Chapter 4
FinTech

Technology has, to different degrees, always been part of the financial world, starting from the 1950s with the introduction of credit cards and ATMs, passing through electronic trading floors and personal finance apps, until present days where technologies such as Artificial Intelligence (AI) , High-Frequency Trading (HFT), and cryptocurrencies are widespread. The prominent role of technology in finance has become so important as to obtain a specific term to describe the intersection between the two—that is, *FinTech*. A portmanteau of "financial technology," FinTech refers to the application of new technological advancements to products and services in the financial industry [204]. The definition is rather broad and also encompasses "innovative ideas that improve financial service processes by proposing technological solutions according to different business situations, while the ideas could also lead to new business models or even new businesses" [205]. Following the previous definitions, FinTech cannot be categorized as a brand new industry but rather as one that has evolved at an extremely rapid pace.

As of today, FinTech represents a huge industry with a momentous growth. Ernst and Young reported for 2015 that there were 1400 FinTech firms, with more than $33 billion in funding. KPMG reports a steadily growing investment trend, which topped in Q3 2019 with a record sum of $77.1 billion, as shown in Fig. 4.1. Also the scale of venture capital activities, private equity deals, as well as mergers and acquisitions is rapidly skyrocketing, as depicted in Fig. 4.2 from data collected by KPMG and reported in Table 4.1. The geopolitical distribution of FinTech firms, and of the related investments, is naturally uneven and highly skewed. The strong dependence on technology makes it so that countries featuring higher research and investment in technology, and that are more technologically advanced, also have higher adoption rates for FinTech. Indeed, one of the countries with the highest adoption rates is Hong Kong, with the United States coming second, followed by Singapore.

© The Author(s), under exclusive license to
Springer Nature Switzerland AG 2021

R. Di Pietro et al., *New Dimensions of Information Warfare*, Advances in
Information Security 84, https://doi.org/10.1007/978-3-030-60618-3_4

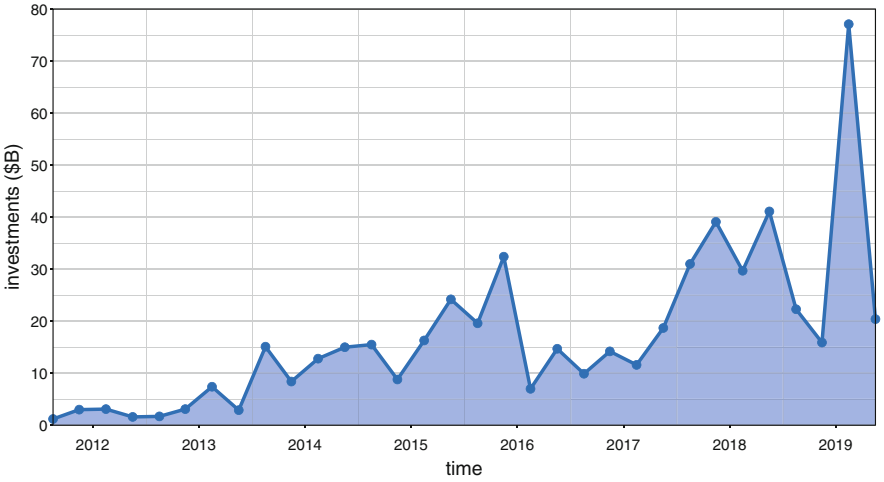

Fig. 4.1 Total worldwide investment activity in FinTech

From the applications side, financial technology is currently used in an increasingly broad array of fields. It is one of the fastest-growing tech sectors, with companies innovating in almost every area of finance. Banking and mobile banking, cryptocurrency and blockchain, investment and savings, trading, payments (e.g., Paypal, Venmo), lending (e.g., new data points and better risk modeling is expanding credit to underserved minorities), insurance (e.g., mobile car insurance, wearables for health insurance) are only some of the fields that are being deeply reshaped by FinTech. In all these scenarios, AI and Machine Learning (ML), big data, and robotic process automation are used to automate tasks and to obtain faster and more accurate predictions. In FinTech, AI algorithms are used to predict changes in the stock market and to give insights into the economy, as well as to provide insights into customer spending habits and to allow financial institutions to better understand their clients. Chatbots are another AI-driven tool that banks are starting to use to help manage customer services. Big data adds to the previous picture and can be used to predict client investments and market changes and to create new strategies and portfolios. Big Data can be employed in conjunction with AI and machine learning to analyze customer spending habits, thus improving fraud detection. Big Data also helps banks create segmented marketing strategies and can be used to optimize the operations of a company. As notable examples of this kind, Bridgewater Associates—the world's largest hedge fund—started a project to automate decision-making to save time and eliminate human emotion volatility. Similarly, Goldman Sachs now has only 2 out of 600 equity traders left in one part of its business. It found that traders can be profitably replaced by computer engineers dedicated to the development of better prediction models. In fact, it is estimated that by 2021, at least 5% of all economic transactions will be handled by dedicated autonomous software.

Fig. 4.2 Top 10 global FinTech deals in 2017, 2018, and 2019

Table 4.1 Top 10 global FinTech deals in 2017, 2018, and 2019. Rows are grouped by year and sorted in descending order by deal value

Firm	City	Country	Value ($B)
Year 2019			
Worldpay	London	UK	42.50
First data	Atlanta, GA	US	22.00
Dun and Bradstreet	Short Hills, NJ	US	6.90
Concardis	Eschborn	Germany	6.00
Assurance IQ	Bellevue, WA	US	3.50
AliExchange	Tallinn	Estonia	2.10
Paytm	Noida	India	1.70
eFront	Paris	France	1.30
Property Exchange Australia	Melbourne	Australia	1.20
Investment Technology Group	New York, NY	US	1.00
Year 2018			
Refinitiv	Eagan, MN	US	17.00
Ant Financial	Hangzhou	China	14.00
Nets	Ballerup	Denmark	5.50
Blackhawk Network Holdings	Pleasanton, CA	US	3.50
VeriFone	San Jose, CA	US	3.40
iZettle	Stockholm	Sweden	2.20
Fidessa Group	Woking	UK	2.10
Ipreo	New York, NY	US	1.90
IRIS Software Group	Datchet	UK	1.70
PowerPlan	Atlanta, GA	US	1.10
Year 2017			
DH Corp.	Toronto	Canada	3.60
Bankrate	Palm Beach Gardens, FL	US	1.44
Vocalink	Rickmansworth	UK	1.10
Intacct	San Jose, CA	US	0.85
BluePay	Naperville, IL	US	0.76
CardConnect	King of Prussia, PA	US	0.75
Trayport	London	UK	0.73
Institutional Shareholder Services	Rockville, MD	US	0.72
Xactly	San Jose, CA	US	0.56
Merchants' Choice Payment Solutions	Shenandoah, VA	US	0.47

Finance is currently seen as one of the industries that are most vulnerable to disruption by technology. In fact, rapid development also has a dark side, and new technologies inevitably bring new vulnerabilities and threats. For example, the interplay between algorithms (e.g., ML and AI) and security in finance is complex. On the one hand, some of the most powerful techniques to ensure cybersecurity in the coming years will be based on machine learning. On the other hand, heavy reliance on machine learning and AI—especially black-box/opaque models [206]—

can make it harder for human analysts to understand system behaviors and the associated security risks. Besides, some of the most powerful techniques leveraged by hackers are also based on ML and AI. Cybersecurity risks are particularly high in FinTech also due to the sensitive data and operations managed by companies in the financial sector. Banks, for example, collect and keep loads of sensitive information about their clients. Since the data has a direct connection to the accounts of the clients, cybercriminals deliberately target that information during attacks, intending to steal personal information of banks' customers if not directly accessing their accounts and seizing their money. The combination of fast-paced technological innovation and extremely sensitive data and operations thus creates the perfect storm for cyberattacks. In fact, a growing number of studies point out that "the unintended consequences of technology-leveraged finance include firesales, flash crashes, botched initial public offerings, cybersecurity breaches, catastrophic algorithmic trading errors, and a technological arms race that has created new winners, losers, and systemic risk in the financial ecosystem" [207]. Some even say that we are living in a "golden age for hackers," testified by gigantic data breaches occurring seemingly on a weekly basis.[1] Examples of this kind are the 2018 attack on HSBC's American operations and the 2019 Capital One data breach that affected 100 million Americans. Or even the infamous 2016 Bangladesh Bank Cyber Heist, when five huge fake money transfers were issued through the SWIFT network, totaling $101 million, of which only $38 million were recovered.[2] In that case, the bank's account at the Federal Reserve Bank of New York was hacked, and the thefts were traced to Sri Lanka and the Philippines—evidence that cybercrime in the financial sector thrives on technology and operates on a global scale. The latest survey conducted in 2019 by the Bank of Lithuania shows that cyberattacks currently pose one of the biggest threats to financial institutions.[3] IBM's 2019 Cost of a Data Breach Report concludes with similar findings and shows that the average global cost of a data breach in the financial sector is as high as €4.9 million per incident.[4]

Given this grim picture, it comes with little surprise that financial services companies are concerned about systems and data security, as well as about the concept of trust, more than organizations in almost any other sector. For instance, cryptocurrencies are completely based on trust, where trust is transferred from centralized and regulated repositories typical of fiat currencies, to trust in technology and decentralization. Without secure and trusted algorithms, FinTech will fail, and any firm developing a new FinTech business should consider how it will implement security and trust through technology [208]. In fact, FinTech entrepreneurs are

[1] https://www.fintechweekly.com/magazine/articles/the-cyber-security-landscape-in-financial-services (Last checked August 2020).

[2] https://en.wikipedia.org/wiki/Bangladesh_Bank_robbery (Last checked August 2020).

[3] https://www.lb.lt/en/news/survey-cyberattacks-and-re-bubbles-pose-the-biggest-threat-to-financial-institutions (Last checked August 2020).

[4] https://www.ibm.com/security/data-breach (Last checked August 2020).

constantly looking to strike the right balance between growing their business and protecting against imminent cyber threats.[5] To this end, academia recently devoted much interest toward the design and development of secure solutions for FinTech. For example, both vulnerabilities, attacks, and possible solutions were discussed for problems such as credit card and online payment fraud detection, as well as for defending against data breaches [209]. However, the majority of existing studies focused on problems of micro-security—e.g., how to enforce security for given apps or how to protect data about individuals using a given service—rather than on macro-security, which concerns with the security and trustworthiness of either whole markets or whole technologies (e.g., HFT). The former is particularly relevant to individual users, while the latter is primarily of concern for governmental entities and nations themselves, since it might have a significant impact on the whole economy.

Nowadays, regulators, innovators, and investors face an increasingly complex environment where computers and infrastructure merge, regulations allow dozens of different exchanges to coexist, and globalized business facilitates round-the-clock deals. The widening gap between innovation and regulation is acute in FinTech and particularly so with respect to cybersecurity. This is the unavoidable result of mixing solutions that are evolving at a rapid pace with regulatory frameworks that change far more slowly. Within the context of economic war, where nations look at the economy as a sharp weapon with which to pursue strategic and political goals, the aforementioned vulnerabilities of FinTech and the weak regulatory frameworks can even be put to "good" use. In fact, the very existence of the many possible ways in which FinTech firms and the underlying technologies can be hacked and tampered with implies that a nation could easily weaponize FinTech itself and use it as an attack vector pointed toward critical economic assets of competing nations. Indeed, one of the main cyber threats to financial institutions already come from state-organized actors, and the news is replete with both facts and conjectures about state-sponsored hacking [210]. In the remainder of this chapter, we investigate the different ways in which FinTech can be weaponized to attack a nation's economic assets, describing the current state of the art with regard to both attacking and defensive means.

4.1 Scenario 1: Stock Market Forecasts

The stock market of a country, also known as the equity market, is one of the most important components of a national free-market economy. It refers to a centralized place where equities or stocks of publicly held companies, bonds, and other classes of securities are issued and traded. In other words, the national stock market is a

[5] https://web.archive.org/web/20160110115516/http://www.whartonfintech.org/blog/protect-assets-cybersecurity-fintech/ (Last checked August 2020).

complex infrastructure that provides companies with access to capital in exchange for a slice of ownership, by offering to investors stock shares and corporate bonds. Through the course of history, several economists and philosophers argued that stock markets are the most effective way of aggregating the pieces of information that are dispersed among individuals within a society. In fact, stock markets give the opportunity to profit from information about a company, by trading that company's shares. Interested traders are thus motivated to acquire and to act on information for personal profit. In doing so, traders contribute to more and more efficient (i.e., accurate) market prices. In the competitive limit, market prices thus reflect all available information, and prices can only move in response to the news. This theory about stock markets and prices is dubbed the *efficient-market hypothesis* in financial economics. A direct implication of this theory is that it is impossible to consistently beat the market, on a risk-adjusted basis, since market prices should only react to new information. As an example, let us suppose that a piece of information about the value of a stock is widely available to investors. If the price of the stock does not already reflect that information, investors will trade on it, thereby moving the price until the information is no longer useful for trading.

Being able to predict where markets are headed is the "Holy Grail" of finance. However, if the efficient-market hypothesis alone could exhaustively model and justify the behaviors of stock markets, there would be little interest in trading because of the limited theoretical possibility to predict the market. Instead, both theoretical and empirical data suggest that markets are not completely efficient. As a consequence, prices might not accurately reflect the true value of stocks. Some economists trace back the imperfections and the irrationalities in financial markets to human factors, by citing a combination of cognitive biases such as overconfidence, overreaction, representative bias, information bias, and various other predictable human errors in reasoning and processing available information. Events such as the Global Financial Crisis (GFC) of 2007–2008 raised additional concerns on the efficiency of financial markets and led even supporters of the efficient-market hypothesis to claim that "poorly informed investors could theoretically lead the market astray" and that stock prices could become "somewhat irrational" as a result.[6]

Resources

The **Billion Prices project**[7] of the Massachusetts Institute of Technology (MIT) is an academic initiative that uses prices collected from hundreds of online retailers around the world on a daily basis to conduct research in macro

(continued)

[6] https://web.archive.org/web/20120406035022/http://fisher.osu.edu/~diether_1/b822/fama_thaler.pdf (Last checked August 2020).

[7] http://www.thebillionpricesproject.com/ (Last checked August 2020).

and international economics. Interested scholars and practitioners have the opportunity to download several datasets related to this research study on the project website.

Perhaps even more worryingly, the same outcome can also be caused by groups of collusive traders, employing market manipulation techniques. In contrast to the properties of an efficient market, the founding idea of market manipulation is a temporary distortion of pricing. Market manipulation is a deliberate attempt to interfere with the free and fair operations of the market and to create artificial, false, or misleading appearances of the price of a product, security, commodity, or currency. Setting artificial prices for financial instruments leads to a redistribution of capital, typically in favor of the small number of participants involved in, and aware of, the manipulation. In turn, this leads to economic imbalances that hinder fair market and unaware investors, undermining confidence in financial institutions. Even though the next generation of market manipulation—e.g., those manipulations that exploit or make use of the latest technological advancements—has the same goal as their traditional counterparts, they can be much more effective due to the unprecedented velocity and interconnectedness of today's digital markets [211]. Market manipulations can take different forms, and while some target low-value and marginal securities, others involve the core of financial markets. These latter manipulations are capable of creating huge shocks in markets, thus making these security issues of concern for nations and of interest for our discussion. On May 6, 2010, the Dow Jones Industrial Average (DJIA) had the biggest 1-day drop in history, later called the *Flash Crash*. After 5 months, an investigation concluded that one of the possible causes was an automated High-Frequency Trading (HFT) system that had incorrectly assessed some information collected from the Web. In 2013, the official Associated Press Twitter account got hacked, and a false rumor was posted reporting that President Obama had been injured during a terrorist attack at the White House, as shown in Fig. 4.3. The fake news rapidly caused a stock market collapse that burned $136 billion. Then, in 2014, the unknown firm *Cynk Technology* briefly became a $6 billion worth company. Automatic Trading algorithms detected a fake social discussion and begun to invest heavily in the company's shares. By the time analysts noticed the orchestration, investments had already turned into heavy losses [212].

The previous anecdotes serve as tangible examples of how market manipulations can cause dramatic shocks capable of affecting even the strongest economy. Previous studies on this subject classified manipulations into two main categories: (1) information-based and (2) trade-based. An information-based manipulation is carried out by issuing false information or by spreading false rumors, while a trade-based manipulation is based solely on buying and selling securities without performing any other publicly observable actions or the spread of false information [211]. As for the whole FinTech domain, some of these manipulative practices

Fig. 4.3 The shocking tweet posted by the hacked AP Twitter account in 2013, which caused a $136 billion worth market crash

have always existed. However, due to the recent technological advancements, they have now become more ubiquitous, and they unfold much faster than before, thus becoming always more indistinguishable from legitimate practices. In the next three sections, we investigate these two traditional forms of market manipulation, as well as new ones that are rapidly emerging. In doing so, we discuss the next generation of manipulation-specific attacks and countermeasures inside those three sections, while in Sect. 4.1.4, we briefly outline generic countermeasures that are not designed to fight a specific type of manipulation.

4.1.1 Threat: Information-Based Manipulation

The first data sources that have been used for predicting future prices of a financial instrument were past prices themselves. In fact, stock market data naturally comes as sequences of values, such as the sequences of opening, high, low, and closing prices at different points in time, and early attempts at stock market prediction were based on the application of statistical (e.g., autoregressive models) and pattern matching techniques to these numeric market data. However, in addition to market data, there exists also a growing wealth of unstructured and heterogeneous big data outside of financial services that have a direct influence on markets. Taken in aggregate, news delivered by companies like Dow Jones, along with the blog posts by experts and analysts, and even online collective user activity gathered by the likes of Google, Facebook, and Twitter, can all be used to understand and improve upon market movements. Because of this, a growing area of FinTech revolves around the acquisition and analysis of "alternative" data. Textual data is a paramount example of alternative financial data capable of expanding the universe of available observations, thus going beyond the merely numerical market indicators [213].

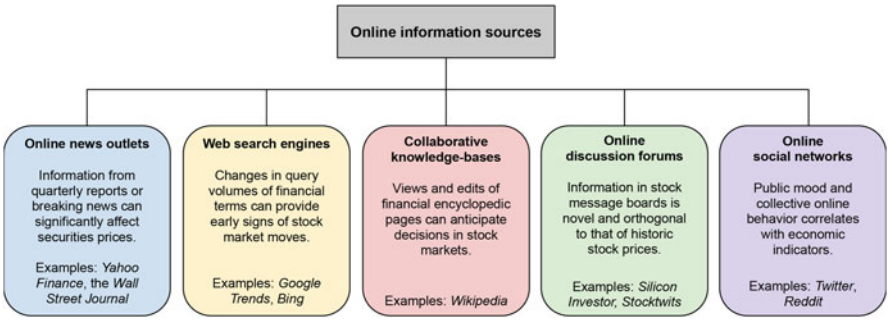

Fig. 4.4 Categorization of the online information sources most used for market prediction

A Large Attack Surface

Recent state-of-the-art market prediction systems exploit all available data, including textual data gathered online, in complex prediction systems based on Machine Learning (ML) and Artificial Intelligence (AI), to come up with more accurate predictions. Figure 4.4 shows an overview of the online information sources that are most widely used for market prediction. For instance, it has been demonstrated that collective patterns of usage of Wikipedia are informative of future stock market moves [214]. Similarly, aggregate data about queries on popular Web search engines, such as Google and Bing, can also be profitably used to predict market movements [215]. The same goes for user-generated information contained in online financial discussion boards [216]. In addition to the aforementioned online data sources, among the textual sources that are currently attracting the majority of research and development are online news (e.g., those shared by the Wall Street Journal, Dow Jones, Reuters, and Yahoo Finance) and user posts in Online Social Network (OSN) data streams (e.g., Twitter above all, but also Reddit, Facebook, and others). In detail, financial news has been proved to convey strong signals for predicting short-term market movements. The system discussed in [217] exploits news articles from Yahoo Finance to predict future prices of S&P 500 stocks. Other studies showed that information extracted from news sources is particularly informative for predicting the direction of assets volatility movement [218]. Still leveraging online financial news, several systems were developed for event-driven stock market prediction [219]. In these systems, events are first extracted from news text, and represented as dense vectors, for instance, via neural tensor networks. Then, deep convolutional neural networks are used to model both short-term and long-term influences of events on stock price movements. Furthermore, text from news articles can also be used to predict intraday price movements of financial assets [220]. Regarding the exploitation of publicly available OSNs data, the *sentiment* score of public stock-related posts has been widely used as a predictor for stock prices and other economic indicators [221, 222]. As an example, the classifier developed in [223] is based solely on the sentiment analysis of tweets and accurately

predicts the next-day trend of the stock values of specific companies. Similarly, the one developed in [224] predicts opening and closing stock market prices with high accuracy. Others have instead proposed to exploit the overall volume of tweets about a company [225] and the topology of stock networks [226] as predictors of financial performance. The role of social media *influencers* has also been identified as a strong contributing factor to the formation of market trends [227]. Finally, the relationships between different companies discussed in OSNs have been found informative [228]. In detail, co-occurrences of stock mentions in online discussions have been exploited to create a graph of companies, which was subsequently clustered. It was found that companies belonging to the same clusters feature strong correlations in their stock prices. This methodology was employed for market prediction and as a portfolio selection method, which outperformed traditional strategies based on company sectors or historical stock prices. The usefulness and predictive power of online and social data are acknowledged not only by academia. In fact, there is a growing number of FinTech start-ups, as well as established companies, that are embracing this business. Companies such as Acuity Trading, Selerity, and iSentium are all harnessing data from social platforms like Twitter to give an indication of investor "sentiment", which, in turn, provides predictive signals of which way to trade. And this information-driven revolution is changing more than the investing habits of individuals or the business of emerging firms. Institutional investors are increasingly subscribing to big data information sources. The more uncommon or uncorrelated is the data source, the more valuable it is, since it is capable of bringing unsaturated information into market models. Each data source then drives a small profit in market allocations. When combined, all of the data sources deliver meaningful profitability to the data acquirers. This uncommon information model of institutional investing has become known as Smart Beta or the Two Sigma model, after the hedge fund that grew 400% in just 3 years after adopting this model.

The previous brief review of the scientific literature on stock market prediction shows that the vast majority of existing systems complement historic market data with additional data collected online, and specifically from online social platforms (e.g., first and foremost OSNs), but also blogs and discussion forums. These online data have been shown to significantly boost the predictive power of market prediction models, hence allowing to achieve larger profits, on average. Thus, if on the one hand, online and user-generated content is increasingly exploited for predicting trends in the stock market, on the other hand, we are running the risk that much of the content those systems rely on is actually fake and possibly artfully created to mislead algorithms and human investors alike. In fact, in Chap. 2 we already showed the extent to which online data can be massively fabricated to artificially support a given narrative, person, or product. Finance makes no exception, and financial spam is rampaging in our online social ecosystems since many years [8]. Because of this, without the adoption of effective defensive techniques, all market prediction systems and Automatic Trading (AT) algorithms are vulnerable to manipulation via targeted fake online content. For instance, automatic systems could be tricked into buying large numbers of stocks simply because they detected a positive sentiment or large

volume of discussions in OSNs, despite the menace that those figures could have been very easily fabricated by hordes of automated accounts (i.e., social bots) or by paid human workers (i.e., trolls). In turn, these manipulative techniques could be used to unfairly favor a nation's company in a stock market against the one of a competing nation or to trick a national fund into making bad investments. In other words, the possibility to control the news and the content that circulates in OSNs currently provides a powerful leverage to influence the stock market. The previously mentioned examples of Cynk Technology and the hacked Associated Press Twitter account testify the tremendous impact that the accidental or intentional spread of false and inaccurate information, combined with AT, can have on financial markets. To our dismay, recent history is replete with such examples. In May 2020, Elon Musk tweeted about Tesla stock price being "too high," which rapidly resulted in a 10% loss by the end of the trading day.[8] To make matters worse, online information—and especially those generated within OSNs—are not only useful in the financial domain for stock market prediction. Even if in this section we specifically focus on this task, there exist many additional scenarios where these data are being actively exploited. For example, social media interactions are used to identify good customers [229]. Similarly, digital footprints are used to assess individual default risk [230], and friendship networks are exploited in peer-lending schemes [231]. As such, information manipulation can have repercussions that go well beyond stock markets. In 2017, Qatar suffered a blockade imposed by Saudi Arabia, United Arab Emirates, Bahrain, Egypt, and few other countries. The spark that ignited the Qatar diplomatic crisis was a tweet by the Qatar News Agency that later claimed to have suffered a hacking by an "unknown entity" that shared a story with "no basis whatsoever." [9]

Attacks and Countermeasures

Information manipulation attacks aimed at influencing the stock market, or other constituents of a national economy, are based on the tools and techniques that we thoroughly discussed in Chap. 2. The same consideration also largely applies to defensive means. In fact, the majority of existing countermeasures for information-based market manipulation are simply based on countermeasures to generic information manipulation, with a few notable exceptions that we discuss in detail in the following.

All research aimed at measuring and thwarting information manipulation, information disorder, strategic information operations, and the like is still relatively young, having sparked only after the shocking results of the Brexit referendum in the UK and Donald Trump's election in the United States, in 2016 [4]. Because of this, the literature on information-based market manipulation—a relatively small

[8] https://www.wired.com/story/elon-musk-tesla-stock-too-high-falls/ (Last checked August 2020).
[9] https://www.bbc.com/news/world-middle-east-40026822 (Last checked August 2020).

subset of all information-based manipulations—is still in its infancy, and thorough studies that investigate this issue are few and far between. Among them, some pioneering works investigated the presence of spam in online financial discussions and the role of social bots in the spread of such spam [212, 232]. The authors collected nine million tweets related to more than 3500 stocks traded in the main US financial markets (NASDAQ, NYSE, NYSEARCA, NYSEMKT). By leveraging anomaly detection algorithms, they identified anomalous discussion spikes in their dataset. Then, they turned their attention to the messages that contributed to the formation of such spikes. Surprisingly, they found that such huge discussions were mainly generated by massive retweets of a few original messages—a technique widely used to artificially increase the popularity of specific pieces of content [13] (i.e., astroturfing). The content of the massively retweeted tweets also revealed something interesting. The majority of such tweets were mentioning a few well-known and highly capitalized stocks, together with many rather obscure stocks with low market capitalization. The latter were all belonging to the OTCMKTS market, a US financial market for over-the-counter transactions with far less stringent requirements than those imposed by NASDAQ, NYSE, NYSEARCA, and NYSEMKT. Finding no apparent explanation for the massive co-occurrence of the unknown OTCMKTS with the other highly capitalized stocks, authors investigated the nature of the accounts that were responsible for a large number of retweets. They leveraged a state-of-the-art bot detection technique [75] and found that as much as 71% of all retweeters were, in fact, social bots. This finding provided the first quantitative and large-scale evidence of the existence of financial spam in OSNs. In a later work, authors went forward and analyzed the characteristics of the financial bots [233]. They found that such accounts were rather simplistic, with few profile information and social relationships. Overall, the bots did not appear as credible sources. Drawing upon these results, they concluded that the mass retweeting of low-value stocks together with some high-value ones, was targeted at Automatic Trading algorithms monitoring social conversations, rather than at human investors. Authors dubbed this practice as *cashtag piggybacking*—by leveraging the notion of piggyback used in computer networks[10] —since fraudsters were likely trying to exploit the popularity of the high values stocks to induce algorithms into buying the low-value ones [212].

Resources

The dataset used in [212, 232, 233] for the analysis of information-based market manipulation is publicly available online.[11] The dataset contains social information (i.e., stock-related online conversations) collected from Twitter and price data collected from Google Finance.

[10] https://en.wikipedia.org/wiki/Piggybacking_(data_transmission) (Last checked August 2020).

[11] https://doi.org/10.5281/zenodo.2686862 (Last checked August 2020).

Open Issues and Future Directions

As introduced in the previous section, research on information-based market manipulation is still at its early stages. As such, many directions of research require contributions, thus depicting a pristine scientific landscape with several low-hanging fruits. One of the most important research questions, which are still unanswered, is the assessment of the impact of such manipulations. Only recently we started demonstrating the existence of information-based market manipulations, but we still have no clue about their goals and their outcomes. Measuring the manipulation impact on the markets is undoubtedly a challenging task. One possible way to achieve it consists in monitoring prices and traded volumes and by investigating the existence of possible correlations between manipulation campaigns and price movements.

Another direction of research that still requires much effort is the one related to uncovering and discussing ongoing manipulations. Until we have a sound idea of which manipulations are going on in our online social ecosystems, how they are organized, who perpetrates them, and ultimately how widespread the phenomenon is, we will not be able to put in place effective countermeasures. In this regard, several informal and anecdotal investigations have been carried out, complemented by only a few full-fledged large-scale scientific works [212, 232]. It is thus crucial to multiply efforts toward this direction in the coming years.

Lastly, we surely need to start developing mechanisms for protecting against these forms of manipulation. As we discussed throughout all this section, basically no market prediction system is currently equipped with information filters capable of protecting it from possible manipulations. As such, all the existing ecosystem of Automatic Trading (AT) algorithms is exposed and vulnerable to information-based manipulations. The quickest way to implement a possible first layer of protection revolves around the adoption of general-purpose techniques against information manipulation. For example, systems that feed on online and social data could be complemented with the latest techniques for detecting fake news and for spotting coordinated, artificial, or automated behaviors. Despite not being specifically tailored for the financial domain and financial spam, several of these techniques are readily available and demonstrated decent performance. In the medium to long term, it would however be advisable to develop specific defensive techniques for financial information-based manipulations. This would allow to obtain more accurate, reliable, and efficient countermeasures. For example, some activities are currently undergoing for exploiting labels automatically assigned to financial posts by generic fake news and bot detectors, to develop market-specific manipulation detection techniques. These next-generation tools promise the timely detection of unfolding manipulations—such as those documented in [212, 232]—at their early stages, to stop automatic systems from ingesting fabricated data.

4.1.2 Threat: Trade-Based Manipulation

A trade-based manipulation occurs when a trader attempts to artificially alter a price simply via buying and selling—that is, without releasing any false information or taking any other publicly observable action. The traditional full-information financial theory, such as the efficient-market hypothesis, asserts that such speculation actually contributes to stabilizing prices since manipulators, like all rational traders, buy when the prices are low and sell when the prices are high. In contrast, by following the theories of incomplete and asymmetric information that challenge the efficient-market hypothesis, it is possible to demonstrate that speculation can destabilize prices and increase volatility. The reason being that uninformed traders oftentimes are unable to distinguish the actions of manipulators from those of informed traders [234]. Under these conditions, it has been demonstrated that trade-based manipulation can be profitable [235]. In practice, groups of coordinated manipulators can make slightly unprofitable initial trades against the direction of available information. This initial investment, possibly complemented by corollary actions aimed at attracting additional external investments, can manage to set in motion a price trend among partially informed followers, thus turning the initial investment into a potential profit. From this point on, manipulators can profitably unwind their position against still less informed market makers and other liquidity providers.

Contrarily to information-based manipulations that thrive on the recent democratization of the information landscape, trade-based manipulations only require an open stock market. In fact, most of the techniques used for carrying out trade-based manipulations are as old as the markets themselves. Having such a long history, many nations have long developed regulations against these forms of manipulation. For instance, Section 10(b) of the Securities and Exchange Act (SEA) of 1934, Rule 10b-5, and Section 9(a)(2) of the SEA prohibit manipulation in the United States. Similar actions are taken in Section 1(2)(a) of the Market Abuse Directive (MAD) 2003/6/EC in the EU and Section 1041A of the Corporations Act (CA) 2001 in Australia. As a result of these and other laws, trade-based manipulations are fairly rare in regulated markets. Such laws, however, only apply to regulated markets, such as the main financial markets where trading takes place. However, there exist other important markets that are not subject to any of the above-cited regulations: cryptocurrency exchanges. As already introduced in Chap. 3, a cryptocurrency exchange is a business that allows customers to trade cryptocurrencies or digital currencies for other assets, such as conventional fiat money or other digital currencies. Exchanges can be brick-and-mortar businesses or strictly online businesses. The latter often operate outside Western countries so as to avoid regulations and prosecutions. Nonetheless, they do handle Western fiat currencies and maintain bank accounts in several countries to facilitate deposits in various national currencies. The largely unregulated nature of cryptocurrency exchanges represents fertile ground for all manipulations that once permeated traditional stock markets and that progressively got banned over time. In this regard,

cryptocurrency exchanges are a fresh start for modern manipulators, as testified by the large number of cryptocurrency frauds. These frauds, if targeted at state cryptocurrencies such as those discussed in Chap. 3, could even endanger a national economy and be considered alike a direct attack to a nation.

Attacks

The basic mechanisms used to perpetrate trade-based manipulations did not change through the years. Here, we explain such mechanisms for a few notable manipulations.

Pump-and-Dump Pump-and-dump (P&D) is a form of security fraud that involves artificially inflating the price of an owned stock, in order to sell it at a higher price. Participants in P&D schemes collectively aim to artificially inflate a currency price through coordinated, simultaneous buying (i.e., the "pump" action). Once outside unaware investors notice the surge in price and start investing in the asset, the participants sell to them (i.e., the "dump" action), thus making a profit and causing a price collapse. Figure 4.5 sketches the typical price trend of a successful P&D operation that, interestingly, also largely follows Jean-Paul Rodrigue's phases of a financial bubble [236]. Generally, there are orchestrators behind the curtain who profit even at the expense of the witting participants themselves, let alone of the other unaware investors [237]. Historically, P&D schemes took place using email spam campaigns, through traditional media channels via fake press releases, or through telemarketing from "boiler room" brokerage houses. Often the stock promoter claimed to have "inside" information about impending news in order to lure investors into buying. In other cases, newsletters that purportedly offered unbiased recommendations then touted a company as a "hot" stock for their own

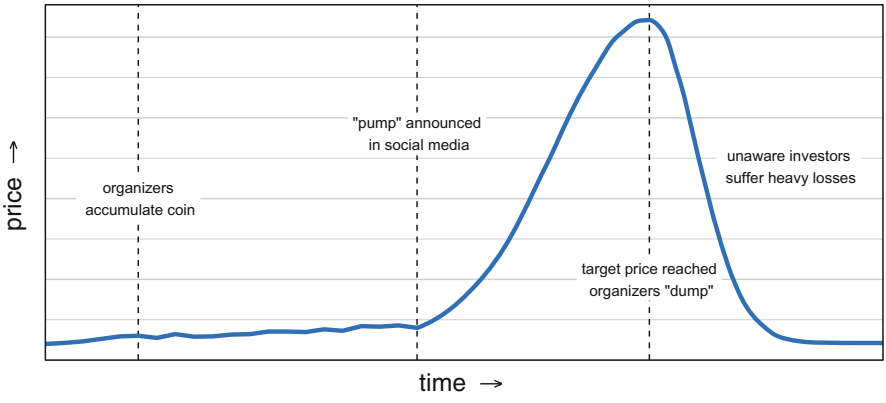

Fig. 4.5 Sketch of a successful P&D operation where organizers manage to create a surge in price, which they later exploit to make an unfair profit at the expenses of unaware investors

benefit. Other times, promoters posted messages in chat rooms or stock message boards urging readers to quickly buy the pumped stock. Nowadays, techniques for luring unaware investors and algorithms have moved from traditional media and largely involve social media, OSNs, and messaging apps. If orchestrators are successful, they will entice unwitting investors to purchase shares of the target company. The increased demand, price, and trading volume of the stock will likely convince even more people to believe the hype and to buy shares as well. When the orchestrators sell their shares and stop promoting the stock, the price plummets, and other investors are left holding a stock that is worth significantly less than what they paid for. In traditional markets, manipulators typically use this ploy with small, thinly traded companies known as "penny stocks", generally traded over-the-counter, since it is easier to manipulate a stock when there is little or no independent information available about the company or little activity anyway. The same principle also applies to pumped coins in cryptocurrency exchanges, which are typically low-value coins with little-to-no activity.

Cornering the Market Cornering the market consists of obtaining sufficient control of a particular stock, commodity, or asset so as to be able to manipulate its market price. To some, it can be defined as having the greatest market share in a particular industry, without having a monopoly. This form of market manipulation can be attempted through several mechanisms. The most direct strategy simply involves buying a large percentage of the available commodity offered for sale in some market and hoard it. However, this manipulation exposes the perpetrator to significant risks. For instance, the "cornerer" is typically vulnerable due to the size of its position, which makes it highly susceptible to market risk. Besides, by definition, cornering a market requires to purchase assets at artificial prices, thus effectively opening profit opportunities for other investors—e.g., through arbitrage. Moreover, if the price starts to move against the cornerer, any attempt to sell would cause the price to drop further, subjecting the cornerer to heavy losses. Indeed, the famous American business journalist Edwin Lefèvre once wrote that "very few of the great corners were profitable to the engineers of them" (*Reminiscences of a Stock Operator*—1923).

Wash Trade A wash trade is a form of market manipulation in which an investor simultaneously sells and buys substantially the same financial instrument, to create a misleading and artificial impression of high trading activity around that instrument. In turn, this typically raises the price of the instrument. In practice, a manipulator will first place a sell (buy) order and then immediately place a buy (sell) order at a specific price so as to buy from itself. This may be done for several reasons, such as to artificially increase trading volume, giving the impression that the instrument is more in demand than it actually is, or also to generate commission fees to brokers in order to compensate them for something that cannot be openly paid for. This form of manipulation can also be referred to as **churn**, especially when carried out to generate commission fees to brokers.

Ponzi Scheme We conclude this brief overview of market manipulation techniques with the introduction of Ponzi schemes. Although they are not strictly trade-based manipulations, Ponzi schemes benefit from the same technological advancements of the aforementioned manipulation techniques (e.g., social media for attracting investors) and often target the same financial assets, such as cryptocurrencies. A Ponzi scheme is a fraudulent investing scam that promises high rates of return with seemingly little risk to investors. The Ponzi scheme generates returns for early investors by acquiring new ones. To this end, it is similar to a pyramidal marketing scheme in that both are based on using funds from new investors to pay the earlier backers (Fig. 4.6). Both Ponzi schemes and pyramid schemes eventually bottom out when the flow of new investors isn't enough to sustain the scam. At that point, the schemes unravel. Manipulators that engage in a Ponzi scheme typically focus all of their efforts into attracting new clients to profit from their investments, hence the renewed interest in this fraud since the rise of social media that makes this task way easier. The Ponzi scheme is named after a swindler named Charles Ponzi, who orchestrated the first one in 1919 in the United States, promising returns of 50% in 45 days or 100% in 90 days. Due to his (legit) success in previous investment schemes, investors were immediately attracted by his new business. However, instead of actually investing the money, Ponzi just redistributed new incoming funds to old investors, telling them that they made a profit. The scheme lasted until August 1920 when The Boston Post began investigating Ponzi's company. As a result of the investigation, Ponzi was arrested by federal authorities and charged with several counts of mail fraud.

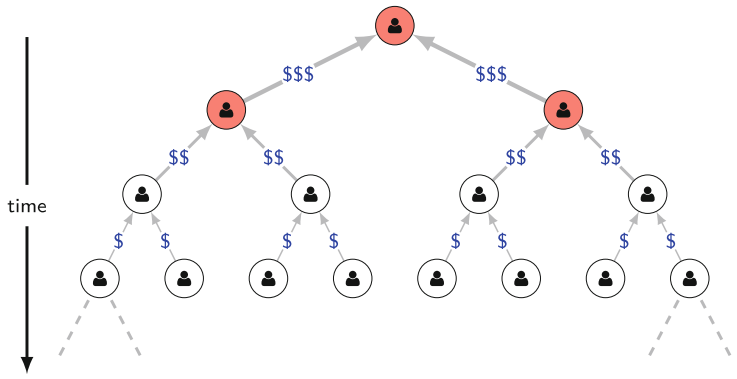

Fig. 4.6 Pyramidal structure of a typical Ponzi scheme. Organizers (red-colored) at the top of the pyramid make a profit at the expenses of subsequent investors (white-colored)

Countermeasures

All previous forms of market manipulation are well-known to regulators and have been banned from regulated markets for many years. For some of them, specific mechanisms have also been put in place in order to automatically prevent fraud. For example, several markets now make it impossible to buy or sell one's own stocks, thus making it more difficult to carry out wash trading and churning manipulations. However, as anticipated, the relatively new and decentralized cryptocurrency exchanges still lack deployed technological tools and regulatory frameworks to prevent some of these frauds. In particular, the combination of cryptocurrency exchanges with the speed and anonymity offered by new media, such as OSNs and encrypted peer-to-peer messaging apps, set the stage for the comeback of frauds, such as P&D and Ponzi schemes. A new stream of research is tackling this recent wave of trade-based market manipulations.

Given that modern trade-based market manipulations deeply leverage social media, some preliminary studies aimed to characterize online social media discussions about cryptocurrencies. In detail, the study in [238] investigated Reddit discussions about a few notable coins—namely, Bitcoin, Ethereum, and Monero. Authors obtained interesting results particularly with regards to Monero, a coin that is often used for shady transactions in the Dark Web. They find that information cascades about Monero are longer and wider than those of the other coins, implying a larger and prolonged interest. In turn, this stressed that one of the main reasons for online interest and debate around cryptocurrencies are indeed cryptocurrency frauds. Furthermore, modeling the discussion patterns around legitimate cryptocurrency transactions, a byproduct of the study in [238] also paves the way for future systems capable of detecting suspicious online discussions, thus possibly leading to the development of automatic fraud detection techniques. Another pioneering study is discussed in [237]. Here, the authors collected a large unbiased sample of online cryptocurrency conversations on Twitter, Discord, and Telegram. Contrarily to the majority of other works, the study did not focus on specific coins or specific frauds, but is aimed at mapping the online cryptocurrency ecosystem considering both legitimate discussions as well as possible manipulations. In addition, they also mapped the interplay between the different platforms by collecting and analyzing invite links to Discord and Telegram closed groups that were shared on Twitter. The analysis in [237] ultimately uncovered the existence of many users and channels/groups involved in P&D and Ponzi schemes. In detail, they estimated that around 20% of all Telegram channels related to cryptocurrencies are in fact involved in either P&D frauds or Ponzi schemes. On the contrary, Discord appears to be a rather healthy online ecosystem, for what concerns cryptocurrencies, with only one Discord channel involved in manipulative activities. Authors also found that more than 56% of the Twitter users that shared invite links to Telegram and Discord were in fact social bots. Such bots were used as a cheap, large-scale, expendable way of luring unaware investors into the secluded Discord and Telegram groups, thus efficiently fueling the P&D and Ponzi frauds [237].

Resources
The dataset used in [237] for documenting P&D fraunds and Ponzi schemes is publicly available online.[12] The dataset contains cryptocurrency-related messages collected from Twitter, Telegram, and Discord, as well as network information about invite links shared on Twitter for joining Telegram and Discord channels.

In contrast to the two previous large-breadth studies, the majority of other works focused on investigating given manipulations for a given set of coins, typically on a single social media. As an example, the work presented in [239] provided a detailed analysis of how P&D schemes occur in Telegram. The authors also developed a basic machine learning model for predicting the likelihood of a coin being the target for manipulation. Then, they used the trained model to inform a simple trading strategy. Interestingly, they empirically demonstrated that the strategy based on early investments on coins that are likely to be pumped allowed to generate a return as high as 60%, within a period of 2 and a half months. This result goes a long way in demonstrating the effectiveness of P&D frauds. Other studies observed that P&D phenomena are widespread on both Discord and Telegram [240] and evaluated the effects of such fraud on the liquidity and price of cryptocurrencies. Among the main findings of the study, large importance is played by coin ranking in predicting the success of the pump operation. In fact, while there are attempts to pump coins spanning a wide range of popularity, pumping obscure coins gave the pump scheme the potential for greater success at the expense of increased risk, such as volatility. The study in [241] is among the few ones that adopted a predictive approach, which focused on detecting the presence and on estimating the success of P&D scams. Based on Twitter and Telegram data, the authors evaluated a computational approach to automatically identify P&D scams as they unfold. Besides, they also developed a multimodal (i.e., textual and visual) approach for predicting whether a particular pump attempt will likely succeed or not, in terms of meeting the expected price target. Finally, they also analyzed the prevalence of bots in cryptocurrency-related tweets, uncovering a significant presence of bots during coin pumping operations, thus corroborating results obtained in [237]. Another predictive study is described in [242]. The authors analyzed trading data related to P&D manipulations, identifying anomalous behaviors. Based on this finding, they proposed the application of anomaly detection techniques to locate points of anomalous trading activity, thus flagging potential P&D attempts.

P&D manipulations are not the only financial fraud under scrutiny. In [243], the authors investigated online Ponzi schemes publicized on the *Bitcointalk* discussion forum. They used survival analysis to identify factors that affect the success of Ponzi scams, finding that credible and active scammers are more likely to succeed than

[12] https://doi.org/10.5281/zenodo.3895021 (Last checked August 2020).

newer and less active ones. In [244] is proposed an automatic approach for detecting Ponzi schemes. The authors analyzed the Bitcoin blockchain and extracted features related to known Bitcoin Ponzi schemes. Then, they used the known schemes as the ground-truth for training a machine learning classifier.

Open Issues and Future Directions

As testified by the previous literature review, the majority of existing studies about modern trade-based market manipulation are descriptive in nature. This approach is motivated by the recency of these frauds and by the need to understand how manipulations are orchestrated, in order to start designing defensive mechanisms. In this regard, current studies are building the knowledge that is required for subsequent predictive works. In addition, results of these studies have important implications for regulatory policies, which are still lagging behind, and that greatly benefit from sound evidence of the different forms of market manipulation, including manipulated assets, targeted markets, involved groups, and platforms.

Despite the generalized focus on descriptive and observational approaches, some scholars already moved forward and tried proposing automatic techniques for detecting widespread frauds such as P&D and Ponzi schemes. Although preliminary, these endeavors contribute to the development of early warning systems that are capable of promptly detecting coins and other assets that are likely to become targets of manipulation, for instance, to temporarily suspend trades involving those assets, thus negating any benefit to fraudsters. For the future, additional efforts will likely follow this research direction. More predictive studies are to be expected, putting to good use the preliminary findings of previous works.

Finally, we also highlighted that the majority of existing studies focused on specific frauds (e.g., P&D and Ponzi schemes), platforms (e.g., Twitter and Telegram), and coins (e.g., Bitcoin). Given this narrow focus, we currently lack a thorough understanding of the whole extent of these market manipulations. For the future, it would be advisable to broaden the scope of subsequent studies or to concentrate on those frauds, platforms, and coins that have not been investigated before, thus contributing to the sketch of a complete picture.

4.1.3 Threat: Algorithm-Based Manipulation

In the two previous sections, we covered the main types of market manipulation, which are either information-based or trade-based. Over the course of the years, market manipulation techniques belonging to these two families represented the main threats to fair and free markets. However, the recent advancements in algorithmic trading, as well as the growing adoption of Artificial Intelligence (AI) in almost all corners of FinTech, anticipate the advent of unprecedented forms of

manipulation. Before digging deep into the new perils posed by automated and AI-driven markets, we first take a step back by discussing the current penetration of algorithms and AI in financial markets and in FinTech at large.

Algorithms Taking Over

As discussed throughout this chapter, the FinTech revolution gradually brought more data, algorithms, and computational power into the hands of financial analysts, which, in turn, led to the development of always more complex data-driven models. *Quant* (short for *quantitative*) traders and portfolio managers were the first to deploy data analysis to improve financial operations in an algorithmic framework. Using mostly daily data and armed with the latest inferences from statistics and physics, the quants sought answers to challenges associated with portfolio risk management, derivative pricing, and diversification. Their early work paved the way for modern Exchange-Traded Funds (ETFs): passively managed, yet actively traded indexes. Motivated by the early success of quantitative approaches, algorithms spread through all the corners of finance. Counterparty risk computation deals with the quantification of the risk of payment by a money-sending party. As an example of the algorithmic takeover, in early 2000s counterparty risk was managed by human traders, and all settlements took at least 3 business days to complete, as multiple levels of verification and extensive paper trails were required to ensure that transactions actually took place as reported. Fast-forward to today, fast technology and fully automated systems enable transfer and confirmation of payments in just a few seconds, fueling a growing market for cashless transactions. Algorithms are making their presence felt in wealth management as well, as *robo-advising* (or robo-investing) is taking over the job of traditional portfolio managers. Although the idea has been around for a while, since around 2015 the momentum has started to grow. The core concept of robo-advising is that a computer, equipped with cutting-edge predictive algorithms, is capable of delivering portfolio optimized solutions faster, cheaper, and at least as good as its human counterparts (i.e., the old-school portfolio managers). Robo-advising now enables investors to use technology to place their money in well-diversified asset pools, at a much lower cost. Given a selected input of parameters to determine the customer's risk aversion and other preferences (say, the customer's life stage and its personal aversion to given stocks), algorithms then output an optimal and personalized investing plan. Automation of investment advice enables fast market risk estimation and the associated custom portfolio management. For example, investors of all stripes can now choose to forgo expensive money managers in favor of investing platforms such as Motif Investing. For less than $10, investors can buy baskets of ETFs preselected based on particular themes. Moreover, companies such as AbleMarkets offer real-time AI-driven risk evaluation of markets, aiding the judgment of market-making and execution traders with real-time inferences from the market data, including the proportion of high-frequency traders and institutional investors present in the markets at any given time. Algorithms and AI are also permeating the operations of hedge funds. BlackRock

is replacing human stock pickers with machine algorithms, using deep learning neural networks. Sentient Technologies is a hedge fund run entirely using AI. It is supposed to have a proprietary algorithm with adaptive learning that uses thousands of machines for computation. Numerai is a hedge fund that makes ensemble trades by aggregating trading algorithms submitted by anonymous contributors. Prizes are awarded to contributors in a cryptocurrency called Numeraire, which resides on the Ethereum blockchain. The latest data shows that funds using AI outperform others quite handily,[13] with a gap that is only bound to widen. Consequently, increasing shares of money are marching into AI-driven funds, as Numerai raised more than $1 million in short order.[14]

Loose Cannons on the Automation Deck

The previous examples, which are just a few out of an immense spectrum of financial applications, highlight the usefulness of algorithms and AI for solving a plethora of real-world tasks. This usefulness is well motivated by the exceptional—even beyond human-like—performance of recent ML/AI algorithms. Especially with the recent progress in AI, which relies on increasingly complicated artificial neural network architectures, the predictive power of algorithms advances in contrast with their interpretability [206], as summarized in Fig. 4.7. On the one hand, we have a large set of different neural network architectures, configurations, and learning algorithms that allow us to effectively automate and speed up many challenging tasks. On the other hand, however, oftentimes we have no clue as to why our latest intelligent system managed to obtain state-of-the-art performance on a specific benchmark dataset or why it yielded *that* particular prediction in the face of a given input instance. This issue is particularly acute in finance, where the trade-off between predictive power and interpretability is completely in favor of the former. Our understanding of the inner functioning of these complex techniques is such that, to many eyes, AI just *automagically*[15] solves increasingly challenging tasks.

Moreover, the performance of AI algorithms is far from being easily generalizable. In fact, almost all such techniques manage to obtain exceptional results only when they operate in their best conditions (i.e., in their sweet spots), such as when analyzing data that is very similar to those used for training them. Since the early days of ML and AI, it is known that instances analyzed at "test" time (i.e., after the model has been trained, when it is deployed) might have somewhat different statistical properties than those used for training the model. This issue

[13] https://www.preqin.com/insights/research/blogs/the-rise-of-the-machines-ai-funds-are-outperforming-the-hedge-fund-benchmark (Last checked August 2020).

[14] https://www.sec.gov/Archives/edgar/data/1667103/000166710316000002/xslFormDX01/primary_doc.xml (Last checked August 2020).

[15] A colloquial portmanteau used in computer science to indicate something effective, yet unintelligible.

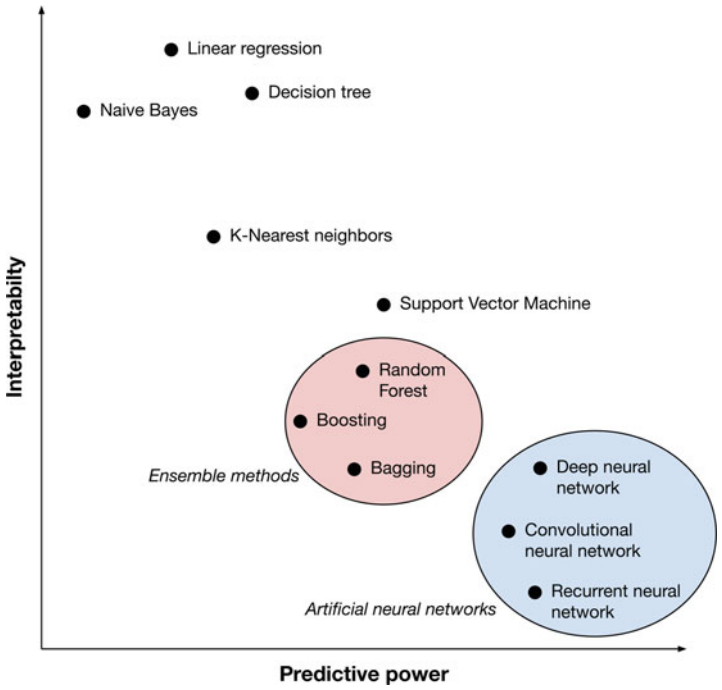

Fig. 4.7 The trade-off between performance (i.e., predictive power) and interpretability in machine learning and artificial intelligence algorithms. Latest advances (e.g., deep neural networks) provide unparalleled performance at the cost of limited interpretability

naturally occurs in many applications and results in degraded performance. For example, in a computer vision system (i.e., such as those used for automatically "reading" and "interpreting" road signs in self-driving vehicles), two different cameras could be used to take pictures at training and test time, thus leading to a trained model achieving degraded performance on test images. Generalizing the previous example, this detrimental effect on performance occurs because the vast majority of algorithms are designed to operate in stationary and neutral (if not even in benign) environments. In practice, however, there are many situations in which the working environment is neither stationary nor neutral. When the previous assumptions on the environment are violated, algorithms operate in sub-optimal conditions and even the best ones start making big mistakes, thus yielding inaccurate and unreliable predictions. In the previous computer vision example, using two different cameras violated the stationarity assumption, since the properties of the images changed when switching from training to test data.

An even more worrying scenario, however, occurs when an attacker deliberately manipulates training or test instances, in order to cause the algorithm to make mistakes. In fact, one can easily imagine several situations in which attackers could have an interest in fooling a machine learning system. Think for example

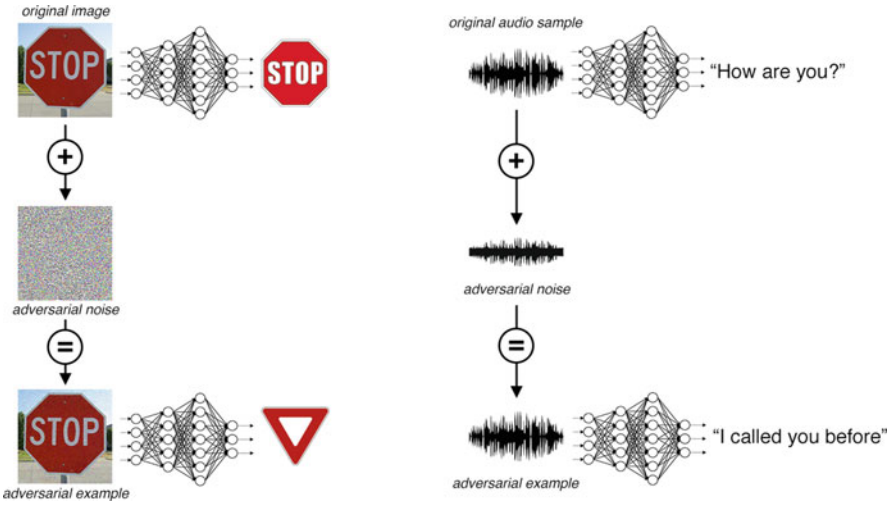

Fig. 4.8 Adversarial examples used to fool computer vision and speech recognition systems. The deceptive adversarial examples are obtained by adding a carefully crafted adversarial noise to the original data instances

of all security applications of machine learning. By definition, the goal of security systems is to keep at bay unwanted accesses (i.e., to data, systems, etc.). Attackers—or adversaries, as they are often called—are thus obviously motivated to fool security systems. One straightforward way to reach their goal is by modifying their data footprint in such a way that security systems misclassify them as legitimate and genuine users. In other words, adversaries make it so that the neutrality assumption of machine learning algorithms is violated. Moreover, they often also violate the stationarity assumption by modifying their behavior through time in order to continually evade detection, thus creating data instances with evolving characteristics. Such intentionally modified data instances are called "adversarial examples", since they are generated by an adversary. Figure 4.8 shows two simple examples of adversarial attacks that can be mounted against computer vision and speech recognition systems. In each case, the original data instance is modified by adding a carefully crafted, subtle adversarial noise, so as to induce classification errors. Interestingly, it is often possible to produce adversarial examples that appear to humans as almost identical to the original data instances, but that nonetheless produce completely wrong classifications in automated systems, as in the examples shown in the figure. As demonstrated by the previous examples, the presence of adversaries is not only related to security systems but rather to all systems that could be gamed in order to gain an advantage, such as an economic or an information advantage. Quite obviously, trading algorithms are a paramount example of this kind of machine learning systems.

The current situation with AI and ML is one where algorithms are like *loose cannons on the automation deck*. On the one hand, when working in the best

conditions, they are capable of providing unparalleled performance. On the other hand, however, their power is "unstable." What's more, it can be weaponized, since it can be easily manipulated by knowledgeable adversaries. The pervasiveness of ML/AI and of algorithmic decisions—both in finance and elsewhere—combined with their vulnerabilities, thus represent an explosive mixture for our increasingly automated world.

Attacks

Because of the huge range of existing situations where adversaries are potentially motivated in fooling a machine learning system, a recent branch of research started focusing on the development of robust ML and AI algorithms, capable of withstanding attackers. Adversarial machine learning, also known as machine learning in hostile environments, is a paradigm that deals with the development of algorithms designed for withstanding the attacks of an adversary. It focuses on the study of possible attacks (e.g., how to modify an image in order to fool a computer vision system) and on the design of countermeasures to those attacks (e.g., smoothing the decision boundary of a classifier so as to make it more robust against adversarial examples). That is, adversarial machine learning studies possible attacks to build more robust and more secure systems. The idea of considering the presence of adversaries for improving the robustness of existing systems has been embraced by the ML community only recently. In fact, despite the original ideas date back to 2004–2005, adversarial machine learning is considered to be still in its infancy [245]. Nevertheless, it is currently regarded as one of the most promising directions of research on ML and AI. The reasons for the great interest around this novel paradigm are partly due to the pervasiveness of the weaknesses it aims to overcome. Indeed, it has been demonstrated that almost every ML system is potentially vulnerable to adversarial attacks, since adversarial examples often transfer from one model to another. In fact, adversarial attacks can be successful even when adversaries have only limited knowledge of the model's characteristics.

In the remainder of this section, we provide a high-level overview of the types of attacks that can be mounted against ML and AI systems. For a detailed analysis of the technical and mathematical foundations of both attacks and countermeasures, we point interested readers to any of the extensive surveys on this topic [246, 247]. Figure 4.9 sketches the landscape of the existing classes of attacks that can be categorized according to the attacker's goals, capabilities, and information. Regarding attacks to machine learning models, an adversary can attempt to manipulate either the collection or the processing of data to corrupt the target model, thus altering the intended output. A first taxonomy categorizes adversarial attacks according to the capabilities of an attacker, in growing order of attack complexity, as detailed in the following.

Exploratory Attacks Given a "black-box" access to the model, exploratory attacks aim at gaining as much knowledge as possible about the learning algorithm of the

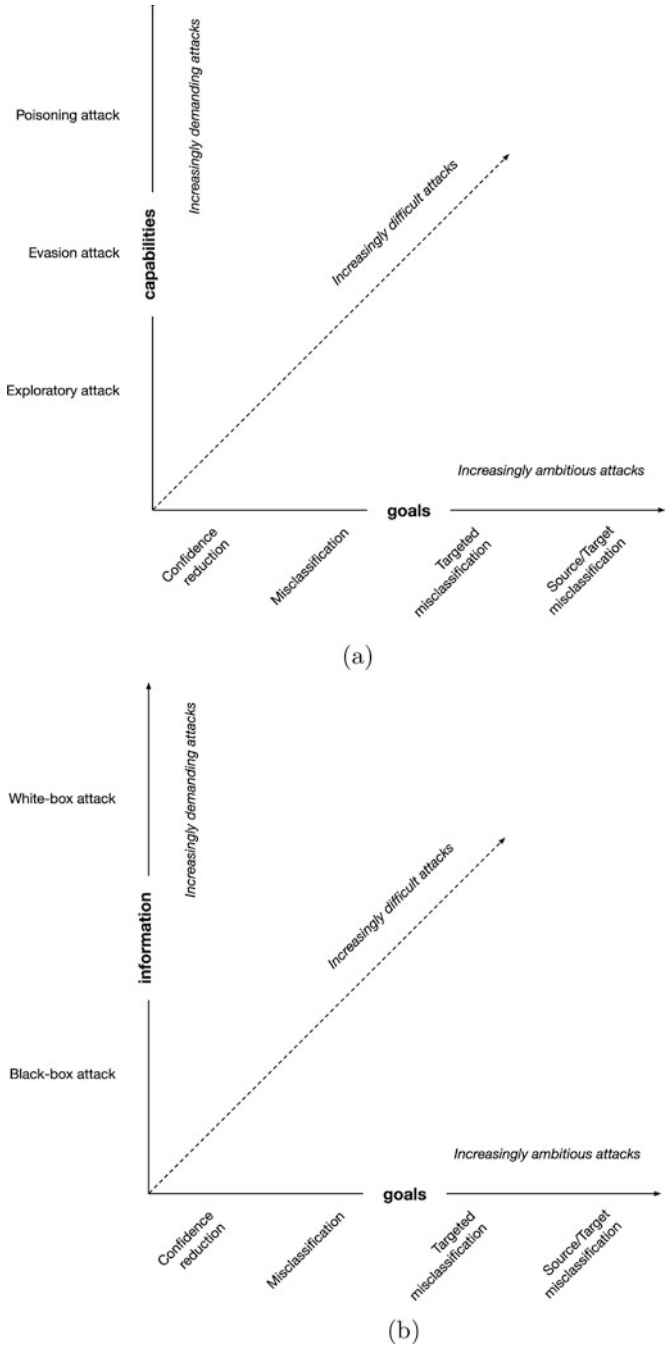

Fig. 4.9 Categorization of adversarial attacks according to the attacker's (1) goals, (2) capabilities, and (3) information. Attacks that are more ambitious and that require better capabilities or more information are considered to be more difficult. (**a**) Goals vs. capabilities. (**b**) Goals vs. information

underlying system and the data patterns exploited by the model. For example, this can be done by feeding a diverse set of inputs to the model in order to reverse engineer its functioning. In turn, acquiring knowledge about the model can help the attacker to circumvent detection. Exploratory attacks require modifying neither the training nor the test data.

Evasion Attacks The adversary tries to evade detection by the system by maliciously adjusting its data footprint at testing time. Evasion attacks evade the ML model by passing an adversarial example so that the model misclassifies it. This setting does not assume any influence over the training data. However, it assumes influence on the test data, and it represents the most common type of attack in the adversarial setting.

Poisoning Attacks This type of attack, also known as contamination of the training data, takes place during the training phase of the machine learning model. An adversary tries to poison the training data by injecting carefully designed data samples, so as to compromise the whole learning process. The attacker adds such adversarial examples to the training data so that the model's decision boundary can be manipulated. For example, the malicious training samples could cause the model to learn a wider decision area for the legitimate class, thus possibly including also some malicious input that the attacker could later exploit. This type of attack represents the most demanding one, since it requires the attacker to be able to tamper with training data—that is, with the training process of the model.

Another categorization, orthogonal to the one presented above, organizes attacks according to the information that the attacker has access to.

White-Box Attacks In white-box attacks on a machine learning model, an adversary has total knowledge of the model (e.g., the type of neural network along with the number of layers and the internal weights). The adversary utilizes this information to identify the feature space where the model may be vulnerable, such as those areas of the feature space for which the model has a high error rate. Then, this information can be used at test time by feeding the model with artfully crafted adversarial examples that lay in the high-error area, thus triggering inaccurate predictions. The possibility to access internal model weights for a white-box attack represents a very powerful adversarial attack.

Black-Box Attacks Black-box attacks, on the contrary, assume no knowledge about the model and use information about the settings or past inputs to analyze vulnerabilities. For example, in an oracle attack, the adversary exploits a model by providing a series of carefully crafted inputs and observing the corresponding model outputs.

Finally, attacks are also categorized based on the goals of the adversaries, as shown in the following.

Confidence Reduction The adversary aims to reduce the confidence of the prediction of the target model. For example, a modified instance (i.e., an adversarial

example) of a given class can be correctly predicted, but with lower confidence than the unmodified one.

Misclassification The adversary aims to alter the output classification of an input to *any* class different from the correct class.

Targeted Misclassification The adversary produces adversarial examples that force the system to yield *a given* target classification. In other words, this goal aims at forcing a target output class, without putting restrictions on the input data.

Source/Target Misclassification The adversary aims to force the output of the classification for a given input to be of a given target class. This is the most ambitious goal since it is the one having the strictest requirements, involving both input data and the output class.

Countermeasures

Given the previous attacks to ML and AI systems, several defense strategies have been developed. While some of them are high-level, generic guidelines for training robust models, others are low-level and tailored for specific types of models, such as Support Vector Machines and deep neural networks [247]. In the reminder of this section, we briefly describe the most widely adopted of such defense strategies.

Adversarial Training The primary objective of the adversarial training is to increase model robustness by injecting adversarial examples into the training set. Adversarial training is a generic and broadly applicable brute force approach where the defender simply generates many adversarial examples and augments the targeted model with these examples during its training phase. The augmentation can be done either by feeding the model with both the original data and the crafted data or by learning with a modified objective function. This defense strategy can be easily implemented, since it only requires the generation of adversarial examples.

Gradient Hiding A large family of adversarial attacks exploits the known characteristics of the attacked model. Among the most important of such characteristics is the gradient of deep neural network models, which is used during the training phase of the model to tune the model's parameters. However, the same information can also be used at test time to tune adversarial inputs so as to trigger wrong classifications by the model. The gradient hiding technique represents a natural defense against gradient-based attacks and attacks using adversarial crafting methods (e.g., FGSM [248]) and simply consists of hiding the information about the model's gradient from the adversary. For instance, if the model is non-differentiable (e.g., decision trees, nearest neighbor classifiers, random forests), gradient-based attacks are rendered ineffective.

Defensive Distillation Let us assume that we already have a neural network that classifies a training dataset into the target classes. The final softmax layer of the neural network produces a probability distribution over the target classes. Let us

also assume that we want to train a second neural network on the same dataset, achieving the same performance. Now, instead of using the target class labels of the training dataset, we use the output of the first network to train the second neural network. Notably, the second network has the same architecture as the first one and uses the same input dataset. The new labels used for training the second network thus contain more information about the membership of input data instances to the different target classes, compared to the simple crisp labels used by the first network. The advantage of training the second model using this approach is that of obtaining a smoother loss function, which better generalizes for an unknown dataset, with high classification accuracy even for adversarial examples.

Feature Squeezing Feature squeezing is another model hardening technique. The main idea behind this defense is that it reduces the complexity of the data representation so that the adversarial perturbations disappear because of the low sensitivity of the new data representation. Though these techniques work well in preventing adversarial attacks, they might have the collateral effect of worsening the accuracy of the model on the true examples as a consequence of the coarse-grained data representation.

Basis Function Transformations This broad family of techniques investigates various defense mechanisms like Principal Component Analysis (PCA), low-pass filtering, JPEG compression (for images), and soft thresholding that alter input data based on basis function representations. All these mechanisms are applied as a preprocessing step on both adversarial and legitimate data instances. The efficiency of each technique is evaluated in terms of its success at distinguishing between adversarial and real inputs.

NULL Labeling One of the main reasons behind the defeat of most of the well-known defense mechanisms is due to the strong transferability property in neural networks. This property implies that adversarial examples generated on one classifier are expected to cause another classifier to perform the same mistakes. This property holds true even if the classifiers have different architectures or even if they have been trained on disjoint datasets. Hence, one profitable way for protecting against black-box attacks is to block the transferability property of the adversarial examples. NULL labeling is a method organized in three steps to prevent the adversarial examples to transfer from one network to another. The main idea behind the proposed approach is to add a new "NULL" class in the dataset and to train the classifier to reject the adversarial examples by classifying them as NULL. The advantage of this method is the labeling of the perturbed inputs to the NULL class, instead of classifying them into their original labels. This method is accurate to reject an adversarial example while not compromising the accuracy of the clean data [249].

Definitions

Generative Adversarial Network (GAN) Machine learning framework where two competing deep learning networks are jointly trained in a game-theoretic setting. A GAN is composed of a *generator* network that creates data instances and a *discriminator* network that classifies data instances. The goal of the generator is that of creating synthetic data instances that resemble the properties of real organic data, while the goal of the discriminator is to classify input data instances as either synthetic or organic. The discriminator is evaluated based on its binary classification performance, while the generator is evaluated in terms of its capacity to induce errors in the discriminator, hence the competition between the two networks.

Defense-GAN Another family of defensive techniques is based on the application of Generative Adversarial Networks [250]. In particular, the core idea of defense-GANs is to project each input instance onto the range of the generator of a GAN by minimizing the reconstruction error, before feeding the input instance to the classifier. Due to this preliminary step introduced by defense-GANs, all adversarial perturbations are greatly dampened and with reduced efficacy.

Open Issues and Future Directions

The previous attacks and countermeasures represent one of the most novel advancements in ML and AI. As such, their application to diverse practical scenarios is still limited. For example, many of the discussed techniques have been originally developed for fields such as computer vision [251], speech recognition [252], and Natural Language Processing (NLP) [253]—that is, those areas where deep learning and AI thrive the most. Recently, the application of adversarial machine learning has also started in fields such as automation detection on the Web (e.g., social bot detection) [254–256] and fake news detection [257]. Some recent works have also seen the application of GANs as data generators for the task of stock market prediction [258–261]. However, in all these studies, the focus was on improving the performance of market predictions systems, rather than on assessing their vulnerabilities or improving their robustness. Unfortunately, as previously anticipated in this section, in FinTech the trade-offs between predictive power and security, robustness, and interpretability are totally in favor of the former. Therefore, we can confidently claim that adversarial machine learning has never been applied to ML and AI systems employed for market prediction, nor for other finance-related tasks, for improving their robustness and reliability. On the one hand, it might appear reassuring that no vulnerability has still been found in market prediction, or in other similar, systems. However, on the other hand, the reason for this is to be attributed to the lack of investigations rather than to the robustness of such systems. In fact,

all theoretical and practical results obtained so far suggest that basically all ML and AI systems are vulnerable to adversarial attacks, unless adequately defended. It is thus just a matter of time before some knowledgeable attacker will exploit the algorithmic vulnerabilities of Automatic Trading systems for its own profit, or worse, for starting an economic war.

Given this situation, the main effort to be expected for the coming years revolves around the assessment of the vulnerabilities to adversarial attacks of the existing algorithm-driven FinTech systems. Based on the results of this assessment, it would then be of the utmost importance to devote efforts toward the deployment of existing defensive techniques, such as those we previously discussed, or toward the development of new, specific ones—all with the ultimate goal of safeguarding our financial systems from targeted algorithmic manipulation.

4.1.4 Other Countermeasures

In the previous sections, we separately addressed different types of market manipulation, describing the main attacks and the corresponding specific countermeasures. Here, we conclude our discussion on financial market manipulation by briefly surveying generic and context-agnostic defenses to market manipulation.

The first, and simplest, generation of generic market manipulation detection methods leverages only raw market data. These are simply anomaly detection techniques implemented via unsupervised algorithms or via rule-based systems that analyze the time series of market data. In case of an anomalous deviation from the expected values or the permissible intervals allowed by the model, an alert is triggered. This alert signal can be used to automatically halt transactions around a given financial instrument or to solicit manual investigations by human analysts. These methods are very generic and simple, and, as a result, they are easy to calibrate and to apply for any market, time granularity, and financial instrument [211]. However, this simplicity is counterbalanced by a limited capability of detecting sophisticated manipulations. The second group of methods still leverages market data, but this time the core idea is that of building a model of a specific market or instrument and using that model to spot anomalous behaviors. In particular, the model is used to forecast the market (or part of the market). Contrarily to the majority of market prediction systems, these forecasts are not used for driving investments but as a baseline against which to compare the actual market movements. As such, these models need not be exceptionally accurate. Then, at any given point in time, the actual market behavior is compared with the prediction given by the model for the same time. If the observed behavior diverges significantly from the expected one, as predicted by the model, an anomaly is detected. Among the algorithms that can be used to create market models are many well-known supervised machine learning techniques such as decision trees, neural networks, regression models, and Support Vector Machines. Contrarily to the first family of manipulation detection techniques, some of these methods require a large number

of parameters to be correctly tuned, in order for the algorithm to be effective. This means that the increased detection performance of these methods is counterbalanced by the need for continuous and accurate model calibration [211].

Market manipulation often requires the coordinated activity of several different actors that work in conjunction to alter the value of a financial instrument. Leveraging this notion, another broad group of detection techniques is based on similarity analyses. In particular, trade networks have been analyzed with the goal of spotting pools of manipulators via graph clustering algorithms [262]. Similarly, other graph clustering techniques, such as Markov cluster algorithm, have been used to detect cyclical trading [263]. Finally, the individual trading activities of different investors have been compared to one another in order to spot suspicious similarities, and the detection of colluding investors was carried out via spectral analysis [264, 265].

The previous techniques leverage raw market data and coordinated behaviors to detect market manipulation. In addition to these, other interesting techniques also exploit human psychology. In fact, it has long been demonstrated that catastrophic market crashes, such as those that are potentially dangerous for a national economy, are driven by panic as much as by economic factors. Because of this, predicting panic is of critical importance not only in many areas of human behavior but also in the context of market dynamics. For the latter case, studies demonstrated that panic may be due to specific external threats but also to self-generated nervousness among investors [266]. It has been shown that both long-lasting economic crises as well as sudden single-day crashes were preceded by extended periods of so-called *market mimicry*—an abstract concept related to the extent to which investors look at one another for cues. When the mimicry is high, the market becomes less influenced by external news and more so by internal dynamics, which are often driven by uncertainty and nervousness. In addition, many stocks follow each other's movements, thus setting the stage for extensive panic to take hold. Ultimately, it has been shown that high levels of mimicry represent a quite general indicator of the potential for self-organized market crises [266]. As such, continuous monitoring of market mimicry can allow to anticipate and, possibly prevent, large market crises.

4.2 Scenario 2: High-Frequency Trading

"Move fast and break things" was an early motto at Facebook, intended to push developers to take risks. The inspirational phrase appeared on office posters and even featured in a letter from Mark Zuckerberg to investors in 2012. Over time, it came to be embraced as a mantra broadly applicable to all technological disruption, and, as such, it was adopted by countless entrepreneurs. As we already investigated in the previous sections of this chapter, the unfolding technological revolution that propels FinTech brought massive changes to the way paying, trading, and investing are done. FinTech thrives on technology, and, in fact, looking back at the last

decades of technological advancements, one can confidently say that this revolution disrupted and reinvented the FinTech landscape multiple times.

> **Definitions**
> **Automatic Trading (AT)** occurs whenever a computer algorithm automatically determines when to initiate or cancel orders, with limited or no human interaction.
> In addition to the previous characteristics, **High-Frequency Trading (HFT)** also demands an infrastructure that minimizes network and other types of latencies using specific facilities as co-location, proximity hosting, or high-speed direct electronic access. HFT is thus a subset and a specialization of AT [267].

One crucial dimension along which technological advancement unfolds is *speed*. Nowadays, thanks to High-Frequency Trading (HFT) and Automatic Trading (AT), transactions happen in microsecond timeframes and unprecedented volumes. Trading strategies based on HFT and AT were even dubbed "flash trading," in recognition of the sheer speeds involved, which at times, can be orders of magnitude faster than the blink of a human eye. To reach such inhuman peaks in trading speed, FinTech giants moved mountains—almost literally. In his bestselling book *Flash Boys* [268], Michael Lewis describes a 2011 $300 million project for the construction of an 827-mile (1331 km) tunnel, hosting fiber optics cables, that cuts straight through mountains and rivers from Chicago to New Jersey. This once titanic endeavor allowed to reduce the roundtrip time from 14.6 to 14.1 ms. This was later rendered obsolete by the construction of a microwave link, which follows an even straighter route. The new air route also takes advantage of the faster speed of signal travel that is possible through the air, as compared to signal travel speed through glass fibers, which slows light down. With these two advantages, this new link shaved 4.5 ms off the fiber optics cables. As copper cables got replaced by fiber optic cables and as radio waves did the same by microwaves, also the latter was eventually rendered obsolete by laser beams. In the EU, the laser beam link between London and Frankfurt is currently being rerouted via Dunkerque and away from Calais, in an effort that is bound to save a few nanoseconds [269]. If you think these tiny gains are not worth the huge investment, think twice.

> **Definitions**
> **Arbitrage** is the legit practice of capitalizing upon a price difference between two markets. The profit results from the difference between the market prices at which the unit is traded.

(continued)

Front running (or **tailgating**) is the illicit practice of issuing a trade based on nonpublic knowledge of other pending transactions that will influence the price of the traded security.

By leveraging (1) high-performance computers; (2) co-location services, as well as individual data feeds, to minimize network and data latencies; and (3) extraordinary high-speed algorithms for generating, routing, and executing transactions, HFT is capable of simultaneously analyzing different global markets and establishing and liquidating positions in very short timeframes, based on real-time market conditions. HFT is thus an advanced technology that opens up new trading possibilities for its adopters, by profiting from lightning-fast analyses and transactions with respect to slower traders. This results in the possibility to profit from even minor price differences across different markets. In fact, high-frequency traders typically benefit more from a large number of minor transactions than from a few particularly significant ones, as manual traders do. Obviously, this is only made possible by the opportunity to monitor markets at large and to immediately benefit from the slightest price differences. Because of this advantage over traditional traders, HFT is often used to perform arbitrage. In addition to the legit practice of arbitrage, HFT has also been used to obtain unfair advantages. One way to do so is via front running, sometimes done even at the expense of one's own clients. This occurs when an HFT firm races ahead of a large client order, scooping up all the shares on offer at various other exchanges (if the client is issuing a buy order) or hitting all the bids (if it is a sell order) and then turning around and selling them to (or buying them from) the client and pocketing the difference [268].

Resources

The existing literature on HFT is based on datasets of four different types [267]:

- Data for equity trading on NASDAQ.
- Data on trading in the E-Mini.
- Data used by CFTC and SEC to release their report on the market disruption that occurred in the case of the 2010 Flash Crash.[16]
- Datasets made available to researchers by exchanges and regulators. These require proxies for identifying HFT activity.

Given the disruptive changes brought by HFT and its uses, sometimes shady, it comes with little surprise that there exists a vast academic literature and a heated

[16] https://www.sec.gov/news/studies/2010/marketevents-report.pdf (Last checked August 2020).

debate on AT and HFT and on their effects on markets. Many papers discuss their roles in capital markets as well as their trading strategies and consequences for market quality. Similarly, market regulators have expressed concerns about the growing participation of HFT and the costs associated with monitoring their activities. Among the existing literature, it is striking that the majority of the most cited papers actually link HFT to positive market effects. In fact, it has been found that HFT tends to reduce information asymmetry between buyers and sellers over time. Despite frequent accusations of arbitrage, many empirical studies demonstrated a general improvement in market liquidity, measured by the reduction of spreads or the increase in depth, and a general reduction of the intraday price volatility [210]. Some examples are in the following studies, showing evidence that HFT stabilizes markets [270], improves market quality and reduces bid-ask spreads [271], and reduces trading costs [272]. These results ultimately seem to suggest that any regulatory action introduced to curtail this activity may have serious negative implications for liquidity and market participants, as also demonstrated recently both in France and in Italy [267].

The previous results reported overall positive effects of HFT on markets, when markets operate under "normal" conditions. However, radically different results have been obtained when studying the role of HFT in markets during distressed times, such as in the case of flash crashes. The signs that HFT could play a negative role in the emergence of systemic crashes surfaced since the infamous 2010 Flash Crash. Years of investigations led to trace back a possible cause of the crash to the London-based trader Navinder Singh Sarao, who used an automated trading program to manipulate the market via spoofing—offering $200 million worth of fake bets that drove prices down, modifying them 19,000 times, and then canceling all of them before they could be completed. As the market fell, he sold futures contracts. When the market began to recover, he bought futures back and sold them again at a higher price [273]. On a large scale, this behavior can have potentially catastrophic effects if a chain reaction of instability sets in. In the years following the 2010 Flash Crash, several studies reported that HFT is not beneficial to the stock market during flash crashes and actually consumes liquidity when it is most needed. HFT exacerbates the transient price impact, unrelated to fundamentals, typically observed during a flash crash [274]. Other studies report that HFT functions essentially as an accelerator and a catalyst to already existing market dynamics, such as bubbles and crashes. More flash crashes, involving additional markets and instruments, can be expected in the future, resulting from the increasing interdependences between various financial instruments and asset classes. Within this evolving scenario, the technological race involving AT and HFT is not expected to provide a stabilization effect. On the contrary, the most recent results support the hypothesis that HFT can in fact lead to catastrophic market crashes and it could do it even more frequently in the future [275]. In light of these findings, the old Facebook motto *Move fast and break things* appears to suit HFT particularly well.

4.2.1 Threat: Technological Bias, Divide, and Monopoly

The preceding paragraphs highlighted the role that HFT can have in the formation of market crashes. They also discussed the strong dependence of HFT on the underlying technologies, which are responsible for its exceptional speed and performance. Similarly to our discussion about possible algorithmic manipulation of financial markets, the combination of technology in HFT and flash crashes opens up the possibility for state actors to perpetrate targeted manipulations. If the most performing (i.e., the fastest) HFT technologies are publicly available and all actors in the financial market rely on them, the single agent advantage is negligible. However, if one actor manages to develop or to acquire a system that is much more performing than those available to the other actors, it will gain a huge and unfair advantage. In fact, the major open problem with respect to the possible weaponization of HFT is related to technological bias and divide.

Here, we define technological bias as an asymmetry or an imbalance in the technology that is available to different economic actors. Similarly, technological divide represents the gap between the technology available to different actors. Although a certain degree of technological bias has always existed, if the resulting gap exceeds a certain threshold, its effects on financial markets may become significant. This asymmetry can even widen up to a point where the leading actor finds itself in a position of monopoly, derived from its enhanced capabilities of driving the market. To complicate matters, the fight for obtaining the upper hand in this situation of technological bias is highly dynamic. In fact, the technology behind HFT, and FinTech at large, is in constant evolution and those who adapt faster to new technology will inevitably hold a higher ground.

4.2.2 Attacks and Countermeasures

Until now, to the best of the existing knowledge, the technological bias in HFT has never been exploited to carry out massive manipulation operations or attacks to national assets and economies. In recent years, despite its growing significance, technological bias failed to attract much interest from academics. Nonetheless, it has worried other stakeholders that are more directly exposed to market dynamics, such as market traders and state decision-makers. For example, "slow traders" have been actively trying to avoid markets that are polluted by high-frequency traders, to not succumb to their greater firepower. Many finance professionals are constantly debating market structure and whether a new exchange can help slow traders to avoid high-frequency ones. Some firms even based their business on providing this sort of information, as AbleMarkets that delivers daily estimates of aggressive high-frequency traders acting across different markets. The negative effects of technological bias have been reported also for other areas of FinTech, in addition to HFT. For example, the improvements in market forecasting opened up by

AI and deep learning are often regarded as another potential factor for technological bias. In turn, this high-tech progress may create challenges for market efficiency, along with information asymmetry and irrationality of decision-making. Skilled traders can leverage this technological division for netting excess returns, at the expense of traders that adopt more traditional technologies [276]. Results reported in [276] for Forex trading are in contrast with the efficient-market hypothesis. The author concludes that with the progressive enhancement in computational methods and software, trading strategies will be vastly improved, with the inevitable result that some traders will be more successful than others. This process contradicts the classical definition of a market with perfect competition. However, it complies with the so-called adaptive-market hypothesis [277], according to which markets are not seen as efficient ecosystems, but rather as fiercely competitive ones. Since the market *ecology* changes over time, adaptation mistakes can occur as a consequence of the different degrees of adaptation of the participants. As a result, some of them will obtain more significant returns than others. Within this context, the technological shift is considered as a primary driver for change in market ecology [276].

While the previous considerations account for direct harms (e.g., immediate financial loss) caused by AT and HFT, others also raised attention on the indirect consequences (e.g., diminished confidence in financial markets). The latter might even have a bigger and worse impact than the former. In particular, HFT has changed those whom trading can harm, how they might be harmed, and the scale of the harm [278]. Therefore, a generalized loss of confidence derived from systemic crashes and failures may even reduce investor appetite for risk, thereby stalling economic growth [278]. To support this grim hypothesis, the authors considered the case of Knight Capital Group. On August 1, 2012, the firm lost $440 million in less than 30 min due to its new AT software flooding the market with orders and forcing the temporary closing of the New York Stock Exchange. The direct harm to the firm and its shareholders was catastrophic and almost led to bankruptcy. The accident, however, also had an indirect impact on the investing public's confidence in the structure of financial markets.

Countermeasures to the aforementioned issues are still being debated, and existing proposals come mainly from the regulatory and ethics communities, rather than from the computer science and engineering ones. This is also reflected in a lack of papers discussing security issues of HFT from a technical standpoint. Regarding regulations, some of the proposed solutions aim at reducing the effectiveness of HFT by changing how markets evade pending orders. In detail, some have argued that the priority rules which determine the execution sequence of submitted orders are designed to prioritize speed. Here, the regulatory conundrum is whether time-price priority unduly rewards high-frequency traders and leads to risky overinvestments in the technology arms race [279]. The greatest benefit of current priority rules is that they treat every order equally. However other priority rules have been proposed, such as one where every order at a price gets a partial execution, independent of time [280]. Others have proposed to replace the continuous trading model by periodic auctions, which can be designed to minimize the advantage of the speed

and to mitigate other negative outcomes of continuous trading such as manipulative strategies [281]. The main benefit of periodic auctions would be a reduction of the speed of trading and the elimination of the arms race for speed discussed throughout this section. Many markets already have auctions at the open and close times and are now considering the introduction of midday auctions, in addition to the continuous trading segment [280].

In addition to the previous regulatory countermeasures—anyway not bound to be deployed at large anytime soon—some politicians also hinted at the possibility to introduce more radical initiatives. For example, Hillary Clinton once suggested the introduction of a small tax on all cancellation orders, in an attempt to quell the practice of spoofing.[17] Introducing comprehensive financial transaction taxes would however face great difficulties, also in light of the potential risks and undesirable consequences that this might cause [282]. On the contrary, specific taxes aimed at thwarting HFT are seen as a more sensible and desirable possibility, although difficult to implement [280].

4.2.3 Open Issues and Future Directions

The growing stack of papers on the topic of AT and HFT indicates a developing academic interest in the potential contributions and limitations of HFT activity. However, many open questions remain unanswered. First and foremost, it is still not clear if the systemic risk of financial markets is embedded in electronic trading or if it is really caused by HFT. Existing studies showed that High-Frequency Trading is a multifaceted, complex, and secretive practice and that it has surely been implicated in nefarious market events. However, correlation does not necessarily imply causation, and isolating causal mechanisms from interconnected automated trading is highly challenging for regulators and scholars alike, seeking to monitor AT and HFT across multiple jurisdictions and markets [283]. For both ethical and practical reasons, markets must remain fair and orderly. Deciding how best to ensure this—in light of the huge growth of HFT in the last decade, also expected to continue in the next—requires additional endeavors of careful thoughts and discussions.

The growing interconnected and automated nature of financial markets also raises additional technology-driven questions. For instance, given that corporate disclosure is progressively moving toward machine-readable reports [267], can firms anticipate HFT trading strategies at the time of disclosure? Then, assuming that such strategies could actually be anticipated, would it be possible to exploit this mechanism to drive HFT toward specific strategies, thus somehow manipulating the market? More in general, despite the recent development of a new generation of network facilities, high-performance computers, and powerful forecasting algorithms, it seems that

[17] https://mechanicalmarkets.wordpress.com/2015/10/14/could-hfts-benefit-from-a-cancellation-tax/ (Last checked August 2020).

academia is still not paying enough attention to technological bias and its potential consequences. Moreover, the technological shift is likely to become even more significant in the coming years. Because of this, the key role of technological bias toward market efficiency deserves to be examined in more depth in future research efforts.

By looking back at where it all started—the 2010 Flash Crash—it's easy to understand that the AT software that initiated the market collapse simply executed the tasks requested by Navinder Singh Sarao. Though suggestive of the kinds of harm that AT and HFT might cause, the crash itself could only partially be blamed on such technologies. Containing, stopping, or even promptly detecting such unfolding market failures exceeds the boundaries of existing scientific and regulatory tools. In addition to calling for renewed efforts, this also inevitably escalates the problem to an ethic and philosophic plane, forcing a debate over "whether it is more important to move fast, or to avoid things being broken" [273].

4.3 Scenario 3: Remote Stock Market

The New York Stock Exchange (NYSE) is the world's largest exchange by market capitalization of its listed companies. Since its founding in 1792, it was the living heart of New York City's financial district, and it gradually became the core of American and world finance. For centuries, the NYSE and all other exchanges that contributed to the rise of Wall Street as an icon of trading and finance have been a place rooted in vast physical rooms. Trading floors have been thriving on hectic human activity and have always been filled with tightly packed rows of desks with hundreds of workstations and monitors, specialized phones called "turrets," and other ad hoc trading equipment. This streamlined organization survived each and every calamity—man-made or otherwise—to happen to New York, the United States, and the world. The shift to decimalization in stocks, the 9/11 attacks, the financial crises, the rise of passive investing, and Hurricane Sandy are just some of the notable shocks that affected Wall Street in the last decades.[18] However, none of this fundamentally changed the way trading was done. More importantly, none of these events caused trading floors to be closed while markets were opened, not even for a single day, not even in the case of a World War. This was only until Monday, March 23, 2020. On that day, the NYSE closed its trading floor—at that time, indefinitely—in the wake of the spreading COVID-19 pandemic. Two workers in the premises had tested positive to the virus, and the NYSE decided to switch to electronic-only trading for the first time since the current trading floor opened in 1903 and since its very founding in 1792. Traders and market makers started working remotely, as concerns of contagion continued to disrupt every facet of our

[18] https://www.cnbc.com/2020/04/30/goldman-sachs-trader-says-wall-street-never-the-same-after-coronavirus.html (Last checked August 2020).

society and everyday life, financial markets included. The US Financial Industry Regulatory Authority (FINRA) confirmed that traders could work remotely and that firms might need to implement alternative supervisory systems to support this switch. FINRA also temporarily waived some record-keeping requirements and opened to some flexibility for firms facing difficulties in meeting other filing obligations. The same decisions were taken worldwide by other exchanges, banks, and by all sorts of financial operators. Amid fears that COVID-19 could further spread, dealers and employees of brokerages have been temporarily permitted to log into trading systems from remote locations. In the United Kingdom, the Financial Conduct Authority said it had no objection to brokerage staff working from home if certain standards—like the recording of conversations and prompt execution of orders—could be met.[19] The London Metals Exchange, despite already having most of its traffic handled electronically, moved its once physically driven price-setting mechanisms to its electronic trading system as well, for the first time after 143 years.[20] Banks sent traders home in the second week of March as the pandemic was wreaking havoc, causing a historic surge in stock volatility and dislocations across credit markets, and IT departments worked around the clock to equip thousands of traders for the task. Similarly, brokers of the National Stock Exchange of India (NSE)—the leading stock exchange in India, in Mumbai—have been allowed to access the market from their homes.

As all activities suddenly moved online while markets stayed opened, traders had little time to come up with a new "normal." This is when technology came to the rescue, with the latest apps and appliances that allowed traders to feel so connected to coworkers and clients alike, as to rarely miss trading floors. Dedicated messaging platforms for investment banks such as Symphony, which is similar to the widely known Slack, became the *de facto* standard for creating chat rooms for internal teams and clients. Each trader had about tens of different chats simultaneously opened, efficiently connecting them with hedge funds and asset managers. The teleconferencing application Zoom also recorded a surge in use, as it became widely adopted for setting up calls with clients and with other traders and for coordinating entire teams. Then, telecommunication companies, such as Cisco, promptly delivered dedicated phones that conveniently provided the possibility to record calls, as demanded by most regulatory agencies. With all these workplace tools at their disposal, it comes with little surprise that many traders even experienced a rise in productivity. And in fact, the five biggest US investment banks reported their best trading quarter in nearly a decade, as both bond and stock desks handily beat expectations. For the first time in history, stock markets had gone fully remote. In the end, the COVID-19 pandemic raised the question as to whether exchanges should retain trading floors at all. Some consider them to be costly and

[19] https://fortune.com/2020/03/22/coronavirus-nyse-trading-floor-closed-stock-market/ (Last checked August 2020).

[20] https://www.economist.com/finance-and-economics/2020/05/25/covid-19-forced-trading-floors-to-close-theyll-be-back (Last checked August 2020).

overhead heavy businesses. Moreover, also human brokers are considered by some to be slower and more error-prone than algorithms. In addition, some researchers also argued that fully electronic markets are more efficient than those featuring a certain degree of human intervention [284]. Researchers analyzed NYSE's hybrid auction structure, which normally allows floor traders to submit their last orders of the day up to 10 s before the market's close, whereas those coming through electronically have to make theirs 10 min before its end. This difference gives floor traders an advantage in end-of-day auctions. Then, to evaluate the effect of an electronic-only market, scholars compared market efficiency while the NYSE was operating only remotely, with its normal (i.e., hybrid) mode of operation. Interestingly, they found that auctions have become more efficient since the NYSE moved entirely online [284].

Despite these results, once the curve of COVID-19 contagions had been "flattened," all exchanges slowly and gradually started to revert back to pre-pandemic work conditions. At the same time, however, it was clear from the beginning that some of the changes caused by the pandemic were bound to remain. In fact, in the few months of extensive lockdown, many businesses had to adapt, sometimes even for the better. With respect to electronic-only trading and remote stock markets, the pandemic only accelerated a trend that was already in place. Many exchanges already had the majority, if not all of their trades, handled electronically. Examples of this kind are the stock markets in Hong Kong, London, Tokyo, Toronto, and Mumbai that have long scaled back the respective trading floors. The NYSE was undergoing the same process as well, with roughly 80% of its trade volume already managed electronically through a data center in suburban New Jersey. On top of this remotely driven scenario, when the NYSE reopened, only a few floor traders actually came back, and they were allowed to do so only after abiding by several requirements including social distancing norms and the need to wear face masks. With the gloomy prospect that COVID-19 was just one in a series of pandemics to test our societies for the coming years,[21] it is quite natural that worldwide stock markets are accelerating their transition to remote- and online-only. An outlook that raises renewed concerns for the security of stock markets.

4.3.1 Threat: Attacks Against Availability

As we have already seen in Chap. 3 with cryptocurrencies that are replacing physical fiat currencies, the switch from physical to virtual inevitably introduces several security challenges. Online and remote stock markets make no exception to this rule. The first concern about a fully online stock market is related to potential attacks on its availability. Stock markets play an important role in modern economies, easing

[21] https://theconversation.com/coronavirus-is-a-wake-up-call-our-war-with-the-environment-is-leading-to-pandemics-135023 (Last checked August 2020).

the access to capital and allowing its allocation from those who have a surplus to those who are in need, and contributing to stabilizing security prices. They represent a hub for a multitude of financial services. As such, limiting or denying access to these services would have tremendous repercussions on a national economy. In the past, when it has been suggested the possibility of a market holiday in the United States, or when other countries have suspended trading, people reacted with widespread panic. In the United States, continuous liquidity is a hallmark of equity markets and assuring investors that they will have the capability to access their money is mandatory for maintaining confidence in the system. It is thus crucial that people continue to have confidence that the markets will be open to both express their investment thesis and to access their savings.[22]

Among the most common types of cyberattacks that are designed to limit the availability of a resource to its intended users are Denial of Service (DoS) attacks. These are typically carried out by flooding the target resource with bogus requests, thus overloading it and preventing some or all legitimate requests from being fulfilled. Some of the most effective DoS attacks are carried out in a distributed fashion—that is, by using multiple machines, typically in the region of thousands, to send the bogus requests. Large-scale, Distributed Denial of Service (DDoS) attacks are more difficult to defend against, since the victim would need to shut down a significant portion of all the machines involved in the attack, instead of a single one.

Attacks and Countermeasures

Typically, DoS and DDoS attacks are perpetrated for profit, as in the case of groups of hackers blackmailing their victims; for obtaining an industrial advantage on a competitor; or for ideological reasons by groups of activists. However, we have also seen a few cases of state-backed actors involved in DDoS attacks for political and economic reasons. One of the biggest DDoS attacks in history occurred in 2007 and targeted government services, media outlets, and financial institutions in Estonia. This had a crushing effect on Estonia, since the country was an early adopter of e-government and was practically paperless at the time, to the point that even national elections were held online. This attack is considered by many to be the first case of cyberwarfare and came in response to political conflict between Estonia and Russia, which is suspected to be responsible for the attack.[23] More recently, during the 2019 Hong Kong protests that are part of the long-lasting Hong Kong-China conflicts, the messaging app Telegram suffered a large-scale DDoS attack aimed at preventing protesters from coordinating their efforts. Investigations by Telegram found that the attack was carried out by a State-sized actor and via IP addresses originating from China.[24]

[22] https://www.iflr.com/Article/3926218/Inside-nyses-response-to-the-Covid-19-crisis.html (Last checked August 2020).

[23] https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/ (Last checked August 2020).

[24] https://www.pcmag.com/news/chinese-ddos-attack-hits-telegram-during-hong-kong-protests (Last checked August 2020).

The previous examples demonstrate how cyberattacks can be weaponized by a nation as a mean to fulfill its political and economic goals. Past attacks mainly targeted governmental Web services and communication facilities, and, until now, there have been no records of state-driven attacks targeting national stock markets. This was partly due to the physical component that stock markets have always had—e.g., their frenzied trading floors. However, with the progressive dehumanization of stock markets resulting in all-electronic trading, this scenario might rapidly change. Since markets are so sensitive to uncertainty, halting trading even for a limited amount of time could send stock prices plummeting. In fact, the online components of stock markets and other financial institutions (e.g., online banks) have a long history of small- to medium-scale DDoS attacks carried out by hackers and fraudsters. In a 2013 survey of 46 stock exchanges, the International Securities Commission Association (IOSCO) reported that more than half had already suffered a DoS cyberattack that year [285]. The majority of such attacks were considered to have had no effect on the proper functioning of the market and only resulted in minimal costs (less than $1 million) for the targeted market. As an example, in 2012, a wave of DDoS attacks led by activists hit the NYSE, NASDAQ, and BATS stock exchanges in the United States, but trading systems were not affected. Instead, in the fall of 2019, the Hong Kong Stock Exchange (HKEx) admitted to have suffered a series of DDoS attacks. The attacks overwhelmed HKEx's website and affected its ability to display prices and publish filings. More importantly, at the same time, the exchange also suffered an extended shutdown of derivative trading. This was later attributed by HKEx's chief executive to an unspecified "software bug," which however did not lift suspicions that the halt was due, or at least linked, to the cyberattack.[25] This scenario could even become more complex than this, if we consider that an attacker could preemptively sell (buy) some shares on a market, only to subsequently carry out a targeted attack aimed at lowering (increasing) the value of the manipulated shares, thus obtaining a direct unfair profit from its attack. This could be achieved by targeting a specific company and manipulating the price of its stocks or even by targeting a market and inducing a flash crash.

While there are only a few recent cases of attacks against the availability of a stock exchange, the scenario is completely different for cryptocurrency exchanges. Having always been online-only, cryptocurrency exchanges naturally attracted all sorts of cyberattacks. In [286], scholars analyzed the impact of DDoS attacks on the volume of Bitcoin traded on the Bitfinex cryptocurrency exchange. The study shows that for most attacks, Bitfinex suffered a significant impact in the aftermath of the attack, but also that it was able to recover within a single day. This proves that a long-lasting DDoS attack can severely cripple the revenues of any given exchange. Also, the study in [287] investigated the impact of DDoS attacks on the Mt. Gox Bitcoin currency exchange. The study concluded that on days where DDoS attacks or other shocks occur, the distribution of daily transaction volume

[25] https://www.finextra.com/newsarticle/34352/hong-kong-exchange-suffers-cyber-attack (Last checked August 2020).

shifts in such a way that fewer large transactions take place, with detrimental effects for the exchange itself. DDoS attacks have also been investigated with respect to the competition between different mining pools. In particular, the study in [288] applied a game-theoretic approach and found that pools have a greater incentive to attack large pools than small ones. They also observed that larger mining pools have a greater incentive to attack than smaller ones. These results suggest that small-scale attacks are unlikely to be profitable, while large-scale ones are instead more profitable for the attackers—a result that raises concerns on the overall stability of stock and cryptocurrency exchanges. The theoretical result of [288] was also later confirmed in [289], which empirically measured that big mining pools are much more likely to be DDoS-ed than small pools. The extensive study in [289] also found that currency exchanges, mining pools, gambling operators, eWallets, and financial services are much more likely to be attacked than other services. They also found that currency exchanges and mining pools are much more likely to have DDoS protection, such as that provided by CloudFlare, Incapsula, or Amazon via AWS Shield.

Resources
Main DDoS protection solutions:

- **Project Shield**[26] by Jigsaw, a technology incubator created by Google
- **Microsoft Azure**'s DDoS solution[27]
- **AWS Shield**[28] by Amazon Web Services
- IBM's DDoS protection as part of **IBM Cloud Internet Services**[29]
- **Cloudflare**,[30] one of the most popular DDoS solutions to date
- **Imperva Incapsula**.[31]

Regarding countermeasures to DDoS and other availability attacks, there already exists an extensive literature on the topic [290, 291]. These attacks are not new and have long been studied by academia. Accordingly, also many industry leaders are already providing their DDoS protection solutions. In Sect. 6.2.2 we present, among other solutions, an overview of several DDoS protection strategies. The challenge related to securing the availability of remote and online stock markets is thus more toward applying the countermeasures that already exist, rather than devoting research and development efforts to propose new ones.

[26] https://projectshield.withgoogle.com/ (Last checked August 2020).
[27] https://azure.microsoft.com/ (Last checked August 2020).
[28] https://aws.amazon.com/shield/ (Last checked August 2020).
[29] https://www.ibm.com/cloud/cloud-internet-services (Last checked August 2020).
[30] https://www.cloudflare.com/en-gb/ (Last checked August 2020).
[31] https://www.imperva.com/products/ddos-protection-services/ (Last checked August 2020).

4.3.2 Threat: Work-from-Home Perils

The COVID-19 spreading and the switch to fully remote stock markets forced organizations around the world to enact remote, work-from-home policies. While some organizations have maintained robust remote work practices for years, many others have had limited experience in this regard. Even for organizations that have long maintained a remote workforce, the breadth and depth of remote work have dramatically increased. Business units and functions that have never been done remotely are now required to operate in a fully remote mode. As a consequence, many that have never experienced remote work—think, for example, of trading floor workers—are now suddenly forced to do so. During these rapid, unplanned, and unprecedented changes, security experts are pondering what new risks may have been introduced.

For instance, many office workers have been originally only provisioned a desktop computer from their employer. How did these workers connect to the network after the switch? How many of them were forced to utilize an unmanaged personal system or an insecure connection in order to keep up their work? On the workers' side, this raises a plethora of security concerns. Some of these concerns are related technical security risks, such as the inevitable increase in computer-mediated communications that could be sniffed by an attacker, or even the vulnerability of the system and the applications (teleconferencing and messaging apps) that employees need to use for their remote work. Other additional concerns are related to the psychological effects of work from home. First of all, operators that start working remotely immediately become more valuable targets for attackers, which exposes them to increased risks. At the same time, they might feel safer at home or even be more carefree as a consequence of the informal work environment. The combination of increased risks and diminished attention represents the recipe for security breaches. The situation also became more complicated on the companies side. In fact, the security perimeter previously established by the company, weakened or even vanished. In addition, security teams are now left with limited visibility and control over remote workers' appliances and actions. If not adequately addressed, these work-from-home security risks might cause a rapid surge in attacks.

Attacks

Here, we discuss in more detail some of the risks that this new remote connectivity paradigm brings. To do so, we consider a sample of the remote access implementations that are typically being used.

Direct Access The simplest and least secure remote access method revolves around directly exposing networking protocols, such as Microsoft Remote Desktop Protocol (RDP), to the Internet. This scenario represents a baseline case, since most mature organizations prohibit direct access through proper firewall configurations and other restrictions. However, a few cases of this type nonetheless still exist. Here,

the main attacks consist of the traditional means of gaining access to externally facing services. Network scanning of external ports and exploitation through brute-forcing, credential spraying, and spear phishing are among the most widely used attacks. Further increasing the risk of this direct network access is that these services likely allow unmanaged devices direct access, providing little visibility into the hosts that are connecting to the services.

Given the lack of control and the risks introduced by the previous model in exposing RDP and other remote protocols to the Internet, organizations have centralized remote access to a few technologies. Two widespread implementations are based on Virtual Private Network (VPN) or virtualized desktops and the so-called zero trust model. These solutions allow for improved access management, logging, and security controls.

VPN/Virtual Desktop In the first solution, which also represents the most widely used one, a VPN or a virtualized desktop interface such as Citrix or VMWare is placed within the organization's Demilitarized Zone (DMZ). Threats to this model include unauthenticated attacks, compromised credentials, and compromised systems. Moreover, attackers often chain-control deficiencies together, by exploiting the initial access they obtain to the VPN or virtualized desktop, to gain further access.

Endpoint Remote Access Endpoint remote access is one of the vulnerabilities of this model, for which mail filtering, endpoint hardening, and reduced administrator privileges and visibility should be enforced. In addition, security teams should validate that endpoint visibility remains consistent for users that switch from in situ to remote and for any new users or third parties.

Multifactor Authentication Bypass Many organizations have implemented Multi-factor Authentication (MFA) to reduce the success of brute-forcing or credential spraying. However, carefree users might still accept push notifications, thus enabling remote access. Against this issue, remote employees should be adequately trained to identify and report unauthorized push notifications.

Unmanaged Device Access Organizations often conduct limited validation checks to identify unmanaged devices, including attacker systems connecting to remote access solutions. Oftentimes, the checks performed by VPN solutions can be bypassed by modifying VPN software responses or registry key settings. In addition to attacker systems connecting to the network, security teams should also consider users connecting from unauthorized systems.

Tunneling Configuration To handle the increase in remote connectivity caused by the switch to remote markets, some organizations passed from a full-tunnel configuration to split tunneling. With a full tunnel, all traffic traverses the VPN, allowing Web proxies to filter traffic and security teams to identify unauthorized activity. Instead, split tunneling reduces this visibility unless appropriate endpoint agents are installed to provide sufficient visibility and control.

Remote Access DoS Entire organizations are moving toward a remote access model. As such, the potential impact of a DoS attack on these remote access portals has significantly increased. An attacker could be able to generate multiple failed password attempts on an account and lock the user out. If this attack runs at scale, it might even cause widespread account lockouts, thus impairing the organization's activities.

Zero Trust Model In contrast with VPNs and virtual desktops, the emerging zero trust model leverages an identity provider to provide access to the applications. Then, authorization rights are determined based on both the user and device, via a series of identity checks. Given that this model is based on device trust as a component of authentication and authorization, issues such as the protection of certificates establishing device trust and access limitations to unmanaged devices, are of the utmost importance. These add to some of the previously mentioned concerns that still apply, such as endpoint visibility and hardening, MFA bypasses, and DoS attacks.

Countermeasures

In light of these threats and in order to adapt to a remote and distributed workforce, organizations need to create a strong set of defenses at the edge of their networks for protecting both identities and applications, regardless of whether they are in the corporate network or in the cloud.

Regarding the problem of authentication, organizations must implement MFA on all external corporate resources to reduce the effectiveness of attacks such as credential spraying, password stuffing, and phishing—which are bound to become increasingly common. However, as previously outlined, MFA alone is not sufficient to ensure secure authentication. Because of this, it is also becoming increasingly important to validate the device establishing connectivity, for instance, by relying on device identity certificates. Then, regarding endpoint control, endpoint visibility should be enforced, and employee endpoints should be hardened in order to reduce the ability for an attacker to gain access to systems and to escalate privileges. Moreover, default configurations of virtualized interfaces should be avoided as they may allow to break out of virtual sessions, thus opening access to the underlying operating system. The switch to remote work also implied an increased dependence on third-party cloud services, as thoroughly analyzed in Chap. 6. Here, contrarily to the current trend, organizations should strive to develop suitable corporate alternatives for cloud services. In all those (alas, many) cases in which this is not possible or not yet available, organizations should at the very least ensure that security teams regularly receive logs from cloud providers so as to review them for unauthorized access and data exfiltration. Moving on, regarding network control and visibility, off-network communications from virtual desktops should be limited to whitelisted resources, in order to reduce potential exposure. Similarly, shifts from full-tunnel to split-tunnel VPN should be limited and possibly also

complemented by augmenting network visibility with a cloud proxy or with a similar solution. Finally, special care should be devoted to providing adequate security awareness training for remote workers, especially for those that are new to this paradigm. In addition to computing hygiene topics such as phishing and password guidance, employees should be trained on physical security topics such as using a privacy screen, limiting work on confidential material in public spaces, and securing physical computing assets. Still, regarding physical security, the possibility for appliances to be lost or stolen should be also taken into account, for instance, by ensuring that all employee computing resources have full disk encryption.

4.3.3 Open Issues and Future Directions

Given the recency of the switch to online- and remote-only markets, the full extent of new security risks and their impact on the market and financial stability is yet to be fully understood. Nevertheless, the analysis that we developed in this section still allows to draw some preliminary conclusions. On the one hand, we have a rapidly evolving situation where multiple voices are advocating the advantage of fully online markets. Among them are those in favor of reducing costs and delays imputable to trading floors and those striving for greater market efficiency, which appears to require less human intervention in favor of more algorithmic decisions. These forces, combined with current digitalization trends and with the possibility of future pandemics or other shocks impairing physical human interactions, are bound to progressively reduce the importance of trading floors and human brokerage. On the other hand, we already collected extensive evidence of the volume and scale of cyberattacks that continuously affect online exchanges, such as cryptocurrency exchanges. For the future, it is thus foreseeable that an increased number of cyberattacks will be targeting stock markets. Solutions for defending against or for mitigating the majority of existing attacks already exist. What we are actually missing is however a greater understanding of the vulnerabilities of fully remote markets and the assessment of the practical impact that a large-scale attack, such as one carried out by a state-sized actor, could cause. Moreover, development and deployment endeavors for applying current solutions to stock markets are still lagging behind.

4.4 Scenario 4: Complex Financial Networks

In the previous sections of this chapter, we discussed many novel technological means that could be used to attack an individual financial entity (e.g., a company) or a single component of a complex financial system (e.g., a market). We surveyed previous attempts at each of those attacks, and we highlighted the impact they could have on the financial ecosystem at large. Despite showing the potential to

endanger a national economy, such attacks would nonetheless be less dangerous and detrimental than a full-fledged systemic failure—that is, a severe widespread financial crisis affecting many components of a financial system. The latter would, in fact, cause widespread and long-lasting dramatic effects, possibly including prolonged recession, severe lack of liquidity, country defaults, and even small- to medium-scale humanitarian crises.

This was exactly the case with the Global Financial Crisis (GFC) that started in 2007–2008. As it happens with all systemic problems, also this crisis had deep roots. In the aftermath of the 2001 9/11 attacks, the US economy faced a mild recession. To contrast the downward trend, the Federal Reserve lowered interest rates up to a minimum of 1% in 2003, the lowest ever reached in 45 years.[32] This injection of liquidity allowed subprime (i.e., risky) borrowers to have mortgages and, with them, houses. House prices raised and homeownership reached a peak of 70% in 2004. At that point, interests started rising again, and few people were asking for houses, due to the high homeownership. By the end of 2005, home prices started to fall. As interests raised and home value decreased, for a growing share of subprime borrowers, it simply became more advantageous to give the home back instead of repaying the mortgage. Between 2006 and 2007, as borrowers started defaulting on their loans, so did an increasing number of subprime lenders. In April 2007, the well-known New Century Financial filed for bankruptcy. Since then, problems spread also outside of the United States, with the British Northern Rock eventually taken into public ownership in 2008. At that time, issues were emphasized also by the number and pervasiveness of securities that were backed by the now massively failing subprime mortgages, such as the extremely popular Collateralized Debt Obligations (CDOs) that also imploded as a result of the subprime mortgage crisis. These massive losses either resulted in investment banks going bankrupt or being bailed out by governments. In the United States, all this led to Lehman Brothers filing for bankruptcy, Indymac bank collapsing, Bear Stearns being acquired by JP Morgan Chase, Merrill Lynch being sold to the Bank of America, and Fannie Mae and Freddie Mac being put under the federal government control, to name but a few notable cases. Figure 4.10 shows the number of failed banks in the United States, comparing aggregate counts for 10 years before the GFC (1997–2007), with yearly data after the crisis. The devastating effects of the GFC are strikingly evident, and, in fact, it took almost 10 years to revert the situation to pre-crisis conditions. The spillovers of the global financial crisis beyond the US borders were equally catastrophic.

Given the devastating impact of global crises, many scholars have long studied the risk of systemic failure that affects financial ecosystems. In many of these studies, global financial systems are modeled as *large complex networks* in which banks, hedge funds, and other financial institutions are interconnected to one another through a series of financial links (e.g., those representing money that has

[32] https://www.investopedia.com/articles/economics/09/financial-crisis-review.asp (Last checked August 2020).

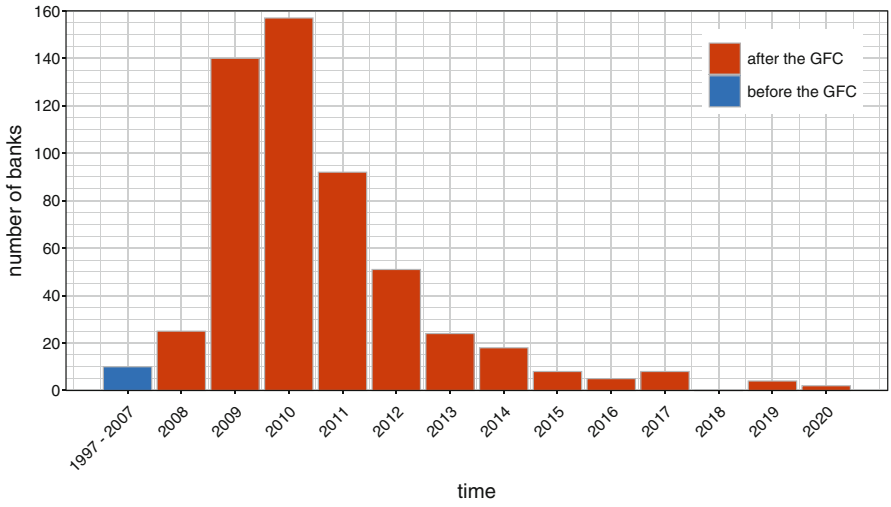

Fig. 4.10 Number of US banks failed per year, before and after the 2007–2008 Global Financial Crisis

been borrowed and lent). Then, grounding on this model of financial ecosystems, attention was directed toward the understanding of those mechanisms capable of triggering network breakdowns, such as those that led to the 2007–2008 GFC. For instance, this can happen when the financial links turn from being a means of risk diversification to channels for risk propagation [292]. Under these circumstances, the failure of a single entity (i.e., a node in the network) can start a cascade of failures that propagates through the financial network by traversing the existing links between entities. Depending on the properties of the financial network and the starting conditions, the shock can rapidly spread and engulf the majority of the network. Until now, manipulation attacks targeting individual entities—such as those discussed in the previous sections of this chapter—and complex financial networks have been studied separately and by different communities. However, it is crystal clear from the previous description that they are nonetheless related. A single targeted attack to a weak node of a financial network could in fact trigger a cascading effect on the network, with much bigger—and likely, much worse—consequences than those expected for the individual entity under attack.

4.4.1 Threat: Systemic Risk and Cascading Failures

In the wake of the 2007–2008 GFC, there has been increasing recognition of the need to address risk at the systemic level, as opposed to the traditional way of focusing on individual banks or single financial entities [293]. *Systemic risk* refers to the risk of a breakdown of an entire system rather than the failure of

individual parts of the system. In a financial context, it denotes the risk of *cascading financial failures*, caused by linkages within the financial system and resulting in a severe economic downturn. The basic mechanics of distress propagation that lead to cascading failures are very simple: when a financial entity suffers a loss, distress propagates to its creditors who, in turn, suffer losses, which propagate further on in the network. Systemic risk has some universally accepted characteristics. It is a risk that has (1) a large impact, (2) is widespread, and (3) creates a ripple effect that endangers the viability of the whole financial system. Systemic risk is an attribute of the financial system and not that of a single entity. However, financial institutions individually contribute to the overall systemic risk and those that provide large contributions to the overall risk are deemed "systemically important." Thus, a key question for policymakers is how to limit the build-up of systemic risk to contain crises if and when they do occur. Instead, the key question for scholars is how to accurately measure and quantify systemic risk in the first place. This is the focus of the subsequent section.

4.4.2 Measures of Systemic Risk

Policymakers, scholars, and practitioners have yet to reach a consensus on how to precisely define systemic risk. Given that systemic risk is not yet fully understood, its measurement is obviously challenging and many competing and contradictory definitions of threats to financial stability exist. As such, a large number of diverse measures for systemic risk have been proposed so far. Indeed, a single agreed-upon measure of systemic risk may be neither possible nor desirable. As in all cases where there exists the need to measure a complex phenomenon, we are often better off leveraging multiple indicators than a single one. In fact, more indicators give higher guarantees to capture the multifaceted and adaptive nature of complex financial systems [294].

Economics and Finance Approaches

Early and traditional approaches to measuring systemic risk quite naturally arose from the economics and finance literature. In the years following the 2007–2008 crisis, a plethora of measures was proposed, which are mainly based on game theory, finance, and macroeconomic modeling. These can be classified according to the data required for computing the measure, to their supervisory scope (e.g., micro- vs. macroprudential measures), or to the relative time when the measure is available with respect to a crisis (e.g., ex ante, contemporaneous, and ex post measures) [294]. For example, according to this latter categorization, many systemic risk measures were proposed for being adopted as ex ante measures of early warning. Among them, there are measures based on predictions of costly aggregate asset price boom/bust cycles [295], macroeconomic early warning indicators for banking sector

crises [296], statistical models for the timing of banking defaults [297], and many others. Other ex ante measures are based on stress tests, such as the 10-by-10-by-10 approach [298] or the marginal and systemic expected shortfall [299]. Then, among the measures for nowcasting systemic risk (i.e., contemporaneous measures), there are measures aimed at assessing the fragility of the system and those aimed at providing tools for monitoring an ongoing crisis. Finally, ex post measures are based on forensic analysis [300] or on the "risk topography" of the financial system [301], derived from a data acquisition and dissemination process that informs policymakers, researchers, and market participants about systemic risk.

Network Approaches

More recently, the assessment of systemic risks involved the scientific communities beyond economics and finance. Complex system scholars, in particular, were attracted by the interconnected and multifaceted nature of financial networks. As a result, a novel stream of research on systemic risk modeling led to the proposal of several new metrics grounded on network science and graph analysis principles. The financial networks needed by these studies can be obtained and built in different ways. In these networks, nodes typically represent banks or other financial institutions, and directed edges represent lending relations weighted by the amount of the outstanding debt. Some simple networks are based on public data released by the FED and assume a star structure with the FED positioned at the center of the network [302]. Other networks also model the many financial dependencies between institutions themselves, leading to more complex structures such as multiplex, bipartite, core periphery, and time-varying networks, depending on the properties of the considered financial links. One possible way to build these latter networks is by applying text mining techniques to SEC filings or financial news, to automatically extract inter-bank loan transactions representing the weighted edges of the financial network [303]. For example, it has been shown that the equity investment network of transnational corporations has a bow-tie structure with financial firms in the core forming a tightly connected component [304]. In addition, other studies focused on the characteristics of bipartite networks between banks and bank assets [305].

After having built a network model of the financial system with any of the approaches just described, the goal consists of designing a metric for quantifying the systemic risk of the network. This is typically achieved by a careful combination of graph theory and economic principles. Among the most widely used metrics of systemic risk is DebtRank. In detail, the DebtRank score of a node (i.e., an institution) in a financial network is a number measuring the fraction of the total economic value in the network that is potentially affected by the distress or the default of that node [302]. The original DebtRank assumes that losses are propagated linearly between connected nodes. However, in other subsequent studies, this assumption was relaxed with the introduction of nonlinear propagation functions, which led to even more accurate estimations of systemic risk [306]. Other studies found that the measure of the scalar assortativity of a financial network

correlates well with the level of systemic risk. In particular, network structures with high systemic risk are scalar assortative, meaning that risky banks are most exposed to other risky banks. Conversely, network structures with low systemic risk are scalar disassortative, with interactions of risky banks with stable banks [307]. Others adopted a model of shock propagation to investigate the bipartite network of US mutual fund portfolios and their assets, to identify a systemic risk component stemming from the similarity of portfolios [308]. By following the evolution of the 2007–2008 GFC, they showed that portfolios became more diversified during the crisis. Nonetheless, a large overlap in portfolios was measured to be far more likely than expected from random baseline models, demonstrating a strong correlation between fund investment strategies. The results of the study ultimately showed that diversification and similarity should be jointly taken into account to properly assess systemic risk. Also the work discussed in [309] focuses on overlapping portfolios and expectation feedbacks to study systemic risk. The model obtained from the study showed that risk expectations play a crucial role in the systemic stability of the financial system. In particular, wrong risk expectations may create panic-induced reduction or over-optimistic expansion of balance sheets. Other works studied Granger-causality tail risk networks to identify periods of distress in financial markets and possible channels of systemic risk propagation [310]. As part of the study, a novel market turbulence indicator is proposed, based on a measure of connectedness of the networks. By retrofitting the indicator to data about the 2009 European Sovereign Debt Crisis, authors showed its informativeness demonstrated by a peak of the indicator at the onset of the crisis, thus signaling the instability of the financial system.

4.4.3 Countermeasures

The multitude of methods for measuring systemic risk that we briefly surveyed in the previous section provides useful indicators for the stability of financial systems. As such, the analysis of the dynamics of financial models carries potentially far-reaching implications for the design and implementation of public policy [293]. Such tools thus represent a convenient control panel for regulators and policymakers interested in the healthiness of our economies. In other words, measures of systemic risk can inform decisions by regulators and policymakers, who can monitor the evolution of such indicators to avert financial crises and to have precious feedbacks on their regulatory activity. In fact, if academia is actively involved in proposing estimations of systemic risk, policymakers and regulators are deputed for acting upon such estimations to come up with effective countermeasures for avoiding financial breakdowns. The majority of policy interventions are aimed at reducing the existing systemic risk, thus reducing the possibility of cascading failures to occur. In the following, we briefly discuss a series of intervention strategies that have been proposed to reduce systemic risk in financial ecosystems.

Setting Capital and Liquidity Thresholds Ratios of capital have been in secular decline in banking institutions for at least the past 150 years, despite their strategic importance as buffers for absorbing external shocks. Reversing these trends by setting higher required thresholds for capital ratios would strengthen the absorptive capacity for each of the nodes in a financial network. As importantly, it would also lessen the risk of idiosyncratic defaults cascading around the system. Broadly, the same arguments apply to the setting of regulatory requirements on bank liquid assets.

Adjusting Requirements to Superspreaders There has been a significant rise in the size and concentration of the financial system over the past two decades. This leads to the emergence of superspreader institutions in financial networks, which are too big, too interconnected, or simply too important to fail. As a testimony of the damage that can be caused by the failure of a superspreader, it is enough to think of the consequences of the Lehman Brothers failure in October 2008. Protecting financial systems from future such events would thus require that superspreader nodes obey to much higher regulatory thresholds—in proportion to the system-wide risk they contribute—with respect to those applied to smaller players.

Regulating Derivatives Markets As discussed at the beginning of this section, the rapid growth in the size and complexity of the derivatives market (e.g., CDOs) was partially responsible for further destabilizing the already heavily stressed system on the onset of the 2007–2008 financial crisis. This inevitably questions the underlying structure of the derivatives market. One means of simplifying the complex Web of interactions between banks in derivatives markets is to centralize the trading and clearing of these instruments.

Shaping the Topology of Financial Networks While discussing measures of system risk, we highlighted the importance of diversification strategies for disentangling network nodes, thus reducing the possibility of cascading failures. However, if overall little effort has been devoted to assessing the system-wide diversity of balance sheet and risk management models, even less effort has been put into providing regulatory incentives to promote such diversity. In rebuilding and maintaining the financial system, the objective of systemic diversity is to be given much greater prominence by the regulatory community. Similarly to the previous point, modularity measures the extent to which a network can be partitioned in different communities. Modular network configurations contribute to preventing cascades from infecting the whole network in the event of an individual node failure. That is the reason for the long-lasting debates on the advantages and drawbacks of splitting banks, either to limit their size—to curtail the strength of cascades following failure—or to limit their activities—to curtail the potential for cross-contamination within firms [293]. In any case, network modularity should also be given greater importance in regulatory activities.

4.4.4 Open Issues and Future Directions

In pretty much all fields of science, it generally takes three subsequent and increasingly difficult steps for any given area of study to become a mature scientific field. The first step mandates to understand and *model* the mechanisms behind a real-world phenomenon. The second step concerns the *forecast* of the future dynamics of the phenomenon. The third and last step challenges to *control* the phenomenon so as to avoid undesired events. Most of the efforts discussed in this section for studying systemic risk are still at the first stage [292]. As such, avenues for future research and experimentation are manifold. On the one hand, we still need more accurate indicators for measuring systemic risk and more efficient policies for regulating financial systems. On the other hand, we also need to push the boundaries of current research by experimenting with forecasting models and by using existing and future models for simulating possible intervention strategies. Only through this iterative process of modeling and simulation we are likely to obtain more secure and stable financial systems.

As shown in our brief survey, the majority of recently proposed indicators for measuring systemic risk are based on network and graph theory approaches, applied to financial networks. However, recent works also demonstrated the importance of social and informational networks, and their interplay with financial networks. For example, pioneering studies showed that house price experiences in OSNs can drive contagious risk-taking in the US housing market [311], and others exploited epidemiological models to examine the interplay between social and informational contagion in driving financial cycles [312]. Despite these compelling results, the study of the interplay between social dynamics and financial risk remains almost completely unexplored. Consequently, much theoretical and empirical work is still needed to gauge the extent to which social contagion matters for aggregate financial system risks. In this regard, early results seem to suggest that "social network approaches could potentially have as much influence on cyclical macroprudential policies over the next decade as financial network approaches have had on stress-testing and structural macroprudential policies over the past decade" [312].

As the concluding section of Chap. 4, here we aimed to bridge the gap between the micro-security level discussed in Sects. 4.1, 4.2, and 4.3, and the macro-security level discussed in this section. In particular, throughout the previous sections of this chapter, we discussed several technology-enabled ways to manipulate companies and markets. Here instead, we took a more systemic point of view, and we showed how a local shock, such as one caused by one of the aforementioned attacks, could result in dramatic system-wide consequences by propagating through financial networks. Defending against such striking events thus involves both increasing the robustness of individual components of financial networks, by reducing the leverage and effectiveness of manipulation techniques, as well as shaping financial networks that are intrinsically more resistant to cascading failures. These challenges currently represent the battlegrounds that will decide the fate of worldwide economic warfare in the coming years.