CHAPTER 1

Introduction

Contents

The main purpose of this book is to provide a broad overview of how to build financial networks used in investment management. In general, finance is a discipline in which all aspects of real-life networks can be found. Collaboration and exchange of services, financial transactions, and contractual obligations represent connections. These connections and the interacting entities can be considered as networks. Finance is not a system that resides in a state of chaos but is characterized by organized complexity. These networks represent the underlying structures and organization. Applying statistical, econometric, and mathematical models when building financial networks helps reveal the often-hidden underlying architecture of financial markets.

In this chapter, we provide a comprehensive introduction to the various types of networks and their relevance to financial markets. We begin with a general overview of networks, covering technological and infrastructure, in addition to information, social, and biological networks. The chapter emphasizes the unique characteristics of financial networks, particularly their proprietary nature, which necessitates specific modeling and estimation techniques.

The chapter further describes how different network properties, such as dyads and triads, apply to financial markets. It explains the importance of understanding the interactions among economic variables, individuals, and financial institutions to capture and explain complex market behaviors. The chapter focus then shifts to data, detailing the types of datasets used in network analysis within finance, including assets and factor returns, hedge fund and mutual fund returns, global equity market returns, and country bond market index returns.

Finally, the chapter delves into the application of network theory in asset management. It discusses how networks can be used for risk and portfolio management, data analysis, trading, risk management, and performance evaluation. The chapter concludes with a discussion on the design of financial networks and the key takeaways from understanding and applying network theory to finance.

NETWORK THEORY, GRAPH THEORY, AND TOPOLOGY

Network theory is a field of mathematics with roots tracing back to the works of some of the world's most famous mathematicians, such as Leonhard Euler, Bernhard Riemann, and Henri Poincaré. It offers a framework for understanding and analyzing the relationships and interactions among interconnected entities. This theory extends the graph theory principles, founded by Euler's solution to the Königsberg bridge problem.[1] Graph theory involves studying graphs (or mathematical structures) used to model pairwise relations between objects. In graph theory, graphs consist of vertices (or nodes) and edges (or links) that connect pairs of vertices. Euler's work laid the groundwork for network theory by introducing these basic components and demonstrating the connection between topology and network theory.

Network theory builds on these concepts to study more complex and large-scale networks, which can be found in various fields that will be discussed later in this chapter. It looks at how nodes (which can represent anything from people in a social network to computers in a technological network) are interconnected and how these connections affect the behavior and properties of the entire network.

Topology, another branch of mathematics, studies how shapes and spaces can change by stretching or bending without breaking apart or sticking together. At its core, topology is a study of how objects are connected, shaped, and arranged in space. Topology shares similarities with network analysis, especially in understanding how different elements are connected and how these connections affect the overall structure (i.e., the arrangement and organization of the elements) and behavior (i.e., how the system operates, reacts, or changes based on these connections).

WHAT IS A NETWORK?

According to Wasserman and Faust (1994), nodes often represent actors in social sciences. The term nodes and actors can be used interchangeably in network science because actors represent social (investors, individuals, corporate or institutional units, countries, and economic entities) that interact. Interaction might be a transfer of goods and services, physical connection, relation, association, or any type of material or non-material movement. When interacting, actors are linked by ties, edges, or links. The basic properties of a network, denoted by $G = (n,m) \equiv G(V,E)$, include the number of nodes (or vertices) $n(V)$ and the number

[1]The Königsberg bridge problem asked if it was possible to walk through the city of Königsberg and only cross each of its seven bridges exactly once. Euler's clever solution introduced the idea of using points, called vertices, to represent landmasses, and lines, called edges, to represent bridges. See Newman (2010).

of links (or edges) (E), where $n \in \mathbb{R}^n$ and $m \in \mathbb{R}^{n \times n}$ $\left(m_{i,j} \neq m_{j,i}\right)$. Throughout this book, we will use the terms "nodes" and "vertices" interchangeably.

In an *undirected network,* $G^* = (n^*, m^*)$ represents the number of nodes n and links or edges (connections) m, where $n^* \in \mathbb{R}^n$ and $m^* \in \mathbb{R}^{n \times n}$ $\left(m_{i,j}{}^* = m_{j,i}{}^*\right)$. The nodes or actors in a network are connected by relational ties, links, or edges, which we also use interchangeably. The edges between the nodes govern the relations, which might be directional or unidirectional. These relations can arise from social connections, information exchange, biological ties, transportation, financial transactions, economic intermediation, social interactions, physical laws, or other forms of interaction. The term "catallactics," meaning exchange, can be used to describe networks in general. In this book, we will explore various types of relationships within networks,summarized into several major categories.

Increased computer power and data processing capabilities have enhanced personal and collective data recording. For example, data collection by smartphones and apps extends the datasets available for analysis. Data gathering using satellites, web scraping, text analysis, social networks, transactions, market data, and the like has grown tremendously. The term "alternative data" refers to novel methods of gathering data. Financial market research is evolving by applying not only traditional methods of analysis, but also tools capable of analyzing and describing a huge amount of data and possible relationships within the data. A shift from traditional methods of analysis like calculus, differential equations, and regression models to new techniques is inevitable and desirable for investigating underlying economic relationships, as stressed by Focardi and Fabozzi (2012).[2]

While traditional methods and procedures rely broadly on statistical correlations, the era of Big Data and significant computing power necessitates new pragmatic tools grounded in both mathematics and economics to enable efficient financial research. Modern finance is evolving, with financial markets, individual actions, and investor behavior and sentiment continually changing, leading to the emergence of new structures. As new data emerges, patterns, transmission channels, and causality also shift. Graph theory aids in identifying and exploring the underlying structure in the data, revealing relationships and connections between economic players and their interactions.

Network theory is dedicated to the science of connections between objects, and as Focardi and Fabozzi (2012) argued, economics is the science of relations. Therefore, network theory is an essential tool for analyzing relationships among actors, institutions, and instruments in financial markets. Within this framework, the nodes in a graph, applied to financial markets, represent various entities such as asset classes, individual security (i.e., stocks, bonds, and funds), countries, factors, financial institutions, trading partners, derivative contracts, or similar financial entities or instruments. The primary focus of graph theory is to identify relationships between these nodes and model the edges based on robust statistical, economic, and mathematical principles. This book aims to model financial networks and the edges between the nodes, which are typically known, but the links between them are often not observable.

Real-world networks might be observed, but networks in financial markets are mostly model-based as opposed to being sample-based. Having said that, it is worth discussing the common characteristics of the major types of real networks, how their properties relate to financial networks used in portfolio and risk

[2]Shannon (1948), Simon (1962), Haken (1978), Kauffman (1993), Loistl and Landes (1989), Watts (1999, 2003), Wasserman and Faust (1994), Mandelbrot and Hudson (2004), Watts (1999a, 1999b, 2003), Laughlin (2005), Klir (2006), Holland (2012), and Newman (2010), among others, highlighted the idea that there is connectedness in various organizational, biological, informational, financial, social, economic, technological, and physical fields. Their work suggests that reductionist approaches in science should be replaced by more holistic methods that preserve and explain complexity.

management, and what the necessary model specifications are to model such networks. The most interesting and decisive property of real-world networks like biological, informational, technological, or social networks is the underlying flow process of transmitting information between the nodes.[3] In the upcoming sections, we provide details of some of the existing and well-known real networks, highlighting the principles that apply to financial markets and asset management.

Technological and Infrastructure Networks

Technological networks include power grids, infrastructure, industrial, and water supply networks.[4] The most interesting property of such networks is the existence of numerous Micro-Nets, which are locally autonomous networks yet form part of an entire global network. According to Newman (2010), examples of technological networks include the Internet, power grids, telephone networks, and other networks underlying physical infrastructure.

Technological networks, such as subway systems in cities, connect stations of other transport routes. Package delivery networks exemplify networks using the shortest path between nodes for efficient delivery. In these networks, information is not duplicated. Rather, each package or delivery is fixed, with a known receiver, which ensures that the goods arrive at the correct node. Additionally, the information or goods are indivisible. Borgatti (2005) illustrates that the delivery of goods follows the shortest possible path between two points within a given space or network, which minimizes the number of edges transversed (referred to as a geodesic path), ensuring the most efficient route is taken. This is a critical concept for optimizing routes and efficiently connecting points in various types of networks, including transportation, communication, and logistics systems. Minimizing distance is the goal, thereby reducing the time and resources needed to move from one node to another. A similarity in financial markets can be observed in transactions.

Although physical delivery no longer applies to all financial market transactions as it did in the early twentieth century when physical delivery was customary, the financial assets bought by an investor still follow a direct path, arriving from the seller to the buyer. Additionally, financial markets might contain local networks that are part of the entire financial landscape. Consider, for example, a network comprising highly connected private equity managers who build a local network. They might be part of a global network of limited and general partners within their private equity network.

Information Networks and Financial Markets

Many types of information networks exist.[5] The most important and well-known are citation networks and the World-Wide Web (WWW). Whereas the citation network links all available bibliographic works already published and there is a specific order when establishing the links between the nodes, the WWW is a network linking the pages available on the Internet. Newman (2010) stresses that the WWW is different from the Internet, which is a physical network connecting computers. The WWW connects pages, enabling information to travel without any specific order. Among the most interesting examples of information networks are email communications and the spread of electronic information. Emails, for example, can reach

[3]Several of the important works on networks include Milgram (1967), Merton (1968), Rapoport (1968) Freeman (1996), Albert and Barabási (2000, 2002), and Newman (2000, 2002).

[4]Seminal works among others are Watts (2002, 2003), Watts and Strogatz (1998), Amaral et al. (2000), and Almaas (2002).

[5]Some of the important papers on information networks, including citation networks, are Price (1965), Seglen (1992), and Redner (1998). Key papers dedicated to the World-Wide Web are Bornholdt and Ebel (2001); Kleinberg and Lawrence (2001); Albert, Jeong, and Barabási (1999); and Krapivsky and Redner (2002).

numerous places and numerous receivers both simultaneously and concurrently. The information is spread through replication.

The relationship of such networks to financial markets highlights two important aspects of modern social and economic life: information processing and information efficiency as well as the exchange of goods, services, or financial instruments. While information processes refer to how information is processed, the transmission or movement of financial assets involves a different dynamic. As we shall see, information can be copied or replicated, but financial assets are indivisible.

Grossman and Stiglitz (1980) investigated the information process in financial markets and discussed the impossibility of achieving information efficiency. They demonstrated that markets cannot be fully efficient because new information is constantly arriving. As a result, equilibrium can only be momentarily achieved until new information arrives or the decision-making context changes.[6]

Importantly, information travels by replication, functioning as a "copy" mechanism, in contrast to a "move" mechanism. The move mechanism applies to money transfers, where the money transfer or payments have a fixed receiver – the buyer of the goods or services.

Social Networks and Financial Markets

Social networks are perhaps the most compelling types of networks that attract interest due to the rise of social platforms that connect individuals.[7] The nodes in these social networks are called actors and are connected by ties, or links. Where the actors are connected or the basic property of social networks, is the dyad, which represents the connection between two actors establishing a social relationship.

An important characteristic of social relations is the existence of triads, which involve the relationships among three actors. A triad consists of three dyads – one between each pair of actors. Unlike dyads, which are simply connections that either exist or do not exist between two actors, triads are more complex due to their potential directions of connections. With 16 possible configurations, triads enable the modeling of intricate social interactions, revealing patterns and local structures.

Wasserman and Faust (1994), Robins et al. (2007), and Robins et al. (2005) emphasized that in social sciences, local structures shape the global structure. The knowledge of these local structures provided by analyzing triads helps to understand the overall network dynamics. Thus, triads represent the global structure of an actor network.[8] A real-life example of a triad this would be the social interactions between individuals on social media platforms.

Facebook, X (formerly Twitter), LinkedIn, and Instagram are not only social network platforms but also successful companies. However, the underlying mechanism is for two actors to be connected. The major principle is that "friends of my friends are also my friends." Social networks are central to how people are influenced by others. An essential property is that connections coincide simultaneously rather than sequentially from individual to individual. However, social networks also possess the distinctive property of directedness, meaning that the links between nodes have specific directions. This directedness implies that an edge between two nodes does not always translate into an edge between a third node and the other

[6]See Monk et al. (2019).

[7]Newman (2010) argues that social networks are among the most studied networks. Notable examples of network studies refer to corporate business, social, sexual, and individual relationships on social networks including those discussed by Mariolis (1975), Mizruchi (1982), Merton (1968), Wasserman and Faust (1994), Glaskiewicz and Marsden (1978), Newman (2001), Girvan and Newman (2002), and Bramoulle et al. (2009).

[8]The triads are directly related to the clustering coefficient or transitivity of a graph. Transitivity of node connectedness means that when there is an edge between node i and j, and there is also an edge between node i and h, then there is an edge between j and h. See Newman (2010) and Robins et al. (2005).

two nodes. In other words, the principle "friends of my friends are also my friends" is often violated, and a direct connection between nodes may not necessarily exist.

A simple example provided by Konstantinov, Aldridge, and Kazemi (2023) illustrates the difference. Consider the social network Facebook. Once two people are connected, the link is symmetrical in any direction; if person A connects to person B, then the link is symmetrical, and individual B is a connected "friend" of individual A. Alternatively, consider the social networks X or Instagram. An individual might be linked to or following another, but that does not mean the latter person is linked to the former. This principle is not limited to social networks but also applies to financial markets where transactions, orders, and money flows follow a certain direction because of causal principles within a market framework.

Of essential interest are the flow processes in social networks and their application to financial markets. Unfortunately, the principle "friends of my friends are also my friends" does not work in financial markets. To model financial markets influenced by social networks, it is necessary to use a statistical approach that employs probability models. The reason for this is that in financial markets with heteroscedastic volatility, the edge connectivity is associated rather with large degree of uncertainty than depending on trajectories or steady-state conditions. While the set of nodes is fixed, the connections are modeled probabilistically. In this context, the edges in social networks applied to financial markets arise from realizing probabilities related to the "social aspect" of a network. In other words, the interaction between nodes is governed by probabilities rather than specific node and edge properties, known as covariates, trajectories, or attributes.[9]

A flow principle from social networks that might apply to financial markets is information replication, which underlies the influence approach. Influencing financial markets and investors is a phenomenon that can be traced back to the early years of investment history. The result of social influence observed in financial markets is the phenomenon of crowdedness, as documented by Charles Mackay (1980).[10] Convincing investors has a long tradition in financial markets and can significantly influence investment behavior.[11]

Biological Networks and Financial Markets

Biological networks are among the most interesting and widely studied types of networks.[12] Many examples have attracted research, including food networks, ecology, protein–protein interaction networks, metabolic networks, signaling networks, and intraspecies and interspecies networks. The most widely used

[9]The topic of node and edge properties is extensively analyzed in Chapter 9 where the Exponential Random Graph model is introduced and node and edge covariates (i.e., attributes) are discussed and modeled.

[10]The crowdedness results when a mass exhibits the same behavior. Three expressions are used in both the academic and practitioner literature to describe the same phenomenon: crowdedness and herding. Whereas crowdedness and herding apply to the same phenomenon, the origin of crowdedness can be traced back to behavioral economics. An overview of the literature provides an informative picture of this important phenomenon. Holland (2012) defined crowded effects as the decreasing payoff for individuals in the case of an increasing number of players in a strategy. An increasing number of players in a strategy is the consequence of increasing payoff. In such a case, the players are crowding the strategy. The concept of crowding itself is not restricted to specialists.

[11]The literature on the negative impact of crowdedness in financial markets is extensive. Several studies measured the effect of common exposure to the same investment vehicle – fund, style, factor, or asset classes – for equities, for example, Nofsinger and Sias (1999); in trading Cahan and Luo (2013); for volatility Marmer (2010); for alternative risk premia Baltas (2019); for currencies Pojarliev and Levich (2011); and for hedge funds Bussiere, Hoerova, and Klaus (2014); Bayraktar, Doole, Kassam, and Radchenko (2015); and Konstantinov and Rebmann (2020), among others. A positive view of crowdedness is provided by Surowiecki (2005).

[12]Some of the important works include works on protein interactions, biological systems, food webs, and the spread of viruses and diseases. Important works include Jeong et al. (2000), Fell and Wagner (2000), Ito et al. (2001),

biological networks in finance are neural networks. However, the primary purpose of discussing biological networks in the context of the topological structure of financial markets in this book is to understand the underlying information transmission process between the nodes.

According to Borgatti (2005), the process of spreading an infection in biological networks is characterized by duplication. Consider, for example, an infection impacting a specific population. The infection spreads by duplication but does not immediately reinfect the initial nodes because they might have developed immunity. The relevance of infection spread in a biological context to financial markets can be seen in the example of duplication of information such as gossip. Specifically, negative information about a company with an elevated risk of default can be transmitted through the network via news, gossip, company announcements and press releases, negatively impacting holders of that risky asset or company. Consider a portfolio initially impacted and then fully hedged against, say, the depreciation of a specific exchange rate. In this case, the spread of negative information cannot affect that portfolio, to how an infection cannot reinfect immunized nodes.

FINANCIAL MARKETS AND FINANCIAL NETWORKS

After a brief overview of the general types of networks that exist and are detectable in nature, it is logical to focus on the primary type of networks covered in this book: financial networks that model, explain, model, and simulate financial market interactions using time series data. Perhaps the major difference between all other networks and financial markets is that financial networks are often not observable. Structures emerge to stay persistent or transitory based on intrinsic and external conditions. The dynamics of the financial market and financial networks are not governed by nodes in isolation but by the interactions that never cease. Financial markets are an unstable system operating by interactions coined by the arrow of time, whose dynamics operates at the statistical level. Therefore, probabilistic thinking applies to edge formation in financial networks. They must be specified, estimated, or simulated because financial markets generate substantial amounts of hidden or proprietary data. Consider, for example, a stock exchange that generates enormous transaction data in milliseconds, but these data are not available to the public because asset prices are formed by the transactions that occur. In contrast, data availability estimates a bipartite network – a type of network in which the set of nodes can be divided into two distinct groups – possible for scientific publications, citation networks, or actor–movie records.

Once data are gathered, a network can be observed. Financial networks can be easily estimated using various data, such as payments between countries. Generally, the availability of data and directional exchange of information between the nodes makes network construction possible. However, financial market data become increasingly proprietary, and the availability of time series data are one of the most common sources of information.

Financial networks share many properties with biological, technological, informational, social, and industrial networks. Yet, modern finance uses returns derived from asset price data for analytical purposes. This is not surprising, as financial models are applied using time series data. With the rise of computing power, large storage processes, and the ability to analyze large data sets, alternative data are becoming more prevalent in asset management. However, economic variables and portfolio management still use traditional data formats. To remain grounded in traditional financial models that use time series data, in this book we primarily use datasets that comprise time series returns and prices. Consequently, while the

Jeaong et al. (2001), Maslov and Sneppen (2002), Sole and Pastor–Satorras (2003), Huxham et al. (1996), Fath et al. (2007), Dorogovtsev and Mendes (2002), and the seminal work of Kauffman (1993) dedicated to the interaction of individuals and the adapting landscape.

models used and applied in this book are not necessarily restricted to time series data, time series models are central to the analysis.

Nevertheless, networks can be constructed using traditional datasets as well as proprietary data, including the exchange of money flows (e.g., loans), market transaction data (e.g., order flow data), or any datasets reflecting interactions between market participants (see Chakrabarti, Pichl, and Kaizoji [2019] and Lee, Liu, and Lin [2010]). Interaction is the key concept and building principle and block of networks. Therefore, networks can be built using any data form measuring interactions between entities.

In finance, the connections among economic variables, individuals, financial institutions, and their relationships are what matter (see Easley and Kleinberg [2010] and Kalyagin, Pardalos, and Rassias [2014]). However, the collective sum might differ from the pure sum of individual parts. Stated differently, the interconnectedness among all hedge funds – based on their risk factor bets, return and risk exposure, stock holdings, etc. – is greater than the sum of all hedge funds active in the market, as merely accounted for by their assets under management and the number of active and dissolved funds. The interactions among economic variables and entities are essential as market complexity increases. Financial research requires techniques that can capture and explain complex behavior and relationships by focusing on the underlying structure rather than design. The most basic property of networks, including financial networks, is the definition of dyads. As explained earlier, dyads are simply the links or connections between two nodes. Dyadic relationships must be modeled and present a challenging task in financial markets. These node relationships can be directional. For example, they can involve the transfer of money or securities from the buyer to the seller or vice versa. Dyadic relationships can also be bidirectional if they are mutual.[13] For instance, the connectivity might be more abstract, such as the contractual obligation between two banks to settle transactions, which is unidirectional. The specific value and amount of the transaction determine the nature of directionality.

DATA

Behind every network, there is underlying data. Data are the central assets in financial data science, which might belong to financial networks. The purpose of this section is to provide an explanation of the data used in this book and why such data – funds, country indices, local markets, currencies, equities, bonds, alternative investments, and other asset classes and data sources –are used. The main idea is to use data that are representative of a wide range of financial analytical allocation processes. Data should provide more insights into how different networks can be modeled and used in asset management. Whereas data in most real networks are directly interpretable to build a network, in finance data are the backbone of almost all types of analytical undertakings. Time series data are the most widely used in finance, and in this book, we will primarily focus on time series data to estimate networks.

Datasets are necessary to illustrate the versatility of networks for portfolio and risk management. Like other books on networks and asset management, we use broad data to provide a better picture regarding the application of graph theory to asset management. For this purpose, we use datasets comprising stocks, assets, factors, country indices, fund returns, and other data used in financial analysis. This intuition behind using different datasets is versatile. For example, an asset network might be of particular interest to portfolio managers. Similarly, a network constructed from index data, such as the S&P 500 Index – a major equity market index – becomes important.

[13]The following intuition applies consider two points (nodes) in space. They can be connected by a line with an arrow pointing a direction from one point to another. The two nodes can be connected by two arrows, pointing to bidirectional connectedness.

On the other hand, a fund network is of primary interest to fund-of-fund managers, performance analysts, investment officers, and investors. Analyzing the fund network enables decision-makers to visualize and investigate the risk and return-based connectedness of specific funds in which they might be interested in investing. More central funds, or funds in the network periphery, provide different risk factor characteristics that might explain performance behavior. Elements such as risk, returns, fund size, capital transition, and other properties can be used to represent a fund network. Dynamically representing funds as a network can be particularly beneficial to database users, as the databanks change continuously. With funds representing nodes, data vendors can gain insights into where a liquidated, delisted, or newly added fund might appear in the database with its connections.[14]

The use of equity market data serves many purposes. Equity markets are among the most important markets in asset management with mutual funds, pension funds, hedge funds, and other entities heavily invested in them. These markets are perhaps the most dynamically changing markets. Equity market networks might be built from index data, representing the nodes as the individual stocks comprising an index or as country-specific equity markets (i.e., indices) representing global markets, such as the US, German, Japanese, and French stock markets, to name a few.

The data used in this book comprise datasets used in several publications. The motivation behind using different datasets is to provide more detailed and multidimensional examples of assets to which network analysis applies.[15]

In the upcoming sections, we provide details about the different datasets and briefly highlight their importance in representing networks.

Assets and Factor Returns

The dataset used for factors and asset returns is the sample set from the study by Konstantinov, Chorus, and Rebmann (2020). Similar dataset has also been utilized in factor research by Blitz (2018); Aked, Arnott, Bouchey, Li, and Shakernia (2019); and by Konstantinov and Rebman (2020), who investigated the hedge fund market.

It is worth briefly describing the asset and factor dataset in detail. Assets and factors constitute two broad datasets representing different economic and financial frameworks, each comprising 15 factors and 16 assets, spanning from January 2001 to March 2021. While Kritzman (2021) regards assets as primary economic units and factors as portfolio enhancements, the rise of factor investing has empowered portfolio management and the construction of factor portfolios.[16]

Factors have gained traction in asset management, with Harvey, Liu, and Zhu (2016) identifying over 300 factors available in financial research. However, both their statistical and economic significance can be questioned. When explaining equity returns using a multifactor regression model, Fama and French (1992, 1993) found the most prominent factors in the equity space to be market, value, and size. Carhart (1997)

[14]Note that a newly added fund in a database does not represent a newly set up and launched fund, because data providers often have criteria to adda a new fund like minimum track record and other statistics like assets under management. To maintain the backfill bias, which arises when adding previous performance, database providers allow for a newly added fund to a database to add up to certain number of past returns. Another issue with database providers is the survivorship bias, which centers that defunct funds should be considered with their past performance. In a dynamical network context this can be challenging, as funds are the nodes and a node ceasing to exist no longer appears in a network. However, because networks might represent the current stage of a fund network, the survivorship bias is not an issue when considering funds in a network.

[15]Sample datasets used in this book can be downloaded from https://finance-resolution.com/

[16]See Fama and French (1992, 1993, 2005), Jagadeesh and Titman (1993), Moskowitz et al. (2012), Asness et al. (2013), Frazzini and Pedersen (2014), and Ilmanen et al. (2021), among others.

added momentum as an explanatory factor for stock returns, while Fama and French (2015) discovered profitability and investment factors in equities.

We extend these traditional equity factors to include other asset classes available in financial research, such as fixed income, commodities, and currencies. To introduce the concept of categorization, which we will use in chapters in this book to construct networks, the monthly data can be divided into two subsets. The first subset is a pure factor set, widely considered in both academic and practitioner research, and consisting of the following factors:

- The equity premium is the US equity market excess return, denoted by Fama and French (1992, 1993) as Mkt.RF.
- The standard Fama and French (1992, 1993, 2015) factors for value (High minus Low), size (Small minus Big), profitability (Robust minus Weak), and investment (Conservative minus Aggressive) denoted by HML, SMB, RMW, and CMA, respectively. Data are obtained from Kenneth French's data library.[17]
- The one-month–lagged equity premium, denoted by MV(t−1), is the total market value of equity. This metric serves as an illiquidity factor and is relevant in various domains such as equity, credit, and duration.
- Asness et al. (2013) and Moskowitz et al. (2012) stressed that value and momentum are everywhere. For this reason, we use the return of time series value and momentum strategies on commodities, currencies, bonds, and equities provided by AQR Capital Management, LLC. The fixed-income long-short value and momentum are denoted as VALLS_VME_FI, MOMLS_VME_FI, respectively. The FX value and momentum factors are denoted as VALLS_VME_FX, and MOMLS_VME_FX. The commodity factors for value and momentum are denoted as VALLS_VME_COM, and MOMLS_VME_COM, respectively.
- The excess returns of long/short equity Betting-Against-Beta factor (BAB) factor, which consists of portfolios that are long low-beta securities and short high-beta securities, as developed by Frazzini and Pedersen (2014) and obtained from AQR Capital Management, LLC.

The second subset comprises assets and is closely related to asset datasets widely used in financial research to explain fund returns and construct multiasset portfolios.[18] We suggest that investors are interested in broad diversification, encompassing traditional and alternative assets. This dataset includes emerging markets, international equities, corporate bonds, private equity, commodity returns, currency, and hedge fund returns. More specifically, this dataset includes:

- The S&P 500 returns, denoted as S&P500
- MSCI Emerging Markets Index, denoted as EM
- The ICE BofAML U.S. Corporates Index, denoted as US.Corp.
- ICE BofAML U.S. High Yield Corporate Bond Index, denoted as US.HY
- J.P. Morgan U.S. Treasuries Index, as UST
- ICE BofAML U.S. Inflation-linked Index, denoted as INF.L
- Russell 2000 Index for the U.S. Small Cap Equities, denoted as Russell.2000
- HFRX Hedge Fund Global Index, denoted as HFRX
- Gold Index, denoted as Golds

[17] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[18] For example, see Bender et al. (2019), Nystrup et al. (2018), Aked et al. (2019), Rafinot (2018), and Martellini and Milhau (2018).

- Thomson Reuters Private Equity Index, denoted as PE
- GS Commodity Index, denoted as Comdty
- U.S. Dollar Spot Index, denoted as the average international exchange rate (DXY)
- DJ Euro Stoxx 50 Index for European equities, denoted as EStoxx50
- Nikkei 225 Index for Japanese equities, denoted as Nikkei225
- ICE BofAML Euro Corporate Bond Index, denoted as EU.Corp
- ICE BofAML U.S. Treasury Bills Index, denoted as UST.Bills)

Hedge Fund and Mutual Fund Returns

Funds are significant financial assets whose managers actively engage in trading to implement trading strategies and adjusting their portfolio, interacting extensively with counterparties in the market. From a financial market microstructure perspective, funds are interconnected entities. Their interactions manifest through trading activities, and their exposure to specific asset classes links them with other funds sharing similar market exposures. This exposure creates a network of interconnected funds, often identified through crowdedness analysis. Understanding fund interactions is crucial for investors, portfolio managers, and risk managers, as it helps in assessing systemic risk, identifying market trends, and making informed investment decisions. To provide a comprehensive overview of network analysis, we utilize monthly fund returns from several published studies. Specifically, we incorporate 239 single hedge fund returns from January 2006 to December 2016, and 442 single hedge funds in the period from April 2011 to December 2016, as analyzed by Konstantinov and Simonian (2020). The study covers with the following hedge fund categories: Global Macro (GM), Alternative Risk Premia (ARP), Active Currency (FX), Commodities (C), Credit Long Short (CLS), Managed Futures (MF), Event-Driven (ED), Liquid Alternatives (LA), and Multi-strategy (MS), Equity Long Short (ELS), Equity Market Neutral (EN), Distressed Debt (DD) and Tail Risk Hedge (TH). Additionally, we include 193 bond and equity fund returns from September 2006 to September 2016, as studied by Konstantinov and Rusev (2020).

Applying various network algorithms to fund returns enables us to compute descriptive statistics and other network measures, visualize fund networks, and investigate their properties. This analysis is particularly valuable for fund managers, fund-of-fund managers, investors, chief investment officers, and regulators, as it enhances understanding of fund interactions, risk exposure, and market dynamics.

Global Equity Market Returns

Research on multi-asset portfolios inherently involves equity markets. Leibowitz (2011) stressed that every asset class contains an equity component. Equity markets are a fundamental part of portfolios, and global asset allocation decisions are impossible without them. For this purpose, we use 24 country-specific equity market returns based on the study of Frazzini and Pedersen (2014), provided by AQR Capital Management, LLC. The networks estimated for this dataset cover the period from June 2000 to October 2023.

The countries included in this dataset are

Australia (AUS)
Austria (AUT)
Belgium (BEL)
Canada (CAN)
Chile (CHL)
Germany (DEU)

Denmark (DNK)
Spain (ESP)
Finland (FIN)
France (FRA)
United Kingdom (GBR)
Greece (GRC)
Hong Kong (HKG)
Ireland (IRL)
Israel (ISR)
Italy (ITA)
Japan (JPN)
Netherlands (NLD)
Norway (NOR)
New Zealand (NZL)
Portugal (PRT)
Singapore (SGP)
Sweden (SWE)
United States (USA)

Using these country-specific equity market returns, we apply network estimation techniques to analyze and visualize the interconnectedness and interactions within global equity markets. This comprehensive analysis supports portfolio management and global asset allocation strategies.

European Multi-Asset and Bond Market Data

European markets represent a distinct subset of assets whose development can be traced back to the establishment of the European Monetary Union (EMU) in 2000, which mitigated currency risk within member countries. Despite this unified currency risk, EMU markets exhibit specific dynamics driven significantly by internal monetary policies set by the European Central Bank (ECB), domestic demand, and other macroeconomic factors. These dynamics highlight the interconnectedness of EMU markets and offer insights into how intra-EMU markets are interconnected or diverge over time.

The European bond and multi-asset market data used in this book draw from studies conducted by Konstantinov (2021) with data comprising monthly total return indices denominated in Euro in the period from January 2000 to December 2020, and Konstantinov and Fabozzi (2021), focusing on European bond market dynamics, manager performance, and portfolio allocation strategies. The sample comprises monthly returns for the period from January 2001 to December 2019. The dataset includes bond yields and spreads from core countries such as Germany, Netherlands, Finland, France, Austria, and Belgium, as well as peripheral countries including Spain, Italy, Portugal, and Ireland.[19] Additionally, the dataset encompasses time series data from major European bond and equity markets from January 2000 to December 2020.

[19]Konstantinov and Fabozzi (2021) highlighted that the separation between core and periphery countries has been the result of the macroeconomic deterioration of economic fundamentals – debt level, fiscal deficits, growth rates, and current account deficits – and European market structure. EMU's launch shaped the economic and financial dynamics and even more transformed during the 2007 to 2008 Global Financial and the European debt crises. The government bond

This comprehensive dataset provides valuable insights into the behavior and performance of European markets, aiding in the analysis of portfolio management and investment strategies within the EMU framework.

Country Bond Market Index Returns

Asset allocation decisions often underscore the significance of country risk, which can extend to include geopolitical risk.[20] Global bond markets constitute a dedicated asset class in asset management. They reflect the interplay among currencies, interest rates, and liquidity. Bond index data facilitate understanding these dynamics.

Bond markets globally encompass both developed and emerging markets. Country-specific bond indices serve a dual purpose: they act as proxies for country risk, encapsulating risks such as interest rate fluctuations, political instability, currency volatility, macroeconomic conditions, and default risks. Simultaneously, these indices represent the fixed-income market's liquidity and integration into the global financial landscape, highlighting substantial connectivity in both dimensions.

Emerging market bonds are particularly appealing to investors due to their potential for high returns, albeit accompanied by heightened risks during global market crises, liquidity challenges, geopolitical uncertainties, and other factors specific to emerging markets. The exploration of emerging market bond portfolios as a distinct asset class traces back to seminal works by Erb, Harvey, and Viskanta (1999) and Chiang, Wisen, and Zhou (2007). Systematic investment strategies in local currency bonds, focusing on styles and factors, have been investigated by Brooks, Richardson, and Xu (2020), with factor investing gaining attention in emerging market bond research (Brooks et al., 2018). Konstantinov (2022) has proposed a network approach to systematic investing in emerging market fixed income.

In financial literature, indices often represent emerging markets due to limitations in investigating individual securities that are sufficiently liquid and tradable. The dataset used includes bond market country index returns, allowing for the analysis of investable performance from risk and return perspectives. Konstantinov (2022) utilized monthly data for emerging market bond indexes denominated in USD from January 2000 to July 2021. Notably, the J.P. Morgan EMBI Global Diversified Index, which consists of USD-denominated bonds, operates as a total return index.

This index encompasses several countries and includes high-yielding government bonds from emerging markets. Given the prevalence of country risk in emerging market hard-currency bonds, empirical analyses focus on the countries constituting the core of the J.P. Morgan EMBI Global Diversified Investment Grade (IG) Index. Specifically, the analysis includes only those emerging markets with investment grade bonds. The 18 country bond indexes covered are

- Brazil (BR)
- Chile (CL)
- China (CN)
- Colombia (CO)

---

yield widening, which began after the former crisis, has marked the definition of core and peripheral countries. The core category comprises the countries of Germany, Austria, Netherlands, France, Belgium, and Finland. The peripheral category refers to the countries with weak economic fundamentals, including Italy, Greece, Spain, Portugal, and Ireland. Beirne and Fratscher (2013) argued that there are substantial and sustained differences in the pricing of fundamentals for sovereign risk among European countries. De Haan et al. (2014) argued that bond yield fluctuations in the Eurozone are less driven by macroeconomic fundamentals, which inadequately explain the bond dynamics of peripheral countries.

[20]For a detailed explanation of geopolitical risk and an investment management approach to addressing such risks and incorporating them into asset allocation, refer to Simonian (2024).

- Hungary (HN)
- Indonesia (ID)
- Malaysia (MY)
- Mexico (MX)
- Philippines (PH)
- Poland (PL)
- Panama (PA)
- Peru (PE)
- Russia (RU)
- South Africa (ZA)
- Turkey (TR)
- Uruguay (UY)
- Ukraine (UA)
- Vietnam (VN)

Payments and Transaction Data

As emphasized in earlier sections, most financial networks used in asset management are model based. By utilizing specific country-flow data, we can generate a sampling-based network derived from exchanging country flows. This approach employs straightforward payment and transaction data to directly represent a financial network, thereby minimizing the need for extensive modeling.

However, the use of such data in asset management is quite limited and typically restricted to hedge funds, brokerage houses, and other entities with privileged access to large databases. Constructing networks directly applicable to asset management necessitates access to appropriate data obtained through direct observation of market transactions, such as data sourced from stock exchanges. This direct data approach enhances the accuracy and applicability of the networks in reflecting real-world financial interactions and transactions.

APPLICATION OF NETWORKS IN FINANCE

Networks in finance encompass a broad spectrum of applications integrating various types of networks discussed in the preceding sections. The application of networks in finance extends well beyond the scope covered in this book, spanning disciplines such as economics, accounting, insurance, finance, geopolitics, and market microstructure. However, several key aspects related to asset management deserve particular emphasis. Future-oriented quantitative asset management focuses prominently on the following topics:

- *Data analysis and processing*: Leveraging advanced techniques to analyze and process vast amounts of financial data
- *Portfolio construction*: Using quantitative graph-based methods to construct portfolios that optimize risk and return profiles based on investor objectives
- *Portfolio management*: Actively managing network-based portfolios to meet investment goals while considering interrelated market conditions and changing interconnected risk factors
- *Trading*: Executing trades efficiently to capitalize on market opportunities while managing transaction costs involving the strategic use of notional sizes and microstructure data

- *Risk management*: Employing strategies to identify, assess, and mitigate risks associated with investment portfolios in interconnected financial markets
- *Performance and strategy evaluation*: Evaluating the performance of investment strategies using quantitative graph-based metrics and benchmarks to inform future decisions

These aspects highlight the critical role of networks and quantitative methods in modern asset management, driving decision-making processes and enhancing investment outcomes.

Network Application in Investment Management

The application of networks in asset management can be broadly categorized into two major areas. First, networks play a crucial role in the following ways:

- *Risk* and *portfolio management, research, and data analytics*: These are integral parts of the investment process within asset management organizations. Networks are used to model relationships between individual securities, assets, factors, countries, or any financial entities considered as nodes. The links between these nodes are modeled to understand dependencies and optimize portfolios for risk and return profiles.
- *Visualization, representation, and structuring*: Networks provide powerful tools to simplify and visualize the complex underlying structures of financial data. By graphically representing relationships between nodes, networks make it easier to interpret and analyze data dependencies, thus enhancing the understanding of intricate financial connections.

Persistent homology, a mathematical discipline applied to networks discussed in Chapter 6, plays a central role in identifying and measuring network structures that persist over time.[21] It utilizes Betti numbers, which are simple numerical values that count the number of distinct features in data (e.g., clusters or holes) that remain consistent across different time periods to measure the data features that remain consistent across various time periods.

In trading, the application of networks dates to the mid-twentieth century with the works of Weaver (1948), Shannon (1948), and Simon (1962), and Loistl and Landes (1989) followed by Lo (2010, 2017). With advancements in quantitative finance and computing power, modern approaches leverage graph theory to analyze market microstructure. For instance, analyzing bid-ask spreads, notional amounts traded, and distances between trades using network concepts provides insights into market dynamics and trading opportunities.[22]

In the asset management industry, evaluating strategy performance is critical to determining success. Fabozzi and Lopez de Prado (2018) advocate for honest reporting of backtest results, suggesting that networks can enhance this evaluation. Networks of backtest results enable visualization of strategy relationships and centrality scores, helping to identify strategies with diverse performance characteristics.

---

[21] See Cohen–Steiner et al. (2007); Edelsbrunner, Kirkpatrick, and Seidel (1983); Perea and Harer (2015); Aktas, Akbas, and Fatmaoui (2019); Baitinger and Flegel (2021); Otter et al. (2017); and Patania et al. (2017); and Edelsbrunner and Morozovy (2014) for technical details on the persistent homology.

[22] In fixed-income analysis, the distance measures have been successfully applied by Howell (2018) to measure curvature of bonds, and Heckel et al. (2020) who measured the distances between bonds. However, the authors do not apply network analysis to measure mutual distance of all bonds in the investment universe. Mizruchi (1982), Chakrabarti, Pichl, and Kaizoji (2019), König (2014), and Krautz and Fuerst (2015) for modeling market interactions.

This approach complements traditional statistical adjustments of p-values and t-statistics used in strategy evaluation, as discussed by Harvey and Liu (2014, 2015), Lopez de Prado and Fabozzi (2018), and Lopez de Prado (2019).

Moreover, networks can quantify the success of strategies using established performance metrics like the Probabilistic Sharpe Ratio. Figure 1.1 illustrates a network of backtest results, demonstrating how

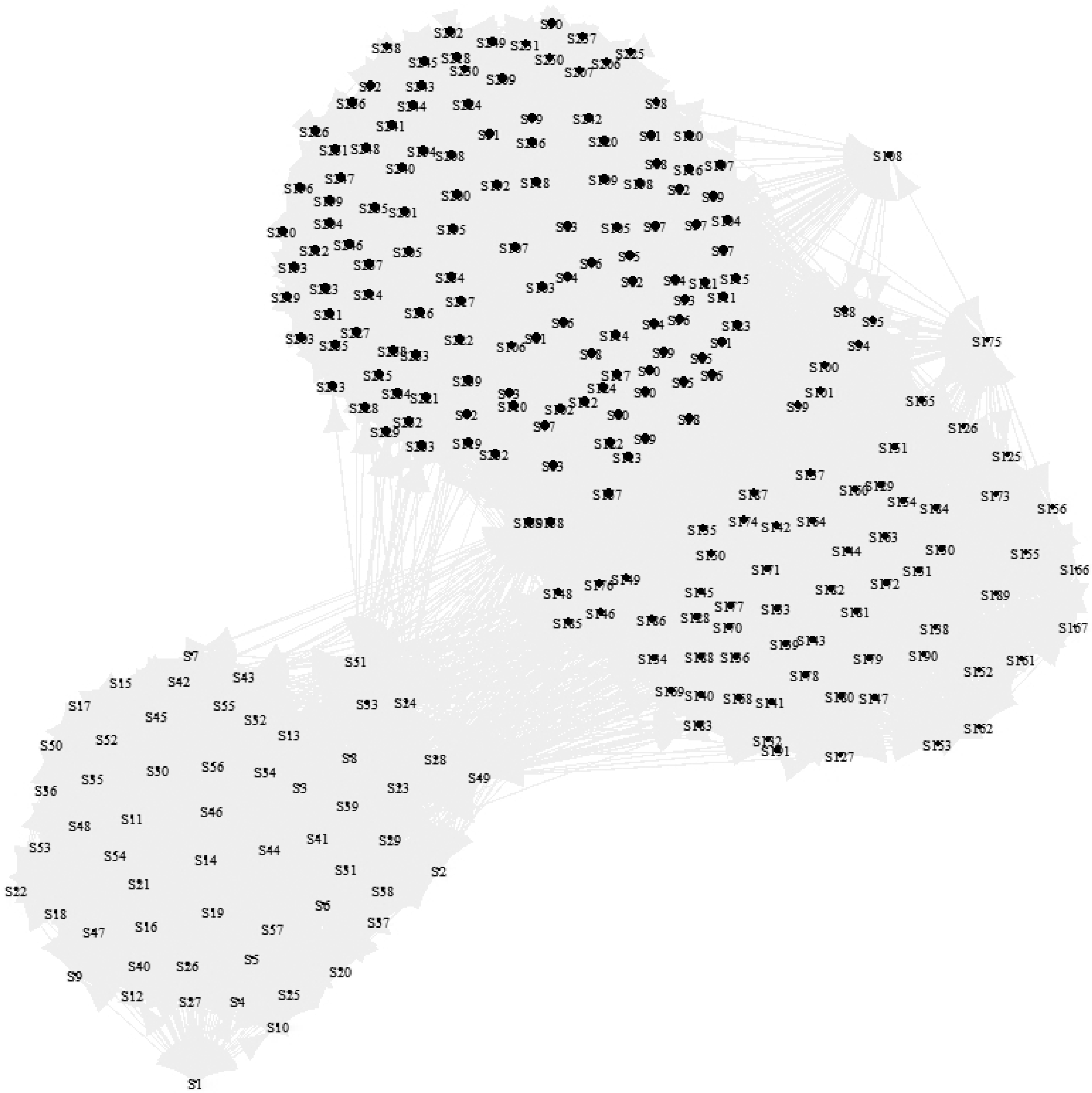

FIGURE 1.1 A Network of Tested Strategies.
Note: In Chapter 6, we explain how to estimate such networks based on the Wasserstein distance. In Chapter 10, we provide information on how to evaluate different networks. The nodes are weighted according to their importance score, which is described in Chapter 4.

strategies can be visualized and evaluated based on their performance characteristics within a network framework. This visualization helps in identifying strategies with diverse and robust performance profiles, providing a more comprehensive evaluation than traditional methods alone.

The application of networks extends beyond strategy evaluation. Another field where networks are applied is risk management, where nodal and edge risks can be considered separately. For example, nodal risk pertains to the risk of default by a financial institution. In contrast, focusing on edge risk identifies the channels through which risk spreads across the network, helping investigate potential paths of risk transmission.

Additionally, networks play a significant role in security selection and asset allocation. In this context, nodes represent financial assets or individual securities, and edges are modeled to represent the links between these financial instruments. Utilizing the tools provided by network theory, it is possible to maintain the entire interconnectedness and construct portfolios that are diversified and capable of risk management, considering the assets in their holistic interconnectedness.

Inference, Structure, Process, Implementation, and Visualization

Network analysis finds application as an asset allocation tool and a risk analysis instrument. Several investment banks apply network theory to asset pricing, specifically in order book and liquidity analysis for electronic trading, order settlement, and order matching. This approach focuses not only on asset price interconnectedness but also on the risks and returns at the industry and macroeconomic levels. Important works in these fields are Vandewalle, Brisbois, and Tordoir (2001); Bonanno et al. (2004); Tumminello et al. (2006); Battistone et al. (2012); Kalyagin et al. (2014); Glasserman and Young (2015); Peralta and Zareei (2016); and Mayoral, Moreno, and Zareei (2022), among others.

The major advantage of network theory in finance is its ability to explain complex interactions and handle large data amounts. The relationship between the three major functions of networks in asset management – implementation, inference, and structure – can be effectively summarized and visualized. Figure 1.2 illustrates this relationship, highlighting how these functions interconnect to support comprehensive financial analysis and decision-making.[23]

Network Modeling Within the inference stage, network analysis applies to raw data by mapping information on various economic variables, fund networks, private equity investors, and similar entities. Graph-theoretical algorithms can identify relationships within this data. In the structure stage, specific network metrics are used to determine the underlying data structure. Understanding how nodes are interconnected is essential in network theory, unlike traditional analytical tools that isolate a variable to analyze its impact. Networks analyze the entire structure to determine the importance of individual entities, such as nodes or links.

However, financial networks are not directly observable, necessitating modeling as a central task for researchers. While nodes may be considered fixed in the short term, links are dynamic and require explicit modeling efforts. The essential task in network modeling is to uncover these mechanisms that connect disparate data points systematically. Inference aims to uncover these mechanisms, while structure identifies how data points are related to each other.

Network modeling employs four groups of models, broadly categorized into mathematical and statistical models. Mathematical models describe the theoretical nature of network theory, while statistical models and tests have a broader application scope. Models such as link prediction, association models, and probabilistic models serve the purpose of network modeling, with statistical models particularly used for

[23] Shearer (2000).

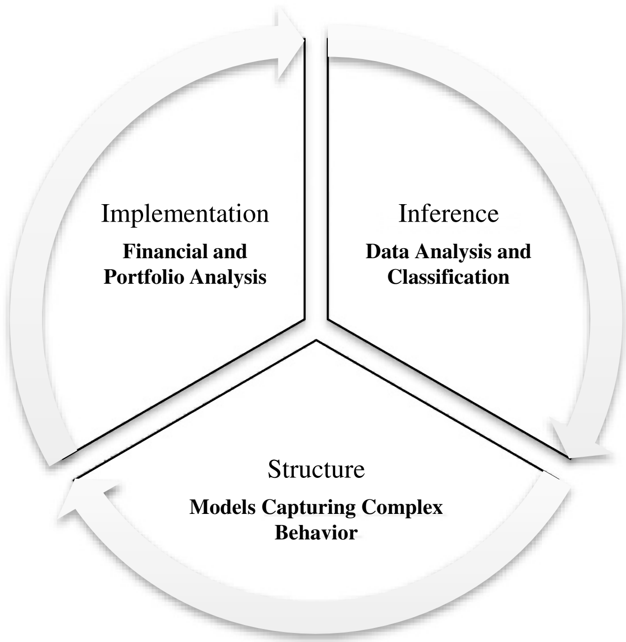

FIGURE 1.2 Network Analysis Applications in Inference, Structure, and Implementation.
Note: The three stages are related to the six-stage Cross Industry Standard Process for Data Mining (CRISP) Investment Process using alternative data.

estimating the probability of edge formation between nodes. These models are central to constructing network structures based on estimates rather than sampling.[24] In other words, other networks underlie different statistical inferences. Statistical inference refers only to the underlying, estimated network structure. Chapter 9 extensively discusses the role of statistical models in network inference.

For instance, in the analysis involving supervised machine learning on a large hedge fund database, techniques may reevaluate data clustering due to strategy shifts.[25] Baitinger and Maier (2019) demonstrated that network analysis effectively identifies hedge fund style shifts based on returns, holdings, or additional attributes, adjusting strategy affiliations or risk factor exposures accordingly. Clustering provides insights into relationships between funds within specific sectors, revealing nonobservable information at first glance. Refer to Figure 1.1 for a comprehensive overview of network analysis in financial data analysis.

[24] See Frank and Strauss (1986), Handcock (2003), Snijders (2002), Snijders et al. (2006), Robins et al. (2005), Robins et al. (2007), and Robins and Morris (2007) for a detailed treatment of statistical models and the exponential random graph model, which we discuss in Chapter 9.

[25] A hedge fund style is a critical property of a hedge fund, which states the focus and exposure of a particular strategy. For example, the equity market-neutral strategy postulates that the exposure of a hedge fund should be market, beta, or currency neutral to a market index. Strategy shift is one of the major issues in hedge fund investing and associated with the (unexpected) change in the style exposure of a hedge fund to common factors that drive its returns closely related to that strategy. Equity market-neutral funds have a beta to common equity and bond indices close to zero. In the current example, a strategy shift might be if the hedge fund shifts to a long-equity style. In this case, the beta of that fund might become significantly positive (e.g., 0.5) to the S&P500 index.

Analyzing the inference and structure of financial markets aims to identify the systemic importance of investors, institutions, and assets. Kinlaw et al. (2012) examined the systemic importance of assets within portfolio structures, emphasizing the critical role of interconnectedness between assets and their utilization in asset management. Network statistical measures and the construction of networks effectively serve these purposes.

Fundamentally, the primary application of networks in asset management lies in estimating and constructing portfolio networks for asset allocation and risk management. However, these networks extend beyond portfolio construction and serve well in performance measurement, market analysis, trading, and strategy evaluation.

Processes and Implementation Networks exhibit natural dynamics. Although plotting a network may appear static, networks possess an intrinsic dynamic nature as edges can change with new information and both exogenous and endogenous processes. The network process centers on the evolution of edges over time. The emergence and disappearance of network structures are fundamental to network science and the mathematical framework of persistent homology, which measures the persistence of network structures in space.[26] The role of network processes is to determine how nodes and edges emerge and develop over time. More precisely, the interactions are the essential property of interest in network analysis in financial markets.

Network formation and evolution are continuous processes. These processes can be modeled using either stochastic or deterministic tools, which modern econometric analysis frequently applies in finance. The need to model network processes stems from the desire to identify specific properties that help explore, investigate, and model the network under specific regimes.

For example, consider a portfolio or risk manager who wants to investigate graph-based asset allocation influenced by projections of asset class volatility. The effects change the network structure, and the information might capture specific properties usually not observed under normal conditions.

In implementation, the selected model for network estimation can be used to leverage the interconnectedness underlying a graph. This underlying network might be used to predict network metrics based on the changing number of links and nodes and their interactions. Specifically, in the implementation phase of network analysis in portfolio management, the most compelling statistical measures that explain and gauge relationships can be computed and monitored. These measures and metrics are explained in detail in Chapters 2 and 3. This application pertains to both risk management analysis and investment management decisions. For example, investigating hedge fund exposure might reveal increased factor risks in the equity market (e.g., the S&P 500 Index) or specific risk factor bets.

Visualization Visualizing data is a natural property of understanding inherent in human cognition. Just as a picture can convey more than a thousand words, network visualization provides an insightful overview of the underlying data. Consider a subway map in an underground station: it captures all relevant information about the stations, cross-stations, lines, and endpoints within a public transport system. Network theory's nature enables modeling relationships as edges and nodes, making visualization a central aspect. Visualization can convey complex interactions while also simplifying less relevant data.

Graph theoretical research has identified numerous algorithms to draw and effectively visualize networks, which we discuss in Chapter 2. The main idea of this section is to provide several clues about the proper visualization of networks, including proper edge and node design, size, structure, and groupings. Node size and edge width, along with the use of corresponding colors identifying node and edge properties

[26] Baitinger and Flegel (2021) and Baitinger (2021) provide a detailed analysis of the persistent homology in financial markets. In an advanced reading, Do Carmo provides a detailed explanations about the geometry of curves, arcs, surfaces and connectedness in general.

like community affiliation, are among the basic properties that help to visualize a network effectively. It is convenient to use color schemes when plotting networks that highlight the specific attributes. Depending on the network size, which is determined by the number of nodes, some algorithms for network drawing might be more suitable than others.

Figures 1.3a and 1.3b captures the visualization of hedge funds' interconnectedness using the dataset of Konstantinov and Simonian (2020), and in Figure 1.4, we provide a simple network identifying highly connected nodes.

In Panel A of Figure 1.4, simulated weighted networks are shown as interacting securities funds. The most important securities (i.e., nodes) are highlighted as larger dots. Panel A differs from Panel B only in layout.

Programming Codes and Implementation

Throughout this book, we use many codes written in the R programming language. Several packages have been used and found applications. Note that the codes are provided within dedicated boxes throughout the text. However, the responsible package is not explicitly mentioned within every box; thus, it is the researcher's responsibility to install and eventually detach any package that is not needed for a specific code.

FIGURE 1.3A A Single Hedge Funds Network with 442 Nodes and Optimized Edge Structure.
Note: A weighted single hedge fund network with 442 nodes and 194,918 edges minimized by a minimum spanning tree and the Kamada–Kawai graph layout algorithm. The edges result from an unsupervised Wasserstein distance algorithm, as detailed by Konstantinov and Simonian (2020).

FIGURE 1.3B A Sample Network.

This section describes the packages used in the book. A brief overview of the packages is given. It is crucial for the reader to be familiar with these packages and their descriptions to effectively implement and understand the provided codes:

- The main package used in this book is the “igraph” developed by several graph-theoretical experts, notably Gabor Csardi. The igraph package is a powerful tool for conducting all relevant information and visualization of networks with the corresponding algorithms. It serves as the backbone of the book, supporting the statistical analysis and programming needed to estimate and draw networks.
- The package “glmnet” is used for conducting LASSO and other regularization models like linear, logistic and multinomial regression.
- The package “foreach” is used for looping structures, enabling efficient execution of competitive tasks.
- The package “transport” computes optimal transport plans and Wasserstein distances, which are needed for network construction and evaluation.
- The package “vars,” developed by Bernhard Pfaff (2008), is used to estimate Vector Autoregression (VAR) models and perform diagnostics. It assists in estimation, lag selection, diagnostic testing, forecasting, causality analysis, and forecast error variance decomposition models.
- The R package “lmtest” is used for conducting Granger causality tests and additional statistical tests.

- The package "entropy" is used to compute the network nodal-based entropy. Note that the network entropy is based on the Shannon entropy, providing a measure of the unpredictability or information content within the network.
- The package "huge" is used to model high-dimensional undirected networks, enabling for the analysis of complex, large-scale network data.
- The "bnlearn" package is used to model Bayesian networks. It requires the inclusion of the "network" package, which models relational data. Note that attaching the "network" package requires the detachment of the "igraph" package due to compatibility issues.
- One of the widely used packages is "eigenmodel", which enables Monte Carlo Markov Chain simulations for networks developed by Hoff (2007). This package provides the background for implementing statistical models for network processes, which we cover in Chapter 9. The "eigenmodel" provides semiparametric factor and regression models for symmetric relational data.
- The package "ppcor" serves to estimate partial and semi-partial correlations used in the undirected networks.
- The package "ROCR" is used to evaluate the performance of scoring classifiers for network estimation. It plots two-dimensional performance curves using cross-validation, which is crucial for assessing model accuracy.
- The "ergm" package is dedicated to estimating the exponential random graph model, which uses node and edge attributes of existing (observed) networks to model new features of simulated networks. This enables for the sophisticated modeling of network structures based on observed data.

Panel A

Panel B

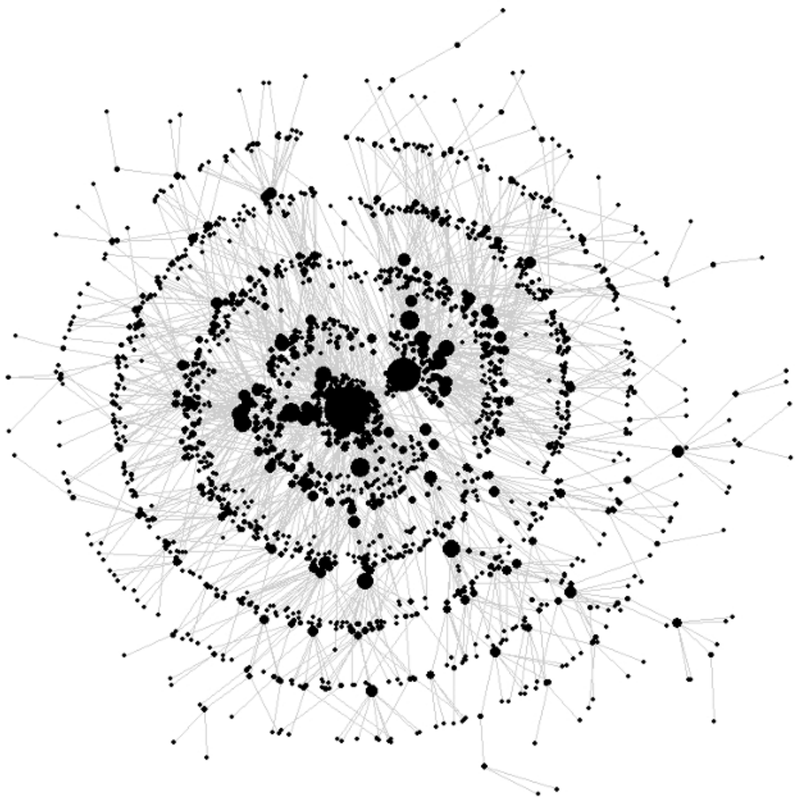

FIGURE 1.4 A Different Network Design with Identification of Highly Connected Nodes and Probabilistic Modelling of Edge Structures.
Note: This is an example of a network that comprises 2,000 nodes simulated using the Barabási –Albert algorithm. See Chapters 4 and 5 for more details on the identification of the node properties (i.e., degree score) and the simulation model defined as "Preferential Attachment."

The book's main idea is to highlight the role of network science in finance and to provide practical examples of network application in asset management. The codes provided can be directly applied to different datasets; however, specifications and adjustments are subject to the specific dataset. Moreover, experienced coders with advanced programming skills might optimize these codes for better performance. As our book's objective is to educate quantitative experts in network science, optimizing the code serves to enhance understanding and proficiency in this area.

NETWORK DESIGN

Nodes and edges are the building blocks of networks. The relationship between nodes is represented by edges, with the adjacency matrix, edge list, and incidence matrix being the most widely used representations to gauge these relationships. The adjacency matrix is a symmetric matrix where the rows and columns represent the nodes, and the elements represent the edges. Usually, the adjacency matrix is the most widely applied matrix that captures relationships, with elements being zero if two neighboring nodes are not connected and one if the two nodes are connected.

The design of a financial network for portfolio and risk management purposes, with data in a time series form, requires careful specification. This design involves modeling the entries of the adjacency matrix, which is the focus of the models presented in this book. The adjacency matrix is the basic unit used in network algorithms to visualize the relationships between nodes. Direct application of the adjacency matrix is provided in Chapter 2.

In Table 1.1, we show an example of a financial network that comprises assets and factors. There are 31 nodes in this network, and the mutual links are represented as 1 if a link between any two nodes exists, and zero otherwise. The purpose of the network models presented in this book is to specify the design experiments for estimating edges between the nodes in financial networks. In the present example, the design experiment is discussed in detail in Chapter 6.

The inclusion of node and edge attributes is not just a design choice but an essential property of complex networks. Node attributes are often assigned based on network model specification and the network statistics that measure node interconnectedness. However, node properties can also originate from external

TABLE 1.1 An Adjacency Matrix for the Undirected Asset and Factor Network.

MktRF	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1
SMB	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
HML	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
RMW	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
CMA	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
RF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
EuroStoxx50	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0
SPX	0	0	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
RTY	0	0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
EM	0	0	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
HF	1	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	1	0
DXY	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0
Golds	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0
UST	1	0	0	0	0	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	0	0	0	1	1	0	0	1	0
USCorp	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	0	1	1	0	0	0	0
USHY	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1
InfLinked	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	0	1	1	0	0	0	0
EUCorp	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	0	0	0
Commodity	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	0	0	1	0	1	0
PE	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
USTBills	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
NKY	0	0	0	0	0	0	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
BAB	1	1	1	0	0	0	1	1	1	0	1	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1	0	1	0
FXValue	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
FXMOM	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
FIValue	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0
FIMOM	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0
COMValue	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0
COMMOM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Liquidity	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1
MOM	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

sources, incorporating additional information into the network design. This means that network design might involve network process formation or inputs from external factors. Researchers must decide how to model the network to incorporate these attributes.

The network design helps evaluate local structures and identify sectors of particular interest, such as clusters, communities, groups of specific nodes, or the relative importance of nodes with systemic relevance. The statistical processes used in network design are then applied to the specific networks designed by financial researchers. This approach enables a detailed analysis of the network's structure and dynamics, providing valuable insights for portfolio and risk management.

KEY TAKEAWAYS

- The basic properties of networks are nodes and the links between them. Real-world networks gauge social interactions, infrastructure, technological complexity, power grids, water supply systems, transportation systems, and biological chains, among others.
- Financial networks consist of properties found in several other types of networks. They are characterized by the interactions of nodes, which can be individuals, enterprises, institutions, and their contractual, monetary, informational, technological, and organizational relationships.
- The use of financial networks in asset management is versatile. They are used in portfolio management for asset allocation, security selection, strategy evaluation, fund selection, and manager evaluation, among other uses.
- The broad application in investment management extends beyond traditional fields to include risk management, risk evaluation, enterprise risk analysis, and the transmission of risk through interacting entities within a financial network.
- Different data types, such as time series data, are crucial for constructing and analyzing financial networks.
- Persistent homology is applied in identifying and measuring network structures that persist over time, using Betti numbers to quantify these features. Stochastic and deterministic tools are used in modeling network processes, which are relevant to financial network analysis.
- Networks possess an intrinsic dynamic nature, with edges changing due to new information and various processes. Carefully specifying network designs is significant for portfolio and risk management, focusing on the modeling of adjacency matrix entries.