



DUBAI • MUMBAI • SINGAPORE • SYDNEY

Portfolio Construction and Analysis of Global Exchange-Traded Funds Using Network Science and Feature Selection Methods

Nathaniel Christopher

DBA 26054

Under the supervision of

Dr Priti Bakhshi

Principal Supervisor

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Business Administration (DBA)

S P Jain School of Global Management

November 2022

Abstract

Using the complex system approach of network science as the main tool for portfolio construction, market portfolio construction, allocations, and frameworks in financial markets is a difficult task. The research methodology employed a complex system approach from network science for portfolio construction in financial markets. Using attribute selection methods like genetic algorithms, Naïve Bayes, and neural networks, link predictions were made among 47 ETFs from the US markets. The network measures of betweenness, closeness, and PageRank were used as ranking tools for constructing portfolios. Eighteen portfolios were constructed based on the top- and bottom-ranked ETFs according to these network measures. This innovative approach aimed to provide a comprehensive perspective on portfolio compositions, demonstrating the efficacy of network science in optimizing financial portfolios.

The primary aim of this research was to explore and validate the application of network science, specifically using the complex system approach, in portfolio construction and allocation within financial markets. By employing attribute selection methods such as genetic algorithms, Naïve Bayes, and neural networks for link prediction, combined with network measures like betweenness, closeness, and PageRank for ranking, the researcher sought to determine the efficacy of these methods in constructing robust portfolios. The study also aimed to understand the relationships between various asset classes and the dynamics of an overall portfolio. The research aimed to provide a comprehensive perspective on portfolio compositions and offer an innovative framework for understanding and optimizing portfolios, presenting a holistic view beneficial for both researchers and practitioners in the financial domain.

We also discovered that the average betas across all clustering and PageRank portfolios were consistently lower for the top set of ETFs, at 0.37, and higher for the bottom set of portfolios, at 0.48 and 0.56. The research revealed that the 18 portfolios had a skewness value of -0.28 and a positive excess kurtosis value of 0.50, indicating that they had a lower risk than the market index. When compared to the market index, the correlation of all 18 portfolios was nearly 0.8. We were able to construct robust portfolios with low risk due to low skewness and excess kurtosis, and the network models were useful in finding and building such portfolios based on a given set of data. The best-performing portfolio based on risk-to-reward metrics was the selection of the top-ranked ETFs of the betweenness centrality portfolios, which returned an annualised average of 8.98%.

The easy identification of the beta values of the various portfolios facilitated an intuitive solution to portfolio analysis and selection. Based on the betweenness centrality measure, the top-performing ETFs had a higher beta value of 0.52 than the bottom set of ETFs. The average beta across all portfolios was consistent at 0.47 when compared to the overall market index. At only 8.3%, the overall standard deviations of all portfolios were lower than the market's, at more than 15.7% on an annualised basis. The study also revealed that the overall Sharpe ratio of the various portfolios was 0.91, while that of the market was 0.55.

The portfolios developed using the network metric framework and analysis provide a low-risk solution to understand portfolios using the 47 ETFs in the research data set. Based on the relationships between the various attribute selection methods, it was found that stronger relationships are caused by the different allocation patterns of the various ETFs when using Markowitz portfolio optimisation. The top and bottom sets of ETF portfolios for each centrality measure provide unique optimisation for robustness in terms of diversification, facilitating an intuitive alternate method.

This holistic approach to understanding portfolios and providing a robust methodology has been a cornerstone of using network science for building networks. Based on the research, the relationships between various ETFs found using the network science method provide an intuitive knowledge discovery process for understanding the relationships between asset classes and the dynamics of an overall portfolio, providing a holistic view that can be useful for researchers and practitioners.

Acknowledgements

First and foremost, I want to thank God almighty for the grace and mindset to complete this thesis. I also want to thank Prof. Dr Seetharaman and Prof. Dr Preethi Bakshi for providing critical insights into the approach to the thesis in terms of the various aspects of the methodology, the problem, and other factors. I also want to thank Prof. Dr Madhulutty for helping me to understand the various aspects of the thesis presentation and addressing the doubts that I had relating to my thesis submission.

I thank my father, Prof. Dr Christopher Abraham, and my mother, Ms. Elizabeth, for their support in evaluating my thesis and for providing insights that were helpful in the assessment of the paper.

I also thank my mother-in-law, Prof. Dr J. Daphy Louis Lovenia, and my father-in-law, Mr J. Devadass Inbasekar, for continually supporting and motivating me and helping me in the evaluation of various thesis formats and criteria.

I also want to thank my beautiful wife, Jemima, for providing me with the needed motivation when I faced difficulties in completing my thesis.

Lastly, I want to thank my friends in India and abroad for motivating me so that I could achieve my goals and complete my thesis on time.

The thesis writing process is a daunting but rewarding experience. During this process, not only did I learn about the many aspects of research, but I also came to understand my strengths and weaknesses. I was able to understand how to manage my time, assess various goals, and be more efficient in framing ideas and creating new frameworks. The most important thing I realised in this journey is that there are wonderful people in my life who will motivate, guide, and support me in all my future endeavours.

I wish for success and bright opportunities for all those who were with me throughout this journey, and I wish everyone good luck in their future endeavours.

Declaration

This is to certify the following:

- This thesis contains no material that has been accepted for awarding the candidate any other degree or diploma except where due reference is made in the text of the examinable outcome.
- To the best of the candidate's knowledge, this thesis contains no material previously published or written by another person except where due reference is made in the text of the examinable outcome.
- Where the work is based on joint research or publication, this thesis discloses the contributions of the respective authors.
- The content of this thesis is the result of work that has been carried out since the official commencement date of the approved research program.
- This thesis is less than 70,000 words in length, including the bibliography and appendices.
- This thesis has been edited by Proofed. The editing addressed only style and grammar and not its substantive content.
- This thesis has met all the requirements of ethics approval from the ethics committee of the S P Jain School of Global Management (see Appendices 1–9).



Nathaniel Christopher (Author)

For S P Jain School of Global Management



Dr Priti Bakhshi (Principal Supervisor)

Table of Contents

Abstract.....	2
Acknowledgements	4
Declaration.....	5
Table of Contents	6
Detailed Table of Contents	7
List of Tables	11
List of Figures.....	12
Chapter 1: Introduction	13
Chapter 2: Literature Review.....	44
Chapter 3: Research Methodology	85
Chapter 4: Results and Findings	118
Chapter 5: Conclusion	160
References.....	177
Appendix 1: Important Portfolio Characteristics of the 18 Portfolios.....	195
Appendix 2: Summary of Turnitin Plagiarism Check.....	239

Detailed Table of Contents

Abstract.....	2
Acknowledgements.....	4
Declaration.....	5
Table of Contents	6
Detailed Table of Contents	7
List of Tables	11
List of Figures.....	12
Chapter 1: Introduction.....	13
1.1 Introduction	13
1.2 Background of the Research	13
1.3 Research Problem	14
1.4 Purpose of the Research.....	16
1.5 Aims of the Study	17
1.6 Research Significance.....	17
1.7 Research Methodology	18
1.7.1 Data and research framework.....	18
1.8 Literature Review	19
1.9 Definitions	20
1.9.1 Complexity theory review.....	20
1.9.2 Systems	21
1.9.3 Interactions or networks.....	22
1.9.4 Emergence	22
1.9.5 Dynamics	22
1.9.6 Self-organisation	23
1.9.7 Adaptation.....	23
1.9.8 Interdisciplinary	24
1.9.9 Methods	24
1.9.10 Pattern formation.....	25
1.9.11 Nonlinear dynamics	26
1.9.12 Game theory.....	26
1.10 Overview of Financial Markets	26
1.11 Market Biases	29
1.11.1 Objective and subjective reasoning	29
1.11.2 Science versus philosophy.....	30
1.11.3 Chaos versus random behaviour.....	31
1.11.4 Certainty versus uncertainty	33
1.11.5 Value and momentum	34
1.11.6 Extreme optimism and irrational exuberance	36
1.12 Market Perceptions and Perspectives	37
1.12.1 Stock market crashes.....	38
1.12.2 Long-term versus short-term expectations	38
1.12.3 Factors influencing stock markets	39
1.12.4 Risk prediction	39
1.12.5 Business cycles	40
1.12.6 Monetary policy	40

1.12.7 Fiscal policy	41
1.12.8 Thematic versus index investing	41
1.13 Delimitations of Scope and Key Assumptions, With Justifications.....	42
1.14 Outline of the Report	42
1.15 Conclusion	43
Chapter 2: Literature Review	44
2.1 Introduction	44
2.2 Networks, Spillovers, and Portfolios.....	44
2.2.1 Network Spillover Effects using Statistical Models.....	45
2.2.2 Network Spillover Effects and Risk between Asset Classes.....	53
2.3 Networks and Statistical Methods.....	60
2.3.1 Networks using Correlation Techniques.....	61
2.3.2 Networks, Machine Learning and AI Techniques.....	65
2.4 Networks and Complex Systems.....	66
2.4.1 Networks and Financial Shocks.....	66
2.4.2 Networks, Clusters and Codependence in Financial Markets.....	68
2.4.3 Networks Constructed from Other Scientific Methods.....	69
2.4.4 Network Modelling for Various Data Sources.....	71
2.5 Networks, Causality, and Various Financial Markets of Countries	78
2.6 Research Propositions	82
2.7 Research Gap Identification and Conclusion	84
Chapter 3: Research Methodology.....	85
3.1 Introduction and Structure of the Research Methodology	85
3.2 Motivation, Problem Statement, Hypotheses, and Research Questions.....	86
3.3 Research Philosophy	88
3.4 Data	89
3.4.1 Sector composition of the markets	90
3.4.2 Overview of asset classes involved in the exchange-traded fund data set.....	91
3.4.3 List of all exchange-traded funds	92
3.5 Data Preprocessing	98
3.6 Instrumentation and Tools	100
3.7 Ethical Considerations	100
3.8 Data Analysis Model.....	100
3.9 Network Theory Metrics – Definitions	101
3.9.1 Betweenness centrality	101
3.9.2 Closeness centrality.....	101
3.9.3 Clustering coefficient	102
3.9.4 Harmonic centrality.....	103
3.9.5 PageRank algorithm	103
3.9.6 Eigenvector centrality	103
3.10 Portfolio Measures – Definitions	103
3.10.1 Markowitz Efficient Frontier.....	104
3.10.2 Sharpe ratio	104

3.10.3 Beta.....	105
3.10.4 R-squared	106
3.10.5 Systematic versus unsystematic risk.....	106
3.10.6 Skewness and excess kurtosis	106
3.11 Models and Elaborations.....	107
3.11.1 Method and model 1: Feature engineering method using genetic algorithms.....	107
3.11.1.1 Detailed presentation of model 1 – genetic algorithms.....	108
3.11.1.2 Genetic algorithms.....	109
3.11.2 Model and method 2: Feature engineering method using Naïve Bayes	111
3.11.2.1 Introduction and establishment of the methodological connection	111
3.11.2.2 Detailed representation of model 2 – Naïve Bayes	112
3.11.2.3 Naïve Bayes.....	113
3.11.3 Model and method 3: Feature engineering method using neural networks.....	114
3.11.3.1 Introduction and establishment of the methodological connection	114
3.11.3.2 Detailed representation of the neural network model.....	114
3.11.3.3 Neural networks	116
3.11.3.4 Evolutionary search	117
3.12 Conclusion	117
Chapter 4: Results and Findings.....	118
4.1 Model and Method 1: Results and Findings From the Attribute Selection Method – Genetic Algorithms.....	118
4.1.1 Results and findings from the network measures	120
4.1.2 Allocations based on the three portfolios of the top 20 stocks based on network measures	125
4.1.3 Results and findings from the portfolios constructed using network measures for the top 20 stocks based on network metrics – betweenness, closeness, and PageRank.....	126
4.1.4 Allocations based on the three portfolios of the bottom 20 stocks based on network measures	129
4.1.5 Results and findings from the portfolios constructed using network measures for the bottom 20 stocks based on network metrics – betweenness, closeness, and PageRank.....	130
4.2 Model and Method 2: Results and Findings From the Attribute Selection Method – Naïve Bayes.....	132
4.2.1 Results and findings from the attribute selection method – Naïve Bayes	132
4.2.2 Results and findings from the network measures	134
4.2.3 Allocations based on the 3 Portfolios of the top 20 stocks based on Network Measures.....	137
4.2.4 Results and findings from the portfolios constructed using network measures for the top 20 stocks based on network metrics – betweenness, closeness, and PageRank.....	139
4.2.5 Allocations based on the three portfolios of the bottom 20 stocks based on network measures	141
4.2.6 Results and findings from the portfolios constructed using network measures for the bottom 20 stocks based on network metrics – betweenness, closeness, and PageRank.....	142
4.3 Model and Method 3: Results and Findings of Attribute Selection Method – Neural Networks.....	145
4.3.1 Allocations based on the three portfolios of the top 20 stocks based on network measures	149
4.3.2 Results and findings from the portfolios constructed using network measures for the top 20 stocks based on network metrics – betweenness, closeness, and PageRank	151
4.3.3 Allocations based on the three portfolios of the bottom 20 stocks based on network measures	154
4.3.4 Results and findings from the portfolios constructed using network measures for the bottom 20 exchange-traded funds based on network metrics – betweenness, closeness and PageRank.....	155
4.4 Results and Findings Compared With the Portfolios Constructed Using the Three Models of Genetic Algorithms, Naïve Bayes, and Neural Networks	157
4.5 Conclusion	160
Chapter 5: Conclusion	161
5.1 Introduction	161

5.2 Research Process.....	161
5.3 Validation Feedback Received From Experts From Industry and Academia	163
5.4 Hypotheses.....	166
5.4.1 Hypothesis 1.....	166
5.4.2 Hypothesis 2.....	166
5.4.3 Hypothesis 3.....	167
5.4.4 Hypothesis 4.....	168
5.5 Research Contributions.....	169
5.5.1 Contribution to Theory.....	169
5.5.2 Contribution to the Literature.....	170
5.5.3 Contribution to Practice	172
5.6 Implications and Recommendations Based on the Findings.....	173
5.7 Limitations and Scope for Future Research	175
5.8 Conclusion.....	176
References.....	178
Appendix 1: Initial Research Application for Ethics Approval.....	200
Appendix 2: Research Application for Ethics Approval at Completion	208
Appendix 3: Research Ethics Committee	216
Appendix 4: Suggestions /Comments are given by experts	219
Appendix 5: Important Portfolio Characteristics of the 18 Portfolios	222
Appendix 6: Turnitin Report	266
Appendix 7: Proofreader Certificate	267

Declaration.....	5
Allen, F., & Gale, D. (2000). Financial contagion. <i>Journal of Political Economy</i> , 108(1), 1–33. https://doi.org/10.1086/262109	179
Bardoscia, M., Battiston, S., Caccioli, F. et al. (2017). Pathways towards instability in financial networks. <i>Nature Communications</i> , 8, 14416. https://doi.org/10.1038/ncomms14416	180
Gai, P., & Kapadia, S. (2010). Contagion in financial networks. <i>Proceedings of the Royal Society of Mathematical Physical and Engineering Sciences</i> , 466. https://doi.org/10.1098/rspa.2009.0410	185
Glasserman, P., & Peyton Young, H. (2016). Contagion in financial networks. <i>Journal of Economic Literature</i> , 54(3), 779–831.....	185
Grilli, R. Iori, G., Stamboglis, N., & Tedeschi, G. (2017). A networked economy: A survey on the effect of interaction in credit markets.	186
Iori, G. & Porter, J. (2012). <i>Agent-based modelling for financial markets</i> (Working paper 12/08). Retrieved from Department of Economics, City University London. https://openaccess.city.ac.uk/id/eprint/1744/1/iori_porter_2012.pdf	188
Iori, G., & Mantegna, R. (2018). Empirical analyses of networks in finance. Handbook of Computational Economics, Elsevier, Volume 4,Pages 637-685, https://doi.org/10.1016/bs.hescom.2018.02.005	188
Kim, M., & Sayama, H. (2017). Predicting stock market movements using network science: An information theoretic approach. <i>Applied Network Science</i> , 2, 35.	189
Maman, A., & Djauhari. (2012). A robust filter in stock networks analysis. <i>Physica A: Statistical Mechanics and its Applications</i> , 391(20), 5049–5057. https://doi.org/10.1016/j.physa.2012.05.060	191
Namaki, A., Shirazi, A. H., Raei, R., & Jafari, G. R. (2011). Network analysis of a financial market based on genuine correlation and threshold method. <i>Physica A: Statistical Mechanics and its Applications</i> , 390(21–22), 3835–3841. https://doi.org/10.1016/j.physa.2011.06.033	193
Somin, S., Altshuler, Y., Gordon, G. (2020). Network dynamics of a financial ecosystem. <i>Scientific Reports.</i> , 10, 4587. https://doi.org/10.1038/s41598-020-61346-y	196
Suggestions /Comments are given by experts and response/corrective action taken.....	220
Dr Ramamurthy Venkatesh.....	221

List of Figures

Figure 1.1: A visualisation of open systems	21
Figure 1.2: Complex adaptive systems	24
Figure 1.3: Standard and Poor's 500 historical annual returns from 1920 to 2021	27
Figure 1.4: China's gross domestic product GDP in numbers	27
Figure 1.5: BSE Sensex price	28
Figure 1.6: United Kingdom Stock Market Index	29
Figure 1.7: Random versus deterministic systems.....	32
Figure 1.8: Frameworks of comparisons between systems.....	33
Figure 1.9: Performance of factor strategies from 2004 to 2020	35
Figure 1.10: Anatomy of a typical bubble (source: Jean-Paul Rodrigue)	36
Figure 1.11: Business cycle	40
Figure 1.12: Outline of the report.....	44
Figure 2.1: Outline of the LR gap identification.....	84
Figure 3.1: Flowchart and structure of research methodology	86
Figure 3.2: Research philosophy design of the Project (by the author)	90
Figure 3.3: Stationarity versus non-stationary time series	99
Figure 3.4: Flowchart of the data analysis structure	101
Figure 3.5: Author's model using genetic algorithms	109
Figure 3.6: Flowchart of a genetic algorithm (Selivanov, Poezjalova, & Gavrilova, 2014)	110
Figure 3.7: Model of the networks using Naïve Bayes.....	113
Figure 3.8: Model of the networks using neural networks	115
Figure 4.1: Top five exchange-traded funds and their causal relationships.....	118
Figure 4.2: Clustering coefficient metric applied to the network chart	123
Figure 4.3: PageRank algorithm applied to network chart	125
Figure 4.4: Betweenness centrality applied to the network chart	136
Figure 4.5: PageRank applied to the network chart – Naïve Bayes.....	137
Figure 4.6: Betweenness centrality applied to the network chart – neural networks	147
Figure 4.7: Clustering applied to the network chart – neural networks	148

Chapter 1: Introduction

1.1 Introduction

The complexity and competitiveness of financial markets have consistently increased since the dawn of the contemporary computing era. Growing market complexity, such as high-frequency trading and cryptocurrencies, has been brought about by increasing technological advancements, particularly in the financial sector. These market complexities have all contributed to market volatility and new regulatory regimes. Machine learning has recently gained popularity due to its ability to predict outcomes in various business case applications and other fields. Financial markets are intricate systems with various characteristics affecting price shifts and regime changes. The ability to differentiate our thinking due to our understanding of the complexity of financial markets can help us design models in a competitive way and where alpha production becomes the most crucial component for a fund or manager. A market thinking model that frequently looks for, interprets, and derives diverse insights from the models that are generated emerges with complexity. This kind of thinking enables us to undertake predictive analyses of stock prices using a variety of variables and understand the intricate system of financial markets. Intuition has always played a significant part in our sense of market values, and it is mostly based on probabilities in financial market modelling.

It is now simpler to create prediction models with a more intuitive understanding, thanks to the rise in popularity of machine learning and artificial intelligence (AI). Details of stocks are analysed in greater detail and with higher accuracy using logic, probability, and forecast insights. A good model frequently includes logic, probability, and the ability to access several alternatives, which can provide an in-depth manner to examine and interpret various results. Applying the notion of complexity to financial markets aims to integrate numerous concepts, theories, and methodologies. Theoretical support aims to address the fundamental question of why prices fluctuate and propose solutions. In many aspects, network science models can offer a profound and full understanding for risk and return; in other words, how risk affects returns in a portfolio allocation process. Understanding risk and return from a portfolio selection process is crucial, since one aims to maximise and the other to eliminate differences.

1.2 Background of the Research

The research analyses various investment options, and the motivations for taking risks and being risk-averse are frequently important. All investments, regardless of their composition – whether in equities, commodities, or bonds – have a certain amount of risk that must be described by a powerful paradigm that comprehends both correlation and causality. Grasping how risk and return will affect the other depends critically on our understanding of correlation and causation. In order to forecast future predictions, it is crucial to comprehend what-if scenarios. Considerations include what will happen to x if y has an impact and vice versa when considering two relationships. When addressing uncertainties and potential opportunities in the future, this way of thinking helps us deal with probability.

A probabilistic framework is vital and frequently creates the groundwork for approaching financial markets using probability-based predictive assessment, preventing intuition alone from dictating our views or thinking about financial markets. The understanding of financial markets becomes more intuitive and predictive as a result. The theoretical foundation for financial markets is the idea that these complex systems may be predicted, and models can be built by connecting numerous correlations between various variables. With an unpredictable future, it is challenging to precisely assess the price of a specific stock or commodity. Future price prediction is frequently a contentious topic that looks to the past to estimate the future. However, prediction accuracy significantly increases when research is conducted to probabilistically deduce a model.

The overall goal of this essay is to provide a theory of markets that recognises the complexity of markets and comprehends the multiple linkages between the various variables. The theory itself will explain why using a variety of approaches and a coordinated strategy improves our knowledge and perspective of the markets from a network theory perspective. A market thinking model is absolutely necessary because prices fluctuate for various reasons, and considering each would be exceedingly time-consuming and difficult. However, modelling a few variables can assist us in identifying significant variables that may have an impact.

1.3 Research Problem

The primary issue raised in this study is that efficient market portfolio constructions of financial markets using prior methods only took into account correlation, other time series techniques, and other techniques and did not fully utilise network science as a means for portfolio construction and allocation by utilising network metrics. The goal of this research is to utilise the attribute selection problem for network building by employing machine learning and AI algorithms. Using the

attribute selection process, it is also possible to successfully build robust portfolios based on the established network's network measures. The researcher has a strong conviction that the attribute selection method, employed as a tool for link prediction, can offer a unique method for creating networks that are reliable enough to create portfolios.

The stock selection or rating approach, which is crucial for creating strong long-term portfolios for market analysis, was not clearly identified or highlighted by prior methodologies utilised for portfolio construction. In order to build a thorough portfolio development process, it is important to utilise a significantly more modern approach that incorporates networks, machine learning, and AI techniques. The process of building all these networks and the convergence of AI and machine learning to the primary framework of portfolio analysis can offer a fresh perspective on understanding portfolios.

The primary problems of the study methodology are that it mainly relies on correlation and network formation. Even if newer techniques for building these networks are available, they do not offer a reliable answer to the portfolio allocation problem. Some of the most intriguingly distinctive components of the overall framework of the modern portfolio theory (MPT) are the allocation and creation of portfolios. Correlation, one of the most prevalent themes for building networks, is one of the ways used to construct portfolios that are not robust enough. The majority of studies focused on correlation prior to the current framework. However, we cannot objectively evaluate or assess the significance of the entire network structure of exchange-traded funds (ETFs) from a dynamic and complex viewpoint without adding causality or, at the very least, knowing the causative perspective of the individual assets.

Our main contributions primarily address the larger gap in understanding, namely the notion that portfolios can be built using a causal perspective and the integration of AI and machine learning into the framework for building networks and comprehending link prediction. Neural networks, genetic algorithms, and Naïve Bayes are the approaches employed in the models. For time series models and data, link prediction can offer a distinctive technique to evaluate and construct interrelated perspectives of diverse networks and their causal properties. This makes link prediction a promising and interesting new area of research. This knowledge can offer an intuitive way to simulate market dynamics, can help build vital portfolios, and may aid in overall allocation, which can be helpful for investors to prepare for various market scenarios. By employing an easy method for feature selection, feature engineering, and the analysis and creation of these portfolios,

our approach offers an engaging framework for understanding network construction, ranking, and selection of ETFs. These advanced techniques offer a fascinating new way to build portfolios with minimal risk and excellent risk-to-return characteristics, which can assist investors in various market conditions, particularly during the bear market of 2022.

1.4 Purpose of the Research

One limitation of previous research has been the focus on correlation models for better network construction for Network Construction. Even the conventional pricing model, which is based on the efficient market hypothesis, is unable to grasp the complexity and dynamic character of asset or stock values. At this point in the modelling process, we need models that can help us understand the features of correlation and causality more thoroughly. According to the current paradigm of the MPT, the variance of a portfolio is used to examine and evaluate it. Additionally, the MPT model does not present an accurate depiction of the genuine stock selection process using networks which involves capturing non linear intermarket relationships between the various asset classes. Studying this is beneficial, since it is crucial to understand how relevant stock price swings are to the overall modelling process. The MPT, which solely takes historical data into account, does not consider market swings that are observed under varied market conditions. The MPT model does take into account the fact that all investors are risk-averse and knowledgeable and that the projected returns can be realistic in terms of their expectations. It is crucial to take into account the emotional components of financial markets. To model financial markets in terms of asset allocation and portfolio management, nonlinear approaches are crucial. The majority of models used in earlier studies primarily highlighted correlation models as a framework for creating complex networks of financial markets. Because allocation to multiple assets can broadly spread returns over a lengthy period, diversification is also a highly essential factor in creating these models. The researchers were able to objectively evaluate the critical elements required for developing a strong portfolio when designing models using networks. The stock selection process, which will be necessary for us to develop the model, are crucial. Previous studies have not consistently stressed how crucial it is to account for the effects of diversification in asset allocation models.

Examining why there has already been some work in the fields of network science, machine learning, and AI is necessary, particularly with regard to portfolio allocation and selection, in order to overcome the limits of effectively building a portfolio selection model. When conducting and studying research, the majority of the work has been on creating networks based on stocks and

attempting to extrapolate facts or information from these networks. However, the distinctive aspect of this study is that we are looking for a connection between creating these networks and being able to use them successfully for portfolio management. Over the past 20 years, one of the most intriguing areas of finance research has been portfolio allocation and selection. A new set of ideologies and viewpoints on market pricing and their models have been developed, particularly due to new technology and increasing knowledge in many fields, such as computer science and engineering. There are a few areas where network science and related measurements can be used to allocate portfolios. We can also use correlation-based algorithms to assign weights to different assets, such as stocks, bonds, commodities, and other financial instruments. This study focuses on the key components of comprehending market prices and the technical skills required to create a more reliable portfolio model.

1.5 Aim of the Study

The purpose of this study is to evaluate the significance and utility of various ETF networks built using the attribute selection approach to creating a strong portfolio, along with portfolio weightings, by employing the ranking of network measures. The objective is to see whether these strategies help create reliable market portfolios by looking at the relationships between various ETFs within the financial markets of different countries.

1.6 Research Significance

This study is significant because it raises the prospect of creating a model and framework that examines the market from a connected viewpoint. Studying markets from various angles provides a more realistic framework to comprehend and lay the groundwork for portfolio analysis and selection, which is why it is crucial. The MPT and other theoretical frameworks have previously been the focus of research, but they do not adequately capture the dynamics of price changes and redundancies. This is significant because any research must examine these markets from a distinctive and intricate viewpoint that considers and investigates markets as complex dynamical systems. Many methods, including machine learning, AI, and other contemporary nonlinear methods, are used as the framework in this research.

This research will assist in developing a new approach and framework that can be used to provide portfolio weightings for different equities in a scenario with a varied portfolio. It begins by examining these markets and concentrating on the nuances of risk and return from a distinct and varied viewpoint. It investigates whether this approach can be used to infer stock prices and other

asset values. From this framework, a solid model that aids in bridging the risk–reward divide can be created. Weights for stocks and other assets in a portfolio can be determined using network science metrics, such as betweenness centrality and closeness centrality.

1.7 Research Methodology

The main research question or problem addressed here involved studying the various methods and methodologies of using network metrics for stock or asset prices. This researcher rigorously attempted to establish a pattern of connections between stocks within a network. The data analysed involved secondary data. It was important to look at secondary data because we were trying to analyse different price and volume metrics for various stock or asset prices. The data chosen were broad and covered many markets, such as United States (US) ETFs, with a broader focus on various asset classes, including bonds, stocks, and commodities, and various sectors.

1.7.1 Data and research framework

The data collected for the research included secondary data on stock, ETF, and commodity prices from US ETF markets from diverse perspectives. The data were collected from Yahoo Finance for the various assets. The sample size of the data used was approximately five years of data for the period 2017–2022. The data also included daily timeframes.

Steps Involved in the Network Process	Type of Data	Source of Data	Method of Data Collection	Method of Data Analysis
Assess the key factors involved in selecting various attributes	47 exchange-traded fund data items	Five years of data: daily data from Yahoo Finance	Secondary data: stock price data as quoted by the New York Stock Exchange and collected by Yahoo Finance	1. Genetic algorithms 2. Evolutionary algorithms using Python
Assess the selected attributes for network building	Selected attributes with names	Selected attributes using a feature selection process	Python used for the final results	May vary depending on network size (each network can have different, more, or fewer attributes selected)
Assess the importance of various network measures	Various network measures data	Network data measures from Python	Python used for the data collection of network measures	Ranking the top 20 and bottom 20 based on the

				network measures
Assess the various factors involved in the portfolio construction process	Ranking of network measures based on betweenness, closeness, and PageRank	Network data measures from Python	Microsoft Excel used to rank the metrics	Betweenness, closeness, and PageRank
Determine the risk and return characteristics of the portfolio	Risk-to-return metrics	Portfolio Optimization in Portfolio Visualizer	Portfolio Visualizer used	Markowitz Portfolio Optimization

Table 1.1: Table of Research Design

Understanding investor and trader dynamics and behaviours was the primary goal of using the daily period. To assess the validity of the network theory and its metrics, such as betweenness and closeness, and their correlation properties, it was necessary to investigate a variety of markets, including the Asian, European, and US markets, by combing data from ETFs. The data performance metrics were thoroughly examined for any missing data. This study is an example of experimental research, which uses quantitative methodologies to quantify a particular result. Quantifying the source of the measurement played a crucial part in determining how successfully our theory was supported or disproved, which was the significance of the experimental study. The basic idea behind experimental research is that it attempts to satisfy tests that validate or invalidate cause-and-effect linkages between various model variables. This is crucial, especially with regard to stock or equity prices. Since numerous factors could have affected how a stock price moved, our major goal was to define and quantitatively measure these factors.

1.8 Literature Review

Based on our research questions and research problem, the most significant themes were found to concentrate mostly on four primary features after carefully analysing the literature. The first aspect focused on the investigation of network construction techniques, particularly in relation to portfolio and spillover effects. The use of networks and statistical approaches that combine correlation and other statistical models for the purpose of establishing networks was the second major topic of the literature review. The third component of the literature study emphasised the connections between networks and complex systems, as well as the many methods used in the current framework for complex systems. The impact of network causation and the examination of numerous financial markets from various nations were the final two areas of the research. The identification of the research gap concerned the impact of knowing the methods that have been

employed in the past and the data that have been used, particularly in relation to the types of marketplaces and data sets. The technique utilised to predict and comprehend these ETF relationships is one of the primary research gaps and the subject of this study. The focus on ETF markets with a diversified and globally oriented approach, focusing on various markets, including crude oil, gold, global stock markets, global bond markets, commodity markets, and global sectoral indices from an interconnected holistic perspective, is another research gap identified in this paper. Another gap that the researcher tried to address is that the network research on financial markets has not placed a strong emphasis on the portfolio component and the creation of portfolios, which requires a careful examination of the risk and return criteria. There has never been a reliable technique or framework for deploying networks in a formal, research-oriented way using the appropriate methods for understanding the profitability of such network portfolios. In the end, the research gap is concentrated on the complexity approach, which emphasises using neural networks, genetic algorithms, and other key techniques to understand the framework for complex system optimisation and using evolutionary algorithms as a search technique for many attribute selection methods.

1.9 Definitions

Complexity theory is a branch of mathematics concerned with the study of complex systems, often characterized by non-linear dynamics and emergent behaviors. It seeks to understand the behavior of such systems by exploring the relationships between their individual components and the patterns that arise from their interactions. Complexity theory has applications in a wide range of fields, including computer science, physics, biology, and economics. Its insights have helped to shape our understanding of the natural world and the complex systems that underpin it.

1.9.1 Complexity theory overview

According to complexity theory, the world can be studied via the force of interactions, in which each component in the universe interacts with every other component in a dynamic global context. This ‘new science’ is frequently referred to as complexity science by Ma and Osula (2011). According to the Sante Fe Institute who are the leading research organisation on complexity theory, defines complexity science, which is also known as complex systems science, studies how a large number of components can spontaneously self-organise to display nontrivial global structures and behaviours and large scales, frequently without the assistance of central leaders or authorities. A complex system is made up of these groups of interactions and interdependencies

because of its complexity and unpredictable nature. It is crucial to remember that complex systems are challenging to anticipate and that obtaining a complete understanding of a system frequently calls for original and imaginative models.

Globalisation, innovation, and global rivalry are driving innovative ways of reinventing organisational growth, where complexity science is being increasingly applied (deMattos, Miller, & Park, 2012). The markets' interconnection on a worldwide scale is referred to as a complex system (Lavin, Valle, & Magner, 2021). Cross-border relationships and interdependencies are being discussed. Studying complexity theory requires consideration of seven factors: interactions, emergence, dynamics, self-organisation, adaptability, interdisciplinary, and complexity techniques.

1.9.2 Systems

The primary objective of complexity theory is the study of systems. A system as a whole can be described as an entity that creates a certain interaction, relationship, or reliance on other system components. These components interact within the system's boundaries.

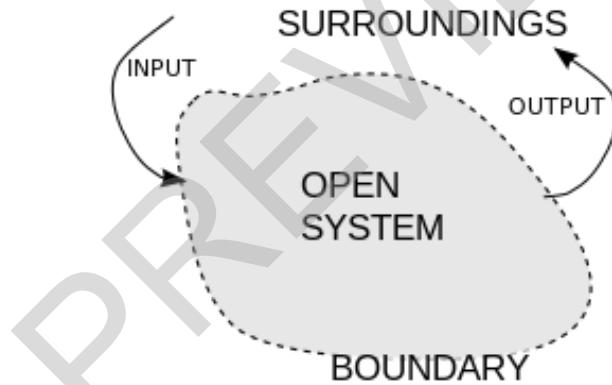


Figure 1.1: A visualisation of open systems.

Systems theory includes a subcategory called complex system research. Systems theory examines how complex systems generally interact behaviourally. It has made a significant contribution to studying interactions and dependencies between system components on a wide scale.

It is crucial to remember that all objects outside a system are referred to as the environment or surroundings of the system. Other similar systems also display many interactions with other systems, creating a shared understanding of these interactions.

A system that receives feedback has an input mechanism for feedback from the environment and an output mechanism to provide feedback to the environment. Systems frequently display

distinctive characteristics, which may result in distinctive behaviours that are particular to a system. The interactions and behavioural patterns of systems are what cause their traits to manifest in their surroundings. In order to link these behaviours with processes over time, time is crucial element that plays an important role in understanding the system behaviours and process in a continuous manner.

1.9.3 Interactions or networks

The world around us is made up of complex networks, thanks to the many interconnected complex systems that interact with one another. High-order interactions and the likelihood that high-order interactions will cause collective behaviour were investigated by Battiston et al. (2021). Interacting neurons in the human brain, a network analysis of the internet and webpages, social networks of relationships and friendships, the structure of and interdependencies and interactions between global financial networks and economic and banking structures are examples of interactions in complex systems.

1.9.4 Emergence

It is challenging to comprehend or forecast the characteristics of the whole, and emergence refers to the knowledge that is gained from comprehending these components and phenomena. The emergence phenomenon aids in the comprehension of interacting dynamics in complex structures and broad-scale behavioural patterns. Much larger scales are used to model the structures. In other words, the expression the whole is more than the sum of its parts can be used to describe emergence. Gene regulatory networks are used as a study to understand the actual phenomena of Emergence while focusing on the emergent behaviour of complex systems (Zheng, 2021). The study also assesses how important network structure is in producing collective oscillation. Examples of emergence in nature include the construction of a tornado by molecules, the cellular structure of a living being, the function of neurons, and the interacting qualities that give rise to consciousness, intellect, and awareness.

1.9.5 Dynamics

Because of the dynamic nature of the system's states, which frequently change, it can be described as the process of being unable to forecast the system's long-term behaviour. The ideas of linearity and nonlinearity are crucial for understanding the dynamics of the system. It can be measured as a variable that alters over time in a linear and orderly manner while maintaining a consistent change. Long-term forecasting of a linear system's states is possible. By contrast, nonlinear systems have

a large number of states that alter rapidly. Although prediction over time becomes complicated, nonlinear systems can also have cycles where changes can be understood over time. For instance, prices on the stock market may shift based on the state of the market from a trending phase to a stable phase or a declining phase. Long-term stock market predictions are challenging due to these sudden phase transitions and complex dynamics. Systems can have bifurcations, phase transitions, or what are referred to as tipping points, which can alter the behaviour of the system over time. In other words, even minor system perturbations can dynamically change the system's state, making it highly challenging to model. Some systems can be exceedingly unpredictable and chaotic, which is frequently referred to as the butterfly effect. The weather system, stock markets, and the complex economic framework system are examples of real-world systems that display dynamics-related properties.

1.9.6 Self-organisation

Distinctive and intriguing pattern formations, which can be referred to as self-organisation, are global patterns of behaviour in complex systems that can be understood by the interactions between their components. Through their interactions with other system components, the control mechanism of a self-organising system is dispersed and integrated over the entire system. Global patterns of behaviour form over time and on a large scale as a result of the complex dynamic nature of interactions, raising the risk of more complexity entering the system and resulting in a more ordered and organised system. Systems that can holistically establish a balance between randomness and regularity as they organise themselves into a critical state. These crucial state patterns have a tendency to produce distinctive self-similarity and power law distributions. Examples of these include the formation of an organism from a single egg, the expansion of cities and their populations, and the flocking behaviours of birds.

1.9.7 Adaptation

One of the most crucial aspects of complex systems is evolution and adaptation. Complex systems actively respond to dynamic environmental changes and adapt to them by moving forward gradually and quickly. The process of adaptation can be carried out using various techniques and scales. It is possible from a cognitive standpoint through ongoing learning and psychological development. From a social standpoint, it is possible through social networks by using knowledge and information about its connections. It can be accomplished via the processes of genetic variation and natural selection from an evolutionary perspective (Holland, 1992). The process of adaptation

inside a system can be very quick, or it can be started slowly so that the system can heal, its components can be revitalised by a better process, and it can adapt to its surroundings quickly.

This level of robustness can be achieved because of a systems ability to withstand perturbations, resilience, and ability to bounce back to original levels even after a massive effect or a downfall within the system – complexity scientists often call this phenomenon the largest perturbation. Adaptation simply means the ability to survive and be able to replenish and rejuvenate even after a large or dissipating perturbation. Some complex systems with these properties of perturbations can be called complex adaptive systems (Sole, Ricard, & Elena, 2018). Examples of adaptation include an immune system learning about various pathogens and a stock market that recovers after a crisis or market crash.

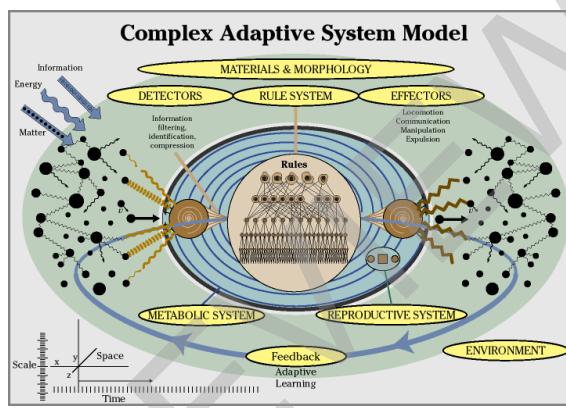


Figure 1.2: Complex adaptive systems.

1.9.8 Interdisciplinary Dynamics

Complex systems' interdependence and distinctive characteristics make them valuable for managing a variety of systems across various disciplines. It is feasible that many qualities from various systems could be connected or share natural properties. This is advantageous because it can be used in various contexts, and knowledge can be deduced from these models in a unique way. (2018) Stefan Thurner Complexity science is currently being used in various fields, including biology, social sciences, psychology, business and management, information technology, and finance. AI, machine learning, deep learning, and autonomous self-driving cars are just a few of the many modern technologies that use complicated technology to build complex system ecosystems and develop emergent features inside those ecosystems. It is also critical to stress that universality is a phenomenon that arises in complex systems. Because many of these systems may exhibit common patterns or qualities, universality indicates that many use cases could benefit from