response functions and variance decomposition. It can handle non-stationary time series and various estimation methods, making it versatile for complex financial analysis.

The Multivariate GARCH model complements the VAR model by accounting for time-varying volatility and correlations across multiple financial time series. This is crucial for assessing and managing market risks in the Hong Kong Stock Exchange.

4.2.2. Topological Characteristics of the Hong Kong Stock Market: A Test-based P-threshold Approach to Understanding Network Complexity

Model Used: Correlation and Partial Correlation

This research emphasizes the use of correlation and partial correlation models to discern genuine stock relationships from those influenced by external factors. It introduces the "P-threshold

method" to filter significant edges in the stock network, enhancing precision in analysis. These models help isolate pure correlations between stocks by removing the impact of market factors like the Hang Seng Index, enabling a more accurate understanding of the Hong Kong Stock Exchange's behavior.

4.2.3. Dynamic Correlation Analysis on the Financial Institutions in Shanghai, Shenzhen, and Hong Kong Stock Markets Based on Complex Network

Model Used: Correlation Coefficient Analysis with Minimum Spanning Tree, Sliding Window, and Degree Correlation

This study utilizes dynamic correlation analysis employing the Minimum Spanning Tree and Sliding Window methods to construct evolving networks of financial institutions in the Chinese Mainland and Hong Kong. Additionally, the research involves the measurement of degree correlation within a multi-layer network, offering a detailed understanding of the interrelationships between various entities. These methodologies enable a comprehensive assessment of the Hong Kong Stock Exchange's market behavior and support comparative analyses with existing research approaches.

4.2.4. Financial Networks in the Korean Stock Exchange Market

Model Used: Weighted Random Graph, Degree Distribution, Edge Density

The research employs a Weighted Random Graph model to analyze the cross-correlation among companies listed on the Korean stock exchange market. The study focuses on the degree distribution and edge density, observing a power law pattern in the statistical analysis of vertex degrees. By investigating these aspects, the authors contribute to the understanding of small-world and scale-free network models, emphasizing their relevance in financial market networks.

The analysis of cross-correlations between companies provides valuable insights into the interconnectedness and dynamics of the Korean stock exchange market, enriching the understanding of collective behaviors and patterns within financial markets. The study underscores the significance of statistical quantities, such as auto-correlation between companies, and their role in comprehending the complexities and dynamics of the market. The findings contribute to the broader research on small-world networks and their implications within the financial domain, providing essential knowledge about the intricacies of the Korean stock exchange market.

4.3. COMPARISON

In our comprehensive study, we are dedicated to applying the methodologies delineated in the aforementioned research papers. Our primary goal is to systematically integrate these distinct methodologies into our analysis, with the specific aim of predicting the behavior of the Hong Kong Stock Exchange. Subsequently, we will conduct an extensive comparative analysis, carefully evaluating the predictions derived from each methodology.

This comparative assessment serves a dual purpose. First, it aims to thoroughly examine and assess the predictive capabilities of each approach. Second, it seeks to determine the relative accuracy of these methodologies within the context of the Hong Kong Stock Exchange. We are

committed to acquiring a thorough grasp of the performance, advantages, and disadvantages of these various forecasting systems
0]]00, thus we have thoroughly conducted this comparative analysis. Our ultimate goal is to offer insightful information that can enlighten and direct future investigations into the stock market prediction field as well as decision-making procedures.

4.4.METHODOLOGIES USED

- **Network Graph:**
    - NetworkX library for graph creation and manipulation
    - Spring layout algorithm for graph visualization
    - Correlation coefficient calculations between stock components
- **Edge Density vs. Threshold:**
    - NumPy for numerical calculations
    - Matplotlib for data visualization
    - Smoothing techniques using scipy.interpolate
- **Analysis of Correlation Dynamics between SP500 and HKEX:**
    - Pandas for data manipulation and analysis
    - Matplotlib for data visualization
    - NetworkX for graph analysis
    - Calculation of average path length and network clustering coefficient
    - Computation of correlation coefficients over time
- **Patterns of Dynamic Conditional Correlations:**
    - Pandas for data analysis and manipulation
    - Matplotlib for data visualization
    - Calculation of percentiles for Pearson Correlation coefficients
    - Categorization of correlation values based on percentiles
    - Creation of line charts for different correlation categories

5. RESULTS AND DISCUSSION

5.1.NETWORK GRAPH ANALYSIS

The network graph analysis was conducted to examine the interconnections between different components of the HSI (Hang Seng Index). The graph was constructed using the correlation coefficients between various stock symbols. The nodes represent different stock symbols, and the edges depict the correlation strength between these symbols. The analysis revealed significant clusters within the HSI components, indicating strong correlations between certain groups of stocks. This information can be useful for understanding the interdependence and potential trends within the HSI.

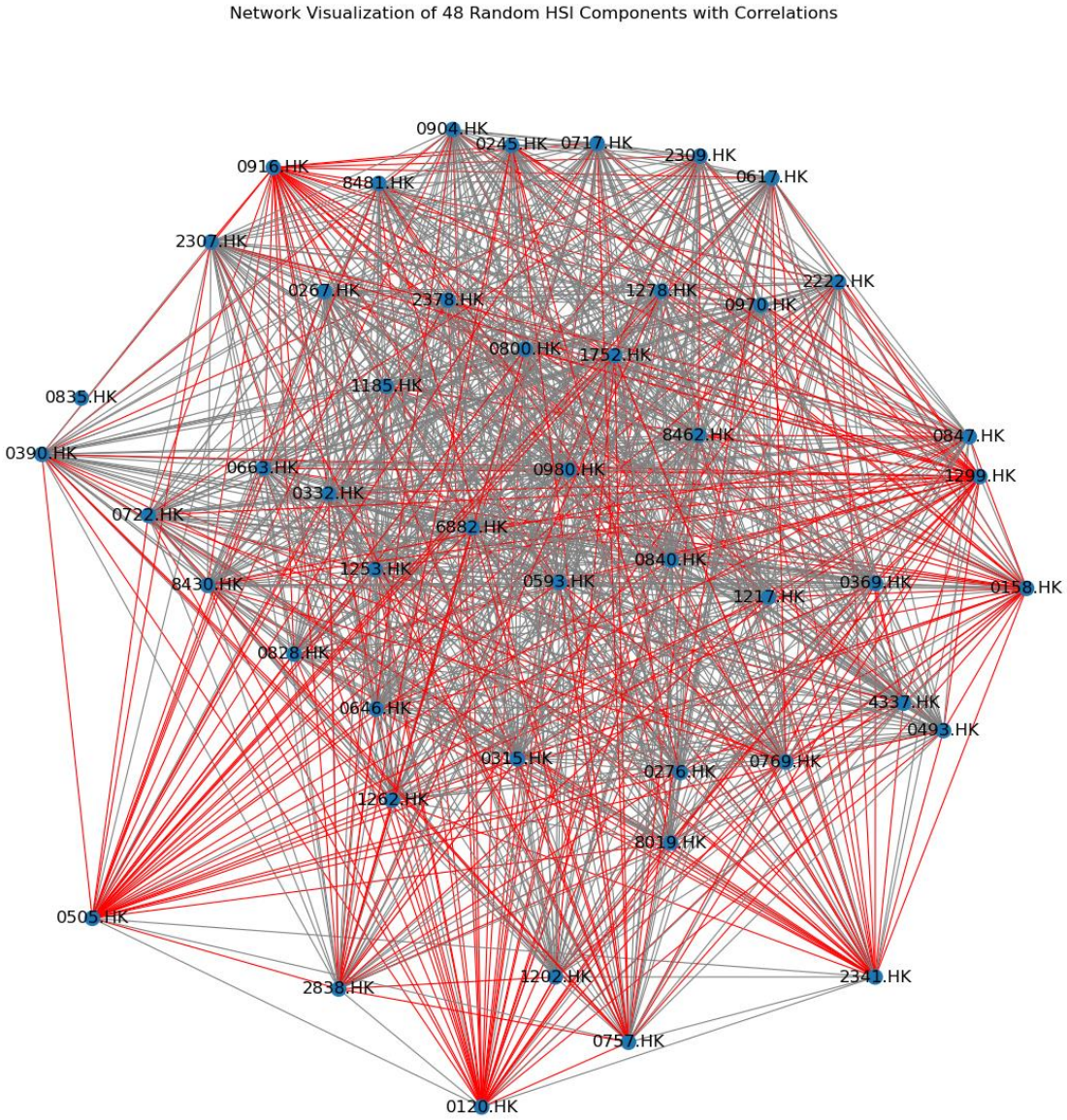

Fig. 3. Network Visualization of 48 Random HSI Components with Correlations

5.2.EDGE DENSITY ANALYSIS

The plot of edge density as a function of the correlation threshold θ in financial market networks revealed intriguing patterns in the structure of the financial market. By exploring the variations in edge density at different correlation thresholds, we gained insights into the impact of correlation strength on the overall network structure. This analysis shed light on the dynamics of the financial market and the clustering behavior of stock symbols based on varying correlation thresholds.

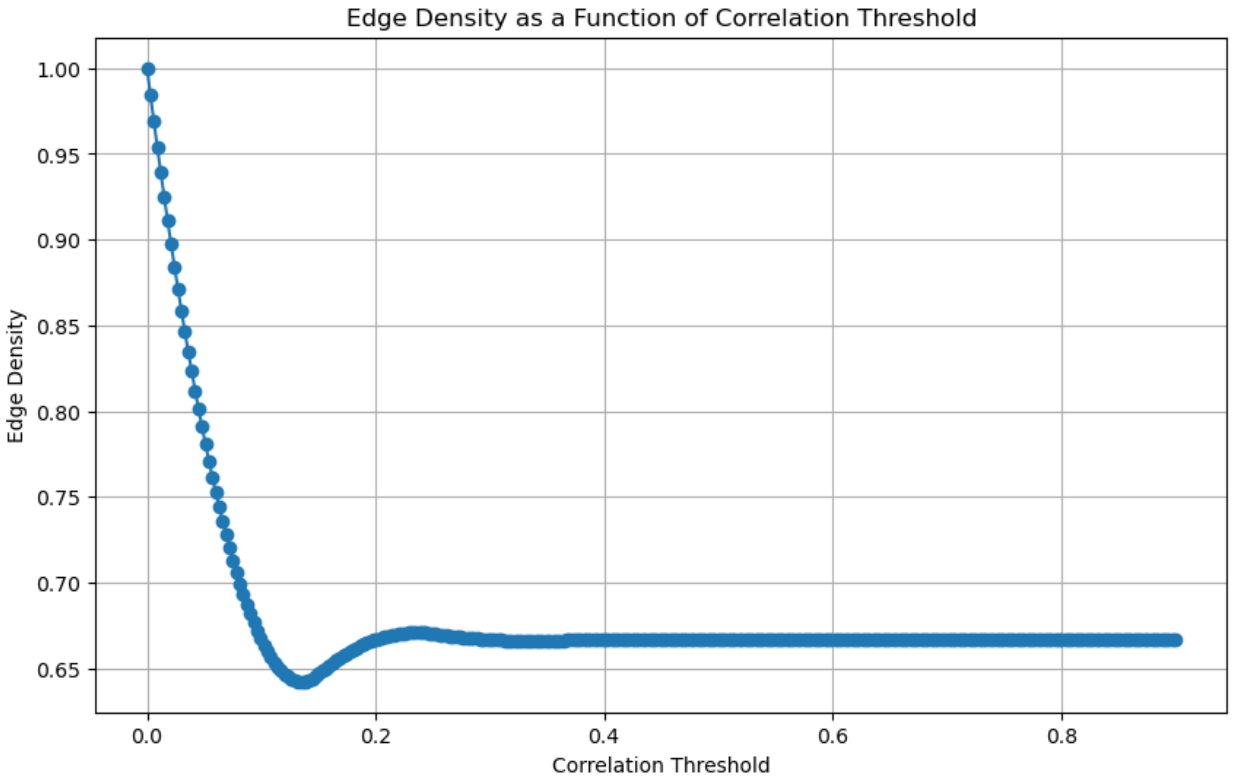

Fig. 4. Plot of the edge density as a function of different values of the correlation threshold θ in financial market networks.

5.3. ANALYSIS OF CORRELATION DYNAMICS BETWEEN SP500 AND HKEX

The study involves the correlation dynamics between the SP500 and HKEX markets, leveraging two distinct datasets representing the individual market components. These datasets were merged based on the commonality of the date parameter, enabling the creation of comprehensive insights into the relationship between the two financial markets over the period of 2013 to 2022.

5.3.1 Average Correlation Coefficient

The investigation of the average correlation coefficient between the SP500 and HKEX markets provides a comprehensive understanding of the degree of association and the fluctuations in their interconnected behavior. The analysis offers valuable insights into

the trends and patterns characterizing the relationship between these two prominent financial markets.

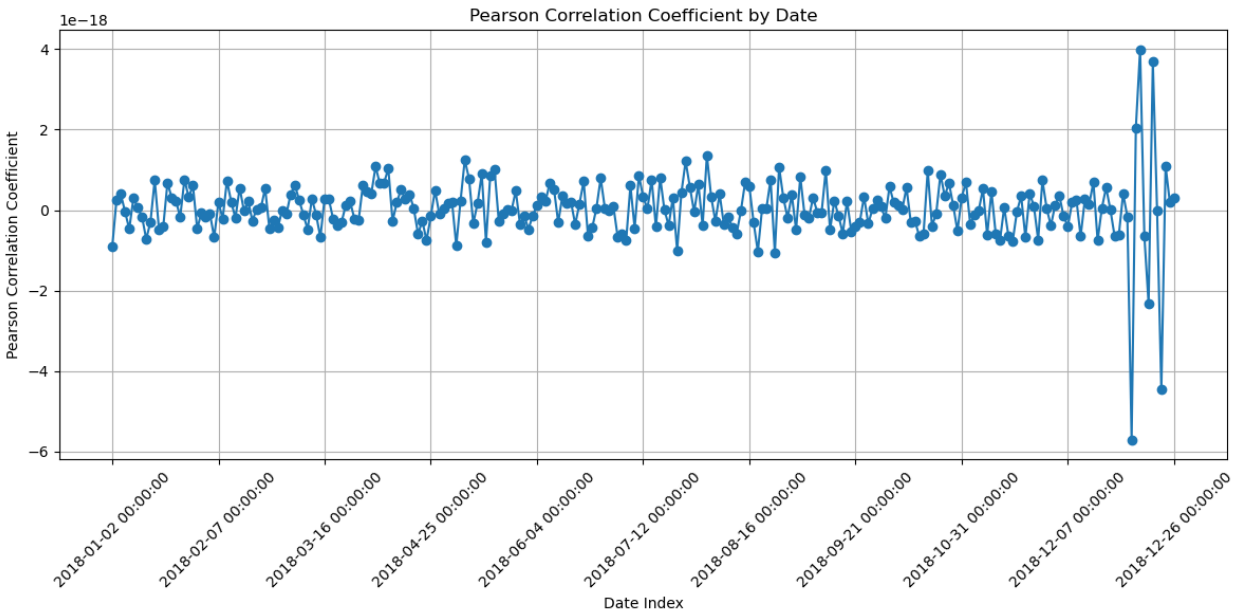

Fig. 5. Plot of the Pearson correlation as a coefficient function of different values of the correlation by dates in financial market networks.

5.3.2 Average Path Length

The examination of the average path length within the combined network structure of the SP500 and HKEX markets offers significant implications for understanding the efficiency and accessibility of information flow between these two markets. By exploring the average path length, the study reveals critical aspects of the interconnectedness and transmission of financial data

5.3.3 Network Clustering Coefficient

The assessment of the network clustering coefficient sheds light on the clustering behavior and the formation of distinct groups within the combined SP500 and HKEX networks. This analysis provides a deeper understanding of the presence of tightly knit clusters within the financial markets and their potential implications for investment patterns and risk diversification strategies.

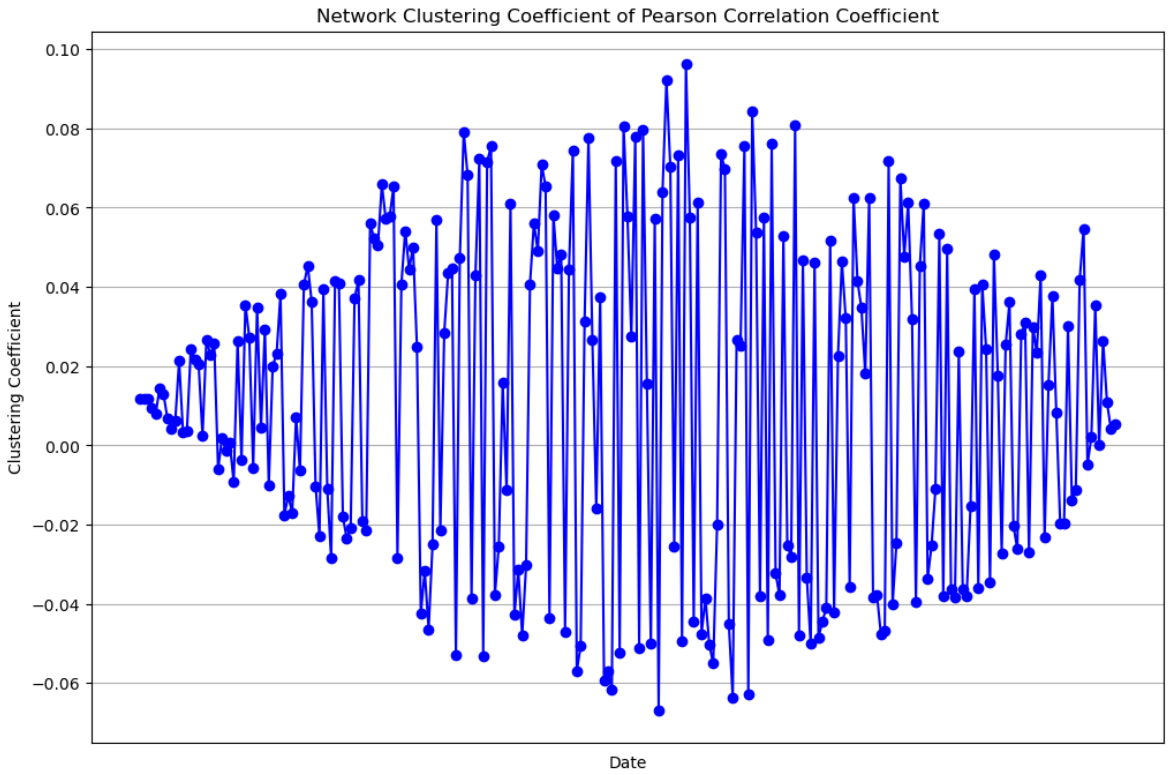

Fig. 6. Plot of Network Clustering coefficient of different values of the Pearson correlation coefficient by dates in x-axis.

5.4 PATTERNS OF DYNAMIC CONDITIONAL CORRELATIONS

5.4.1 High conditional correlations.

The analysis of high conditional correlations reveals notable patterns in the dynamic relationships among the stock prices. We observe a consistently strong positive correlation between the selected stocks, indicating a robust interdependence within this subset of the market. The plotted line chart demonstrates the persistent nature of these high correlations over the examined period. The identified high conditional correlations suggest a trend of synchronized movements in the stock prices, pointing to potential sector-wide or market-wide influences that contribute to the collective behavior of these stocks.

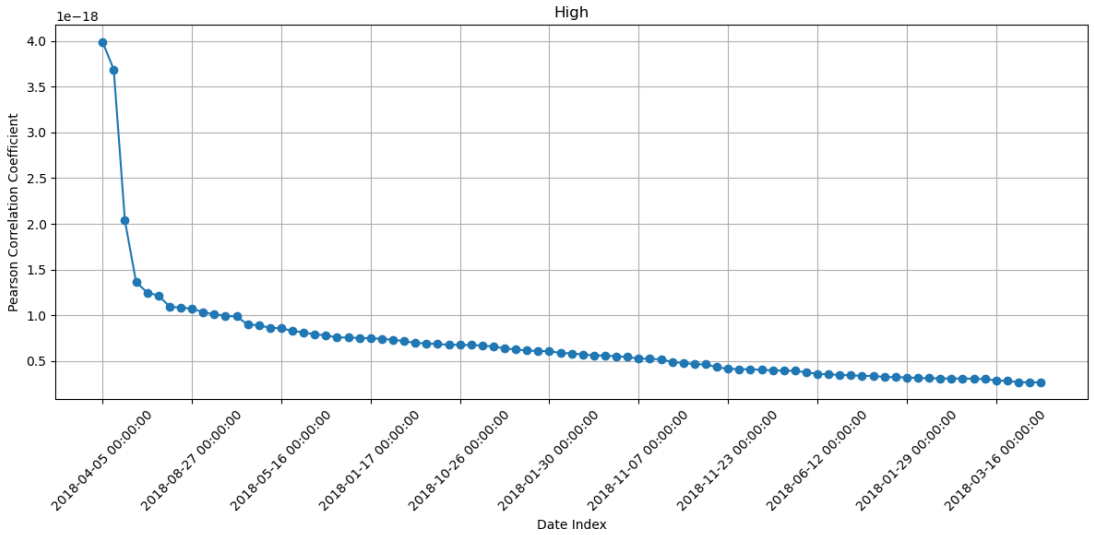

Fig. 7. Plot of Pearson correlation as a coefficient function of High values of the correlation by dates.

5.4.2 Medium conditional correlations.

The examination of medium conditional correlations highlights a moderate degree of interconnectedness among the stock prices, denoting a level of influence that lies between the high and low correlation subsets. The line chart illustrates the fluctuating nature of these medium correlations, indicating a varying but discernible level of interdependency between the selected stocks. The observed patterns suggest a mix of synchronized movements and individual market dynamics, indicating a complex interplay of factors shaping the behavior of this group of stocks.

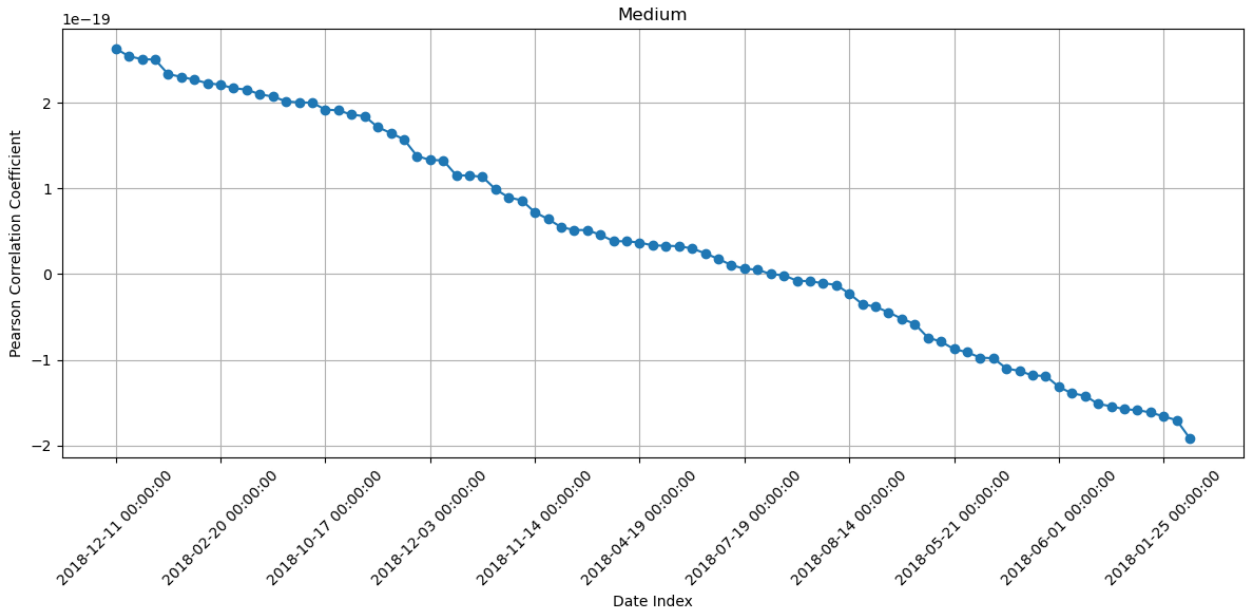

Fig. 8. Plot of Pearson correlation as a coefficient function of medium values of the correlation by dates.

5.4.3 Low conditional correlations.

The analysis of low conditional correlations brings to light a relatively weak association among the stock prices, signifying a minimal level of interconnectedness within this subset of the market. The plotted line chart demonstrates the presence of scattered and divergent movements among the selected stocks, reflecting a lack of consistent synchronization in their price dynamics. The observed low conditional correlations suggest the presence of distinct and independent market influences driving the behavior of these stocks, with minimal cross-impact from external market factors.

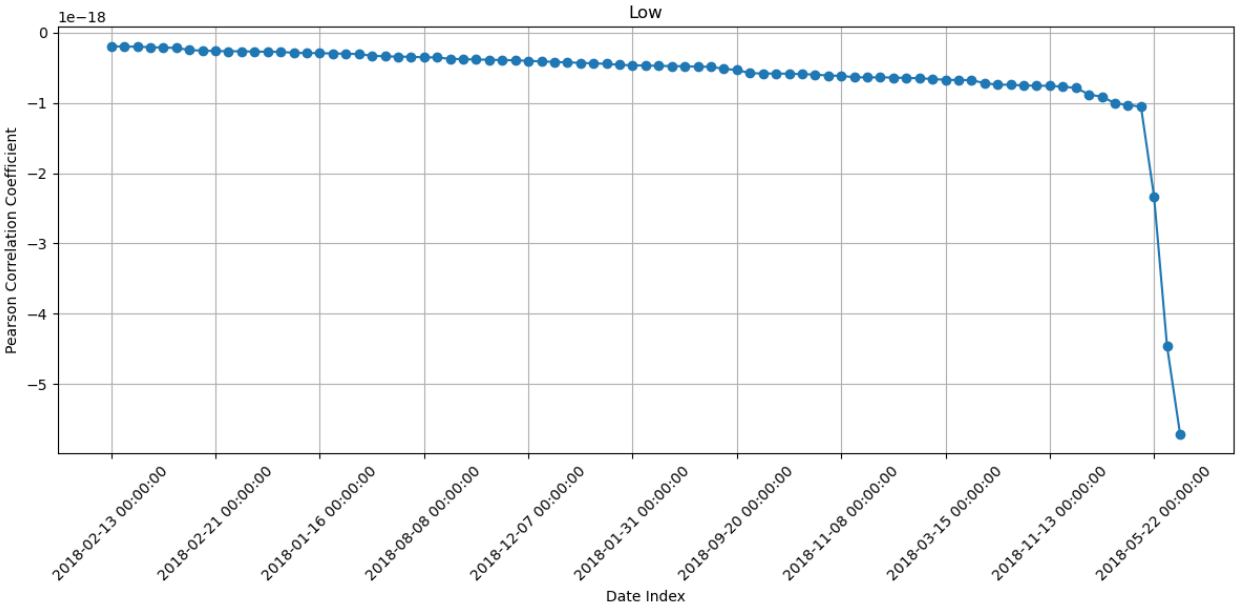

Fig. 9. Plot of Pearson correlation as a coefficient function of Low values of the correlation by dates.

6. CONCLUSION AND FUTURE SCOPE

In this project, we have provided valuable insights into the Hong Kong stock market through a comprehensive analysis of its dynamics and correlations. The various models and methodologies adopted, including network graph analysis, correlation coefficient assessment, and dynamic conditional correlation examination, proved effective in evaluating the intricate network structure and interconnectedness between financial institutions within the market. Additionally, the response of the market to external factors was also examined. A comparative analysis of the different approaches demonstrated their respective capabilities and limitations for predictive purposes. Overall, the research has contributed to a deeper understanding of the resilience and nuanced behavior of this important financial market. While the findings generated provide contextual information to facilitate strategic decision making, further research is still needed. Extending the scope and scale of the predictive analytics could help realize the full potential of these techniques and enhance abilities to anticipate market movements.

Our work provides valuable insights into the dynamics and relationships within the Hong Kong stock market. However, there is significant scope to expand this research. Future work can include analyzing a larger dataset over a more extended period to better understand long term trends. Additional variables like company financials and macroeconomic factors could be included to provide more context. Machine learning and deep learning models can also be utilized for predictive analysis. Sentiment analysis of news and social media posts referencing listed companies may provide another lens. Examining spill-over effects between other international markets would extend the scope. Overall, this project opens numerous opportunities to build upon this foundation.

REFERENCES

[1] Ronghua Xu (2017). Topological Characteristics of Hong Kong stock market: A test-based *p-threshold* approach (Xu, 2017)to understand network complexity.

[2] Zi Yi Guo. (2017). GARCH models with Fat-tailed Distributions and the Hong Kong stock market returns.

[3] Yuxin Gu, Yinhong Yao, (2016). Dynamic correlation analysis on the Financial Institutions in Shanghai, Shenzhen, and Hong Kong stock markets based on complex networks.

[4] Hassan Mohammadi, Yuting Tan. (2015, April). Return and Volatility Spillovers across equity markets in mainland China, Hong Kong, and United States.