The Role of Networks in FinTech Applications

Ye Liu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:
Yong Tan, Chair
Mingwen Yang
Amandeep Singh
Yingfei Wang

Program Authorized to Offer Degree:
Foster School of Business

University of Washington

Abstract

The Role of Networks in FinTech Applications

Ye Liu

Chair of the Supervisory Committee:
Yong Tan
Department of Information Systems and Operations Management

Networks are increasingly central to how individuals make investment decisions, how prices are formed, and how platforms design user experiences in the context of modern financial technologies (FinTech). Rather than viewing market participants in isolation, this dissertation examines how surrounding social structures shape economic outcomes across three FinTech applications: social trading, cryptocurrency, and non-fungible tokens (NFTs). Each essay highlights how network position, peer behavior, and relational dynamics can influence attention, adoption, and valuation — often in ways that rival or exceed the impact of individual intrinsic characteristics. First, rather than focusing solely on individual attributes such as financial performance, I examine how strategy similarity among traders affects their popularity. I construct dynamic similarity networks and use a coevolution model to analyze how traders' positions within these networks shape follower growth. I find that as more peers adopt similar strategies to a focal trader, that trader tends to attract fewer followers over time. This substitution effect is particularly strong for high-performing traders. Counterfactual simulations suggest that promoting a more even distribution of followers and reducing network density can mitigate substitution and increase overall platform engagement. This suggests that platform design should consider not only individual performance but also the network positioning of users to foster diversity and sustain long-term engagement. Second, I investigate how social learning drives cryptocurrency adoption. Using a large dataset from a

social trading platform, I analyze how investment opinions and behaviors of network neighbors influence an individual's decision to trade Bitcoin. The results show that people are more likely to adopt Bitcoin when their peers' actions align with their stated views. Under higher uncertainty, individuals rely more on optimistic opinions, even if behaviors are not consistent. This suggests that social media plays a dual role — as a source of behavioral signals and emotional reinforcement —particularly when individuals face ambiguity. It also highlights the importance of credibility and alignment between opinion and action in peer influence. Third, I study the role of transactional networks in shaping NFT value. Drawing on data from a leading NFT platform, I construct monthly networks linking NFTs based on shared traders and utilize a permutation-invariance-based method to circumvent the dimensionality issue of network data. The results show that incorporating transactional network information substantially improves NFT price prediction, beyond what image or metadata features can explain. This chapter highlights the importance of social dynamics in digital asset valuation and offers implications for both pricing models and platform design. Together, these chapters demonstrate that financial decision-making and asset valuation cannot be fully understood without accounting for the social context in which they occur. Whether through imitation and substitution among traders, peer-driven learning under uncertainty, or value formation via transactional ties, networks serve not just as conduits of information but as fundamental structures shaping market outcomes. By integrating structural modeling, behavioral analysis, and machine learning on large-scale digital traces, this dissertation offers new insights into how financial technologies operate when embedded in social systems. These findings contribute to emerging conversations at the intersection of finance, platforms, and computational social science, and underscore the need for theories and tools that reflect the relational nature of modern digital markets.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Foremost, I am deeply thankful to my advisor, Professor Yong Tan, for his support and mentorship. He is the best advisor one can imagine for a PhD journey. His wisdom, patience, and encouragement have profoundly shaped both my academic path and personal growth. I am incredibly fortunate to have had the opportunity to learn from him, not only in conducting meaningful research but also in leading a life with integrity and purpose.

Second, my sincere thanks go to my coauthor and committee member, Professor Mingwen Yang. Collaborating with her has been one of the most formative experiences of my PhD journey. Beyond research techniques, I have learned the value of resilience, intellectual rigor, and the sincere attitudes necessary for becoming a scholar.

Third, I would also like to express my gratitude to my other committee members: Professor Yingfei Wang, Professor Amandeep Singh, and Professor Quan Wen. Their thoughtful feedback and constructive suggestions have played a key role in helping me refine my research and elevate its quality. I am also grateful to Nuzulita Budhiari, Xuan Wang, Beau Kirkeby, Shawna Reimers, and the many staff members at the Foster School of Bzusiness for their generous support throughout my studies.

Lastly, I want to thank my parents, Hong Liang and Canhong Liu, for their unconditional love and support. They have always believed in me for every life decision I made. I owe my deepest gratitude to my wife, Shuang. From the very beginning, she supported my decision to pursue a PhD and stood by me through every challenge along the way. Her wisdom often helped me find clarity when I felt stuck, and her thoughtful care in the rhythms of daily life carried me through the most difficult moments. This achievement is as much hers as it is mine.

DEDICATION

To my wife, Shuang

Chapter 1

INTRODUCTION

In the last two decades, financial technology (FinTech) has reshaped how individuals access, evaluate, and participate in financial markets. FinTech applications have evolved from tool-based services, such as online banking and digital wallets, to platform-based innovations, such as crowdfunding, peer-to-peer lending, and social trading. In the last decade, new asset classes have emerged that are built upon decentralized blockchain technology, including cryptocurrencies, decentralized finance (DeFi), and non-fungible tokens (NFTs). As FinTech continues to evolve, the role of networks across these applications has become increasingly important for understanding economic outcomes and user behavior in these contexts.

First, FinTech platforms increasingly embed social features — such as following, copying, and commenting — that foster direct and indirect interactions between users. These peer-to-peer interactions can significantly influence individual behavior, particularly in environments characterized by high uncertainty and low institutional oversight. By observing others' actions, individuals may learn about market conditions, infer asset quality, or form heuristics for decision-making. However, such social learning can also give rise to herding, path dependence, and information cascades. Understanding the mechanisms through which peer signals propagate and impact financial decision-making is critical for designing more effective and equitable financial platforms.

Second, digital platforms vary widely in the type and granularity of information they disclose, ranging from detailed transaction histories and ranking systems to curated summaries or social signals. The design of these information environments can systematically shape user decisions by highlighting certain signals (e.g., price history, peer activity, or creator reputation) while obscuring others. Increased transparency may reduce uncertainty and improve decision quality, but it can also lead to strategic behavior, information overload, or overreliance on socially constructed signals. This theme examines how platform-level information architectures influence user interpretation, behavior, and welfare outcomes.

Third, in blockchain-based markets such as NFTs or cryptocurrencies, many assets lack clear intrinsic value or well-established pricing benchmarks. Instead, valuation often depends on subjective interpretations of scarcity, aesthetics, provenance, and social signaling. This ambiguity opens up a rich space to study how individuals assign value when traditional financial metrics are unavailable or unreliable. Factors such as visual rarity, social proximity to high-status users, historical performance, and network centrality may all contribute to perceived value. Understanding these drivers helps illuminate the behavioral foundations of price formation in novel digital markets.

Leveraging rich and transparent data in FinTech, this dissertation explores the role of networks in three settings. In Chapter 3, I examine how traders' similarity networks affect follower dynamics, showing that greater overlap in strategies leads to substitution effects, especially for high-performing traders. In Chapter 4, I study social learning in cryptocurrency adoption and find that individuals are more likely to adopt Bitcoin when peers' investment opinions align with their trading behavior, particularly under uncertainty. In Chapter 5, I analyze transactional networks in the NFT market and demonstrate that network information significantly improves price prediction beyond traditional image features. Together, these chapters reveal how financial outcomes emerge not just from individual decisions but from interdependencies encoded in networks, with important implications for platform design, user experience, and market efficiency.

Chapter 2
LITERATURE REVIEW

In this literature review, I first examine the existing research across various FinTech applications and highlight how these applications intersect with digital platform dynamics. Next, I discuss two key issues in network-related literature: social learning and models that capture the dynamic evolution of networks.

2.1 FinTech Applications

2.1.1 Social Trading

One key feature distinguishing social trading from traditional brokers is its high information transparency, as social trading platforms display every trade of traders and other information to the public. This high level of information transparency could reduce the principal-agent dilemma that commonly exists in the traditional financial market and thus benefit investors (Doering et al. 2015). On the other hand, prior studies on the general context of information transparency show that information transparency could be a double-edged sword, as it increases competitive risks (Zhu 2004).In social trading, the scopic regime can increase herding and intensify competition. For example, Gemayel and Preda (2018) compare the trading behavior between traders in a social trading platform and traders in a traditional foreign exchange broker. The level of information transparency increases herding among traders, especially when information is scarce (Gemayel and Preda 2018). In addition, what information and how it is disclosed affect trading behavior. For instance, Apesteguia et al. (2020) find that in social trading, providing information on other investors' success can lead to excessive risk-taking behavior among investors, especially when the investors are allowed to copy from others automatically. Whether the social trading platform displays the ranking or social features of traders may affect the disposition effect (a tendency of investors selling too early in an upmarket while holding too long in a downmarket) of traders (Liêu and Pelster 2020). Moreover, information transparency can lead to opportunistic free-riding behavior of investors in social trading platforms. Therefore, the platforms face a trade-off between the level of information transparency and revenue (Yang et al. 2022).

Our study contributes to the fast-growing literature on social trading (Ammann and Schaub 2021, Deng et al. 2024). The high level of information transparency can affect traders and followers on social trading platforms. Potential followers can observe the number of followers and the total amount of money following a trader on the trader's public profile page. This transparency may lead to a situation where many followers start following a trader because other followers follow the trader instead of conducting their own thorough analysis based on the trader's trading record. On the other hand, traders can observe other traders' trading records from their public profile pages. This transparency may lead to a situation where traders start imitating the trading strategies of others and free-riding on the information transparency. We acknowledge that two traders may adopt the same strategy from their analysis or imitate a third party. However, the above situations all lead to a high level of similarity among traders, which may affect a trader's popularity. In this study, we investigate the impact of the similarity pattern among traders' strategies on the traders' popularity. The side question investigates whether there is a herding effect among followers when making their following decisions.

2.1.2 Cryptocurrency

Cryptocurrencies as a novel asset class (Gemayel and Preda 2021) have attracted the attention of many scholars in recent years. Fang et al. (2022) and Almeida and Gonçalves (2023) provide comprehensive reviews of the recent research on investor behavior in cryptocurrency markets. Despite sharing some characteristics with traditional assets, cryptocurrencies exhibit unique features.

First, cryptocurrencies, built on the blockchain technology, operate on a decentralized and near-anonymous network and rely heavily on peer-to-peer exchanges with a lower level of financial institution involvement compared to stock, bond, and mutual funds (Fang et al. 2022). This decentralization raises issues regarding the verification and security of cryptocurrency ownership (Catalini and Gans 2020), which may impose challenges on the scalability of cryptocurrencies (Malik et al. 2022). Second, cryptocurrency prices are highly volatile, facilitated by the 24-hour exchange market. The high volatility in prices is attributed to speculative trading (Blau 2017), cryptocurrency mining (Jin et al. 2022), and network effect (Liu and Tsyvinski 2021, Wei and Dukes 2021, Cong et al. 2021). Third, the adoption of cryptocurrency is largely influenced by social factors (Gupta et al. 2021), such as public sentiment (Guégan and Renault 2021, López-Cabarcos et al. 2021). Such social characteristics also result in a pronounced tendency towards herding among investors in cryptocurrency trading (Bouri et al. 2019).

In this study, we focus on the social factors driving cryptocurrency adoption. We exploit a unique combination of individual-level social media and trading data to better understand investor behavior related to cryptocurrency with a more granular lens, while most of the existing literature relies on aggregate market-level data (Liu and Tsyvinski 2021, Lyandres et al. 2022). Closely related to our paper, Kogan et al. (2023) show that retail investors are contrarian in other asset classes, but trade on momentum in cryptocurrencies. The authors argue that investors believe that cryptocurrency returns influence the likelihood of future adoption of other investors, which influences future returns. Han et al. (2023) use Bitcointalk data to shed light on the role of peer sentiment on investors' beliefs in the Bitcoin market. The authors show that sentiment is contagious and investors update their beliefs relative to their peers, which predicts Bitcoin volatility and trading volume. We add to the literature by studying the interplay of neighbors' opinions and trading behaviors on individual investors' cryptocurrency adoption decisions.

2.1.3 NFT markets

The surge in non-fungible token (NFT) markets has prompted a growing body of research examining what drives NFT prices, how users interact with these digital assets, and how valuation mechanisms differ from those in traditional financial markets. Building on foundational work in cryptocurrency and asset pricing, researchers have begun to explore the unique dynamics of NFT markets.

A significant strand of research investigates how prior sales records influence future NFT prices (Borri et al. 2022). Historical transaction data—such as purchase price, frequency of trading, and wallet history—provides reference points that shape investor expectations and contribute to anchoring effects. Similar to the art market or collectibles, NFTs often derive value not from intrinsic features but from their market history and perceived desirability.

NFT valuation is also affected by broader developments in cryptocurrency markets (Ante 2023). Given that NFTs are typically transacted in cryptocurrencies like Ethereum, their prices and liquidity are closely tied to fluctuations in crypto assets. Changes in gas fees, crypto market sentiment, and token price volatility can all influence NFT trading activity. These cross-asset spillovers highlight the interconnectedness of blockchain-based asset classes.

Investor attention plays a critical role in NFT markets (Jiang and Xia 2024). Metrics such as page views, likes, and favoriting behavior serve as proxies for public interest and demand. As shown in related work on cryptocurrencies and stocks, attention-driven demand shocks can drive prices away from fundamentals. In the NFT context, such effects are even more pronounced due to the lack of objective valuation benchmarks and the strong influence of virality and platform visibility. Textual information—such as creator descriptions, collection narratives, and social media engagement—further contributes to how NFTs are valued. These unstructured signals shape the perceived authenticity, uniqueness, and community status of a collection. As in crowdfunding and ICO markets, the language and sentiment expressed by creators and early adopters play a crucial role in attracting investor interest and influencing valuation.

A more recent line of work introduces the use of machine learning to analyze NFT visuals. Borri et al. (2022) demonstrate that features extracted from the image itself—such as composition, color palette, and stylistic elements—can be strong predictors of price, even after controlling for rarity traits and transaction history. These findings suggest that visual appeal, beyond what is explicitly labeled in metadata, plays an integral role in how users assign value in NFT marketplaces.

Together, these studies highlight that NFT valuation is shaped by a multifaceted blend of historical, social, visual, and market-level information. Unlike traditional assets, NFTs combine symbolic, aesthetic, and social value components that challenge classical valuation models. As a result, understanding user behavior in NFT markets requires a hybrid approach that draws from finance, behavioral science, and machine learning.

2.2 Networks

2.2.1 Social learning

Social learning[1] refers to the process by which individuals learn from others when making decisions, first proposed by Banerjee (1992), Bikhchandani et al. (1992). Social learning significantly influences decision-making in many contexts (Huang et al. 2019, Hu et al. 2019, Qiu et al. 2021). It plays a critical role in influencing the adoption of new products, technologies, and services with limited information, such as agriculture technology (Conley and Udry 2010), healthcare products (Dupas 2014), anime (Ameri et al. 2019), new energy systems (Gillingham and Bollinger 2021), and microfinance (Banerjee et al. 2013).

With the increasing accessibility of data, recent literature has evolved its focus from understanding social learning based on location-based proximity or crowd behavior, to using

[1] The concept of social learning is defined differently across various disciplines. In the earlier economics literature, it is particularly associated with learning from the outcomes observed in prior adopters, in contrast to social influence, which involves learning from others' behaviors (Young 2009). Related concepts, such as observational learning and herding behavior (Cai et al. 2009, Chen et al. 2011), also emphasize learning based on others' actions. In this context, we adopt a broad definition of social learning that includes all forms of information acquisition, such as observing behaviors (what others do), listening to word-of-mouth (what others say), and inferring from outcomes (what others achieve).

social network data for a more granular understanding of the social learning process. For example, Banerjee et al. (2013) and Miller and Mobarak (2015) study the learning from offline social networks, while others focus on learning from online social networks (Hu et al. 2019, Ameri et al. 2019, Qiu et al. 2021). Prior literature also explores different types of neighbors and emphasizes the strength of ties (Hu et al. 2019, Qiu et al. 2021).

We contribute to the social learning literature by examining cryptocurrency adoption on social trading platforms. Two unique features set our research apart from the existing literature. First, extending beyond the findings of Ammann and Schaub (2021) that users make investment decisions based on others' posts on social trading platforms, we capture a critical and overlooked aspect of social trading: Users not only view others' opinions but also their trading actions, which can sometimes be contradictory. How the interplay between trading actions and expressed opinions of others, especially their consistency, influences an individual's decision to adopt cryptocurrency remains unclear. Investigating the interplay between opinions and actions is particularly important in our cryptocurrency adoption setting, as users may have incentives to share misinformation that does not align with their own trading actions. Çelen et al. (2010) explore this issue in a broad context through a lab experiment designed to discern the differential effects of advice versus actions. They study a sequential game setting, which differs significantly from the learning process on social trading platforms. Second, unlike the extensive literature on product adoptions where the expected value is fixed, cryptocurrency is characterized by high price volatility. While asset volatility affects investment strategy adoption in traditional financial markets (Han et al. 2022), its moderating role in social learning has not been thoroughly explored. Specifically, the moderating role of cryptocurrency volatility on the interplay between investors' opinions and actions enriches the literature.

2.2.2 Coevolution Model

The causal relationship between node characteristics and network structure is a fundamental question in social network analysis. The characteristics (attributes) of nodes can affect the network structure (through forming or dissolving ties between nodes), while the network structure can also affect the characteristics of nodes. Ignoring the identification of the interplay between node characteristics and network structure can result in endogeneity. Prior studies leverage natural experiments (Hasan and Bagde 2013, Aral and Nicolaides 2017) or randomized controlled experiments (Aral and Walker 2011) to mitigate this potential endogeneity. However, network-related research with observational data poses distinct challenges. One approach to constrain endogeneity is to employ instrumental variables derived from the attributes of a node's neighbors' neighbors (Bramoullé et al. 2009). This method is valid, assuming that the network formation process is exogenous (Bramoullé et al. 2016). In cases where the network is endogenous and dynamic, a more effective approach to mitigate endogeneity is to directly model the network formation process (Jackson et al. 2017). For example, Goldsmith-Pinkham and Imbens (2013) integrate a network formation process into a Manski linear-in-means model (Manski 1993) within a static setting to pinpoint the peer effect. To accommodate dynamic network changes and identify a broader framework of network effects (e.g., degree effects, homophily effect), Snijders et al. (2017) introduce the coevolution model — a statistical model to identify the causality between the dynamic network structure and attributes of nodes. The coevolution model is a stochastic actor-based continuous-time model that simultaneously allows the network structure and individual node attributes to evolve. Snijders et al. (2017) apply the continuous-time Markov process to model the network structure and discrete values of attributes. Niezink et al. (2019) extend the coevolution model to continuous values of attributes, where the network evolution was modeled by a continuous-time Markov chain and the evolution of the continuous attribute value was modeled by an SDE.

The coevolution model has been applied in longitudinal sociological studies; for example, the relationship between the friendship network and individual behavior includes delinquent behavior (Burk et al. 2007), alcohol consumption (Snijders et al. 2017), and individual academic performance (Lomi et al. 2011). In the social networks context, Bhattacharya et al. (2019) apply the coevolution model to capture the relationship between the network structure and user content creation behavior in an online network, where the attribute is modeled to be a discrete value. In this study, we apply the coevolution model to investigate the interplay between the trader similarity network and the trader's popularity (a continuous value, measured by the number of followers in our main analysis).

Chapter 3

TRADER SIMILARITY NETWORK IN SOCIAL TRADING

3.1 Introduction

Social trading has rapidly emerged as a popular form of retail investing in recent years. Leading social trading platforms, such as eToro and ZuluTrade, have attracted over 21 million traders across 179 countries, offering trading in stocks, options, foreign exchanges (Forex), and cryptocurrencies (Curry 2021, Daytrading 2025). By 2021, the annual trading volume on eToro exceeded one trillion U.S. dollars (Shome 2021). A major driver of this explosive growth is the high level of information transparency — social trading platforms allow traders to publicly display their trading activity, and others can automatically follow their strategies through an auto-copy function. Once a follower follows a trader, their trades are automatically executed in the follower's account. In return, followers pay a following fee, part of which is reimbursed to the trader by the platform. This creates strong financial incentives for traders to attract followers, as the number of followers directly affects a trader's extra income.

More importantly, a follower follows a trader and copies all trades of the trader with real money. That means, when a trader loses money, the follower also loses money. Therefore, it is an important decision for potential followers to decide who to follow, and followers usually carefully evaluate traders before deciding whom to follow. Platforms facilitate followers' decision-making process by releasing detailed trading histories and performance metrics for each trader. These metrics, such as win ratio (measured as the percentage of profitable trades out of the total trades), are useful indicators and can influence a trader's popularity (measured by the number of followers following a trader) (Yang et al. 2022). However, these aggregate financial measures often yield similar values across multiple traders, making it hard for potential followers to distinguish among them. As a result, when evaluating traders, potential followers do not rely solely on the aggregated financial metrics displayed on traders' public profiles. Instead, they also analyze detailed trade histories to understand the specific instruments traded.

One particularly notable phenomenon is that traders on social trading platforms often demonstrate more similar trading strategies than do those in traditional financial markets, where traders make their own trading decisions (Gemayel and Preda 2018). This similarity could result from traders accessing similar information, whether intentionally or not. Alternatively, traders might deliberately adopt a strategy similar to that of many others, believing it to be reliable due to its broad acceptance among traders. This broad acceptance can itself act as a reputational signal, making widely adopted strategies appear more trustworthy to traders and potential followers. By aligning with the crowd, traders leverage collective insights that may reflect a form of aggregated expertise. Moreover, choosing a commonly adopted strategy offers protection for reputation: if the strategy underperforms, traders face less blame because their decision aligns with what many others consider reasonable. This phenomenon, known as the *reputation effect*, suggests that deviation from the norm carries greater risk to one's reputation. As a result, following a commonly recognized strategy helps protect a trader's reputation.

However, if many traders use similar trading strategies, they become interchangeable to followers. This can reduce the unique value any individual trader brings, decreasing their popularity. For the platform, such homogenization also poses a longer-term risk: if followers realize that traders are not providing any new information, they may stop following them and even leave the platform. This *substitution effect* is intuitive in competitive environments where many users offer similar content or services; thereby, one user can be replaced or substituted by another. However, a key distinction between our context — social trading — and platforms like social media (Yoo et al. 2019) or crowdfunding (Wei et al. 2022) lies in the presence of direct monetary incentives. Traders earn real money from their followers, while followers pay fees to access strategies and bear the financial consequences of trading

— both gains and losses — of copying a trader's trades.

In this study, we ask one simple but fundamental question: *How does the similarity pattern among traders' trading strategies affect the traders' popularity in social trading?* This question is crucial, as a trader's popularity directly affects their social utility (such as reputation) and their extra financial gain from social trading platforms, which is based on the number of followers the trader attracts. With a lack of financial or social incentives, traders become inactive and even leave social trading platforms. Consequently, similar to other two-sided markets dependent on cross-side network effects, the business model of social trading could collapse if it fails to retain proficient traders.

To answer our research question, we first construct a similarity network among traders, in which each node is a trader, and a tie (link) exists between two nodes if two traders adopt similar trading strategies. Traders can view each other's current and past trades through their public profile pages, which display both current and historical trades. It is possible that similarity among traders may arise either from their independent interpretation of market signals or from observing how others have interpreted those signals (i.e., copycat behavior). However, distinguishing between these two scenarios is inherently challenging. What remains observable — and still highly relevant to traders, followers, and the platform — is whether similarity patterns emerge in trading behaviors and the impact of similarity patterns on followers' decision-making, as well as their broader implications for traders, followers, and the platform.[1] In our analysis, we do not attempt to distinguish between these sources; instead, we focus on the realized similarity patterns that emerge from traders' individual strategy

[1]Followers on social trading platforms are usually novices who lack trading expertise and seek to simplify their trading experience by automatically replicating expert traders' strategies with a single click. However, since a trader's current and historical trades are publicly accessible on their profile pages, it is also possible for strategic users to infer and imitate others' trading behaviors manually (free-riding on others' public information or copycat behavior). For instance, instead of using the platform's auto-copy feature officially, a user might choose to copy another trader's strategy by manually executing trades themselves (often with a time delay). This form of copycat behavior is challenging to detect, confirm, or identify within our dataset. It is equally difficult for the platform as well. However, if such strategic imitation occurs, it manifests as a similarity in trading patterns among traders. To better understand these dynamics, we examine similarity patterns among traders and their potential implications for followers, traders, and the platform.

choices on the platform. Furthermore, each trader may adjust her trading strategy over time; thus, the similarity network dynamically changes. Given that the dynamic network formation is determined by individual trading strategies, we shall effectively address the potential endogeneity to answer how the similarity pattern affects the follower dynamics.

In addition, due to the high level of information transparency, herding behavior is ubiquitous on many online platforms, e.g., crowdfunding (Xiao et al. 2021) and peer-to-peer lending (Zhang and Liu 2012). Herding is the tendency for an individual to follow the actions of preceding peers (Banerjee 1992). In social trading, potential herding behavior among followers can directly affect the measurement of our key variable of interest — the trader's popularity. Therefore, another challenge in answering how the similarity pattern in the similarity network affects the trader's popularity is to account for potential herding behavior among followers. This requires controlling for the autocorrelation of each trader's popularity, thereby adding another layer of complexity to model our research question.

To address the aforementioned potential concerns, we model the dynamic change in the trader similarity networks and the number of followers (attribute) following a trader using the coevolution model (Niezink et al. 2019). The coevolution model consists of two parts: the evolution of the trader similarity network and the evolution of the trader's number of followers. The evolution of the network is modeled by a continuous-time Markov chain, which captures the likelihood that each pair of traders adopts similar trading strategies. The evolution of the trader's number of followers is modeled by the stochastic differential equation (SDE), which captures not only the deterministic component but also the stochastic nature of the attribute. Network dynamics and attribute dynamics mutually influence each other or coevolve over time. In our model, we can separate the *influence* of the similarity network on traders from the traders' decision problem on the *selection* of position within the similarity network. Moreover, it incorporates the effect of network structure on both the network and attribute evolution. Finally, the evolution process of the attribute is modeled by SDE, incorporating the potential herding effect among followers.

Our data are from ZuluTrade.com, a leading social trading platform mainly for the foreign exchange (Forex) market. We first construct a dynamic similarity network based on traders' trading strategies on the platform. After constructing the similarity network among traders, we run the coevolution model with a trader's number of followers as her key attribute. Our results show that when a trader has more similar traders (peers), her number of followers decreases over time. We further confirm the existence of herding behavior among followers from the estimates of the coevolution model. We also conduct a heterogeneity analysis, which reveals that the negative effect of similar peers is stronger for more successful traders. Finally, our counterfactual simulations demonstrate that promoting a fairer distribution of followers and reducing the density of the similarity network could help mitigate the negative substitution effect and increase the total number of followers on the platform. Additionally, we conduct several robustness analyses to further validate our main findings.

Our study suggests managerial implications for traders, followers, and social trading platforms. First, our study highlights the value of incorporating trader similarity networks in the operations management of social trading platforms. A similarity network provides an overview of the interconnections among traders' trading strategies, with richer information than individual aggregated performance metrics. Social trading platforms can leverage this additional information to monitor the overall risk at the platform level. Second, employing a trading strategy similar to that of numerous other traders can adversely impact the focal trader's popularity and financial gain. As the additional payoff of traders declines due to the loss of followers, traders may become less active on or even leave the social trading platform, resulting in long-term losses for the platform. More importantly, our analysis reveals that this negative impact of similar peers is especially pronounced for more successful traders, who are more vulnerable to the substitution effect by similar peers. These findings underscore the need for platforms to actively monitor the similarity network and promote diversity in trading strategies, particularly to retain more successful traders who contribute disproportionately to platform revenue. In addition, our counterfactual analysis shows that encouraging a less densely connected similarity network, alongside a fairer distribution of followers, can mitigate the adverse effects of similar peers and help build a more sustainable trading community while enhancing platform revenue. Third, we observe a significant herding effect among followers, regardless of the traders' performance. Herding behavior among followers can lead to the "richer get richer" problem for traders, although it can also be beneficial for some popular traders. However, herding behavior stifles diversity by concentrating attention on a small group of popular traders on the platform, which increases the platform's vulnerability to market risks. To mitigate the potential herding effect, platforms could implement features highlighting traders with profitable strategies but fewer followers.

3.2 Theoretical Background

3.2.1 Effect of Adopting Similar Trading Strategies

When a trading strategy is widely adopted by retail traders on a social trading platform, it often reflects a collective assessment. This broad adoption serves as a manifestation of the wisdom of the crowd, indicating that the strategy has been judged to be valuable by a large number of traders. In such contexts, the reputation of the strategy may serve as a positive signal to other traders. By leveraging the crowd's aggregated insights and reputational signals, individual traders can make more informed, data-driven investment decisions. In particular, adopting the same strategy as the majority can be beneficial when it reflects a well-evaluated strategy.

Another important aspect to consider is that adopting a widely acknowledged trading strategy can be a safer approach. If the market moves against the adopted strategy, the trader is less likely to be blamed since the strategy was broadly accepted. This phenomenon, known as the *reputation effect*, suggests that poor performance poses a greater reputational risk for traders who deviate from the norm. As a result, following a commonly recognized strategy helps protect a trader's reputation. Previous research (Scharfstein and Stein 1990, Trueman 1994) document a substantial reputational loss (measured by the loss of investors) for traders that do not conform.[2] If such a reputation effect exists on social trading platforms,

[2]Appendix A.1 provides a detailed discussion of the mechanisms of why traders adopt similar strategies

traders who share a similar strategy with many other traders may attract more followers.

On the other hand, when a trader shares a similar strategy with many other traders, it can significantly increase competition, potentially hurting the focal trader due to the *substitution effect*. From an economic perspective, traders who adopt similar trading strategies can be seen as providing substitute goods — they offer similar trading strategies and target the same customer group: followers seeking specific trading patterns. The substitution effect has been observed on many online platforms, such as social media and crowdfunding. This effect is intuitive in competitive environments where many users offer similar content or services; thereby, one user can be replaced or substituted by another.

Specifically, on social trading platforms, traders receive financial compensation when followers replicate their trading strategies and followers pay following fees to follow traders. In contrast, users on social media platforms (e.g., X, formerly known as Twitter) do not receive direct monetary rewards for gaining followers, nor do users pay to follow others. We acknowledge the growing trend that access to social media content is increasingly becoming monetized, with users now required to pay subscription fees to access certain content. Our findings may offer relevant implications for this emerging shift in the social media landscape.[3] Similarly, on crowdfunding platforms (e.g., Kickstarter), users typically do not receive direct monetary returns from pledging to a campaign. Their motivation is often driven by personal interest or a desire to support innovation, based on subjective judgments such as the novelty or appeal of the campaign idea (Wei et al. 2022). In contrast, in the context of trading — particularly on Forex — individuals are unlikely to follow a trading strategy simply because it appears “novel". Moreover, once a crowdfunding campaign is launched, the idea and key attributes (e.g., the fundraising target) of the campaign are almost static. In contrast, trading strategies are dynamic, highlighting the importance of modeling changes in the trader similarity network over time, which has not been addressed in prior literature on similarity

in traditional financial markets.

[3] https://www.hostafrica.com/blog/news/musk-said-twitter-x-will-charge-users-for-subscription-fees/, last accessed on April 17, 2025.

networks in other contexts. The fundamental difference in incentive structures sets social trading apart and plays a critical role in shaping user behavior.

Given these two contrasting effects — *reputation effect* and *substitution effect* — a critical empirical question emerges: is adopting a similar strategy beneficial or detrimental for traders on social trading platforms? Understanding this dynamic tension provides pivotal insights for strategic decision-making among traders in the social trading ecosystem. In addition, it has important implications for platform operations, as the aggregate dynamics of individual traders influence the overall revenue of social trading platforms.

3.2.2 Herding Behavior of Followers

Herding is defined as the tendency for an individual to follow the actions of preceding peers (Banerjee 1992). A classic example of herding behavior is when potential consumers, faced with limited knowledge about the quality of restaurants, tend to choose the one with more customers. Herding has been studied in the fields of economics, marketing, and operations management. For example, Cai et al. (2009) examine the herding effect on dinner dish choices in a restaurant setting and find that when dinners are presented with information about popular dishes ordered by others, the demand for these dishes increases. In marketing, Zhang (2010) identifies one of the behavioral mechanisms behind herding, which describes situations in which people draw inferences from mere observations of others' actions and yields a set of marketing implications. In operations management, previous research focus on the herding in queue selection (Veeraraghavan and Debo 2011, 2009) and how herding affects inventory management (Hu et al. 2016).

Recently, herding has become particularly relevant on online platforms, as the Internet has vastly improved the observability of others' actions. For example, emerging peer-to-peer (P2P) lending platforms facilitate direct interactions between borrowers and lenders without intermediaries. In such cases, if a lender attracts a considerable number of borrowers, they are likely to attract additional funds, showcasing a clear manifestation of herding behavior (Brass 2015, Zhang and Liu 2012). Similarly, on crowdfunding platforms, herding behavior plays a significant role (Wei et al. 2021, Xiao et al. 2021, Burtch et al. 2021, Zhang et al. 2023, Kim and Viswanathan 2019, Jiang et al. 2022). Backers of a campaign often make their funding decisions based on the actions of previous backers. When a campaign receives initial backing, it signals potential success to other users, enticing them to contribute. As more backers contribute, the campaign gains more visibility, further reinforcing herding. In addition, Da and Huang (2020) study the herding problem in a crowd-based corporate earning forecast platform and find that herding leads to a worse consensus.

On social trading platforms, a follower can observe each trader's existing followers. Based on herding theory, they might herd with others' choices by following traders with many followers. Our research provides empirical evidence on whether herding among followers exists.

3.3 Data

Our data are from ZuluTrade.com, a leading social trading platform mainly for the Forex market operating worldwide.[4] The platform has two types of accounts: trader accounts and follower accounts. The social trading platform enables follower accounts to follow trader accounts. After a follower decides to follow a trader, the trader's investment decisions (i.e., trades) are automatically executed or replicated in the follower's account. The trader earns a following fee determined by the number of followers or the amount of money following. Consequently, the number of followers or the amount of money following the trader represents the additional benefit (extra payoff) a trader receives besides the direct profits from their trades.

Each trader account has a public profile page that displays the trader's trading history (at the granular level of trade), financial indicators measured from the trader's trades, the number of followers following the trader, etc. Potential followers can rely on the complete

[4] Our data includes a reasonable proportion of traders from the United States, along with a well-diversified mix of users from various countries. This geographic diversity supports the generalizability of our results across different countries, and our empirical findings are also applicable to the United States. Please refer to Appendix A.2 for more details.

information displayed on the trader's public profile to evaluate the trader and decide whether to follow the trader. A typical trader's profile contains three pieces of information: 1) historical trade-level information; 2) aggregated financial measures calculated from the trader's trading history; and 3) social measures including the number of followers following the trader, the amount of money following the trader, and the number of views that the trader's profile has received.

We obtain the complete trading history and the corresponding financial and social measures of 3,332 traders from May to July 2017. We consider the time when the trader opened her first trade as the time of joining the platform (as a basis to calculate the trader's trading experience). We consider each month as a period. The information of each trade from the trading history includes the currency (e.g., USD/EUR), type (buy or sell), open time (the timestamp when opening the trade), open price (the spot price at the time when the trade was opened), close time (the timestamp when closing the trade), close price (the spot price at the time the trade was closed), and lot size (a common unit in Forex; a standard lot size is 100,000 units of the base currency). From the trading history, we calculate the win ratio ($WinRatio_{i,t}$) of each trader i at period t to measure her financial performance. The win ratio is the number of winning trades divided by the total number of trades of trader i at period t. The win ratio is a key metric to evaluate the performance of an investment strategy, and a consistently high win ratio indicates a higher probability of generating positive returns over time. From the trading history, for each trader i at period t, we also calculate the number of trades ($Trade_{i,t}$) to capture how active the trader is, the number of months since trader i joined the platform ($Experience_{i,t}$) at time t to capture trading experience, and the average lot size of trades ($Volume_{i,t}$) to assess the risk exposure of trader. Traders who consistently trade larger lot sizes may expose themselves to higher levels of risk. In addition, for each trader i at period t, we have the number of views their profile has received ($View_{i,t}$). The number of views reflects the amount of attention the trader has received from potential followers. The platform periodically ranks traders, and this rank is publicly displayed on each trader's profile under the label "ZuluRank," which reflects their standing among all traders on the platform. While the exact algorithm used to generate these rankings is not publicly disclosed, the platform does indicate that its proprietary ZuluRank system considers multiple parameters, such as trading performance and other relevant metrics.[5] For better interpretation, we standardize the rank values ($Rank_{i,t}$) to range from 0 to 1, with higher values indicating better rankings.

The number of followers following trader i ($Follower_{i,t}$) is the key attribute of a trader ($Z_{i,t}$), as the number of followers measures the popularity of a trader and acts as social proof of a trader's expertise. When potential followers see that a trader has a large number of followers, it may give them confidence in the trader's abilities and increase the likelihood of following that trader. Lastly, traders can earn additional income based on the number of followers they have. Therefore, the number of followers impacts a trader's potential payoff on the platform. In the robustness check, we consider the amount of money following a trader ($Amount_{i,t}$) as an alternative key attribute of a trader.

Table 3.1 presents summary statistics of these variables. We observe considerable variation in all the variables, suggesting a diverse sample of traders. On average, a trader makes approximately 31 trades per month, with a standard deviation of approximately 109, indicating a wide spread of trading frequency. The gap between the number of views and the number of followers is also noteworthy. On average, a trader attracts over 26,000 views, and the conversion from viewing to following is relatively low. More than half of the traders do not have any followers, and the average number of followers is approximately 7. This discrepancy suggests that a high number of views does not necessarily translate into a high number of followers. The average win ratio is approximately 0.69, suggesting that many traders on the social trading platform have a good winning record. We present a correlation matrix in Table 3.2. We observe a high correlation between the number of followers and the amount of money following a trader. Apart from this relationship, the correlation between our variables remains moderate. Hence, possible multicollinearity is not a concern in our

[5] https://fxcmglobal.zulutrade.com/zuluranking, last accessed on April 17, 2025.

analysis.

Table 3.1: Summary Statistics of Variables

Variable	Mean	Std. Dev.	Min.	Median	Max.
$Follower_{i,t}$	6.5734	49.0386	0	0	1553
$Amount_{i,t}$	10049.1053	89986.6121	0	0	2858457
$Trade_{i,t}$	30.7491	108.8633	0	0	1754
$View_{i,t}$	26029.9274	78168.9648	133.2500	4944	1323505
$WinRatio_{i,t}$	0.6868	0.2378	0	0.7250	1
$Volume_{i,t}$	275624.6261	4188453.7064	0	0	100000000
$Experience_{i,t}$	22.3648	17.9591	6	15	110
$Rank_{i,t}$	0.6275	0.3157	0	0.6487	1

Note: We apply a log transformation to $Trade_{i,t}$, $Volume_{i,t}$, $View_{i,t}$, $Follower_{i,t}$, and $Amount_{i,t}$ in our analysis due to the skewness of the distribution.

Table 3.2: Correlation Matrix of Variables

	$Follower_{i,t}$	$Amount_{i,t}$	$Trade_{i,t}$	$View_{i,t}$	$WinRatio_{i,t}$	$Volume_{i,t}$	$Experience_{i,t}$	$Rank_{i,t}$
$Follower_{i,t}$	1	0.8760	0.0731	0.3062	0.0583	-0.0023	0.0851	0.0946
$Amount_{i,t}$	0.8760	1	0.1307	0.0757	0.0745	0.0017	0.0089	0.1141
$Trade_{i,t}$	0.0731	0.1307	1	0.0463	0.0540	0.0059	0.0065	0.0673
$View_{i,t}$	0.3062	0.0757	0.0463	1	0.0143	0.0055	0.5980	0.0051
$WinRatio_{i,t}$	0.0583	0.0745	0.0540	0.0143	1	0.0008	-0.0657	0.4885
$Volume_{i,t}$	-0.0023	0.0017	0.0059	0.0055	0.0008	1	0.0184	0.0093
$Experience_{i,t}$	0.0851	0.0089	0.0065	0.5980	-0.0657	0.0184	1	-0.0306
$Rank_{i,t}$	0.0946	0.1141	0.0673	0.0051	0.4885	0.0093	-0.0306	1

Whether the financial market is bullish or bearish can significantly influence investors' trading strategies, which might, in turn, affect the generalizability of our results. However, our dataset includes only Forex trading, as the platform primarily focused on Forex trading during the time covered by our data. Unlike the stock market, the Forex market offers investors opportunities in both bull and bear markets. This is because Forex trading is always in pairs; when the price of one currency is falling (a bear market), the other is rising (a bull market). This allows investors to take advantage of both rising and falling markets, as they can choose to go long (buy) or short (sell) on a currency pair, for example, USD/EUR. We observe a diverse mix of bullish, bearish, and mixed market conditions in our dataset, which supports the generalizability of our findings across different market environments. Please refer to Appendix A.2 for more details.

3.4 Similarity Network

In this section, we first describe how we construct the similarity network among traders. In our dataset, the primary financial instruments traded on the platform are currency pairs in the Forex market during the observed periods. To construct a similarity network of trading strategies, we consider various factors, including the currency pair, trade type (buy or sell; open or close), opening and closing times, trade size, and trading frequency. These elements allow us to develop a comprehensive measure for comparing traders' strategies.[6] To better understand the network, we visualize its structure and present descriptive statistics.

[6] While it might be too complex for potential followers to apply our academically comprehensive approach in identifying traders with similar trading strategies, they can still identify similar strategies through intuitive evaluations based solely on currency pairs. Even when using only the currency pair as a basis for similarity, the identified network is significantly positively correlated with the similarity network derived from our algorithm. The correlation coefficient of the similarity score is 0.401, with a p-value less than 0.001. A comprehensive algorithm for identifying traders with similar trading strategies is practical and vital from the platform's perspective, as it can help regulate and maintain the platform's ecosystem.

3.4.1 Network Construction

In our constructed network, each node represents a trader. For each pair of traders, we compute a similarity score, denoted as S_{ijt}, based on trader i and j's trading activities in period (month) t. We next elaborate on how we calculate the value of S_{ijt} and construct the similarity network step by step.

In **Step 1**, we first create a vector representation, or "embedding," for each trader, denoted as U_{it} for trader i at period t, where $U_{it} \in R^d$ and d is the dimension of the vector embedding. Each element in the vector embedding, denoted by U_{ikt} where $k \in \{1, 2, 3, \ldots, d\}$, presents a combination of a Forex (for example, EUR/USD) and a trade type (buy or sell; open or close). For instance, open a buy position in EUR/USD, open a sell position in EUR/USD, close a buy position in USD/EUR, close a sell position in USD/EUR, etc. Therefore, the dimension of the vector is the total number of unique combinations of financial instruments and trade types, with a value of 380 (that is calculated by 95 (unique currencies) $\times$ 2 (sell or buy) $\times$ 2 (open or close)). The value of an element is set to 1 if there is at least one such trade (for example, opening a buy position in EUR/USD) and to 0 otherwise. The vector enables us to capture the distinctive combination of financial instruments and trade types a trader may engage with. For example, we can tell whether traders have a diversified trading strategy in their instrument choices or focus on a few selected ones.[7]

In **Step 2**, based on the vector in the first step, we use cosine similarity (Salton et al. 1975) to calculate a score for each pair of traders. Cosine similarity is a widely applied metric of similarity in various contexts, such as identifying similar users (Provost et al. 2015), similar

[7] We acknowledge that portfolio holding information is valuable and widely utilized, as evidenced by mutual funds, which typically disclose their portfolio holdings rather than individual trading transactions. In mutual fund disclosures, detailed transaction-level data, such as the exact timing of trade openings and closings at the second level, is generally not publicly available. In contrast, our research focuses on trading transactions, which provide a more detailed and granular view than holdings data. An important characteristic of our dataset is the relatively short duration of asset holdings. Specifically, the median holding time is approximately 7 hours, 10 minutes, and 18 seconds, and even the 90th percentile of holding time remains under one week. Therefore, constructing a similarity network based on trading transactions will likely be more informative.

photos (Zeng and Wei 2013), and similar mutual funds (Koch 2017). It enables us to measure similarity based on the relative distribution of a trader's trading activities across different types of financial instruments. Mathematically,

$$\frac{\sum_k U_{ikt} U_{jkt}}{\sqrt{\sum_k U_{ikt}^2 \sum_k U_{jkt}^2}}. \tag{3.1}$$

The above similarity captures how similar trader i and trader j are in terms of financial instrument and trade type (buy or sell; open or close).

In **Step 3**, in addition to financial instrument and trade type, to account for other trading patterns and behaviors, we adjust the cosine similarity using trading frequency, average transaction price, and trading volume. The additional measures account for crucial dimensions of trading: when and how frequently and at what volume trades are made. The adjustments enable us to capture more nuanced aspects of trading behavior that a pure instrument-type-based approach might overlook.

Specifically, we first consider the frequency of trades, denoted as N_{ikt}, which is the number of trades for the k-th combination of financial instrument and trade type from trader i at period t. This measurement captures how active a trader is. Two traders may engage in similar trades, but their overall trading patterns may diverge substantially if one trader is much more active than the other. To account for this, we compute the relative difference in trading frequencies, denoted as $FreqSIM_{ijkt}$. Mathematically,

$$FreqSIM_{ijkt} = 1 - \frac{|N_{ikt} - N_{jkt}|}{max(N_{ikt}, N_{jkt})}. \tag{3.2}$$

This adjustment ensures that traders engaging in similar trades with similar frequencies have higher similarity scores, thus presenting a more comprehensive view of similarity.

Next, we consider the average transaction price, represented as $AvgPrice_{ikt}$, which is the average price of transactions for the k-th combination of financial instrument and trade type from trader i at period t. The average transaction price captures the timing of trades, a critical aspect given the fluctuating nature of financial instruments' prices. While two traders may trade the same type of instrument, the timing of these trades (reflected in different prices) may reveal differing trading strategies and skills. To capture these nuances, we adjust for the average transaction price, denoted as $PriceSIM_{ijkt}$. Mathematically,

$$PriceSIM_{ijkt} = 1 - \frac{|AvgPrice_{ikt} - AvgPrice_{jkt}|}{max(AvgPrice_{ikt}, AvgPrice_{jkt})}. \tag{3.3}$$

This adjustment enables us to account for the similarities in market timing behaviors among traders.

Additionally, we consider the average transaction lot size, represented as $AvgVolume_{ikt}$, which is the average lot size of transactions for the k-th combination of financial instrument and trade type from trader i at period t. This measure captures variations in trading quantity, which is a crucial element of trading strategy. Although two traders might engage in similar trades with comparable frequency and timing, their investment volumes could vary significantly. To account for this difference, we compute the relative difference in average lot size, denoted as $volumeSIM_{ijkt}$. Mathematically,

$$VolumeSIM_{ijkt} = 1 - \frac{|AvgVolume_{ikt} - AvgVolume_{jkt}|}{max(AvgVolume_{ikt}, AvgVolume_{jkt})}. \tag{3.4}$$

Next, we compute the final similarity score as follows:

$$S_{ijt} = \frac{\sum_k U_{ikt} U_{jkt} FreqSIM_{ijkt} PriceSIM_{ijkt} VolumeSIM_{ijkt}}{\sqrt{\sum_k U_{ikt}^2 \sum_k U_{jkt}^2}}. \tag{3.5}$$

The similarity score, S_{ijt}, captures the instrument, type, frequency, timing, and volume of the trades — these key components that together define a trader's strategy. These adjustments are distance-based. Consequently, it enables the identification of similar traders who employ similar trading strategies and behaviors.

We use a binary vector combined with distance-based adjustments, rather than directly encoding trading frequency, prices, or volume into the vector representation U_{it} (a weighted vector approach), since cosine similarity primarily captures the directional alignment between vectors, rather than similarities in their magnitude. As a result, a weighted vector approach does not effectively reflect differences in trading frequency, prices, or volume between traders.

We provide a detailed example illustrating how our similarity score is calculated and why the weighted-vector approach does not work in Appendix A.3.

In **Step 4**, we choose a threshold value of 0.7.[8] If the similarity score between two traders on a social trading platform is high, it suggests that they are adopting similar trading strategies. That is, if the similarity score S_{ijt} is greater than the threshold value, we set an undirected link between trader (node) i and trader (node) j. We repeat the process for all the pairs of traders and construct the similarity network among all the traders on the platform. We summarize the network construction process in the following pseudo-code.

Algorithm 1 Trader Similarity Network Construction

Require: A list of traders' transactions

Ensure: A similarity network

```
for each trader i at period t do
    create a binary vector representation U_it
end for
for each pair of traders (i, j) at period t do
    compute similarity score S_ijt based on Equation (3.5)
    if S_ijt > 0.7 then
        link trader i and trader j at period t
    end if
end for
return similarity network among all traders
```

To validate the effectiveness of our algorithm, we randomly select a subsample of 3,183 trader pair observations of 1,941 traders from our full dataset. We recruited six college students with finance backgrounds to manually compare the transaction histories of pairs of traders and label them as similar or dissimilar. We compare the results of our algorithm

[8] We try alternative values in the robustness check.

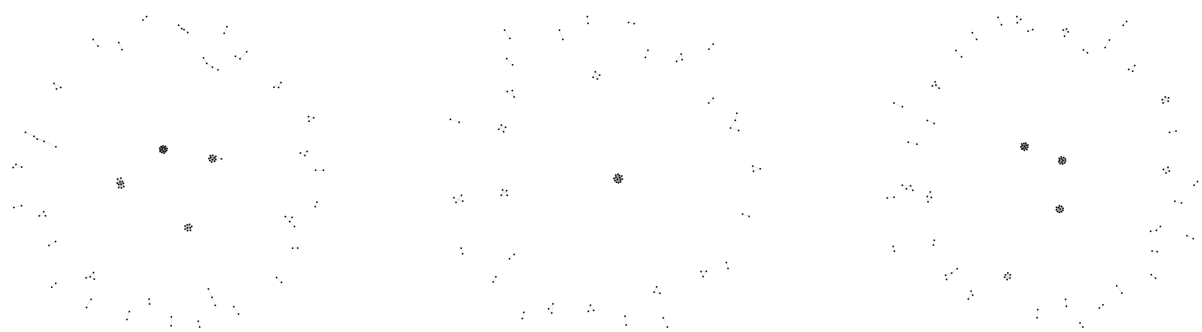
with those obtained from the human evaluation. We find a high consistency between the two, with an accuracy of 97.37%. This indicates that our algorithm can effectively capture the similarities in traders' trading strategies and that the construction of our similarity network is reliable.

The use of human labels as ground truth in training or validating machine learning models is well established in the literature, particularly in design science and empirical studies in OM. As summarized in Bastani et al. (2022), human-labeled data is a common approach in supervised machine learning, and has been used in recent studies (Te'eni et al. 2023, Dong et al. 2024). We acknowledge the potential subjectivity relying on human judgment to establish the ground truth. To address this concern, our final labels are not based on a single individual's assessment; rather, they are the result of aggregating judgments from six human participants. Additionally, we provided them with detailed labeling instructions to ensure consistency throughout the process (see Appendix A.4 for details).

3.4.2 Network Description

Following our proposed detection algorithm, we identify 789 pairs of similar traders over three months. To gain an initial understanding of the network's structure, we visualize the network of each month, as illustrated in Figure 3.1. First, we notice a couple of dominant clusters composed of many traders. These traders engage in similar financial instruments, employ similar trade types, and display similar trading frequency, volumes, and timing patterns. This cluster formation likely indicates a couple of popular trading strategies broadly adopted among a considerable portion of traders on the platform. Second, our network features numerous small clusters, indicating the presence of various trading strategies on the platform. Though small, these clusters represent groups of traders who closely mirror each other's trading behaviors. Third, our visualization does not include isolated traders. These traders, exhibiting unique trading patterns not closely matched with other traders, constitute a considerable 89.6% of the platform's traders. These isolated traders signify a wide array of unique trading strategies. We report the network statistics in Table A.5 in

Appendix A.5.



(a) Period 1 (b) Period 2 (c) Period 3

Figure 3.1: Illustrations of Trader Similarity Networks

3.5 Model and Results

In this section, we first describe how we set up the coevolution model to investigate the interplay between the evolution of a similarity network of traders adopting similar trading strategies (hereafter, trader similarity network) and the continuous trader attribute, namely a trader's number of followers. We then present the estimation results of the main model and heterogeneity analysis.

3.5.1 Coevolution Model

The main aim of the coevolution model is to disentangle the dynamics between the formation of the trader similarity network and the trader attribute, as mentioned in Section 2.2.2. Network dynamics and attribute dynamics mutually influence each other or coevolve over time. In addition, as a network analysis approach, the coevolution model incorporates the network structure capturing the potential tie dependence among nodes. More importantly, it incorporates the effect of network structure on attribute evolution. Finally, the evolution process of the attribute is modeled by SDE, which captures not only the deterministic component but also the stochastic nature of the attribute. Following Niezink et al. (2019), we model the network evolution using a continuous-time Markov chain and the evolution of trader attribute using SDE.

Network Evolution

The network is defined by its node (trader) set $\mathcal{N} = \{1, 2, ..., n\}$. We redefine $t \in R^+$ to represent continuous time, as in the coevolution model, the network and attribute evolve continuously in time. Conversely, we define $m \in \{1, 2, ...\}$ as discrete periods, since the data used for model estimation is recorded in discrete intervals (e.g., months 1, 2, and 3). The binary tie (or link) variable $x_{ij,t}$ represents an undirected relation between trader i and trader j at time t. $x_{ij,t} = 1$ ($x_{ij,t} = 0$) indicates the presence (absence) of a tie between trader i and trader j. The network is represented by the $n \times n$ adjacency matrix X, and the network at time t is denoted as X_t. X_t is the state variable of network evolution, where each element corresponds to $x_{ij,t}$. Note that the network is symmetric as the network is undirected.

We decompose the network evolution process into two subprocesses: (1) modeling the speed of the network change (i.e., the rate at which each trader in the network obtains an opportunity to change its ties) and (2) determining which particular tie is changed when the opportunity arises. We assume that trader i's waiting time until the next opportunity to change a tie is exponentially distributed with a rate λ_i. The rate λ_i is period-dependent (i.e., $\lambda_i = \lambda_m$).[9] Therefore, the waiting time until traders change a tie is exponentially distributed with rate $\lambda_+ = \sum_{i=1}^{n} \lambda_i = n\lambda_m$.

If trader i obtains the opportunity to make a network change, she can choose to change a tie to one of the other traders or maintain the status quo. We use $x^{(\pm ij,t)}$ to denote the adjacency matrix of the state x, in which $x_{ij,t}$ is changed to $1 - x_{ij,t}$, denoting that trader i changes its tie with trader j. (We suppress period m in the network for notational simplicity.)

[9]Period m ($m = 1, 2, ...$) refers to the time between t_m and t_{m+1}.

The utility of trader i choosing to change a tie with trader j, denoted as $f_{i,t}(x_{ij,t}, z_t, \vec{c}_t)$, takes into account the potential new network state $x_{ij,t}$, the key attribute of traders $z_t = \{z_{i,t}\}_{i \in \mathcal{N}}$, and a vector of control variables $\vec{c}_t = \{\vec{c}_{i,t}\}_{i \in \mathcal{N}}$. The probability that trader i selects to change the tie with trader j is modeled using a multinomial logit model as

$$p(x^{(\pm ij,t)}) = \frac{exp^{f_{i,t}(x_{ij,t}, z_t, \vec{c}_t)}}{\sum_{h=1}^{n} exp^{f_{i,t}(x_{ih,t}, z_t, \vec{c}_t)}}, \tag{3.6}$$

where $x^{(\pm ij,t)}$ denotes the adjacency matrix and $f_{i,t}(x_{ij,t}, z_t, \vec{c}_t)$ is the utility of trader i choosing to change a tie to trader j.

The utility of trader i is defined as a weighted sum of various network effects that affect network formation. Mathematically,

$$f_{i,t}(x_{ij,t}, z_t, \vec{c}_t) = \sum_{l} \beta_l s_{il}, \tag{3.7}$$

where s_{il} refers to the l-th network effect and β_l represents the corresponding coefficient. We control the following network effects in the utility.

- Degree effect:

$$\sum_{j} x_{ij,t},$$

  which measures the general tendency of trader i to form ties with other traders.

- Alter effect of the key attribute:

$$\sum_{j} x_{ij,t} z_{j,t},$$

  which measures the tendency of trader i to form ties with other traders with attribute value z_j. We include the alter effect of followers to account for the possible tendency whether traders tend to form ties with traders with a higher number of followers.

- Homophily effect of the key attribute and control variables:

$$\sum_{j} x_{ij,t} \left(1 - \frac{|z_{j,t} - z_{i,t}|}{Range(z)}\right),$$

$$\sum_j x_{ij,t} \left(1 - \frac{|c_{j,t} - c_{i,t}|}{Range(c)}\right),$$

where $Range(z)$ and $Range(c)$ denote the ranges of the key attribute and each control variable, respectively. The homophily effect measures the tendency of trader i to form ties with trader j when they are similar in terms of the key attribute (the number of followers, $Follower_{i,t}$) or control variable. Specifically, for control variables, we include the homophily effect of all control variables, including trading frequency ($Trades_{i,t}$), trading performance ($WinRatio_{i,t}$), social attention ($Views_{i,t}$), trading experience ($Experience_{i,t}$), and rank ($Rank_{i,t}$).

We provide a numerical illustration on how each variable is calculated in Appendix A.6. We summarize all the network effects in the utility of a specific network change of ij of trader i in Equation (3.8).

$$\begin{aligned} f_{i,t}(x_{ij,t}, z_t, \vec{c}_t) =& \beta_1 Degree_{i,t} + \beta_2 FollowerAlter_{i,t} + \beta_3 FollowerHomophily_{i,t} \\ &+ \beta_4 WinRatioHomophily_{i,t} + \beta_5 TradesHomophily_{i,t} + \beta_6 ViewsHomophily_{i,t} \\ &+ \beta_7 VolumeHomophily_{i,t} + \beta_8 ExperienceHomophily_{i,t} + \beta_9 RankHomophily_{i,t}. \end{aligned} \tag{3.8}$$

Attribute Evolution

We use a SDE to model the evolution of the key attribute of the trader — the number of followers. The change in the number of followers in a small time interval from time t to time $t + dt$ ($dZ_{i,t}$) is modeled as

$$\begin{aligned} dZ_{i,t} =& \tau_m(aZ_{i,t} + b_0 + b_1 SimilarPeers_{i,t} + c_1 WinRatio_{i,t} + c_2 Trades_{i,t} \\ &+ c_3 Views_{i,t} + c_4 Volume_{i,t} + c_5 Experience_{i,t} + c_6 Rank_{i,t})dt + \sqrt{\tau_m} dW_{i,t}, \end{aligned} \tag{3.9}$$

where $Z_{i,t}$ denotes the number of followers of trader i at time t, $dW_{i,t}$ is the increment of a Wiener process following a normal distribution with mean zero and variance dt, $SimilarPeers_{i,t}$ denotes the degree (the number of similar peers) of trader i in the similarity network at time

t. a is the feedback parameter, b_0 is the intercept, and b_1 is the effect of similar peers on the number of followers. τ_m is the period-specific coefficient comparable to λ_m and accounts for the speed of key attribute evolution. We also include $WinRatio_{i,t}$, $Trades_{i,t}$, $Volume_{i,t}$, $Experience_{i,t}$, $View_{i,t}$ and $Rank_{i,t}$ to control for time-varying factors that may affect the number of followers. We highlight that a captures the herding effect as it measures the effect of the existing number of followers on the incremental number of followers in the next period. Moreover, the coefficient b_1 captures the effect of the number of similar peers of trader i on her number of followers, which is the main effect to be investigated in our study. We provide a summary and description of all parameters and variables in Appendix A.7.

Coevolution of Network and Attribute

From Equations (3.8) and (3.9), the key attribute $Z_{i,t}$ is incorporated into the network evolution, and the network structure is considered in attribute evolution. This is how we model the interplay between two dependent variables — the trader similarity network and the trader's number of followers. Therefore, the network and attribute coevolve over time. To summarize, in our coevolution model, we have two models (equations), thus two dependent variables and two sets of independent variables.

In the network evolution model, the dependent variable is $x^{(\pm ij,t)}$, a binary variable indicating whether a link between two traders (trader i and trader j) changes at time t in the trader similarity network; and the independent variables include degree, follower alter effect, and homophily effects among a series of characteristics of a trader. The key independent variable in the network evolution model is $FollowerAlter_{i,t}$, which measures how the number of followers of a trader affects the network change $x_{ij,t}$, thus capturing how the evolution of followers affects the evolution of the trader similarity network. $Degree_{i,t}$ and a series of homophily effects are control variables, capturing the standard degree-based (preferential) attachment effect (Barabási and Albert 1999) and homophily effect (McPherson et al. 2001).

In the attribute evolution model, the dependent variable is the number of followers of a trader i at time t, $Z_{i,t}$. Since we model the evolution of the dependent variable using a SDE, we capture how $dZ_{i,t}$ (the infinitesimal change of $Z_{i,t}$ in a small time interval dt) is affected by the independent variables, including two key independent variables: the current number of followers $Z_{i,t}$, capturing the herding effect (the tendency for popular traders to attract more followers over time); and the number of similar peers $SimilarPeers_{i,t}$, capturing whether the number of similar peers benefit (reputation effect) or hurt (substitution effect) the number of followers of a trader.

We use the R package RSiena to estimate the model parameters (Ripley et al. 2021). The estimation uses the simulated methods of moments, with Monte Carlo simulation and the Robbins Monro algorithm (Robbins and Monro 1951) adopted. Details can be found in Appendix A.8. We further explain our choice of the coevolution model over a regression approach and report the corresponding goodness-of-fit statistics in Appendix A.9.

3.5.2 Results

Our data includes 3,332 traders over three periods. As a large number of nodes can cause computational intractability issues, especially for the coevolution model, we sample a smaller set of nodes to run the coevolution model.[10] Specifically, we employ a snowball sampling approach, a widely used method for identifying subsets within large networks (Goodman 1961, Zeng and Wei 2013), to construct the sub-sample. Our main analysis on the coevolution model is based on a sample of 387 traders. The strength of snowball sampling lies in its ability to capture non-isolated traders while incorporating a fraction of isolated traders. Consequently, this enhances the extent to which our sample represents the broader dynamics of the trading network. We detail our sampling strategy and the representation of our sample in Appendix A.5. To mitigate the concern of the representation of our sample, we re-run our model with another random draw of samples as a robustness check.

[10]It is a common practice to sample a smaller set of nodes to achieve computational feasibility when estimating network analysis models (Yan et al. 2015, Lee et al. 2016). For example, Lee et al. (2016) study strategic network formation in a location-based social network. Their network analyses are conducted on three city-level subsamples consisting of 336, 129, and 146 users. Moreover, Yan et al. (2015) examine the driving forces behind patients' social network formation and evolution using a subsample consisting of 1,322 individuals.

The estimation results are shown in Table 3.3. For network dynamics, we first find that a trader with more similar traders (*Degree*) has a lower tendency to form new ties with other traders (negative degree effect). Second, homophily effects in terms of followers, trades, views, volume, and experience are significantly positive in the likelihood of forming ties in the trader similarity network. The results suggest that traders tend to adopt similar trading strategies with others similar to them in terms of trading frequency (*Trades Homophily*), trading volume (*Volume Homophily*), trading experience (*Experience Homophily*), social attention (*Views Homophily*), as well as popularity (*Follower Homophily*). By contrast, the homophily effects regarding the win ratio and rank are insignificant. Third, we find a significant positive effect of the neighbors' total number of followers (*Follower Alter*) on the likelihood of forming ties in the trader similarity network. Thus, traders also adopt trading strategies based on the number of followers of a trader. The alter effect of the number of followers, defined as the total number of followers of a trader's similar peers, also capture whether the observed similarity in trading strategies emerges from imitation of more popular peers (that is, with more followers) or less popular ones (that is, with less followers). Our results indicate that similarity in trading strategies stems from imitating more popular peers. This further validates the necessity of modeling the network formation and follower dynamics as a coevolution process. We provide a detailed interpretation of the magnitude of coefficient estimates in the network evolution model in Appendix A.10.

For the attribute dynamics, the key coefficient (b_1) is significantly negative, implying that when a trader has more similar peers, her number of followers will decrease over time. In other words, the number of similar peer traders hurts the focal trader in terms of her popularity and extra financial gain. The feedback coefficient (a) is also significant. Using Equation (3.9), we can calculate the magnitude of the herding effect among followers. Holding all other variables constant, the effect of the number of existing followers on the subsequent number of followers (herding effect among followers) is calculated as $(1 + a\tau_m) \cdot Z_{i,t}$. Thus, our estimations reveal a significant herding effect among followers with an average coefficient of 0.981. In addition, all control variables are insignificant, suggesting that the number of followers is driven mainly by the herding effect and the trader similarity network structure. This emphasizes the critical role of the trader similarity network in the context of social trading. Finally, we report the estimated rate parameters, which are all significant.

To further interpret the economic meaning of the estimated coefficients, we substitute the estimated coefficients (using the average values across all periods when statistically significant) into the attribute evolution equation and then rearrange the equation. The expected number of followers of trader i at time $t+1$ given the number of followers of trader i at time t is presented as:

$$\mathbb{E}\left(Z_{i,t+1} \mid Z_{i,t}\right) = -0.3 + 0.98 \cdot Z_{i,t} - 0.16 \cdot SimilarPeers_{i,t}. \tag{3.10}$$

Increasing one follower for a trader at time t results in an average 0.98 unit increase in the number of followers in the subsequent period $(t+1)$. In contrast, increasing one similar peer (trader) for a trader at time t results in an average 0.16 unit decrease in the number of followers in the subsequent period $(t+1)$.

3.5.3 Heterogeneity analysis

We further conduct a heterogeneity analysis by introducing a new variable, $Success_{i,t}$, a binary indicator equal to 1 if the number of followers or win ratio of trader i at time t is above the 90th percentile of each respective variable. To be more specific, we have added an interaction term between $Success_{i,t}$ and $SimilarPeers_{i,t}$ in the attribute evolution equation, as specified in Equation (3.11). The network evolution equation remains the same as in the main model. The estimation results are reported in Table 3.4.

$$\begin{aligned} dZ_{i,t} =& \tau_m(aZ_{i,t} + b_0 + b_1 SimilarPeers_{i,t} + b_2 SimilarPeers_{i,t} \times Success_{i,t} \\ &+ c_1 WinRatio_{i,t} + c_2 Trades_{i,t} + c_3 Views_{i,t} + c_4 Volume_{i,t} + c_5 Experience_{i,t} \\ &+ c_6 Rank_{i,t} + c_7 Success_{i,t})dt + \sqrt{\tau_m} dW_{i,t} \end{aligned} \tag{3.11}$$

Our results reveal a significant heterogeneity effect between more successful and less successful traders. Specifically, the negative impact of $SimilarPeers$ is stronger for more

Table 3.3: Estimation Results of the Coevolution Model

		Estimate	Standard Error
Network Dynamics			
β_1	Degree	-16.8189***	5.0293
β_2	Follower Alter	1.1724***	0.1413
β_3	Follower Homophily	4.6951***	1.1274
β_4	WinRatio Homophily	31.8085	23.3513
β_5	Trades Homophily	4.6837***	1.5012
β_6	Views Homophily	8.0207***	1.8159
β_7	Volume Homophily	6.709***	1.4252
β_8	Experience Homophily	11.891***	2.5577
β_9	Rank Homophily	-0.4062	0.3544
Attribute Dynamics			
a	Feedback	-0.2977**	0.1517
b_0	Intercept	-0.302*	0.1604
b_1	SimilarPeers	-0.1579**	0.0802
c_1	WinRatio	0.5773	0.7998
c_2	Trades	0.1073	0.1389
c_3	Views	-0.2341	0.2349
c_4	Volume	-0.0037	0.0463
c_5	Experience	0.0202	0.0169
c_6	Rank	0.2493	0.5083
Rate Parameters			
λ_1	Network evolution (Period 1 to 2)	4.9247***	0.531
λ_2	Network evolution (Period 2 to 3)	0.5979***	0.1422
τ_1	Followers evolution (Period 1 to 2)	0.0839***	0.0057
τ_2	Followers evolution (Period 2 to 3)	0.0441***	0.0027

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

successful traders than for less successful ones. The focal trader's number of followers declines more sharply with a larger number of similar peers for more successful traders compared to less successful ones.

3.6 Counterfactual Analyses

Our model enables us to estimate the influence of various factors on an individual trader's number of followers and their evolution over time. Similar to traditional econometric models, the estimated coefficients from our coevolution model can be interpreted to capture individual-level effects, conditional on other factors. Beyond that, our model also enables counterfactual analyses at the platform level, offering insights into system-wide implications.

One key advantage of our coevolution model is its ability to conduct counterfactual analysis by capturing the dynamic interplay between a trader's number of followers and their position within the similarity network of all traders. In our framework, a trader's choice of trading strategy and the number of followers they attract are mutually interdependent. This structure enables us to simulate counterfactual scenarios to examine how some changes influence a central outcome variable — the total number of followers across all traders. This metric is particularly interesting to platform owners, as it directly relates to their revenue; a higher total number of followers leads to increased following fees collected by the platform. Specifically, we conduct two counterfactual analyses from the platform's perspective to illustrate how the total number of followers across all traders might evolve. Our first counterfactual simulation examines how the initial distribution of followers across the platform influences the subsequent evolution of the number of followers over time. Our second counterfactual simulation investigates how the structure of the similarity network among traders' trading strategies impacts the subsequent evolution of the number of followers over time.

Table 3.4: Estimation Results of Heterogeneity Analysis

		Estimate	Standard Error
Network Dynamics			
β_1	Degree	-16.8690***	3.9387
β_2	Follower Alter	1.1795***	0.1342
β_3	Follower Homophily	4.7077***	1.1312
β_4	WinRatio Homophily	31.8472	17.7731
β_5	Trades Homophily	4.7246***	1.3305
β_6	Views Homophily	8.0407***	1.8907
β_7	Volume Homophily	6.7698***	1.4744
β_8	Experience Homophily	11.9370***	2.4198
β_9	Rank Homophily	-0.3756	0.3195
Attribute Dynamics			
a	Feedback	-0.3593**	0.2032
b_0	Intercept	-0.2365*	0.1639
b_1	SimilarPeers	-0.2885**	0.1124
b_2	Success $\times$ SimilarPeers	-0.8197**	0.4144
c_1	WinRatio	0.2292	0.9203
c_2	Trades	0.1146	0.1429
c_3	Views	-0.1807	0.2329
c_4	Volume	0.0002	0.0462
c_5	Experience	0.0152	0.0165
c_6	Rank	0.1370	0.5149
c_7	Success	1.0765**	0.5886
Rate Parameters			
λ_1	Network evolution (Period 1 to 2)	4.9360***	0.5742
λ_2	Network evolution (Period 2 to 3)	0.5909***	0.1237
τ_1	Followers evolution (Period 1 to 2)	0.0839***	0.0054
τ_2	Followers evolution (Period 2 to 3)	0.0436***	0.0027

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.6.1 A fair distribution of the number of followers

From our data, we observe that the distribution of the number of followers among all the traders on the platform is highly skewed, with more than 75% of traders having zero followers. Such extreme concentration is not unique to our setting — it reflects a broader “long tail" (Brynjolfsson et al. 2003) phenomenon observed across various online platforms, for example, social media (Bakshy et al. 2011), where a small fraction of users — often referred to as the “head" — attracts the vast majority of engagement. In contrast, the remaining many users in the “tail" receive minimal attention.

At first glance, this may not appear problematic for the social trading platform, given that its revenue is generally proportional to the total number of followers across all traders — regardless of how followers are distributed among individual traders. However, as our results demonstrate, a trader’s follower dynamics are influenced not only by their own characteristics and performance but also by their position within the similarity network of all traders on the platform. Moreover, the number of followers a trader has can, in turn, shape the similarity network itself. This mutual dependency highlights a coevolutionary process among traders, suggesting that the distribution of followers across traders can, over time, influence the platform-wide total number of followers and thus impact overall platform revenue.

In our counterfactual, we keep the total number of followers across all traders constant but redistribute them to reduce the initial imbalance in the distribution of followers. Specifically, we assign five followers to each trader who initially has zero followers in our data, and decrease the number of followers for the top 5% of traders with the highest number of followers in our data. On average, each of the top 5% of traders loses 20.3% of their original followers in our data. This rebalancing procedure reallocates followers from highly followed traders to those with no followers, leading to a more equitable or fair distribution of followers across the platform. We then apply our coevolution model to simulate the follower dynamics for all traders with the counterfactual initial distribution of followers. We aggregate the simulation results on how each trader’s number of followers evolves and summarize the overall patterns.

Our results are shown as Counterfactual A in Table 3.5. Under the fair distribution of followers across all traders, we observe an average 2.08% increase in the total number of followers per period, compared to the baseline scenario with the highly skewed distribution of the number of followers in our original data. These counterfactual results suggest that reducing the imbalance in follower distribution can increase overall platform engagement, leading to a higher total number of followers in subsequent periods.

Table 3.5: Counterfactual Results on the Total Number of Followers at the Platform Level

	Period 1 to 2	Period 2 to 3	Average
Counterfactual A: Fair distribution of followers	+2.72%	+1.43%	**+2.08%**
Counterfactual B: Denser similarity network	-4.60%	-0.69%	**-2.65%**

Notes: This table reports the percentage difference in the total number of followers at the platform level between each counterfactual scenario and the current (baseline) setting. For each period transition (e.g., Period 1 to 2), we compute the growth rate in the number of followers under both the counterfactual and current scenarios, and report the difference between the two growth rates. Specifically, for a given period, the reported value is $\left(\frac{F_t^{\text{cf}} - F_{t-1}^{\text{cf}}}{F_{t-1}^{\text{cf}}}\right) - \left(\frac{F_t^{\text{baseline}} - F_{t-1}^{\text{baseline}}}{F_{t-1}^{\text{baseline}}}\right)$, where F_t^{cf} denotes the total number of followers on the platform at time t under the counterfactual scenario, and F_t^{baseline} denotes the total number of followers at time t under the baseline. The final column reports the average percentage difference across the periods. The magnitude difference between two period transitions is due to the differences in the rate parameters (λ_m and τ_m) across the two periods. All results are based on 1000 random draws of simulations.

3.6.2 A denser similarity network

The results of our main model estimations show that a larger number of similar peers has a negative impact on the number of followers an individual trader attracts. While this finding focuses on the trader-level dynamics, it prompts the question of how such effects manifest at the platform level. Through counterfactual simulations, we address the following questions:

How does the overall similarity in trading strategies among traders affect the total number of followers, which is closely tied to platform revenue? Over time, does low diversity in trading strategies diminish the total number of followers — and if so, to what extent? Diversity is defined based on the density of the similarity network among all traders on the platform, where lower diversity corresponds to a denser similarity network.

In this counterfactual, we explore how the density of the similarity network among all the traders influences the total number of followers from all the traders, which relates to the platform's revenue. To simulate a denser network, we assign five random links to each isolated node (trader) in our data — where "random" refers to randomly selecting five connections (links) between the focal trader and other traders on the platform — thereby increasing the overall network density.

Our results are shown in Counterfactual B in Table 3.5. With the denser similarity network, we observe an average 2.65% decrease in the total number of followers per period, compared to the baseline scenario with a sparser similarity network in our original data. This counterfactual analysis demonstrates that reduced diversity in trading strategies across the platform negatively impacts the platform's total follower count, ultimately diminishing its revenue. This adverse effect is observed not only at the individual trader level but also at the platform level.

3.7 Robustness Check

In this section, we conduct a series of robustness checks to validate our analysis. First, one concern of our study is that we choose a specific threshold value when constructing the similarity network. It is possible that our main results would be different if we use other threshold values. If the threshold value is tighter than that chosen in our main analysis, the network would be denser, and each individual would have more similar peers. We use the same sample as in the main model and only change the threshold value for constructing the network. Thus, we consider a tighter threshold (0.75) to label similar traders and rerun the coevolution model. We rerun the coevolution model with the results shown in Column (1) in Table 3.6. The results in terms of network dynamics and the key attribute dynamics are consistent with those in the main analysis.

Second, following the same logic as in the previous robustness check, it is also possible that the threshold value chosen in our main model is tighter than the actual value. Therefore, we consider a looser threshold (0.65) to label similar traders and rerun the coevolution model. We rerun the coevolution model with the results shown in Column (2) in Table 3.6. Again, the results in terms of the network dynamics and the key attribute dynamics are consistent with those in the main analysis.

Third, to enhance the robustness of our findings, we conduct an additional analysis using an alternative measure of the popularity of a trader. Instead of using the number of followers, we utilize the total amount of money following a trader. We present our results in Column (3) in Table 3.6. The coefficient b_1 remains significantly negative. This finding illustrates that the presence of similar peers has a negative impact on the amount of money following the trader. This finding further corroborates our primary conclusions.

Fourth, due to computational constraints, we use a sub-sample of the complete dataset to estimate the co-evolution model. In our main model, we have 387 traders. To mitigate the potential bias from sampling, we re-run our model using an alternative sub-sample of 384 nodes. The estimation results are shown in Column (4) in Table 3.6. Our main results remain qualitatively consistent.

3.8 Managerial Implications and Conclusion

Despite the emergence of social trading platforms, there is little research in this field. Although one prior study empirically found a high level of similarity of trades among traders on social trading platforms (Gemayel and Preda 2018), little is known about the impact of similarity patterns on traders. From the platform's perspective, understanding the effect of similarity patterns among traders' trading strategies can help manage the overall risk and increase revenue. The utility of traders gaining from trading is vital to the platform, as maintaining a pool of competent traders is key to the platform's success. By gaining insights

Table 3.6: Estimation Results of Robustness Checks

		(1) Tighter Threshold	(2) Looser Threshold	(3) Alternative Dependent Var.	(4) Alternative Sample
Network Dynamics					
β_1	Degree	-16.5085***	-16.9872***	-17.3768***	-13.8409***
		(4.7504)	(4.9272)	(7.6585)	(7.0193)
β_2	Follower Alter	1.0772***	0.9812***	0.4769***	0.6015***
		(0.1355)	(0.1094)	(0.0483)	(0.0559)
β_3	Follower Homophily	4.3988***	4.0437***	3.1796***	2.6478***
		(1.1743)	(0.9590)	(0.7338)	(0.7018)
β_4	WinRatio Homophily	31.6368	35.5071	32.7660	32.4507
		(22.4647)	(21.9549)	(34.8748)	(29.9932)
β_5	Trades Homophily	6.0828***	4.1054**	5.5469**	-1.0204***
		(1.9357)	(1.1169)	(1.3754)	(0.4278)
β_6	Views Homophily	5.7942***	7.2703***	8.2716***	1.4452***
		(1.8383)	(1.5745)	(1.8235)	(0.6785)
β_7	Volume Homophily	4.8772***	7.0010***	6.2307***	7.2856***
		(1.6265)	(1.3173)	(1.5340)	(0.7750)
β_8	Experience Homophily	7.1734***	10.8187***	11.0341***	7.5829***
		(2.0445)	(2.1208)	(2.3363)	(1.0720)
β_9	Rank Homophily	-1.0768**	0.3208	-0.2716	0.4129
		(0.4463)	(0.2952)	(0.3168)	(0.3834)
Attribute Dynamics					
a	Feedback	-0.3045**	-0.2957**	-0.0592**	-0.4205***
		(0.1544)	(0.1517)	(0.0088)	(0.1606)
b_0	Intercept	-0.3184**	-0.2878	-0.0138	-0.2512
		(0.1529)	(0.1540)	(0.0278)	(0.1869)
b_1	SimilarPeers	-0.1846*	-0.1608**	-0.0249*	-0.2600***
		(0.1066)	(0.0710)	(0.0133)	(0.0919)
c_1	WinRatio	0.6389	0.5744	0.0310	0.4854
		(0.8005)	(0.7983)	(0.1615)	(0.7748)
c_2	Trades	0.1010	0.1177	0.0239	-0.3903***
		(0.1450)	(0.1398)	(0.0257)	(0.1159)
c_3	Views	-0.2406	-0.2313	0.0702*	-0.0161
		(0.2217)	(0.2290)	(0.0366)	(0.2187)
c_4	Volume	-0.0050	-0.0130	0.0031	0.1029***
		(0.0468)	(0.0466)	(0.0083)	(0.0421)
c_5	Experience	0.0204	0.0198	-0.0026	0.0115
		(0.0163)	(0.0164)	(0.0029)	(0.0160)
c_6	Rank	0.1501	0.2471**	0.2195**	0.1270
		(0.5080)	(0.5251)	(0.1024)	(0.5849)
Rate Parameters					
λ_1	Network evolution (Period 1 to 2)	2.9906***	4.4167***	4.9016***	5.3125***
		(0.4349)	(0.5341)	(0.6174)	(0.5380)
λ_2	Network evolution (Period 2 to 3)	2.4984***	1.1713***	0.6878***	3.9001***
		(0.6406)	(0.2540)	(0.1572)	(1.0108)
τ_1	Followers evolution (Period 1 to 2)	0.0846***	0.0843***	2.4186***	0.0932***
		(0.0152)	(0.0032)	(0.2647)	(0.0047)
τ_2	Followers evolution (Period 2 to 3)	0.0441***	0.0439***	1.4853***	0.0658***
		(0.0026)	(0.0032)	(0.0978)	(0.0072)

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are reported in parentheses.

into the role of trader similarity networks, platforms can make operational decisions regarding fine-tuning the similarity network — nudging traders' decisions on trading strategies and followers' following decisions.

By leveraging a coevolution model, we capture the interplay between the evolution of the trader similarity network and follower dynamics. We find that when a trader has more similar peers, her number of followers will decrease over time. Our results show that the negative substitution effect of adopting a similar strategy dominates the positive reputation effect, and the overall effect of the similarity pattern on the trader's popularity is negative. We also find that the negative effect is more pronounced for more successful traders. In addition, we confirm a significant herding effect among followers. Moreover, our counterfactual simulations reveal that promoting a fairer distribution of followers and reducing the density of the similarity network can help enhance the total number of followers on the platform.

The contribution of our study is on the findings of modeling the dynamic similarity networks on the trader's popularity in emerging social trading platforms as well as providing implications through counterfactual analyses/simulations. A simple econometrics model is neither able to address the identification issue when analyzing the impact of dynamic similarity networks on the trader's popularity, nor can it illuminate implications through counterfactual analyses to answer "what-if" questions.

Our findings provide important managerial and operational implications for traders, followers, and social trading platforms. First, our study underscores the importance of incorporating similarity networks in the operations management of social trading platforms. A similarity network provides a bird's eye view of the interconnections among traders based on their trading strategies, carrying richer information than individual performance metrics. The proliferation of similar trading strategies among many traders poses challenges for platforms. For example, if a particular strategy proves unsuccessful, many traders and their corresponding followers adopting it could suffer simultaneous losses. Social trading platforms should encourage diversification and empower traders to explore different trading strategies to promote a healthy and sustainable trading environment. In addition, the similarity network can be a valuable tool for monitoring risk at the platform level. For instance, if a high-risk trading strategy becomes overly popular and widely adopted, as revealed through the similarity network, it could pose a systemic risk to the platform. Therefore, implementing risk control measures and identifying such risks in advance would allow platforms to take preventative measures to mitigate the associated risks.

Second, using a strategy that is similar to that of many other traders can hurt the focal trader in terms of the number of followers, which directly affects the social and financial utility derived from engaging in social trading platforms. The success of a social trading platform, similar to a two-sided market, depends heavily on having a substantial number of traders whom potential followers consider worthy of following. As the utility of traders declines due to the loss of followers, traders may become less active on or even leave the social trading platform, resulting in long-term losses for the platform. More importantly, our heterogeneity analysis shows that the negative effects of similar peers are particularly pronounced for more successful traders, who are particularly valuable to the platform. To counter the potential risk arising from traders adopting similar strategies, social trading platforms should encourage diversity in trading strategies. For example, the social trading platform could use our proposed algorithm to calculate a similarity score for each pair of traders on the platform. If a trader's score exceeds some designated threshold value, the platform can nudge the trader to diversify their trading strategy by sending automated reminders or notifications. Furthermore, in addition to encouraging more diversified trading strategies, platforms can also promote a fairer distribution of followers to help mitigate these adverse effects, as demonstrated by our counterfactual analysis.

Third, our findings reveal a strong herding effect when followers decide whom to follow, regardless of the traders' performance. While seemingly beneficial for some popular traders, herding among followers can lead to the "richer get richer" problem for traders, which has adverse implications for social trading platforms, followers, and traders. At the platform level, such herding behavior stifles diversity, as it concentrates attention on a small group of popular traders. This narrow focus on a few popular strategies increases the vulnerability of the entire platform to market risks and decreases overall resilience. For followers, the herding effect could lead to suboptimal investment decisions. By blindly imitating (following) popular traders, followers may overlook traders with few followers but superior strategies or unique market insights, potentially leading to missed opportunities for better financial returns. For traders, the herding effect can lead to the “richer get richer” problem, which can breed a sense of unfairness and discourage participation from those traders with superior strategies but few followers. As a result, those traders who possess valuable trading skills but receive less attention might leave the social trading platform, which in turn diminishes the follower base over time. The chain reaction mirrors that of a two-sided market. The long-term impact of a reduced follower base can extend even to popular traders. Therefore, addressing these issues is crucial to enhance revenue, foster diversity, and promote fairness among various stakeholders in social trading platforms. To address these challenges, platforms could implement features that highlight traders with profitable strategies but fewer followers, thus mitigating the herding effect. For instance, on a follower’s page, platforms can recommend skilled traders who have demonstrated good performance but have received less attention. In practice, certain social trading platforms offer rankings for all traders, which can be considered a form of platform recommendation. To mitigate herding, platforms can incorporate the trader’s popularity information (e.g., by assigning a negative weight to traders who receive excessive attention relative to their performance) into their ranking algorithm. Finally, to mitigate the herding effect, platforms can opt to disclose a range of follower numbers rather than the exact count for each trader.

Our study is not without limitations. First, the coevolution model suffers from “the curse of dimensionality" as the number of nodes increases; thus, our estimation is based on a medium-sized sample instead of the full sample. Second, we consider only the linear effect of similar peers on the attribute. There might be a nonlinear effect of similar peers on the number of followers. Third, in our current specification, we take similarity as a binary decision — i.e., we model whether a trader adopts a similar trading strategy as another trader, rather than modeling the intensity of similarity, which would correspond to a weighted network. Fourth, we acknowledge the potential subjectivity relying on human judgment to establish the ground truth, which may introduce bias despite our efforts to ensure consistency and reliability.

Chapter 4

SOCIAL LEARNING IN EARLY CRYPTOCURRENCY ADOPTION DYNAMICS

4.1 Introduction

Social trading platforms have gained increasing popularity in recent years. These platforms integrate the interactive elements of social media and social networks with brokerage services (Pelster and Hofmann 2018). Investors can trade financial instruments, and all transactions and social interactions are shared in a highly transparent environment, appearing as part of a news feed to investors' peers. In such a setting, individual investors can access the opinions shared by their peers to make their investment decisions (Ammann and Schaub 2021)—while transparently observing their trading behavior. Little prior research has examined how this interplay between expressed opinions and observable trading behaviors affects individual investors' investment decisions. In this study, we aim to fill this research gap.

We focus on cryptocurrency adoption because cryptocurrency is a novel asset built on the disruptive blockchain technology. Cryptocurrency has gained global attention since the advent of Bitcoin in 2009 (Nakamoto 2008). Understanding the diffusion of cryptocurrency adoption provides important insights for promoting similar new technologies. Cryptocurrency challenges conventional financial systems and offers a new avenue for investment. What once seemed like a niche concept for tech-savvy enthusiasts has now captured the mainstream audience's attention. In 2023, the U.S. Securities and Exchange Commission (SEC) allowed Bitcoin ETFs for the first time.[1] At the end of 2023, the cryptocurrency market comprised 8,848 active cryptocurrencies and approximately 420 million users. In the

[1] `https://apnews.com/article/bitcoin-exchange-traded-funds-etf-sec-59a5bb81ab891af57a1bd1765024144f`, last accessed on June 26, 2024.

United States, about 8% of the population engages in cryptocurrency trading.[2] In January 2024, the global cryptocurrency market capitalization stood at $1.62 trillion.[3]

Yet, cryptocurrency is characterized by its extreme price volatility and risk. Cryptocurrency markets are among the largest unregulated markets in the world (Foley et al. 2019). Unlike stocks or bonds, cryptocurrency lacks regular financial reports or analyst assessments. Weber et al. (2023) report that many non-adopters know little about cryptocurrencies. Such disparities introduce complexities in leveraging theories in traditional financial assets to understand cryptocurrency adoption behaviors. The lack of authoritative sources of information (Catalini and Gans 2020) necessitates that investors seek alternative channels for evaluating the potential and risks of these digital assets. Social media, which has become an important medium for information sharing and decision-making among investors (Chen et al. 2014, Wang et al. 2024), fills this void. Twitter, Reddit, and forums specifically for cryptocurrencies (e.g., `Bitcointalk`) have become information centers for conversations, news, and investment guidance—also for cryptocurrencies (Han et al. 2023). Unlike traditional information sources, the nature of information distribution on social media hinges on social networks rather than (professional) expertise or authoritative sources. Consequently, the reliability and accuracy of such information are highly variable.

In contrast to traditional assets, where the value is determined by discounting cash flows in standard valuation models, the value of cryptocurrency is based mainly on transactional demand (Cong et al. 2021, Biais et al. 2023), embedding its social nature within the trading process. At the same time, the cryptocurrency market, due to its regulatory void, is susceptible to price manipulations such as the "pump-and-dump" schemes (Li et al. 2021). Investors may spread misinformation in favor of their investments. Such potentially manipulative behaviors make obtaining credible information via social networks a challenge. Consequently, it is very important to understand how individuals learn and act upon the information obtained from their social networks. When both the information on trading be-

[2] `https://explodingtopics.com/blog/number-of-cryptocurrencies`, last accessed on June 26, 2024.

[3] `https://www.coingecko.com/`, last accessed on June 26, 2024.

havior and opinions are available, do users (investors) weigh trading behavior more heavily, or do they predominantly base their investment decisions on the opinions of others—which could be cheap talk? How do users assimilate information from others when faced with alignment or discrepancy between trading actions (behavior) and expressed opinions? Answers to these questions allow us to gain a deeper understanding of the decision-making dynamics that influence cryptocurrency adoption.

We also explore how product uncertainty moderates the interplay of expressed opinions and actions in social learning. The literature on social learning for various product adoptions (Banerjee et al. 2013, Miller and Mobarak 2015, Gillingham and Bollinger 2021, etc.) typically assumes a fixed expected value for the product; yet, many new products come with uncertainty about their true value. In addition, we investigate heterogeneous effects of user characteristics on social learning. Younger investors, particularly millennials and Generation Z, are typically more receptive to social media and new technologies, which may result in different social learning patterns compared to older investors. Understanding how social learning influences cryptocurrency adoption among younger investors is crucial, as nearly $70 trillion of wealth is expected to transfer from baby boomers to these rising demographics over the next 25 years, making it only a matter of time until they drive the market.[4]

We focus on the early Bitcoin adoption during 2016 because Bitcoin was the first cryptocurrency and still accounts for a significant portion of the total market capitalization of all cryptocurrencies.[5] Early adoption helps to diffuse the product among the masses and influences the success or failure of a new product (technology). The implication of Bitcoin's early adoption can offer valuable guidance for other novel technologies.

We use data from one of the leading social trading platforms and explore how the interplay of opinions and behaviors within social networks influences an individual's decision to adopt Bitcoin. We utilize all social activities of the platform to construct a social network that

[4] `https://apexfintechsolutions.com/news-resources/blog/millennial-and-genz-investing-habits/`, last accessed on June 26, 2024.

[5] In January 2024, Bitcoin (BTC) boasts a market cap of $782 billion, accounting for almost 50% of the cryptocurrency market; see `https://www.coingecko.com/`, last accessed on June 26, 2024.

captures how the information is transmitted. We identify the neighbors' adoption status, posting behavior, and trading behavior for each focal user. We apply a two-stage least squares linear probability model with user and time-fixed effects to quantify the impact of social learning on Bitcoin adoption. Using an instrumental variables approach, we address the endogeneity concern of the effect of neighbors on the focal user's adoption, which might be merely a correlation instead of a causal social learning effect. We further study the moderating effect of product uncertainty on social learning and analyze the heterogeneous effects of user characteristics on social learning.

We find that the likelihood of Bitcoin adoption increases when neighbors who have adopted Bitcoin share positive posts about it. Conversely, positive posts from neighbors who have not adopted Bitcoin decrease the likelihood of adoption. Our result highlights the consistency between expressed opinions and trading actions in influencing a user's decision to adopt Bitcoin. We also observe a significant moderating effect of product uncertainty (measured using the asset's volatility) on social learning. High volatility significantly increases a focal user's likelihood of adopting Bitcoin. Positive posts from adopters increase the probability of adoption, enhancing the overall effect. Additionally, even though positive posts from non-adopters during times of high volatility still decrease the probability of adoption, the negative effect typically associated with non-adopters is mitigated. In addition, we observe some heterogeneous effects of user characteristics on social learning. Notably, we find that high-income and older users are more susceptible to opinions from their neighbors, with an increased positive influence from adopters' positive sentiments and a decreased negative impact from non-adopters' positive sentiments. Surprisingly, younger users exhibit more conservative behaviors interpreting positive sentiments from neighbors, despite typically being associated with a receptiveness to social media and new technologies, and a higher risk tolerance (Grable 2000).

To understand the underlying mechanisms of our main results, we scrutinize the textual differences between the posts of adopters and non-adopters. We find that adopters write longer, more informative posts with a higher likelihood of including external validations in their positive Bitcoin posts. Interestingly, non-adopters show a higher likelihood of including forward-looking statements in their positive Bitcoin posts. The trading duration analysis shows that users who make positive Bitcoin-related posts during their holding period hold Bitcoin positions longer compared to the scenarios involving no post, or posts with non-positive sentiments. Our results indicate that users, on average, post in line with their trading activities. Users' posts closely correlate with their actual trading actions. The consistency between user's expressed opinions and trading actions enhances credibility of social communication on the social trading platform. Finally, we explore two possible alternative explanations (awareness and profitability of neighbors) and rigorously test the robustness of our main results.

Our findings have various practical implications. First, our study highlights the role of social learning in the early adoption of cryptocurrency. Aligning expressed opinions with actual trading behaviors provides an effective strategy for promoting cryptocurrency adoption. Utilizing social influencers who trade cryptocurrency can enhance credibility and trustworthiness of their expressed opinions. Disclosing the adoption status of influencers can further reinforce the reliability of the information and positively impact adoption. One analogy is the common practice in e-commerce of revealing reviewer purchase information to enhance credibility and boost the influence of customer reviews. Amazon, for instance, labels reviews from customers who have purchased the products as “Verified Purchase." Second, our findings indicate that social learning is particularly relevant when product uncertainty is high. Our results on cryptocurrency highlight that during times of high volatility, users are more susceptible to positive posts, irrespective of whether they come from adopters or not. Third, our findings show that high-income and older investors are more susceptible to opinions from their neighbors. Cryptocurrency developers and marketers can utilize this insight to generate interest and potentially boost adoption. Finally, our study also emphasizes the critical role of transparency and public disclosure of trading activities in social trading platforms. Social trading platforms may serve as a natural fit for promoting cryptocurrency due to their social features and high level of transparency. Developers and marketers of cryptocurrencies can leverage social trading platforms to amplify positive sentiment around their digital assets.

The remainder of the paper is structured as follows. We review the relevant literature in Section 2.2. We describe our data, the network construction, and our variables in Section 4.2. Section 4.3 describes model specifications and presents the main results. We explore mechanisms in Section 4.4 and provide several robustness checks in Section 4.5. Section 4.6 concludes the paper.

4.2 Data

Our data is from a large global social trading platform that offers its services in multiple countries. In line with typical brokerage services, the platform allows its users (investors) to trade various financial instruments such as stocks, contracts for differences (CFDs), foreign exchange, commodities, and cryptocurrencies. The platform first introduced Bitcoin trading in 2013 and later added other cryptocurrencies to its trading universe.

A distinctive feature of social trading platforms is that all investors' trading activities are public. Investors on social trading platforms can automatically replicate the trading strategies of other investors. Signal providers, i.e., investors who have their trading activities replicated by other investors can earn extra fees based on their performance, the number of copiers, and/or their assets under management (AUM). As such, social trading platforms offer investment opportunities that are similar to mutual fund investments (Doering et al. 2015). While mutual funds offer a traditional, professionally managed investment option with more regulation, social trading platforms provide a more interactive and transparent way for retail investors to trade (Doering et al. 2015).

In addition to trading opportunities, the social trading platform allows its users to interact with each other in a social environment. The social interactions on the trading platforms are similar to those on typical social media platforms such as Twitter/X. Users can post messages and comment and like posts. Importantly, the platform does not allow for any private communication among users. Consequently, all communication on the platform is public and transparent, accessible to every user. Gemayel and Preda (2018) label this transparent environment a "scopic regime."

4.2.1 Data

We obtained all users' trading activities and social interactions on the platform. We focus on the early adoption period of Bitcoin throughout the year 2016. Thus, our sample period ends before Bitcoin became a mainstream investment opportunity. This mitigates concerns about the impact of confounding information sources. For example, after the first Bitcoin bubble in December 2017 and January 2018, Bitcoin permeated mainstream discussions and became a popular topic in investment discussions. Following this first bubble, investors are more likely to learn about Bitcoin from multiple sources, making it almost impossible to rule out alternative factors that influence adoption decisions. Additionally, focusing on the early adoption period of Bitcoin offers relevant insights for promoting other cryptocurrencies in the early stages, especially considering that thousands of new cryptocurrencies are created each year.[6] Figure 4.1 illustrates the search volume for Bitcoin during our observation period and one year after from Google Trends.[7] The search volume stayed relatively stable throughout 2016, but 2017 saw several significant increases.

For each user, we obtain the following information: 1) the user's demographic information including their gender, age, investment experience, and self-reported risk preferences provided by the user at the time of registration; 2) the user's full trading history, including the open time, close time, the instrument, and the profit of all trades executed on the platform during our sample period; 3) the user's social interactions, including their posts, comments, and likes, together with the corresponding timestamps and contents;[8] 4) the leader-copier relationships that are based on the auto-copy function, which allows users to

---

[6] `https://www.statista.com/statistics/863917/number-crypto-coins-tokens/`, last accessed on June 26, 2024.

[7] We use "Bitcoin" as the search term to retrieve the Google Trends data. The index also considers spelling variations, including "bitcoin" and "BITCOIN".

[8] We translate non-English content to English using the Google translate API. On the platform, users have the convenient option to click "translate" for easy translation into English.

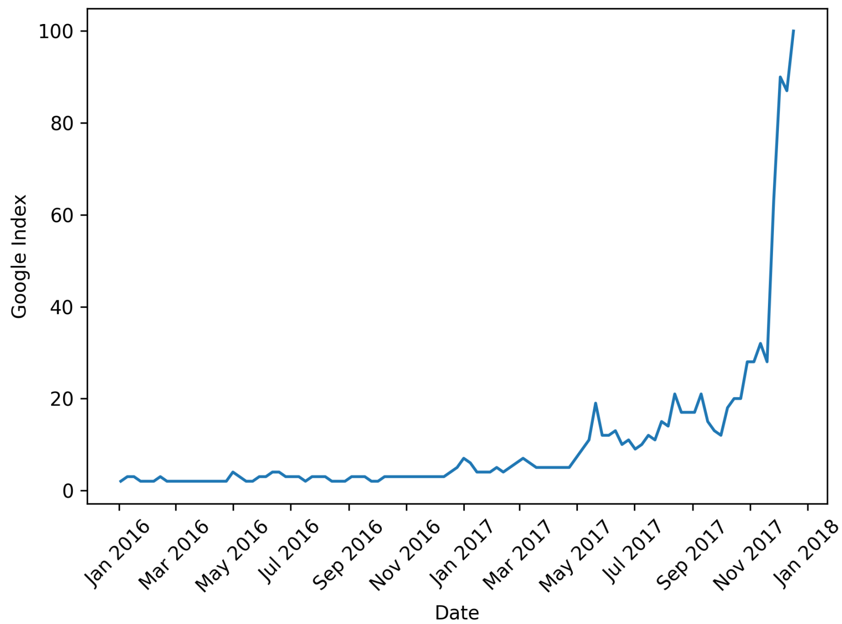


Figure 4.1: Google Trends of Bitcoin from 2016 to 2017

Note: This figure shows the search volume for Bitcoin from Google Trends during 2016 and 2017. The search volume refers to the number of times a specific search query is entered into Google over a given period of time. It is a metric used to gauge the popularity or frequency of the searched query.

automatically duplicate the trading strategies of others.

We define executing a trade (open or close), a social interaction (post, comment, or like), or copying another trader as an activity and a user that executes any activity as an active user. Our data contain all active users (181,351) on the platform during our sample period (from January 2016 to December 2016). We aggregate our data at the weekly level by calculating averages, resulting in a user-week panel data with 2,504,679 observations. In our sample, 24% of users copy trades from at least one other user, and 53% engage in social interactions with others at least once. In our main analysis, we consider all active users. In addition, we conduct a robustness test using a more focused sample consisting of more actively engaged users (see Section 4.5).

4.2.2 Network construction

Given the large number of users and volume of information on the platform, it is not feasible for each user to access and process all available information due to cognitive limitations. Hence, we construct a network in an effort to efficiently capture potential information transmission among users.

We construct our network from the leader-copier relationships and all the users' social interactions, i.e., their posts, comments, and likes. The leader-copier relationship reflects the information transmission in a directed network as copiers receive all activities of the leaders they copy in their news feed. Therefore, the leader is the information sender, while the copier is the information receiver. Since we observe the exact timestamps of formation and dissolution for every leader-copier relationship pair, we can formulate a dynamic network.

Turning to social interactions, we only have access to the timestamps of each interaction (post, comment, or like), rather than data on viewing or browsing activities. It is probable that a user pays attention to a focal user for some periods before engaging with their posts through commenting or liking. Similarly, even if a user does not comment or like subsequent posts of the focal user, it is likely that they continue to pay attention to the focal user after their initial engagement. Therefore, we assume that the network derived from social interactions changes periodically every four weeks. That is, we model a link between two users who interact within a four-week period. We define the user who posts as the (information) sender and the user who comments or likes as the (information) receiver. The directed link runs from a sender to a receiver, representing the direction of information transmission in the network.

We denote the network inferred from the leader-copier relationship in period t as $CopyNet_t$ and the social interactions in period t as $SocNet_t$. We pool these two networks into a combined network, denoted as Net_t. We report summary statistics of the networks in Table B.1 of Appendix B.1. We use Net_t to calculate neighbor-related variables in Section 4.2.3.

4.2.3 Variables

Our dependent variable, $Adopt_{i,t}$, is a dummy variable that takes a value of 1 if user i trades (buy or sell) or holds Bitcoin at period t, and 0 otherwise.[9] We conduct a robustness test (see Section 4.5) that considers only the first adoption of Bitcoin. We use three groups of explanatory variables to investigate the determinants of the adoption decision.

First, we include various neighbor-related variables derived from the network and users' attributes to capture the social learning process. We denote the network as $N_{i,j,t}$, where $N_{i,j,t} = 1$ if user j (information receiver) receives information sent by i (information sender) at period t. Note that the network is directed; hence, $N_{i,j,t} = 1$ does not necessarily imply $N_{j,i,t} = 1$. We examine three channels of social learning: neighbors' adoption status, neighbors' opinions, and neighbors' trading performance. To measure the neighbors' adoption status, we calculate the ratio of Bitcoin adopters among the focal user i's neighbors at period t ($NeighborAdopt_{i,t}$). The total number of neighbors for the focal user i at period t is denoted by $Neighbor\_N_{i,t}$. We next construct the variables capturing the neighbors' opinions, inferred from their posts. We compute the average number of posts made by user i's neighbors at period t ($Post\_N_{i,t}$). Since we aim to capture the interplay between

[9]To emphasize social learning, we consider only users who adopt Bitcoin through self-executed trading, excluding those who trade Bitcoin through auto-copy.

opinions and actions (whether adopt or not), we calculate the percentage of neighbors who post positively about Bitcoin among all adopters ($AdoptBTCPositive_{i,t}$) and non-adopters ($NonAdoptBTCPositive_{i,t}$) in the neighborhood, respectively. We utilize lexicon-based sentiment analysis (Loughran and McDonald 2011) to extract the sentiment of each post (see also Deng et al. 2024). Specifically, a post is labelled as positive if the number of positive words is greater than the total number of negative and uncertain words. Similar to posting-related variables, to capture the neighbors' trading performance, we construct the average profit of neighbors at period t for adopters ($AdoptReturn_{i,t}$) and non-adopters ($NonAdoptReturn_{i,t}$) respectively.

Second, we proxy for individuals' trading behavior. We calculate the average profit of user i by period t ($AccReturn_{i,t}$) and the standard deviations of user i's profit by period t ($AccSD_{i,t}$) to measure the return and risk of a user's trading performance. We count the number of weeks user i has been active on the platform by period t since January 1, 2016 ($Tenure_{i,t}$), and the number of trades user i has executed on the platform by period t ($AccTrade_{i,t}$). We also measure temporal trading intensity using the number of trades a user executes in period t ($NTrade_{i,t}$). We include the total number of unique instruments of user i by period t ($Diversification_{i,t}$) as a measure for trading diversification.

Third, we take a look at individual-level demographic characteristics. $Male_i$ takes a value of 1 if user i identifies as male, and 0 otherwise; $Income_i$ denotes the annual income of user i (categorical variable); $AgeRange_i$ denotes the age range of user i (categorical variable); and $Experience_i$ denotes the trading experience (categorical variable) of user i, before joining the platform. $RiskLevel_i$ denotes user i's stated risk preferences (categorical variable). These variables are reported by the user at the time of registration.

Table B.2 in Appendix B.2 summarizes the definitions of our variables. We provide summary statistics for numerical variables in Table 4.1 and for categorical variables in Table B.3 of Appendix B.2. We log-transform $Neighbor\_N_{i,t}$, $Tenure_{i,t}$, $AccTrade_{i,t}$, $NTrade_{i,t}$, and $Diversification_{i,t}$ to correct the skewness in our regression analyses. Table B.4 and Table B.5 in Appendix B.3 show the correlation matrix and the variance inflation factor

(VIF) scores of variables used in our main model, respectively.

Table 4.1: Summary statistics of numerical variables

Variable	Mean	Std. Dev.	Min	P25	Median	P75	Max
$\text{Adopt}_{i,t}$	0.098	0.294	0	0	0	0	1
$\text{NeighborAdopt}_{i,t}$	0.035	0.154	0	0	0	0	1
$\text{Neighbor\_N}_{i,t}$	1.255	4.967	0	0	0	1	604
$\text{Post\_N}_{i,t}$	6.050	52.923	0	0	0	1	3,321
$\text{AdoptBTCPostive}_{i,t}$	0.004	0.059	0	0	0	0	1
$\text{NonAdoptBTCPostive}_{i,t}$	0.003	0.043	0	0	0	0	1
$\text{AdoptReturn}_{i,t}$	-0.011	2.393	0	0	0	0	1
$\text{NonAdoptReturn}_{i,t}$	0.791	8.156	0	0	0	0	1
$\text{AccReturn}_{i,t}$	-1.585	17.421	-3,668.986	-3.513	0.000	1.2043	2,940
$\text{AccSD}_{i,t}$	11.931	17.767	0	0.450	7.029	16.989	1,848.794
$\text{Tenure}_{i,t}$	17.007	13.815	1	1	14	27	52
$\text{AccTrade}_{i,t}$	193.271	540.588	0	4	42	169	5 3,587
$\text{NTrade}_{i,t}$	19.829	62.4382	0	0	2	15	6,870
$\text{Diversification}_{i,t}$	5.609	3.740	0	0.276	1.273	3.438	148

Note. The table shows summary statistics of all numerical variables. P25 represents the 25th percentile, and P75 represents the 75th percentile. Variable definitions can be found in Table B.2 of Appendix B.2.

4.3 Model and Results

4.3.1 Main model and identification

We begin by studying the impact of various factors on an individual's Bitcoin adoption. We are particularly interested in the impact of neighbors' Bitcoin adoption and neighbors' opinions on the focal user's Bitcoin adoption. We aim to estimate Equation (4.1).

$$\begin{aligned} Adopt_{i,t} =& \beta_1 NeighborAdopt_{i,t-1} + \beta_2 AdoptBTCPositive_{i,t-1} \\ &+ \beta_3 NonAdoptBTCPositive_{i,t-1} + Control_{i,t-1} + \gamma_i + \delta_t + \epsilon_{i,t} \end{aligned} \tag{4.1}$$

The main challenge is to address the endogeneity concern of the effect of neighbors on the focal user's adoption, which might be merely a correlation instead of a causal social learning effect. Specifically, endogeneity could arise from endogenous network formation, correlated unobservables, or simultaneity (Moffitt 2001, Bramoullé et al. 2016).

First, endogenous network formation means that a user may choose to form links with others (neighbors) when they have similar trading preferences. For example, a user may prefer a high-risk trading strategy, thus choosing a neighbor employing a high-risk trading strategy. In this case, the observed correlation could arise from omitted variables that represent the individual's hidden preferences. We include the user-level fixed effect (γ_i) to account for any user-level unobserved characteristics that affect the formation of the network.[10] It also captures other inherent preferences toward adopting Bitcoin. As a result of including the user-level fixed effects, we cannot estimate coefficients for time-invariant covariates such as age, trading experience, risk level, or income in Equation (4.1).

Unobserved exogenous events that influence the adoption decision of the focal user and others at the same time pose a second concern that may explain correlations in Equation (4.1) beyond social learning. For example, an increase in the Bitcoin price may lead to a higher likelihood of adoption. Bitcoin may also become more attractive when it receives more public attention. Such factors change over time and we control them using weekly-level fixed effects (δ_t). We further alleviate this concern using an instrumental variables approach, following the practice in literature (Patacchini and Zenou 2008, Bifulco et al. 2011). We use neighbors' characteristics, that is, gender, age, income, risk level, and trading experience as instrument variables for the neighbors' adoption status and posts ($NeighborAdopt_{i,t-1}$,

[10] We acknowledge the user's preference on network formation is assumed to be fixed over time, following the practice in literature (Nair et al. 2010, Ameri et al. 2019).

$AdoptBTCPositive_{i,t-1}$, and $NonAdoptBTCPositive_{i,t-1}$).[11] The rationale behind our instrumental variables is that the neighbors' characteristics do not directly influence the focal user's adoption decision, as these characteristics are collected by the platform but are not explicitly displayed to the public, thereby satisfying the exclusion condition. Furthermore, given our dynamic networks, even though these neighbors' characteristics are time-invariant, the instrumental variables capture the time-variant changes of neighbors, providing sufficient variations.

Third, the simultaneity means that the focal user's adoption and the neighbors' behaviors may happen concurrently. To mitigate this concern, in our regression model, the dependent variable is measured at time t, and the independent variables are measured at time $t-1$.

Past trading activities can influence an investor's future trading decisions. For example, biased self-attribution of past returns can increase investors' overconfidence and lead to higher trading intensity (Odean 1998, Barber and Odean 2000). This may also influence the investor's perceived knowledge (Merkle 2017) on Bitcoin and influence the adoption decision. Therefore, we control user i's trading performance ($AccReturn$ and $AccSD$), proxies for user i's trading strategy ($AccTrade$, $Diversification$, and $NTrade$), and the length of time user i has been trading on the platform ($Tenure$). We also control the total number of user i's neighbors ($Neighbor\_N$) and the average number of posts made by user i's neighbors ($Post\_N$) in Equation (4.1).

In panel data, the coefficient estimates using traditional probit or logit with fixed-effects may be biased due to the incidental parameter problem (Neyman and Scott 1948). Consequently, we apply a linear probability model (LPM) with individual-level (γ_i) and week-level (δ_t) fixed effects. Since we have a large number of individuals and periods, the linear probability model with fixed effects should yield unbiased estimates (Horrace and Oaxaca 2006).

[11] To aggregate the characteristics of neighbors, we take the average values of the characteristics. For categorical characteristics, we use the fraction of male neighbors to aggregate for gender. For age, trading experience, risk level, and income, we first convert each categorical value into a numerical number and then take the average among the neighbors. For instance, for the categorical age range 18-24, we approximate and convert it into a numerical value of 21.

4.3.2 Main results

We estimate the model using a two-stage least squares (2SLS) regression approach. In the first stage, we regress neighbor-related variables on instrumental variables representing neighbors' characteristics. In the second stage, we regress the user's Bitcoin adoption decision on the fitted values obtained from the first stage to identify the social learning effects from other confounding factors. We summarize the results in Table 4.2. Columns (1) - (3) show the results of first-stage regressions and Column (4) shows the result of the second-stage regression. We conduct various tests for underidentification and weak identification of the endogenous regressors, with the statistics reported in Table 4.2. All statistical tests support the validity of our instrumental variables.[12]

Table 4.2 shows a positive coefficient on $NeighborAdopt$, indicating that a user with more neighbors who have adopted Bitcoin is more likely to adopt Bitcoin. We also find that adopters who post positively about Bitcoin significantly increase the likelihood of adoption of the focal user (indicated by the positive coefficient of $AdoptBTCPositive$), while non-adopters who post positively about Bitcoin significantly decrease the likelihood of adoption of the focal user (indicated by the negative coefficient of $NonAdoptBTCPositive$). This indicates that users who consider the opinions of others (neighbors) also pay attention to their actual trading actions. Users prefer to learn from those whose opinions and actions are consistent. When actions and voiced opinions are inconsistent, this actually has a detrimental effect on the adoption decision.

The coefficient of $Neighbor\_N$ is significantly negative, indicating that individuals with a larger number of neighbors are less likely to adopt Bitcoin. One possible explanation is that exposure to a variety of information sources might lead to encountering conflicting opinions

[12] We report the Sanderson-Windmeijer (SW) first-stage F statistics and Chi-squared statistics (Sanderson and Windmeijer 2016) for each endogeneous regressor. These statistics provide evidence against underidentification and weak identification of these endogenous regressors, respectively. We also conduct additional tests for the joint identifiability of these endogenous regressors. The weak identification test, Cragg–Donald Wald F statistics (Cragg and Donald 1993), is 131.73; the Kleibergen-Paap Wald rk F statistic (Kleibergen and Paap 2006) is 32.54; the underidentification test, Kleibergen–Paap rk LM statistic (Kleibergen and Paap 2006), is 161.12.

Table 4.2: Main model estimation results

	First Stage			Second Stage
	NeighborAdopt	AdoptBTCPositive	NonAdoptBTCPositive	Adopt
	(1)	(2)	(3)	(4)
$\widehat{NeighborAdopt}$				0.237***
				(0.040)
$\widehat{AdoptBTCPositive}$				4.698***
				(0.538)
$\widehat{NonAdoptBTCPositive}$				-2.688***
				(0.347)
Neighbor_N	0.011***	0.013***	-0.001***	-0.044***
	(0.001)	(0.000)	(0.000)	(0.007)
Post_N	-0.00003***	0.000001***	0.00001***	0.00003***
	(0.000)	(0.000)	(0.000)	(0.000)
AccTrade	-0.002***	-0.001***	-0.002***	-0.014***
	(0.000)	(0.000)	(0.000)	(0.001)
Tenure	-0.002	0.001***	0.001***	0.022***
	(0.001)	(0.000)	(0.000)	(0.002)
NTrade	-0.001***	0.0001***	0.001***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Diversification	0.003**	0.001**	0.004***	0.138***
	(0.001)	(0.000)	(0.000)	(0.004)
AccReturn	0.00003***	0.00001***	-0.00001	-0.0002***
	(0.000)	(0.000)	(0.000)	(0.000)
AccSD	0.00006***	0.00002***	-0.00005***	-0.0004***
	(0.000)	(0.000)	(0.000)	(0.000)
NeighborGender	-0.049***	-0.004***	-0.007***	
	(0.003)	(0.001)	(0.000)	
NeighborAge	-0.001***	-0.0003***	-0.001***	
	(0.000)	(0.000)	(0.000)	
NeighborRisk	-0.00000967	0.00004**	0.0001***	
	(0.000)	(0.000)	(0.000)	
NeighborExperience	-0.020***	0.002***	0.008***	
	(0.002)	(0.000)	(0.000)	
NeighborIncome	0.016***	0.000***	0.001***	
	(0.001)	(0.000)	(0.000)	
Constant	0.004**	0.000	-0.000	
	(0.001)	(0.000)	(0.000)	
User-level FE	Yes	Yes	Yes	Yes
Week-level FE	Yes	Yes	Yes	Yes
Observations	2,316,559	2,316,559	2,316,559	2,316,559
SW F Statistics	88.81***	58.70***	57.04***	-

Note: Standard errors clustered at the user level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

about Bitcoin trading, thereby decreasing the likelihood of adoption.[13] We also find that a user with more frequent trading, longer trading experience on the platform, a preference for a diversified investment portfolios, and a history of poorer trading performance is more likely to adopt Bitcoin.

4.3.3 The moderating effect of product uncertainty

Cryptocurrency adoption differs from other products primarily due to its high product uncertainty. Since cryptocurrency has limited intrinsic value, its value is difficult to identify and based on investors' own valuation. The moderating effect of such high product uncertainty on social learning remains under-explored in the literature. We address this gap by analyzing how uncertainty about the value of cryptocurrency moderates social learning.

We quantify the product uncertainty using the volatility of Bitcoin prices. We define the volatility as the standard deviation of daily log-returns of Bitcoin over the past 90 days (following Kumar 2009, etc.). We create a dummy variable $Volatility\_High_t$ that takes a value of 1 if the observed volatility in period t is above the median volatility in our data sample, and 0 otherwise. We formulate the model with the moderator in Equation (4.2).

$$\begin{aligned} Adopt_{i,t} =& \beta_1 NeighborAdopt_{i,t-1} \\ & + \beta_2 AdoptBTCPositive_{i,t-1} + \beta_2' Volatility\_High_{t-1} \times AdoptBTCPositive_{i,t-1} \\ & + \beta_3 NonAdoptBTCPositive_{i,t-1} + \beta_3' Volatility\_High_{t-1} \times NonAdoptBTCPositive_{i,t-1} \\ & + Control_{i,t-1} + \gamma_i + \delta_t + \epsilon_{i,t} \end{aligned} \tag{4.2}$$

We present our estimation results in Column (1) of Table 4.4. Again, our major estimates are consistent with the results from the main model reported in Column (4) in Table 4.2. Notably, we observe a significant moderating role related to volatility. High volatility significantly increases a focal user's adoption of Bitcoin. Positive posts of adopters increase the probability to adopt, thereby increasing the overall effect. Even though positive posts of

[13] We perform additional analysis and validate this argument in Appendix B.4.

non-adopters in times of high volatility still decrease the probability of adopting, the negative effect of non-adopters is mitigated.

Prior research has shown a substantial presence of speculative investors in the Bitcoin market (Wei and Dukes 2021). These investors thrive on volatility, aiming to capitalize on rapid price changes and make quick profits. When volatility is high, speculative opportunities tend to increase. Our results suggest that when the volatility is high, users are more likely to be influenced by positive Bitcoin sentiments from credible neighbors whose actions and opinions are consistent, reinforcing their speculative belief in adopting Bitcoin. As a result, they tend to adopt Bitcoin during periods of higher volatility. Even the impact of less credible neighbors, whose opinions and actions are inconsistent, diminishes, as the focal user's strong desire to speculate and adopt Bitcoin becomes dominant.

4.3.4 Heterogeneous effects on user characteristics

Social learning may differ across users depending on their demographic characteristics and socio-economic backgrounds. To understand how social learning varies across different segments of users, we analyze heterogeneous effects based on various demographic characteristics. Specifically, we focus on the user's age, income, self-reported risk preferences (levels), and trading experience.[14] We create a set of dummy variables that split our sample at the median, with respect to each characteristic. Table 4.3 summarizes the construction of our dummy variables. We formulate our heterogeneous effect model in Equation (4.3).

$$\begin{aligned} Adopt_{i,t} = &\beta_1 NeighborAdopt_{i,t-1} + \beta_2 AdoptBTCPositive_{i,t-1} + \beta_2' Charac_i \times AdoptBTCPositive_{i,t-1} \\ &+ \beta_3 NonAdoptBTCPositive_{i,t-1} + \beta_3' Charac_i \times NonAdoptBTCPositive_{i,t-1} \\ &+ Control_{i,t-1} + \gamma_i + \delta_t + \epsilon_{i,t} \end{aligned} \tag{4.3}$$

where $Charac_i$ represents each characteristic of user i defined in Table 4.3.

[14] We do not study heterogeneous effects of gender, because our data is highly skewed toward male investors and has 9% of female investors.

Table 4.3: Binary categorization of demographic characteristics

Variable	**Definition**
Age_Young	1, if a user's age is less than 45 years old, and 0 otherwise
Income_High	1, if a user's annual income is greater than 50,000 U.S. dollars, and 0 otherwise.
Risk_High	1, if a user's risk preference is above medium, and 0 otherwise.
Experience_Long	1, if a user's trading experience is more than 1 year, and 0 otherwise.

We present our estimation results in Table 4.4. Columns (1) - (4) show the estimation results from heterogeneous effects on the user's age, income, risk preference, and trading experience. Our main estimates are consistent with the results from the main model reported in Column (4) in Table 4.2. Interestingly, the learning effect from adopters is more pronounced among older and high-income users. The interaction with *Age_Young* (see Column (2) from Table 4.4, with a coefficient of -2.330) is significantly negative and about 1/3 of the magnitude of the main effect. Conversely, the (negative) learning from non-adopters is more substantial among younger and low-income users. Here, the interaction with *Age_Young* (Column (2) from Table 4.4, with a coefficient of -1.112) is also significantly negative and about 2/3 of the magnitude of the main effect. We do not find any significant heterogeneous effects concerning a user's risk preference or trading experience.

Our results indicate that high-income and older users are more susceptible to opinions from their neighbors, with an increased positive influence from adopters' positive sentiments and a decreased negative impact from non-adopters' positive sentiments. Surprisingly, younger users exhibit more conservative behaviors interpreting positive sentiments from neighbors, despite typically being associated with a higher risk tolerance (Grable 2000). This can be attributed to a higher peer influence observed in older age groups (Steinberg and Monahan 2007) and is also consistent with the observation of Cheng et al. (2013) who show that older investors trade with lower discipline.

Table 4.4: Results of moderating and heterogeneous effects

	Volatility	Age	Income	Risk	Experience
	(1)	(2)	(3)	(4)	(5)
NeighborAdopt	0.262***	0.211***	0.242***	0.237***	0.237***
	(0.041)	(0.0377)	(0.0405)	(0.0389)	(0.0398)
AdoptBTCPositive	4.268***	6.071***	4.197***	4.394***	4.270***
	(0.572)	(0.894)	(0.530)	(0.590)	(0.569)
$\times$ *Volatility_High*	0.888***				
	(0.306)				
$\times$ *Age_Young*		-2.330***			
		(0.778)			
$\times$ *Income_High*			1.645***		
			(0.633)		
$\times$ *Risk_High*				0.241	
				(0.491)	
$\times$ *Experience_Long*					0.714
					(0.490)
NonAdoptBTCPositive	-5.321***	-1.646***	-2.952***	-3.013***	-2.529***
	(0.802)	(0.505)	(0.420)	(0.538)	(0.418)
$\times$ *Volatility_High*	3.111***				
	(0.632)				
$\times$ *Age_Young*		-1.112*			
		(0.621)			
$\times$ *Income_High*			1.152*		
			(0.612)		
$\times$ *Risk_High*				0.748	
				(0.617)	
$\times$ *Experience_Long*					-0.246
					(0.606)
Neighbor_N	-0.045***	-0.037***	-0.043***	-0.042***	-0.044***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Post_N	0.0001***	0.00003***	0.00003***	0.00003***	0.00003***
	(0.00001)	(0.000)	(0.000)	(0.000)	(0.000)
AccTrade	-0.012***	-0.013***	-0.013***	-0.014***	-0.014***
	(0.001)	(0.00139)	(0.00145)	(0.00144)	(0.00146)
Tenure	0.022***	0.021***	0.022***	0.022***	0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
NTrade	0.003***	0.004***	0.004***	0.004***	0.004***
	(0.0004)	(0.000)	(0.000)	(0.000)	(0.000)
Diversification	0.134***	0.135***	0.137***	0.137***	0.138***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
AccReturn	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.00004)	(0.000)	(0.000)	(0.000)	(0.000)
AccSD	-0.0004***	-0.0003***	-0.0004***	-0.0004***	-0.0004***
	(0.0001)	(0.000)	(0.000)	(0.000)	(0.000)
User-level FE	Yes	Yes	Yes	Yes	Yes
Week-level FE	Yes	Yes	Yes	Yes	Yes
Observations	2,316,559	2,316,559	2,316,559	2,316,559	2,316,559

Note. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is $Adopt_{i,t}$. The definitions of demographic variables are in Table 4.3.

4.4 Mechanisms

In this section, we aim to uncover the mechanisms that may explain our main finding. First, we delve into textual analysis and trading duration analysis to gain deeper insights into the users' opinions. Additionally, we explore two possible alternative explanations for our main findings: awareness and the profitability of neighbors.

4.4.1 Textual analysis

We find that positive Bitcoin posts from neighboring adopters prompt the focal user to adopt Bitcoin, while positive Bitcoin posts from neighboring non-adopters deter the focal user from adopting Bitcoin. We further explore potential textual differences between the posts of adopters and non-adopters. Specifically, we focus on the following aspects that could influence the effectiveness of post content and are potentially associated with users' learning:

- The length of posts. Longer sentences may boost the learning effect because they usually offer a more thorough and convincing argument. After removing the stop words and punctuations, we count the number of words for each post. We analyze the length of posts to examine whether a longer post is associated with a greater learning effect.

- External validation. By citing external resources, people can bolster their arguments. Users may learn more from posts that include links to external sources because these provide additional credibility. We observe that users' posts frequently include URLs directing to external resources like policy guides or news articles (e.g., from coindesk.com). Thus, we examine whether incorporating these external resources is associated with a larger learning effect.

- The informativeness of posts. To measure the informativeness of each post, we apply the Flesch-Kincaid readability tests (Kincaid et al. 1975).[15] A higher score presents higher informativeness. We explore whether the level of informativeness in posts is associated with the learning effect.

- Forward-looking statements. For listed companies, managers often use forward-looking statements to inform market participants about future prospects, though there is ongoing debate (e.g., Hassanein and Hussainey 2015, Wang and Hussainey 2013) regarding how they impact firm valuation. It is unknown how investors perceive forward-looking statements from others, particularly for cryptocurrencies where there is a lack of regulation and limited fundamental values. To examine whether a post with forward-looking statements is associated with a greater/lower influence, we utilize a FinBERT-based (Araci 2019) model to label whether a post contains forward-looking statements.[16]

Table 4.5 summarizes the textual measures based on different categorizations of posts. Panel (a) in Table 4.5 shows that Bitcoin-related posts are longer, are more likely to include external URLs, contain more forward-looking statements, and are less informative, compared to non-Bitcoin posts. From Panel (b) and (c) in Table 4.5, both adopters and non-adopters write longer posts, are more likely to include external URLs, and exhibit more forward-looking tendencies in Bitcoin-related posts than non-Bitcoin-related posts. A notable difference is that adopters' Bitcoin-related posts are more informative than their non-Bitcoin-related posts, while the opposite is true for non-adopters. From Panel (c) and (d) in Table 4.5, adopters write longer, more informative posts with a higher likelihood of including external validations in their positive Bitcoin posts, compared to their negative or

[15]We use the Python package `textstat` to calculate the score of each post. The score calculated by the package ranges from 0 to 100, and a higher score means lower informativeness. We derive our informativeness score by subtracting the score calculated from the package from 100. To provide a benchmark for the informativeness score, we report the relevance scores of some major media outlets. According to Flesch (1979), the New York Daily News scores 40, The Times scores 48, the Wall Street Journal scores 57, the New York Times scores 61, and the Harvard Law Review scores 68.

[16]The model is fine-tuned to detect forward-looking statements on 3,500 manually annotated sentences from Management Discussion and Analysis section of annual reports of Russell 3000 firms.

Table 4.5: Summary statistics of textual measures

	(a) All Posts		(b) All Posts	
	Bitcoin	Non-Bitcoin	Adopter	Non-Adopter
Avg. length of posts	52.48	31.74	26.28	36.12
% Posts w/ URL	22.54%	14.20%	13.03%	15.36%
Avg. informativeness	18.69	19.40	15.99	21.37
% Posts w/ forward-looking statements	9.34%	5.96%	7.95%	4.98%
	(c) Bitcoin Posts		**(d) Positive Bitcoin Posts**	
	Adopter	Non-Adopter	Adopter	Non-Adopter
Avg. length of posts	48.08	79.20	**95.81**	75.26
% Posts w/ URL	20.99%	31.93%	**39.38%**	17.15%
Avg. informativeness	18.35	20.76	**29.74**	25.86
% Posts w/ forward-looking statements	9.25%	9.85%	9.35%	**13.19%**

Note. (a) All posts related to Bitcoin versus non-Bitcoin; (b) All posts from adopters versus non-adopters; (c) Bitcoin-related posts from adopters versus non-adopters; (d) Positive Bitcoin-related posts from adopters versus non-adopters.

neutral Bitcoin posts. Conversely, non-adopters write shorter posts with a lower likelihood of including external validations in their positive posts, compared to their negative or neutral Bitcoin posts. Additionally, we observe that non-adopters have a higher likelihood of including forward-looking statements in positive Bitcoin posts, compared to their neutral or negative Bitcoin-related posts, while adopters have a similar likelihood between positive and neutral or negative Bitcoin-related posts.

To summarize, we first observe that users put more effort into their posts (e.g., writing longer posts, including more external URLs, or providing more information) when the sentiment of the Bitcoin-related posts aligns with their adoption status (i.e., adopters write

Bitcoin-related posts with positive sentiments, while non-adopters write Bitcoin-related posts with non-positive sentiments). This suggests that users whose opinions and actions align tend to put extra effort into their posts to support their actions. Our main results suggest that only when users' opinions align with their actions, the likelihood of focal user's adoption increases. Hence, these additional efforts by users, whose opinions and actions align, are associated with greater learning opportunities for others. Second, we observe that Bitcoin-related posts are more likely to contain forward-looking statements. Unlike financial reports where executives use forward-looking statements for future forecasts, the forward-looking statements in Bitcoin-related posts on social trading platforms lack credible information sources due to the absence of central regulation and a fundamental value associated with Bitcoin. More importantly, non-adopters tend to include more forward-looking statements in their positive Bitcoin posts. Our main results suggest that positive sentiment from neighboring non-adopters decreases the likelihood of adoption. Therefore, the higher prevalence of forward-looking statements in neighboring non-adopters' positive Bitcoin posts is associated with a decreased likelihood of focal users adopting Bitcoin.

We conduct a more integrated analysis to further test these conjectures. We estimate a linear regression model with user-level and week-level fixed effects to control for unobserved individual characteristics and time-specific influences. The dependent variables are the length of post, whether the post contains a URL, the informativeness of the post, and whether the post contains forward-looking statements, respectively. The explanatory variables are the user's adoption status ($Adopt$, takes a value of 1 when the user adopts Bitcoin, and 0 otherwise.), the relevance to Bitcoin (IS_BTC, takes a value of 1 when the post is relevant to Bitcoin, and 0 otherwise), and post sentiment ($Positive$, takes a value of 1 when the post has a positive sentiment, and 0 otherwise). To obtain richer insights, we also incorporate the interaction terms. We summarize estimation results in Table 4.6.

Table 4.6 indicates that Bitcoin posts are characterized by forward-looking, higher informativeness, greater length, and a higher likelihood of including URLs. This observation supports the notion that social media-type information is particularly important for cryp-

Table 4.6: Results of textual analysis

Dependent Variable	Length	URL	Informativeness	Forward-looking
	(1)	(2)	(3)	(4)
Adopt	0.437	0.124***	2.831***	0.046***
	(0.298)	(0.003)	(0.321)	(0.002)
IS_BTC	39.790***	0.219***	1.833***	0.022***
	(2.001)	(0.012)	(0.630)	(0.008)
Positive	9.294***	0.004**	4.856***	0.003**
	(0.249)	(0.002)	(0.154)	(0.002)
IS_BTC × Adopt	-34.240***	-0.208***	-1.911**	-0.015*
	(2.066)	(0.013)	(0.769)	(0.009)
IS_BTC × Positive	-8.447**	-0.192***	-0.251	0.016
	(4.011)	(0.018)	(1.164)	(0.019)
Adopt × Positive	2.375***	-0.040***	0.268	0.005*
	(0.401)	(0.003)	(0.279)	(0.003)
IS_BTC × Adopt × Positive	37.330***	0.286***	1.610	-0.053**
	(4.259)	(0.020)	(1.325)	(0.021)
Intercept	29.480***	0.100***	17.100***	0.043***
	(0.137)	(0.001)	(0.123)	(0.001)
User-level FE	Yes	Yes	Yes	Yes
Week-level FE	Yes	Yes	Yes	Yes
Observations	411,144	411,144	411,144	411,144

Note. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

tocurrencies. We also find that posts from adopters are typically more likely to include URLs and forward-looking statements, and are more informative compared to posts from non-adopters. Adopters who share negative sentiments related to Bitcoin tend to deviate from their typical posting pattern of less forward-looking, higher informativeness and inclusion of URLs, as indicated by the coefficients of $Positive$, $Positive \times Adopt$, $Positive \times IS\_BTC$ and $IS\_BTC \times Adopt \times Positive$ jointly. These posts are more forward-looking, less informative, shorter, and less likely to include URLs compared to neutral or negative Bitcoin-related posts. In contrast, adopters who express positive sentiments about Bitcoin write longer posts and are more likely to include URLs. Thus, adopters who express positive opinions about Bitcoin use less forward-looking statements and put more effort into their posts. Similarly, non-adopters who share positive posts about Bitcoin are more forward-looking and less likely to include URLs compared to their negative posts as indicated by the coefficients of $Positive$ and $Positive \times IS\_BTC$ jointly.

4.4.2 Trading duration analysis

Although social media is not always reliable and people may express strategic opinions on investments,[17] our textual analysis suggests that users on social trading platforms exert more effort when they share opinions that are consistent with their Bitcoin trading actions. We provide additional evidence that users indeed share *credible* opinions related to Bitcoin by examining the consistency between the opinions expressed in users' posts and their corresponding trading actions.

To this end, we examine the relationship between the duration of each trade and the user's expressed sentiment in their posts. The duration (holding time) of each trade is quantified by the time interval in hours between the opening and closing time of the trade. The trade duration reflects the user's actual trading actions of Bitcoin. Our hypothesis is that if users truly express a positive sentiment towards Bitcoin, they are likely to maintain their Bitcoin

[17] `https://www.flyingvgroup.com/socialmediastrategy/social-media-financial-services/`, last accessed on June 26, 2024.

positions for a longer duration.

We analyze all Bitcoin trades, categorizing them into three distinct groups based on the sentiment and timing of posts.

1. Trades during which the focal user posts positive sentiments about Bitcoin between the opening and closing times of the trade.

2. Trades during which the focal user posts negative or neutral sentiments about Bitcoin between the opening and closing times of the trade.

3. Trades during which the focal user does not make any post about Bitcoin between the opening and closing times of the trade.

We plot the cumulative distribution of trade duration in each group in Figure 4.2.[18] Figure 4.2 suggests a tendency for users to retain Bitcoin for a longer duration when they share positive sentiments about Bitcoin. The trade duration of those from Group I clearly dominates the cumulative probability distribution compared to the other groups. These findings provide further evidence that users, on average, put their money where their mouth is.

Again, we provide a more integrated analysis of potentially manipulative trading behavior among traders. We estimate a linear regression model on the holding time of each Bitcoin trade with user-level and week-level fixed effects (the week when the trade was opened). Explanatory variables are whether the user makes a Bitcoin-related post during the holding time of the trade ($NoPost$, takes a value of 1 when the user does not make a Bitcoin-related post during the holding interval, and 0 otherwise), the sentiment of Bitcoin-related posts ($Positive$, takes a value of 1 when the post has a positive sentiment, and 0 otherwise), and the profit of the trade ($Profit$). The binary variables $NoPost$ and $Positive$ capture whether the user posts during the holding interval, and in the case of posting, the sentiment expressed in the post. We summarize the estimation results in Table 4.7.

[18]We take the logarithm of the trade duration due to its high skewness.

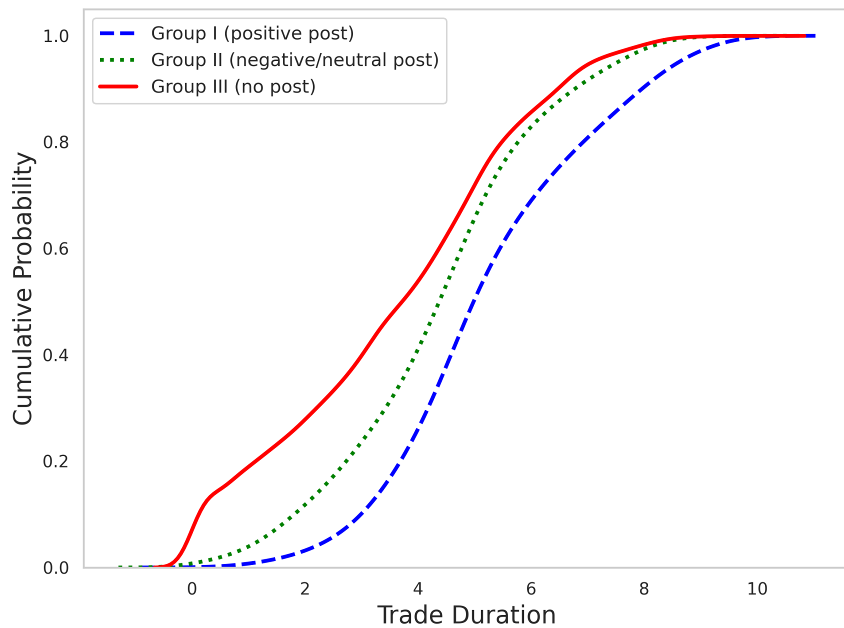


Figure 4.2: Cumulative distribution of trade duration by different groups

Table 4.7: Results of trading duration analysis

Dependent Variable	Holding time
	(1)
NoPost	-1.453***
	(0.056)
Positive	0.556***
	(0.096)
Profit	0.001***
	(0.000)
Intercept	4.952***
	(0.056)
User-level FE	Yes
Week-level FE	Yes
Observations	105,417

Note. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.7 indicates that users who make positive Bitcoin-related posts during the holding interval also hold Bitcoin positions longer compared to the other two scenarios. Interestingly, investors who do not post about Bitcoin hold positions, on average, shorter than investors who post neutrally or negatively about Bitcoin. We also find that higher profits correlate with longer holding durations of trades. Our results indicate that users on the platform, on average, post in line with their trading activities. The consistency between user's expressed opinions and trading actions enhances the credibility of social communication on the social trading platform.

4.4.3 Awareness

Rather than obtaining informative insights from their neighbors, users may simply become *aware* of Bitcoin. Thus, it may be the case that social learning is restricted to knowing about the digital assets. Users might initially be unaware of Bitcoin, and their adoption decision could be driven by their first awareness of Bitcoin. To disentangle whether users simply become aware of Bitcoin or whether they may obtain informative insights from their neighbors, we estimate a variation of our main model using a sub-sample composed of users who were already aware of Bitcoin prior. Specifically, we use those users with prior Bitcoin trading or posting activity, as these clearly are already aware of Bitcoin. The sub-sample accounts for approximately 18% of the full data sample. As users were already aware of Bitcoin prior to the adoption decision, the decision to adopt can no longer be explained by the awareness alone.

We summarize the estimation results in Column (1) of Table 4.8. We observe a significantly positive coefficient on $NeighborAdopt$, highlighting the importance of social learning in the adoption of Bitcoin trading that goes beyond awareness. More importantly, we find that adopters who post positively about Bitcoin significantly increase the likelihood of adoption of the focal user, while non-adopter who post positively about Bitcoin significantly decrease the likelihood of adoption of the focal user. These results are consistent with our main estimation results.

Table 4.8: Alternative mechanisms: awareness and profitability of neighbors

	Awareness	Profitability of Neighbors	
	Traded/posted Bitcoin	Full sample	Sub-sample
	(1)	(2)	(3)
NeighborAdopt	0.103**	0.236***	0.170***
	(0.045)	(0.042)	(0.041)
AdoptBTCPositive	3.793***	4.703***	3.107***
	(0.895)	(0.541)	(0.315)
NonAdoptBTCPositive	-1.990***	-2.698***	-1.210***
	(0.569)	(0.349)	(0.174)
AdoptReturn	–	-0.003***	0.003***
	–	(0.0006)	(0.001)
NonAdoptReturn	–	0.00004	0.000
	–	(0.00003)	(0.000)
Neighbor_N	-0.023	-0.045***	-0.020***
	(0.014)	(0.008)	(0.004)
Post_N	-0.00002	0.00003***	0.00003
	(0.00002)	(0.00001)	(0.000)
AccTrade	-0.114***	-0.014***	-0.011***
	(0.007)	(0.001)	(0.001)
Tenure	-0.131***	0.022***	0.015***
	(0.012)	(0.002)	(0.002)
NTrade	0.029***	0.004***	0.003***
	(0.001)	(0.0004)	(0.000)
Diversification	0.123***		0.122***
	(0.015)	(0.004)	(0.003)
AccReturn	0.0003	-0.0002***	-0.0001***
	(0.0003)	(0.00004)	(0.000)
AccSD	-0.002***	-0.0004***	-0.0002***
	(0.000)	(0.0004)	(0.00005)
User-level FE	Yes	Yes	Yes
Week-level FE	Yes	Yes	Yes
Observations	459,105	2,316,559	2,078,139

Note. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table summarizes estimation results studying alternative mechanisms. The dependent variable is $Adopt_{i,t}$.

4.4.4 Profitability of neighbors

Another explanation for our main finding could be that the focal user's decisions are influenced by their neighbors' profitability, rather than the sentiments expressed in Bitcoin-related posts. Thus, users could be driven by a fear of missing out and try to keep up with the Joneses (see, e.g., Klocke et al. 2022, among others), rather than obtaining informative insights from their neighbors. To address this concern, we control for neighbors returns. Specifically, we control for the average profit of adopters ($AdoptReturn$) and non-adopters ($NonAdoptReturn$). The idea is to differentiate the effects of sentiments expressed in posts from the financial performance of neighbors. We re-estimate our main model and present the estimation results in Column (2) of Table 4.8. We find that the results are consistent with the results from the main model.

We further investigate the impact of neighbors' trading performance and compare the average profit of adopters and non-adopters from the focal user's neighbors. To ensure a fair comparison, we compare adopters and non-adopters who participate in trading activities during the same time period. The mean profit for adopters is -0.0617, while the mean profit for non-adopters is 1.0191. The average profit for non-adopters is significantly ($p < 0.01$) greater than that of adopters. In line with this observation, in 60.8% of users' neighborhood, non-adopters perform better than adopters, on average. In the remaining 39.2% of users' neighborhood, adopters are more profitable than non-adopters, on average. We further exclude those users from our analysis. Specifically, we exclude observations where users' adopted neighbors perform better than non-adopters and re-estimate our main model. We summarize the result in Column (3) of Table 4.8, and the results remain consistent. This further mitigates the concern that “keeping with the Joneses" explains our findings.

4.5 Robustness Checks

In this section, we present a series of robustness checks to further test our main results. First, our main model is based on the full data sample from the social trading platform. We conduct a robustness test using a focused subsample comprising more active users. Specifically, we select those users who have been active for over one month on the platform and have executed at least five trades during our observation period. This criterion ensures that our analysis accurately represents the behaviors of traders who consistently engage on the platform, as opposed to those who participate occasionally or only once. We estimate the model based on this subsample and present the estimation results in Column (1) of Table 4.9. The results align with the major findings from the main model.

Second, it is possible that different sentiment dictionaries generate different sentiment scores, which might affect the estimated effects of sentiment-related variables. In our main model, we implement a sentiment approach commonly used in financial text analysis. Considering the unique combination of financial and social media dynamics inherent in social trading platforms, we now apply an alternative sentiment analysis approach. Specifically, we implement the Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto and Gilbert 2014), a lexicon and rule-based tool that is particularly adept at deciphering sentiments expressed on social media. Then, we estimate the model based on the alternative measurement of sentiment and summarize the estimation results in Column (2) of Table 4.9. We find that the results from this alternative measurement of sentiment are consistent with our main results.

Third, our main analysis includes users who may have adopted Bitcoin at some point in the past, discontinued trading the cryptocurrency, and re-adopt Bitcoin. One may argue that these users do not provide the cleanest identification because of potentially confounding effects of the first adoption. We now address this concern and explicitly focus on a subsample of users who adopt Bitcoin for the very first time. This rules out any potential effects from a prior Bitcoin adoption experience in our data. We summarize the estimation results in Column (3) of Table 4.9. The results are consistent with our main results.

Fourth, in our main model, we include the user-level fixed effects to account for the unobserved heterogeneity of users. However, including user-level fixed effects may introduce information from the future into the model. This may happen during the demeaning of

Table 4.9: Results from robustness tests

	Sub-sample w more active users	Alternative Measurement of sentiment	Sub-sample w/o repetitive adoption	Model w/o User-level FE
	(1)	(2)	(3)	(4)
NeighborAdopt	0.228***	0.060*	0.195***	0.121***
	(0.040)	(0.028)	(0.035)	(0.035)
AdoptBTCPositive	4.862***	3.756***	3.490***	6.070***
	(0.533)	(0.481)	(0.498)	(0.329)
NonAdoptBTCPositive	-2.275***	-3.741***	-1.835***	-3.206***
	(0.297)	(0.468)	(0.313)	(0.270)
Neighbor_N	-0.042***	-0.002	-0.040***	-0.055***
	(0.007)	(0.004)	(0.007)	(0.004)
Post_N	0.00001	0.00002**	0.00002**	0.0001***
	(0.00001)	(0.000008)	(0.000)	(0.000)
AccTrade	-0.021***	-0.010***	0.018***	-0.035***
	(0.002)	(0.001)	(0.001)	(0.001)
Tenure	0.036***	0.015***	0.031***	0.056***
	(0.003)	(0.002)	(0.002)	(0.002)
NTrade	0.004***	0.003***	-0.004***	0.004***
	(0.0004)	(0.0004)	(0.000)	(0.001)
Diversification	0.151***	0.122***	0.061***	0.187***
	(0.004)	(0.003)	(0.004)	(0.003)
AccReturn	-0.0003***	-0.0001***	-0.0001***	-0.001***
	(0.00004)	(0.00004)	(0.000)	(0.000)
AccSD	-0.0003***	-0.0002***	-0.0001***	-0.001***
	(0.00005)	(0.00004)	(0.000)	(0.000)
User-level FE	Yes	Yes	Yes	No
Week-level FE	Yes	Yes	Yes	Yes
Observations	1,988,142	2,316,559	2,074,482	2,323,329

Note. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table summarizes four robustness checks. The dependent variable is $Adopt_{i,t}$.

variables when the mean considers the future behavior of users and thereby potentially contaminates the analysis with information not available at the time of decision-making. To address this concern, we re-run our main model without the user-level fixed effects. The result remains consistent (see Column (4) of Table 4.9).

Fifth, in our main model, the two key variables $AdoptBTCPositive$ and $NonAdoptBTCPositive$ are defined as the percentage of adopters and non-adopters who make positive Bitcoin-related posts. To further validate the robustness of our findings, we employ two alternative measures for these two key variables. The first alternative measure computes the percentage of positive Bitcoin-related posts relative to the total number of posts from adopters ($AdoptBTCPositive$) and non-adopters ($NonAdoptBTCPositive$) in a user's neighborhood, respectively. We calculate the percentage aggregating all posts from users at a collective post-level, while in the main model, we calculate the percentage at the user level. A second alternative measure simply counts the number of positive Bitcoin-related posts by adopters ($AdoptBTCPositive$) and non-adopters ($NonAdoptBTCPositive$) in the neighborhood, respectively. Thus, this measure relies on the absolute number rather than the percentage of positive Bitcoin-related posts. We re-estimate our model using these alternative measures and present the estimation results in Column (1) and (2) of Table 4.10, respectively. We find that the estimation results are consistent with our main results.

4.6 Conclusion

Social trading platforms have attracted increasing interest in recent years. Unlike traditional brokerage services, social trading platforms allow investors to view both the trading actions and opinions of other investors. Although prior literature has studied the impact of others' opinions on investment decisions, little is known about how the interplay between trading actions and opinions influences investment decisions. Our study fills this research gap by exploring the social learning in cryptocurrency adoption on a social trading platform. Cryptocurrency is distinct from traditional financial instruments. Investors rely heavily on social media to inform their investment decisions. Additionally, the volatile nature of cryptocur-

Table 4.10: Results from alternative measures of AdoptBTCPositive and NonAdoptBTC-Positive

	By Percentage at post level	By Count
	(1)	(2)
NeighborAdopt	0.170***	0.264***
	(0.034)	(0.031)
AdoptBTCPositive - % Post	17.290***	
	(1.911)	
NonAdoptBTCPositive - % Post	-7.062***	
	(1.005)	
AdoptBTCPositive - # of Posts		1.645***
		(0.155)
NonAdoptBTCPositive - # of Posts		-0.963***
		(0.168)
Neighbor_N	-0.031***	-0.026***
	(0.006)	(0.004)
Post_N	0.00001	
	(0.000007)	
AccTrade	-0.009***	-0.013***
	(0.001)	(0.001)
Tenure	0.021***	0.016***
	(0.003)	(0.002)
NTrade	0.002***	0.004***
	(0.0004)	(0.0004)
Diversification	0.124***	0.129***
	(0.004)	(0.003)
AccReturn	-0.00009**	-0.0001***
	(0.00004)	(0.00004)
AccSD	-0.0002***	-0.0003***
	(0.00005)	(0.00005)
User-level FE	Yes	Yes
Week-level FE	Yes	Yes
Observations	2,316,559	2,316,559

Note. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table shows the estimation results with alternative measures of AdoptBTCPositive and NonAdoptBTCPositive.

rency sets it apart from previous literature on product adoption, where the expected value of product is typically fixed.

Our paper expands the literature on social learning by exploring the novel context of cryptocurrency adoption and makes several theoretical contributions. First, we enrich the social learning literature by examining how the interplay between actions and opinions influences cryptocurrency adoption. This is crucial because users may express opinions that contradict their actual trading actions. This phenomenon is particularly prevalent in the cryptocurrency market, which has less regulation and where users might have incentives to spread misinformation. Second, we contribute to the literature on social learning by considering product uncertainty, where the value of the adopted product changes frequently. This offers a more nuanced understanding of social learning processes in highly volatile contexts.

Our findings provide practical implications on the who, when, and where of fostering social learning for cryptocurrency promotion (adoption). We first observe a notable learning effect when trading behavior aligns with expressed opinions, highlighting the importance of disclosing trading behavior in promoting cryptocurrency, especially in the early stages. Our findings suggest that engaging actively trading cryptocurrency investors could be more effective in promoting cryptocurrency than relying solely on opinion leaders. Moreover, our results show that high-income and older investors are more susceptible to opinions from their neighbors. Our results also indicate that positive opinions during periods of higher volatility are more effective in promoting cryptocurrency. Finally, our study highlights the essential role of transparency, particularly the public disclosure of trading activities, on social trading platforms in influencing investment decisions. Developers and marketers can leverage inherent social features and information transparency of social trading platforms to foster cryptocurrency adoption.

Our study is not without limitations. First, we focus specifically on Bitcoin. While other cryptocurrencies may exhibit different mechanisms due to their unique characteristics, many insights derived from Bitcoin can still be extended to other cryptocurrencies. Second, we focus on social learning in the early stage of cryptocurrency adoption. The process of product diffusion may differ between the early and later stages. Future research can investigate whether the effects of social learning remain relevant in the later stage and how their magnitudes change. Finally, our study is limited to examining the factors influencing the binary decision of cryptocurrency trading. Future research can provide a more comprehensive understanding by exploring the intensity or extent of cryptocurrency trading.

Chapter 5

TRANSACTIONAL NETWORKS IN NFT MARKETS

5.1 Introduction

The Non-Fungible Token (NFT) market has experienced explosive growth in recent years. Annual NFT sales jumped from roughly US$ 82 million in 2020 to an estimated US$ 17.7 billion in 2021, a 215× surge that outpaced the early growth of cryptocurrencies themselves.[1] NFT marketplaces such as OpenSea and Foundation lower minting costs while embedding social features, including social media badge, user profile, as well as live bidding to motivate participation. Yet the same period revealed fragility: average resale profits shrank, liquidity concentrated in a handful of "blue-chip" collections, and by mid-2022 exuberance had faded amid broader crypto turmoil.[2] These boom-and-bust dynamics make NFTs a natural laboratory for studying digital asset valuation.

An NFT is a unique, indivisible token recorded on a public ledger that confers verifiable ownership of a digital object. Three design choices set NFTs apart from conventional securities or even cryptocurrencies:

1. **Non-interchangeability.** Each token carries a distinct on-chain identifier, precluding fungibility.

2. **Collection membership.** Most tokens reside in branded series (e.g., Bored Ape Yacht Club, CryptoPunks), creating club-good externalities in which value depends on who else belongs.

---

[1] `https://www.axios.com/2022/03/10/nft-sales-17b-2021-report`

[2] `https://www.wired.com/story/the-year-the-nft-died-and-came-back-to-life`

3. **Social price discovery.** Community sentiment, creator reputation, and influencer endorsements often move markets more than intrinsic traits (Dowling 2022, Ante 2023).

These features generate large, rapid swings in valuation that standard asset-pricing models struggle to reconcile. Early research tackles the puzzle by extracting visual attributes, including image embeddings, trait rarity, or textual metadata, to forecast sale prices (e.g., Borri et al. 2022). However, attribute-centric approaches cannot explain why two almost indistinguishable Bored Apes can sell for vastly different prices. We argue that the value of NFTs is socially constructed through transactional history rather than encoded in pixels. To test this hypothesis, we construct a transactional network in which two NFTs are linked if they were traded by the same wallet during a time period and then investigate how the transactional network is associated with NFT pricing.

Utilizing network-structured data imposes some methodological challenges. First, the dimensionality issue is a major hurdle as the size of networks increases, leading to computationally expensive estimation processes. This problem is not merely technical but impacts the feasibility of applying sophisticated models in real-world scenarios where timely data processing is crucial. Second, network models frequently depend on strong parametric assumptions to characterize network effects. For instance, in studies examining peer influence, the effect of the network is commonly approximated by averaging the behaviors of neighbors. This approach, while simplifying the analysis, may overlook the more complex dynamics of how such influences actually manifest. Indeed, the real interactions and influences within networks can be far more varied and intricate than what such simplified models capture. These assumptions can overly constrain the model, hindering its ability to capture the complexities of network dynamics accurately. Third, the issue of limited sample sizes is prevalent; typically, only a few networks are available for study, or in some cases, just a single network. This restriction is problematic in the context of large models, which require substantial data to train effectively and avoid overfitting.

We address these methodological obstacles with a permutation-invariant, set-based architecture inspired by Deep Sets (Zaheer et al. 2017). The local network of each NFT is considered as an element of a large set, while a set represents the full network structure of a market. We first demonstrate the validness of this method through simulation studies, next adapt this method into the real empirical data of the NFT markets.

We construct a comprehensive dataset from OpenSea, the largest decentralized NFT marketplace on the Ethereum blockchain. Our dataset combines three layers of information: (i) collection-level metadata such as creator fees, categories, and social media presence; (ii) item-level attributes including image files and visual traits that drive rarity; and (iii) full transaction histories capturing wallet addresses, timestamps, and sale prices. To focus on the formative phase of the NFT economy, prior to the speculative surge of 2021, we restrict our sample to trades completed on or before January 1, 2021, resulting in 48,192 transactions across 19,670 distinct tokens. We aggregate NFT prices at the monthly level to construct return, and build a transactional network in which two items are linked if they were traded by the same wallet in the same month. Our empirical results show that incorporating network-based features significantly improves the predictive performance of NFT pricing models, compared to using only image-based or metadata features. This finding highlights the critical role of market participants in shaping NFT value.

Our findings offer several important implications for platform design, digital asset pricing, and market monitoring. First, our results suggest that standard hedonic pricing models, which focus narrowly on token traits, may miss economically meaningful variation driven by network position. Incorporating transactional network structures can yield more accurate value assessments, particularly in volatile or thinly traded markets. Second, NFT marketplaces may enhance liquidity and price discovery by surfacing transactional provenance—for example, by displaying badges such as “Previously held by Top 1% trader”, to signal embeddedness and trust. Third, this study contributes to a growing body of work that views digital ownership not as a purely technological innovation, but as a socially embedded process shaped by interaction histories.

5.2 *Data*

We construct our dataset from OpenSea, the largest decentralized marketplace for non-fungible tokens (NFTs). Because OpenSea accounts for the majority of NFT trading volume and user activity on the Ethereum blockchain, it provides a natural laboratory for studying the economic forces that shape NFT value.

Our raw data contain three components:

1. **Collection–level metadata**. For each collection, we observe its total item count, creator fee, category (e.g., *Art*, *Gaming*), verification badge, social media channels (official website, Discord, Twitter, Medium, Instagram), launch date, and a free-text description.

2. **Item–level attributes**. For every NFT we capture the image file and the platform–supplied visual traits (e.g., color palette, accessory type, background style).

3. **Transaction histories**. Each on-chain event records the transaction type (SALE or TRANSFER), block-time stamp, sender and recipient wallet addresses, and the price paid (when a sale occurs).

We scraped all transactions for the top-200 collections listed on OpenSea's ranking pages. To analyze price formation during the market's formative period—and to avoid the extraordinary volatility that began with the 2021 boom—we restrict the sample to trades executed on or before 1 January 2021. This screening leaves 48,192 transactions. Although the headline count of collections is large, trading activity is highly concentrated: 99% of the early transactions occur in just nine collections (see Table 5.1). Within those nine collections we retain only NFTs that were actually traded, because our objective is to model the returns of transacted items. The resulting analysis set contains 19,670 distinct NFTs, of which 3,422 changed hands at least twice.

Collection	Share of Early Transactions
Voxels (formerly Cryptovoxels)	35.03%
CryptoSkulls	29.46%
CryptoPunks	14.20%
Decentraland Wearables	8.52%
MLB Champions	3.18%
BCCG	3.09%
Autoglyphs	3.01%
CryptoSpells	2.89%
Meme Ltd.	0.62%

Table 5.1: Distribution of Pre-2021 Transaction Volume by Collection

We aggregate the item transactions in month, and use the month end price to represent the price of the month and the change in the month end price as the return for each NFT; consider the first month of an NFT as the first observation and the last month it has transaction as the last observation.

We further construct a transactional network among all items, specifically, we consider two items are linked undirectionally if they have overlap in the user wallet who traded them. Figure 5.1 illustrates a network from a month. It is worth noting that there is a largest component in the transactional network with many small components in the network. This suggests, a portion of NFT items are traded by very similar set of users.

Figure 5.1: Illustrations of An NFT Network

5.3 Model and Simulation Results

5.3.1 Model

In this section, we provide a general characterization of network-based models. Consider a scenario where there are N individuals distributed across M distinct networks. Each network is denoted as $\mathcal{G}_m$ and is composed of nodes and edges, represented by N_m and E_m respectively. Here, N_m refers to the set of nodes and E_m to the set of edges within network m. Within each network, nodes are characterized by feature vectors. Specifically, for network m, each node j has an associated feature vector X_{mj}, which encapsulates various attributes relevant to the node's role and function within the network. These feature vectors represent critical data points for analysis, as they contain information pertinent to each individual node.

To systematically analyze these networks, we define X_m as the set containing all feature vectors X_{mj} for the nodes in network m. Additionally, $\mathcal{G}_m$ is defined as the set that includes both the nodes N_m and the edges E_m, collectively outlining the full structure of the network m. Furthermore, we introduce E_{mj}, a binary vector for each node j within network $\mathcal{G}_m$. Each element of E_{mj} indicates the presence or absence of an edge between node j and every other node i in the same network. In this way, all characteristics within $\mathcal{G}_m$ can be written as a set, $\{E_{mj}, X_{mj}, \varepsilon_{mj}\}_{j \in N_m}$, where ε_{mj} denote random errors.

Network models typically focus on two types of outcomes: node-level outcomes and edge-level outcomes. This paper specifically concentrates on node-level outcomes. Accordingly, we observe an outcome vector $Y_m \in \mathbb{R}^{N_m}$ for each network m, which records the outcomes for each node within that particular network.

In the existing literature, the node-level outcome has been extensively studied in various contexts, illustrating the impact of network structure on individual decisions and behaviors. For any research question studying the effect of network structures or the interplay of network structures and individual characteristics on some individual outcome or behavior, the central challenge is to specify a function that captures all effects, encompassing both network and non-network influences. Without loss of generality, all network effects or non-network effects can be represented as a function of $\mathbb{X}_m$, $\mathcal{G}_m$. Therefore, the outcome of an individual could be expressed as

$$Y_{mj} = f(\{E_{mj}, X_{mj}, \varepsilon_{mj}\}_{j \in N_m}), \tag{5.1}$$

where ε_{mj} denote random errors. Notably, with defining the network functions in this way, we essentially define a set-based input instead of a vector-based input. This allows us to use recently developed techniques in set-based input, such as Deep Sets and Transformers.

Next, we specify a few common assumptions.

Assumption 1 (Exogeneity)**.** *The unobserved error term ε_{mj} is independent and identically distributed (i.i.d.) across all individuals. This can be expressed as follows:*

$$\mathbb{P}(\varepsilon_{mj} \mid \mathbb{X}_m, \mathbb{G}_m) = \mathbb{P}(\varepsilon_{mj})$$

Addressing the central issue of endogeneity in network models, prior research has developed three primary methodologies: first, conducting randomized experiments to eliminate concerns about endogeneity; second, employing an instrumental variable approach; and third, incorporating a model of network formation within a structural model framework. Models utilizing the third approach typically assume that endogeneity can be resolved through observed features after specifying certain functional forms. For example, some methods model the network formation process explicitly to address the issue of endogenous network formation, while the network formation is also explained observed data. In our approach, we do not make any functional assumptions on how these features affect the outcome. This flexibility allows for the possibility that some characteristics may first impact network formation, which in turn affects the outcomes.

Assumption 2 (Identity Independence)**.** *For any individual $j \in \mathcal{N}_m$ and any network N_m, we assume the function f does not depend on the identity of the each indvidual ($j \in \mathcal{N}_m$). That is:*

$$f_{mj}(\{E_{mj}, X_{mj}, \varepsilon_{mj}\}_{j \in N_m}) = f(\{E_{mj}, X_{mj}, \varepsilon_{mj}\}_{j \in N_m})$$

This is a very common assumption in network models where each individual has a homogeneous network effect model.

We use σ_j to denote a permutation on all the nodes except the focal node within a network: $N_m \setminus \{j\} \to N_m \setminus \{j\}$, and σ_j^{-1} as the inverse permutation. We next make the assumption below,

Assumption 3 (Permutation Invariance)**.** *The network function f is invariant under any permutation σ_j applied to the all other nodes within the same network, such that:*

$$y_{mj} = f_{mj}(\{E_{mj}, X_{mj}, \varepsilon_{mj}\}, \{E_{m\sigma_j^{-1}(k)}, X_{m\sigma_j^{-1}(k)}, \varepsilon_{m\sigma_j^{-1}(k)}\}_{k \in N_m \setminus \{j\}})$$

This assumption is also common within network data, particularly because the underlying processes that govern network interactions often do not inherently depend on the specific ordering of nodes.[3] Permutation invariance implies that the behavior or properties of the focal node j are determined by the characteristics and relationships of the other nodes, regardless of their specific sequence within the dataset. One important trick to highlight here is that we include each pairwise relationship among nodes, any permutation on top of

[3]Some networks may have a ranking for each individual. In such cases, we can consider the ranking as an observed characteristic of an individual and incorporate into X_{mj}, then it can become

it only changes the sequence of the dataset, thus it does not affect the outcome of the focal individual.

Theorem 1. *For any network $\mathcal{G}_m$ containing nodes $\{1, 2, \ldots, N_m\}$ with each node j having a feature vector X_{mj} and a corresponding outcome Y_{mj}, if the function f representing the node outcomes satisfies assumptions 1, 2, and 3, then there exists suitable functions ρ, ϕ_1, and ϕ_2 such that:*

$$Y_{mj} = \rho\left(\phi_1(X_{mj}, E_{mj}) + \sum_{k \in N_m \setminus \{j\}} \phi_2(X_{mk}, E_{mk})\right),$$

where X_{mj} represents the feature vector for node j and X_{mk} for the nodes k in the network, excluding node j.

This result is a generalization of the results shown in Zaheer et al. (2017) and can be demonstrated using similar arguments. The above result illustrates that the input space of the network function does not scale quadratically with the number of nodes. Instead, the input dimensions of the network function, specifically through the transformations ϕ_1 and ϕ_2, increase linearly with the sum of the number of nodes and the number of features per node. This aspect is a significant advantage of this modeling approach. This characteristic facilitates the management of large networks without the computational overhead typically associated with increasing the number of nodes, which is a critical advantage in network analysis, especially in contexts involving complex and large-scale networks.

5.3.2 Simulation Performance

Data generation

As a baseline, we employ a linear-in-means model to investigate how various factors influence an individual's binary choice outcome within network settings in this simulation. These factors include an individual's own characteristics, their connectivity within the network, and the influence of their neighbors' attributes and behaviors.

To simulate this, we generate M separate Erdős-Rényi networks, each denoted as $\mathcal{G}_m$, containing J nodes. In these networks, each possible link between two nodes forms independently with a probability of $5/J$, meaning the average degree is 5 in these networks. Each node within these networks is characterized by a feature vector X_{mj} of K exogenous features. These features are drawn from a standard normal distribution, reflecting the diverse range of attributes that might influence individual behaviors and interactions within a network.

Additionally, to construct a fully exogenous case, we consider an external agent randomly designates 5% of individuals in each network as seeds. This seeding status is represented by a binary variable S, indicating whether a node is a seed. The influence of seeding reflects typical interventions or initial conditions in network dynamics studies. The outcome for each node, y_{mj}, is modeled to be dependent on: 1) Their own feature vector X_{mj}, 2) Their degree d_{mj}, which is the number of connections they have, 3) The average feature vector of their neighbors $\bar{X}_{mj}$, 4) The average seeding status among their neighbors $\bar{S}_{mj}$.

This dependency is captured through a logistic regression model where expected utility u_{mj} that node j in network m derives from choosing a particular action is defined as:

$$u_{mj} = \beta_0 + \beta_1 X_{mj} + \beta_2 d_{mj} + \beta_3 \bar{X}_{mj} + \beta_4 \bar{S}_{mj} + \varepsilon_{mj} \tag{5.2}$$

By assuming a random error that following the Type-I Extreme Value distribution, the probability that node j in network m chooses a particular action (or exhibits a particular behavior) is then given by:

$$\text{Prob}(y_{mj} = 1) = \frac{e^{u_{mj}}}{1 + e^{u_{mj}}}. \tag{5.3}$$

$$\text{Prob}(y_{mj} = 1) = \frac{e^{u_{mj}}}{1 + e^{u_{mj}}}. \tag{5.4}$$

Model Performance

In the simulation, for each component of our model (ϕ_1, ϕ_2, and ρ), we use a standard 3-layer neural network, and this is implemented without further hyperparameter tuning for number of layers or number of neurons in each layer. Specifically, all layers for ϕ_1, ϕ_2 and all layers except for the output layer of ρ consist of 64 neurons.

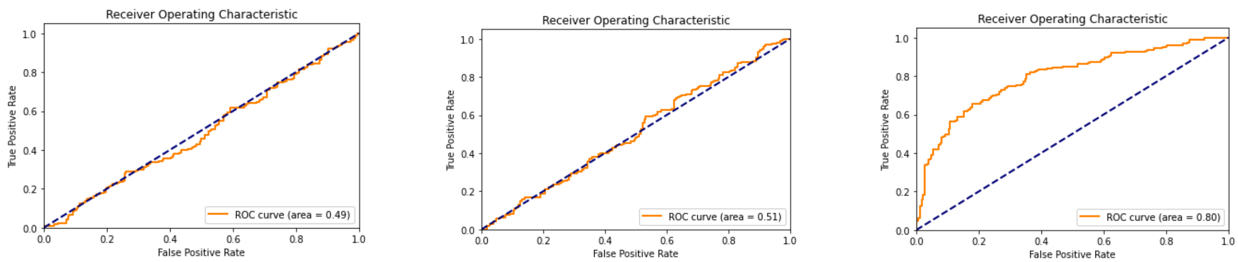
To evaluate the performance of our proposed method, we compare the result of our method with a couple of other standard neural network non-parametric estimators. Specifically, we compare with

- A standard multi-layered neural network, with $(J \times (J + K)) \times 1$ input vector; the number of layers and number of neurons in each layer, as well as the number of epochs and learning rate is well-tuned. We apply a sigmoid activation in the final layer to ensure output values between 0 and 1.

- A standard convolutional neural network, here we handle network data via convolutional layers (specifically LeNet) and node features through fully connected layers. It consists of two convolutional layers followed by max-pooling and three fully connected layers, where the first two process the network and node features separately before combining them. The model outputs node-level predictions, applying a sigmoid activation in the final layer to ensure output values between 0 and 1.

In our simulation, we consider there are in total of $M = 20$ networks and each network has $N_m = 100$ individuals and each has $K = 5$ features. We use 60% of generated data as training data, 20% as validation data and 20% as test data. Note that the sample size for both standard neural network and the standard CNN are both equal to the number of networks (20), which is fairly small, particularly for neural network methods. In contrast, the sample size for our proposed method is $M \times N_m$. We random draw 20 times report the performance in the unseen test data in Figure 5.2. Since we assume a binary outcome, we plot the ROC curve on the performance on the unseen test data for each method. With such small sample size, the standard neural network and CNN barely learn anything from the data. However our proposed method achieve an AUC at 0.80.

Our model distinguishes itself from traditional neural network architectures in two significant ways. First, the number of parameters in our proposed method scales linearly with



(a) MLP (b) CNN (c) Deep Sets

Figure 5.2: Result comparison ($N_m = 100$, $M = 20$)

the number of nodes, rather than quadratically. While CNN-based methods can reduce the number of parameters through the use of convolutional layers and pooling operations, the parameter count in the fully connected layers that often follow still depends on a quadratic function of reduced dimensions. Consequently, CNNs generally possess a higher total number of parameters. Table 5.2 exemplifies this by showing the parameter count for one draw, demonstrating our method's scalability.

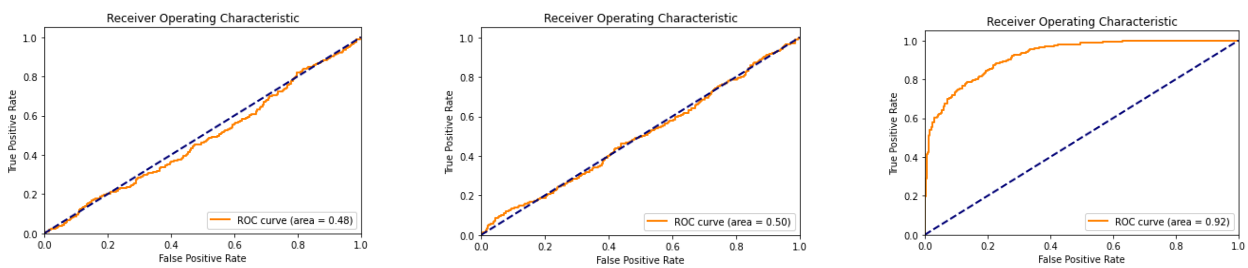
Table 5.2: Comparison of the Number of Trainable Parameters

	MLP	CNN	Our method
Total Trainable Parameters	342,500	1,020,836	51,073

Variant size of networks

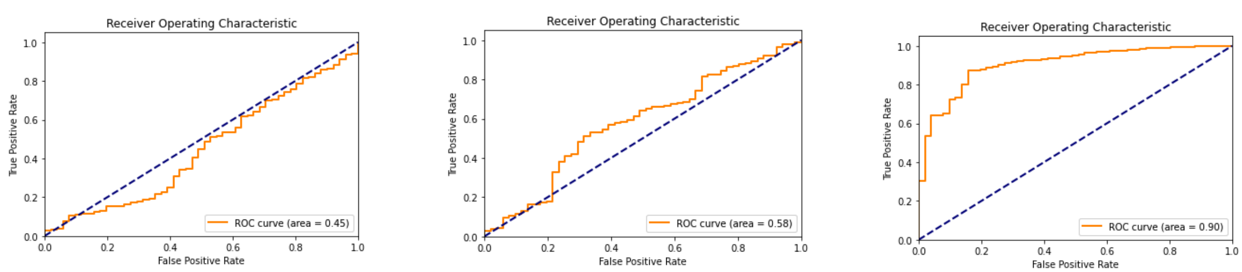
To assess the scalability and effectiveness of our method compared to others, we vary the size of each network while maintaining a fixed number of networks in our simulations. Specifically, we increase the number of individuals per network to 300 and 500, respectively. Figures 5.3 and 5.4 illustrate these results, showing that our method consistently outperforms well-tuned

MLP and CNN methods, despite the fact that our approach does not have any hyperparameter tuning regarding the neural network structure. In addition, our model also demonstrates improvement in performance when the size of the network increases. This is because the sample size of our method is the total number of individuals thus as the size increases, the sample size also increases.



(a) MLP (b) CNN (c) Deep Sets

Figure 5.3: Results comparison ($N_m = 300,\ M = 20$)

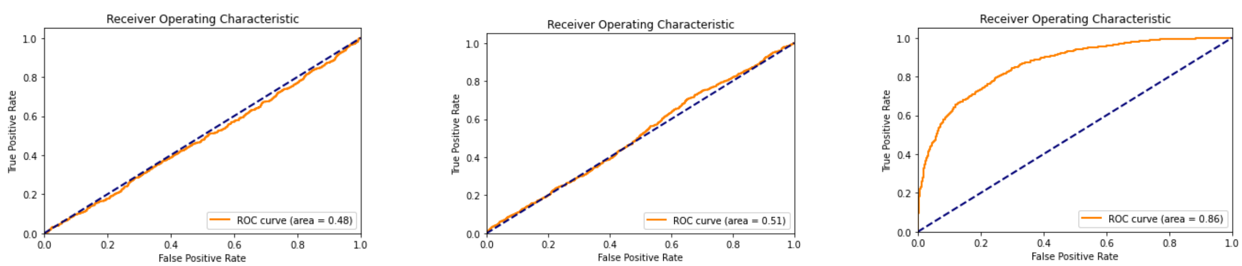


(a) MLP (b) CNN (c) Deep Sets

Figure 5.4: Results comparison ($N_m = 500,\ M = 20$)

Variant number of networks

To further validate the performance of our proposed method, we vary the number of networks in the simulated data. Our model consistently outperforms other methods. In addition, our model shows monotonic improvement as the number of networks increases. This pattern is not observed in the two methods used for comparison. A possible explanation for this disparity is that, while the sample size grows with an increase in the number of networks, it does not adequately compensate for the escalation in the number of parameters required by these other methods. This inability to scale effectively with parameter increases highlights a significant advantage of our approach over traditional methods.



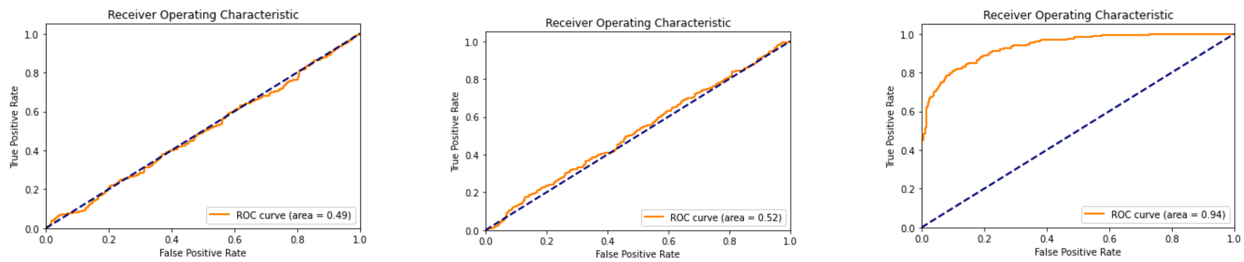
(a) MLP (b) CNN (c) Deep Sets

Figure 5.5: Results comparison ($N_m = 100$, $M = 50$)

5.4 Empirical Results

Building on the simulation evidence, we now investigate whether transactional–network information improves price forecasts in the live market for NFTs. We compare two model specifications:

(a) **Image & Metadata Only**. A 3-layer ReLU multilayer perceptron (MLP) that receives the same visual embeddings and static collection/item covariates used in prior work (e.g., color histograms, , creator fee, social media presence, etc.).



(a) MLP (b) CNN (c) Deep Sets

Figure 5.6: Results comparison ($N_m = 100$, $M = 100$)

(b) **Image + Monthly Network (Deep Sets)**. Our permutation-invariant architecture, which augments the baseline inputs with a set representation of each token's *monthly* network (wallet overlap links formed in the same calendar month).

We construct a rolling-origin evaluation: for every month t in 2020, we train each model on trades up to $t-1$, validate on month t, and test on month $t+1$. The process yields twelve non-overlapping test folds, which we concatenate to form the "future-month" hold-out sample. Continuous features are standardized within each training fold; price targets are ETH values at sale.

Table 5.3 reports mean absolute error (MAE) and root mean squared error (RMSE) averaged across the twelve test folds. We find that adding monthly transactional-network information reduces prediction error significantly.

Specification	**MAE**	**RMSE**
(a) Image & Metadata Only	22.39	176.36
(b) + Monthly Network	**17.25**	**133.52**

Table 5.3: Out-of-Sample NFT Price Prediction (USD)

5.5 Conclusion

This chapter provides empirical evidence that NFT prices are shaped not only by token-specific traits but also by the structure of the transactional network in which the token is embedded. While prior studies emphasize image features, rarity attributes, or textual metadata, our findings highlight the critical role of *who trades what*—not just *what is traded*. Leveraging a pre-boom dataset on OpenSea, the largest NFT marketplace on Ethereum, we show that network position of an NFT in the market is not just incidental; it is economically meaningful. Adding network information significantly reduces out-of-sample prediction errors compared to models based on visual and metadata features alone.

These findings open up novel design levers for NFT marketplaces. Platforms may benefit from surfacing transactional provenance as a signal of value—badges such as "Previously held by Top-1% wallet" or "Linked to Blue-Chip Collection" could help guide users toward higher-quality assets. Similarly, marketplaces could consider rewarding high-centrality traders, either through reduced fees, exclusive drops, or curated exposure, to amplify their positive spillover effects. As NFT markets evolve, tools that make the invisible structure of trades visible may become as important as front-end listing and trait filters.

This study is not without limitations. First, our empirical setting focuses on the formative phase of the Ethereum NFT market. This was a relatively low-volume and low-speculation period, and results may not generalize to later phases marked by exponential growth, speculative bubbles, and increased retail participation. Extending the analysis to multi-chain ecosystems such as Solana or Polygon, as well as collections with embedded utility (e.g., gaming assets, token-gated communities), may reveal distinct patterns of network influence. Second, although our predictive framework demonstrates that wallet overlap carries significant information about future prices, it stops short of establishing causality. It remains unknown whether influential wallets actually generate value through visibility and trust.

Chapter 6

CONCLUDING REMARKS

In this dissertation, I study the role of networks in three distinctive FinTech applications: social trading, cryptocurrency adoption, and NFT pricing.

First, rather than focusing solely on individual attributes, such as financial performance, I examine the impact of strategy similarity patterns on the popularity of traders. I first construct similarity networks among traders, allowing for dynamic changes. Recognizing that being similar to peers could be a part of a trader's strategy, we employ a coevolution model to illustrate the interplay between the formation of traders' similarity networks and follower dynamics. I find that as the number of peers adopting similar strategies to a particular trader increases, the trader's number of followers tends to decrease over time. Moreover, this negative effect is especially pronounced for more successful traders, who are more susceptible to substitution by similar peers. Using counterfactual simulations, I show that promoting a fairer distribution of followers and reducing the density of the similarity network can help mitigate the negative substitution effect, thereby increasing the total number of followers on the platform. I further validate our results with additional robustness checks. This study identifies several managerial and operational implications for the social trading platforms.

Second, I exploit a unique dataset from a large social trading platform and study how the interplay of opinions and behaviors within social networks influences an individual's decision to adopt cryptocurrency. I focus on cryptocurrency adoption, as cryptocurrencies have gained increasing attention over the last few years. Due to the lack of standardized information, social media has become a main information channel for individuals to decide whether to trade cryptocurrencies. This reliance on social media presents a unique challenge: The vast and unregulated nature of social media can often lead to disseminating information of questionable credibility. Using investment opinions and actual trading behaviors, I find a pronounced social learning effect in cryptocurrency adoption. Individuals are more inclined to adopt Bitcoin when their network peers do, especially when these peers' trading actions are consistent with their expressed opinions. In times of high uncertainty, investors tend to rely more on positive opinions, regardless of others' actual trading actions. In addition, high-income and older investors are more susceptible to opinions from their neighbors. Textual analysis indicates that Bitcoin posts are longer and more forward-looking compared to non-Bitcoin posts. I rule out alternative mechanisms such as awareness or "Keeping up with the Joneses."

Third, I investigate the role of transactional networks in shaping the value of NFTs, proposing that value is not solely embedded in visual features or metadata but is also socially constructed through trading behaviors. Using data from OpenSea, I construct monthly transactional networks by linking NFT items that share common traders, capturing the relational embeddedness among items. I develop a novel modeling approach that leverages permutation-invariant property of networks to incorporate network information efficiently, overcoming common challenges in network modeling such as dimensionality. Our simulation results validate the scalability and accuracy of the proposed method, and our empirical analysis demonstrates that incorporating network information dramatically improves NFT price prediction. These findings underscore the importance of social dynamics in NFT valuation and suggest new avenues for asset pricing and market design in digital asset ecosystems.

Together, the three chapters of this dissertation highlight the central role that network structures play in shaping financial behavior and outcomes across diverse FinTech contexts. Rather than attributing attention, adoption, or pricing to intrinsic qualities alone, each study shows how relationships, whether among traders or assets through shared ownership, serve as key mechanisms in decision-making. These findings underscore the importance of designing FinTech platforms with an awareness of how social signals and structural positioning can amplify or dampen user engagement, influence perceived value, and contribute to inequality in visibility or influence.

Building on these insights, a promising direction is to investigate the heterogeneity in network links—not all ties are equal, and factors such as tie strength, recency, or role (e.g., expert vs. novice) may differentially influence perception and behavior. Understanding such heterogeneity is key to evaluating how information and influence diffuse across FinTech contexts. Moreover, as FinTech platforms increasingly shape and reshape user networks through design interventions, an important question is how to translate these insights into algorithmic design—for instance, in curating content, ranking peers, or surfacing signals that foster informed and equitable participation. Research that bridges network theory with algorithmic implementation could provide practical guidelines for building platforms that not only optimize engagement but also support healthy, inclusive financial ecosystems. Finally, this dissertation suggests a broader research agenda that links the study of financial technologies with theories of social behavior and platform governance. As FinTech systems become increasingly complex and participatory, understanding how users interpret peer behavior, respond to structural cues, and act under ambiguity will be crucial for designing markets that are not only efficient but also equitable and resilient. Future studies can further investigate how network features mediate trust, visibility, and value, which are core elements that underpin the success and sustainability of digital financial ecosystems.

BIBLIOGRAPHY

Alloway T (2014) From Icahn to 'I can' social trading takes off. Accessed April 17, 2025, https://www.ft.com/content/5032a046-63d9-11e4-8216-00144feabdc0.

Almeida J, Gonçalves TC (2023) A systematic literature review of investor behavior in the cryptocurrency markets. *Journal of Behavioral and Experimental Finance* 37:100785.

Ameri M, Honka E, Xie Y (2019) Word of mouth, observed adoptions, and anime-watching decisions: The role of the personal vs. the community network. *Marketing Science* 38(4):567–583.

Ammann M, Schaub N (2021) Do individual investors trade on investment-related internet postings? *Management Science* 67(9):5679–5702.

Ante L (2023) Non-fungible token (nft) markets on the ethereum blockchain: Temporal development, cointegration and interrelations. *Economics of Innovation and New Technology* 32(8):1216–1234.

Apesteguia J, Oechssler J, Weidenholzer S (2020) Copy trading. *Management Science* 66(12):5608–5622.

Araci D (2019) Finbert: Financial sentiment analysis with pre-trained language models. *arXiv:1908.10063* .

Aral S, Nicolaides C (2017) Exercise contagion in a global social network. *Nature Communications* 8(1):1–8.

Aral S, Walker D (2011) Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* 57(9):1623–1639.

Aspris A, Foley S, Svec J, Wang L (2021) Decentralized exchanges: The "wild west" of cryptocurrency trading. *International Review of Financial Analysis* 77:101845.

Athey S, Parashkevov I, Sarukkai V, Xia J (2016) Bitcoin pricing, adoption, and usage: Theory and evidence. *Stanford University Graduate School of Business Research Paper* .

Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. *Proceedings of the fourth ACM international conference on web search and data mining*, 65–74.

Bali TG, Cakici N, Whitelaw RF (2011) Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99(2):427–446.

Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science* 341(6144):1236498.

Banerjee AV (1992) A simple model of herd behavior. *The Quarterly Journal of Economics* 107(3):797–817.

Bao C, Bardhan IR, Singh H, Meyer BA, Kirksey K (2020) Patient–provider engagement and its impact on health outcomes: A longitudinal study of patient portal use. *MIS Quarterly* 44(2).

Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.

Barber BM, Odean T (2000) Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance* 55(2):773–806.

Barber BM, Odean T (2001) Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics,* 116(1):261–292.

Bastani H, Zhang DJ, Zhang H (2022) Applied machine learning in operations management. *Innovative Technology at the Interface of Finance and Operations: Volume I* 189–222.

Bensoussan A, Chevalier-Roignant B, Rivera A (2022) A model for wind farm management with option interactions. *Production and Operations Management* 31(7):2853–2871.

Bertrand M, Luttmer EF, Mullainathan S (2000) Network effects and welfare cultures. *The quarterly journal of economics* 115(3):1019–1055.

Bhattacharya P, Phan TQ, Bai X, Airoldi EM (2019) A coevolution model of network structure and user behavior: The case of content generation in online social networks. *Information Systems Research* 30(1):117–132.

Biais B, Bisiere C, Bouvard M, Casamatta C, Menkveld AJ (2023) Equilibrium bitcoin pricing. *The Journal of Finance* 78(2):967–1014.

Bifulco R, Fletcher JM, Ross SL (2011) The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy* 3(1):25–53.

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100(5):992–1026.

Blanchard SJ, Noseworthy TJ, Pancer E, Poole M (2023) Extraction of visual information to predict crowdfunding success. *Production and Operations Management* 32(12):4172–4189.

Blau BM (2017) Price dynamics and speculative trading in bitcoin. *Research in International Business and Finance* 41:493–499.

Blume LE, Brock WA, Durlauf SN, Jayaraman R (2015) Linear social interactions models. *Journal of Political Economy* 123(2):444–496.

Bojd B, Yoganarasimhan H (2022) Star-cursed lovers: Role of popularity information in online dating. *Marketing Science* 41(1):73–92.

Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.

Bollinger B, Gillingham K, Kirkpatrick AJ, Sexton S (2022) Visibility and peer influence in durable good adoption. *Marketing Science* 41(3):453–476.

Borri N, Liu Y, Tsyvinski A (2022) The economics of non-fungible tokens. *Available at SSRN* 4052045.

Boucher V, Fortin B (2016) Some challenges in the empirics of the effects of networks. *Handbook on the Economics of Networks* 45:48.

Bouri E, Gupta R, Roubaud D (2019) Herding behaviour in cryptocurrencies. *Finance Research Letters* 29:216–221.

Bowman KO, Shenton LR (1985) Method of moments. *Encyclopedia of Statistical Sciences* 5:467–473.

Bradley MA, Daly AJ (1997) Estimation of logit choice models using mixed stated preference and revealed preference information. *Understanding travel behaviour in an era of change* 209–232.

Bramoullé Y, Djebbari H, Fortin B (2009) Identification of peer effects through social networks. *Journal of Econometrics* 150(1):41–55.

Bramoullé Y, Galeotti A, Rogers BW (2016) *The Oxford Handbook of the Economics of Networks.*

Brass DJ (2015) Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding. *MIS Quarterly* 39(3):729–742.

Brown NC, Wei KD, Wermers R (2014) Analyst recommendations, mutual fund herding, and overreaction in stock prices. *Management Science* 60(1):1–20.

Brynjolfsson E, Hu Y, Smith MD (2003) Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* 49(11):1580–1596.

Burk WJ, Steglich CE, Snijders TA (2007) Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors. *International Journal of Behavioral Development* 31(4):397–404.

Burtch G, Gupta D, Martin P (2021) Referral timing and fundraising success in crowdfunding. *Manufacturing & Service Operations Management* 23(3):676–694.

Bushman RM, Smith AJ (2003) Transparency, financial accounting information, and corporate governance. *Economic Policy Review* 9(1).

Cai H, Chen Y, Fang H (2009) Observational learning: Evidence from a randomized natural field experiment. *The American Economic Review* 99(3):864–882.

Calvó-Armengol A, Patacchini E, Zenou Y (2009) Peer effects and social networks in education. *The review of economic studies* 76(4):1239–1267.

Catalini C, Gans JS (2020) Some simple economics of the blockchain. *Communications of the ACM* 63(7):80–90.

Çelen B, Kariv S, Schotter A (2010) An experimental test of advice and social learning. *Management Science* 56(10):1687–1701.

Chakraborty S, Swinney R (2021) Signaling to the crowd: Private quality information and rewards-based crowdfunding. *Manufacturing & Service Operations Management* 23(1):155–169.

Chamberlain G (1984) Panel data. *Handbook of Econometrics* 2:1247–1318.

Chan TY, Li J, Pierce L (2014) Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science* 33(4):463–484.

Charness G, Gneezy U (2012) Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization* 83(1):50–58.

Chen H, De P, Hu YJ, Hwang BH (2014) Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27(5):1367–1403.

Chen J, Chen DL, Lewis G (2020a) Mostly harmless machine learning: learning optimal instruments in linear iv models. *arXiv preprint arXiv:2011.06158* .

Chen Y, Wang Q, Xie J (2011) Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research* 48(2):238–254.

Chen YJ, Dai T, Korpeoglu CG, Körpeoğlu E, Sahin O, Tang CS, Xiao S (2020b) OM forum—innovative online platforms: Research opportunities. *Manufacturing & Service Operations Management* 22(3):430–445.

Cheng TY, Ocktaviani S, Li HC, Lin CH (2013) Do gender and age matter in trading discipline? *Working Paper* .

Choi N, Sias RW (2009) Institutional industry herding. *Journal of Financial Economics* 94(3):469–491.

Chung S, Kim K, Lee CH, Oh W (2023) Interdependence between online peer-to-peer lending and cryptocurrency markets and its effects on financial inclusion. *Production and Operations Management* 32(6):1939–1957.

Cong LW, Li Y, Wang N (2021) Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies* 34(3):1105–1155.

Conley TG, Udry CR (2010) Learning about a new technology: Pineapple in ghana. *American Economic Review* 100(1):35–69.

Cragg JG, Donald SG (1993) Testing identifiability and specification in instrumental variable models. *Econometric Theory* 9(2):222–240.

Curry D (2021) Etoro revenue and usage statistics (2025). Accessed April 17, 2025, https://www.businessofapps.com/data/etoro-statistics/.

Da Z, Huang X (2020) Harnessing the wisdom of crowds. *Management Science* 66(5):1847–1867.

Daniel K, Hirshleifer D, Subrahmanyam A (1998) Investor psychology and security market under and overreactions. *The Journal of Finance* 53(6):1839–1885.

Daytrading (2025) Zulutrade review 2025. Accessed April 17, 2025, https://www.daytrading.com/zulutrade.

Deng J, Yang M, Pelster M, Tan Y (2024) Social trading, communication, and networks. *Information Systems Research* 35(4):1546–1564.

Doering P, Neumann S, Paul S (2015) A primer on social trading networks–institutional aspects and empirical evidence. *EFMA annual meetings.*

Dong W, Saar-Tsechansky M, Geva T (2024) A machine learning framework for assessing experts' decision quality. *Management Science* Forthcoming.

Dowling M (2022) Is non-fungible token pricing driven by cryptocurrencies? *Finance Research Letters* 44:102097.

Du L, Hu M, Wu J (2022) Contingent stimulus in crowdfunding. *Production and Operations Management* 31(9):3543–3558.

Duan W, Gu B, Whinston AB (2009) Informational cascades and software adoption on the internet: an empirical investigation. *MIS quarterly* 23–48.

Dupas P (2014) Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica* 82(1):197–228.

Engström P, Forsell E (2018) Demand effects of consumers' stated and revealed preferences. *Journal of Economic Behavior & Organization* 150:43–61.

Fang F, Ventre C, Basios M, Kanthan L, Martinez-Rego D, Wu F, Li L (2022) Cryptocurrency trading: a comprehensive survey. *Financial Innovation* 8(1):1–59.

Farrell M, Green TC, Jame R, Markov S (2021) The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics* .

Flesch R (1979) *How to write plain English: A book for lawyers and consumers*, volume 76026225 (Harper & Row New York).

FM Contributors (2023) Exploring the surging interest of retail investors in cryptocurrency in 2023. URL `https://www.financemagnates.com/cryptocurrency/education-centre/explorin` `g-the-surging-interest-of-retail-investors-in-cryptocurrency-in-2023/`, [Online; accessed 25-Jan-2024].

Foley S, Karlsen JR, Putniņš TJ (2019) Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies* 32(5):1798–1853.

Fry J, Cheah ET (2016) Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis* 47:343–352.

Fujiki H (2020) Who adopts crypto assets in japan? evidence from the 2019 financial literacy survey. *Journal of the Japanese and International Economies* 58:101107.

Garg R, Smith MD, Telang R (2011) Measuring information diffusion in an online community. *Journal of Management Information Systems* 28(2):11–38.

Gemayel R, Preda A (2018) Does a scopic regime produce conformism? herding behavior among trade leaders on social trading platforms. *The European Journal of Finance* 24(14):1144–1175.

Gemayel R, Preda A (2021) Performance and learning in an ambiguous environment: A study of cryptocurrency traders. *International Review of Financial Analysis* 77:101847.

Gilbert B, Datta A, Ogburn E (2021) Approaches to spatial confounding in geostatistics. *arXiv preprint arXiv:2112.14946* .

Gillingham KT, Bollinger B (2021) Social learning and solar photovoltaic adoption. *Management Science* 67(11):7091–7112.

Goldsmith-Pinkham P, Imbens GW (2013) Social networks and the identification of peer effects. *Journal of Business & Economic Statistics* 31(3):253–264.

Goldstein I, Yang L (2017) Information disclosure in financial markets. *Annual Review of Financial Economics* 9:101–125.

Goodman LA (1961) Snowball sampling. *The Annals of Mathematical Statistics* 148–170.

Grable JE (2000) Financial risk tolerance and additional factors that affect risk taking in everyday money matters. *Journal of Business and Psychology* 14:625–630.

Granados N, Gupta A (2013) Transparency strategy: Competing with information in a digital world. *MIS Quarterly* 637–641.

Granados N, Gupta A, Kauffman RJ (2010) Research commentary—information transparency in business-to-consumer markets: concepts, framework, and research agenda. *Information Systems Research* 21(2):207–226.

Greene W, Han C, Schmidt P (2002) The bias of the fixed effects estimator in nonlinear models. *Unpublished Manuscript, Stern School of Business, NYU* 29.

Guégan D, Renault T (2021) Does investor sentiment on social media provide robust information for bitcoin returns predictability? *Finance Research Letters* 38:101494.

Gupta S, Gupta S, Mathew M, Sama HR (2021) Prioritizing intentions behind investment in cryptocurrency: a fuzzy analytical framework. *Journal of Economic Studies* 48(8):1442–1459.

Hackethal A, Hanspal T, Lammer DM, Rink K (2021) The Characteristics and Portfolio Behavior of Bitcoin Investors: Evidence from Indirect Cryptocurrency Investments. *Review of Finance* 26(4):855–898.

Halko ML, Kaustia M, Alanko E (2012) The gender effect in risky asset holdings. *Journal of Economic Behavior & Organization* 83(1):66–81.

Han B, Hirshleifer D, Walden J (2022) Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis* 57(1):390–412.

Han B, Liu H, Sui P (2023) Social learning and sentiment contagion in the bitcoin market. *Available at SSRN 4543326* .

Hanneman RA, Riddle M (2005) *Introduction to social network methods* (University of California Riverside).

Hao H, Padman R, Sun B, Telang R (2018) Quantifying the impact of social influence on the information technology implementation process by physicians: A hierarchical bayesian learning approach. *Information Systems Research* 29(1):25–41.

Harshman J (2021) Social trading vs copy trading. Accessed April 17, 2025, https://www.creditdonkey.com/social-trading-vs-copy-trading.html.

Hasan S, Bagde S (2013) The mechanics of social capital and academic performance in an indian college. *American Sociological Review* 78(6):1009–1032.

Hassanein A, Hussainey K (2015) Is forward-looking financial disclosure really informative? evidence from uk narrative statements. *International Review of Financial Analysis* 41:52–61.

Hasso T, Pelster M, Breitmayer B (2019) Who trades cryptocurrencies, how do they trade it, and how do they perform? evidence from brokerage accounts. *Journal of Behavioral and Experimental Finance* 23:64–74.

Havakhor T, Rahman MS, Zhang T, Zhu C (2024) Tech-enabled financial data access, retail investors, and gambling-like behavior in the stock market. *Management Science* .

Horrace WC, Oaxaca RL (2006) Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters* 90(3):321–327.

Howarth J (2024) How many cryptocurrencies are there in 2024? URL `https://explodingtopics.com/blog/number-of-cryptocurrencies`, [Online; accessed 25-Jan-2024].

Hu M, Milner J, Wu J (2016) Liking and following and the newsvendor: Operations and marketing policies under social influence. *Management Science* 62(3):867–879.

Hu MM, Yang S, Xu DY (2019) Understanding the social learning effect in contagious switching behavior. *Management Science* 65(10):4771–4794.

Huang H, Parker G, Tan YR, Xu H (2020) Altruism or shrewd business? implications of technology openness on innovations and competition. *MIS Quarterly* 44(3).

Huang N, Sun T, Chen P, Golden JM (2019) Word-of-mouth system implementation and customer conversion: A randomized field experiment. *Information Systems Research* 30(3):805–818.

Hutto C, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*, volume 8, 216–225.

Jabr W, Rai A (2022) When trading becomes social: How social trading platforms affect the disposition effect. *ICIS 2022 Proceedings* 16.

Jackson MO (2008) *Social and economic networks*, volume 3 (Princeton university press Princeton).

Jackson MO, Rogers BW, Zenou Y (2017) The economic consequences of social-network structure. *Journal of Economic Literature* 55(1):49–95.

Jiang M, Xia Y (2024) What drives the volatility of non-fungible tokens (nfts): macroeconomic fundamentals or investor attention? *Applied Economics Letters* 31(16):1439–1448.

Jiang Y, Ho YC, Yan X, Tan Y (2022) What's in a "username"? the effect of perceived anonymity on herding in crowdfunding. *Information Systems Research* 33(1):1–17.

Jin C, Lou B, Qiu L, Zhang J (2022) Toward understanding the bitcoin mining and exchange markets. *Available at SSRN 3776923* .

Jin X, Li R, Zhu Y (2021) Could social interaction reduce the disposition effect? evidence from retail investors in a directed social trading network. *Plos one* 16(2):e0246759.

Jonker N (2019) What drives the adoption of crypto-payments by online retailers? *Electronic Commerce Research and Applications* 35:100848.

Jung J, Bapna R, Golden JM, Sun T (2020) Words matter! toward a prosocial call-to-action for online referral: evidence from two field experiments. *Information Systems Research* 31(1):16–36.

Karlan D, Mobius M, Rosenblat T, Szeidl A (2009) Trust and social collateral. *The Quarterly Journal of Economics* 124(3):1307–1361.

Kawai D, Soska K, Routledge B, Zetlin-Jones A, Christin N (2024) Stranger danger? investor behavior and incentives on cryptocurrency copy-trading platforms. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–20.

KI Economics (2021) Social trading: benefits and risks. Accessed October 13, 2021, https://www.kieconomics.com/social-trading-benefits-and-risks/.

Kim K, Viswanathan S (2019) The 'experts' in the crowd: The role of experienced investors in a crowdfunding market. *MIS Quarterly* 43(2):347 – 372, ISSN 02767783.

Kincaid JP, Fishburne Jr RP, Rogers RL, Chissom BS (1975) *Derivation of new readability formulas for navy enlisted personnel* (Institute for Simulation and Training, University of Central Florida).

Kleibergen F, Paap R (2006) Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133(1):97–126.

Klocke N, Müller D, Hasso T, Pelster M (2022) The impact of peer performance on individual investors' risk-taking. *HKIMR Working Paper* 19/2022.

Knüpfer S, Rantapuska E, Sarvimäki M (2022) Social interaction in the family: Evidence from investors' security holdings. *Review of Finance* 27(4):1297–1327.

Koch A (2017) Herd behavior and mutual fund performance. *Management Science* 63(11):3849–3873.

Kogan S, Makarov I, Niessner M, Schoar A (2023) Are cryptos different? evidence from retail trading. *Available at SSRN 4289513* .

Korniotis GM, Kumar A (2011) Do older investors make better investment decisions? *The Review of Economics and Statistics* 93(1):244–265.

Kuchler T, Stroebel J (2021) Social finance. *Annual Review of Financial Economics* 13:37–55.

Kumar A (2009) Who gambles in the stock market? *The Journal of Finance* 64(4):1889–1933.

Kumar S, Mookerjee V, Shubham A (2018) Research in operations management and information systems interface. *Production and Operations Management* 27(11):1893–1905.

Lee GM, Qiu L, Whinston AB (2016) A friend like me: Modeling network formation in a location-based social network. *Journal of Management Information Systems* 33(4):1008–1033.

Lee H, Tang C, Yang SA, Zhang Y (2023) Dynamic trade finance in the presence of information frictions and fintech. *Manufacturing & Service Operations Management* 25(6):2038–2055.

Lee I, Shin YJ (2018) Fintech: Ecosystem, business models, investment decisions, and challenges. *Business Horizons* 61(1):35–46.

Lee W, Ma Q (2015) Whom to follow on social trading services? a system to support discovering expert traders. *2015 Tenth International Conference on Digital Information Management (ICDIM)*, 188–193 (IEEE).

Lee YJ, Hosanagar K, Tan Y (2015) Do i follow my friends or the crowd? information cascades in online movie ratings. *Management Science* 61(9):2241–2258.

Leskovec J, Faloutsos C (2006) Sampling from large graphs. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 631–636.

Lewbel A (2007) Endogenous selection or treatment model estimation. *Journal of Econometrics* 141(2):777–806.

Li T, Shin D, Wang B (2021) Cryptocurrency pump-and-dump schemes. *Available at SSRN 3267041* .

Li Z, Duan JA, Ransbotham S (2020) Coordination and dynamic promotion strategies in crowdfunding with network externalities. *Production and Operations Management* 29(4):1032–1049.

Liêu L, Pelster M (2020) Framing and the disposition effect in a scopic regime. *Quarterly Review of Economics Finance* 78:175–185.

Liu CW, Mithas S, Pan Y, Hsieh JPA (2024) Mobile apps, trading behaviors, and portfolio performance: Evidence from a quasi-experiment in china. *Information Systems Research* .

Liu X, Patacchini E, Zenou Y (2014) Endogenous peer effects: local aggregate or local average? *Journal of economic behavior & organization* 103:39–59.

Liu X, Zhang B, Anjana S, Rema P (2020) Go to youtube and call me in the morning: Use of social media for chronic conditions. *MIS Quarterly* 44(1):257–283.

Liu Y, Sheng J, Wang W (2023a) Technology and cryptocurrency valuation. *SSRN Working Paper* .

Liu Y, Tsyvinski A (2021) Risks and returns of cryptocurrency. *The Review of Financial Studies* 34(6):2689–2727.

Liu Y, Yang M, Tan Y (2023b) Coevolution of trader networks and follower dynamics in social trading. *Available at SSRN 4528456* .

Lomi A, Snijders TA, Steglich CE, Torló VJ (2011) Why are some more peer than others? evidence from a longitudinal study of social networks and individual academic performance. *Social Science Research* 40(6):1506–1520.

López-Cabarcos MÁ, Pérez-Pico AM, Piñeiro-Chousa J, Šević A (2021) Bitcoin volatility, stock market and investor sentiment. are they connected? *Finance Research Letters* 38:101399.

Lou B, Wu L (2021) Ai on drugs: Can artificial intelligence accelerate drug development? evidence from a large-scale examination of bio-pharma firms. *MIS Quarterly* 45(3).

Loughran T, McDonald B (2011) When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance* 66(1):35–65.

Lowry PB, Zhang J, Wang C, Siponen M (2016) Why do adults engage in cyberbullying on social media? an integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research* 27(4):962–986.

Lyandres E, Palazzo B, Rabetti D (2022) Initial coin offering (ico) success and post-ico performance. *Management Science* 68(12):8658–8679.

Malik N, Aseri M, Singh PV, Srinivasan K (2022) Why bitcoin will fail to scale? *Management Science* 68(10):7323–7349.

Manski CF (1993) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3):531–542.

Max Wei Y (2020) The similarity network of motion pictures. *Management Science* 66(4):1647–1671.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1):415–444.

Merkle C (2017) Financial overconfidence over time: Foresight, hindsight, and insight of investors. *Journal of Banking & Finance* 84:68 – 87.

Miller G, Mobarak AM (2015) Learning about new technologies through social networks: experimental evidence on nontraditional stoves in bangladesh. *Marketing Science* 34(4):480–499.

Moffitt RA (2001) 3 policy interventions, low-level equilibria, and social interactions. *Social Dynamics* 4:45.

Mukherjee UK, Ball GP, Wowak KD, Natarajan KV, Miller JW (2022) Hiding in the herd: The product recall clustering phenomenon. *Manufacturing & Service Operations Management* 24(1):392–410.

Muslu V, Radhakrishnan S, Subramanyam KR, Lim D (2015) Forward-looking md&a disclosures and the information environment. *Management Science* 61(5):931–948.

Nadini M, Alessandretti L, Di Giacinto F, Martino M, Aiello LM, Baronchelli A (2021) Mapping the nft revolution: market trends, trade networks, and visual features. *Scientific reports* 11(1):20902.

Nair HS, Manchanda P, Bhatia T (2010) Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research* 47(5):883–895.

Nakamoto S (2008) A peer-to-peer electronic cash system. *Bitcoin* 4(2):15.

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society* 1–32.

Nicolosi G, Peng L, Zhu N (2009) Do individual investors learn from their trading experience? *Journal of Financial Markets* 12(2):317–336.

Niezink NM, Snijders TA, van Duijn MA (2019) No longer discrete: Modeling the dynamics of social networks and continuous behavior. *Sociological Methodology* 49(1):295–340.

Nofsinger JR, Sias RW (1999) Herding and feedback trading by institutional and individual investors. *The Journal of Finance* 54(6):2263–2295.

Odean T (1998) Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance* 53(6):1887–1934.

Odean T (1999) Do investors trade too much? *The American Economic Review* 89(5).

Oh W, Jeon S (2007) Membership herding and network stability in the open source community: The ising perspective. *Management Science* 53(7):1086–1101.

Pan W, Altschuler Y, Pentland AS (2012) Decoding social influence and the wisdom of the crowd in financial trading network. IEEE, ed., *International Conference on Privacy, Security, Risk and Trust and International Conference on Social Computing*, 203–209 (Washington DC).

Pan Y, Mithas S, Po-An Hsieh J, Liu CW (2023) Do risk preferences shape the effect of online trading on trading frequency, volume, and portfolio performance? *Journal of Management Information Systems* 40(2):440–469.

Park A, Sabourian H (2011) Herding and contrarian behavior in financial markets. *Econometrica* 79(4):973–1026.

Patacchini E, Zenou Y (2008) The strength of weak ties in crime. *European Economic Review* 52(2):209–236.

Pelster M (2017) I'll have what s/he's having: A case study of a social trading network. *Proceedings of the International Conference on Information Systems 2017* .

Pelster M, Breitmayer B, Hasso T (2019) Are cryptocurrency traders pioneers or just risk-seekers? evidence from brokerage accounts. *Economics Letters* 182:98–100.

Pelster M, Hofmann A (2018) About the fear of reputational loss: Social trading and the disposition effect. *Journal of Banking & Finance* 94:75–88.

Provost F, Martens D, Murray A (2015) Finding similar mobile consumers with a privacy-friendly geosocial design. *Information Systems Research* 26(2):243–265.

Pu J, Liu Y, Chen Y, Qiu L, Cheng HK (2022) What questions are you inclined to answer? effects of hierarchy in corporate q&a communities. *Information Systems Research* 33(1):244–264.

Qiu L, Chhikara A, Vakharia A (2021) Multidimensional observational learning in social networks: Theory and experimental evidence. *Information Systems Research* 32(3):876–894.

Rhue L, Sundararajan A (2019) Playing to the crowd? digital visibility and the social dynamics of purchase disclosure. *MIS Quarterly* 43(4):1127–1142.

Ripley R, Snijders TA, Boda Z, Vörös A, Preciado P (2021) Manual for Siena version 4.0. Technical report, r package version 1.3.0. https://www.cran.r-project.org/web/packages/RSiena/.

Robbins H, Monro S (1951) A stochastic approximation method. *The Annals of Mathematical Statistics* 400–407.

Röder F, Walter A (2019) What drives investment flows into social trading portfolios? *Journal of Financial Research* 42(2):383–411.

Ryan SP, Tucker C (2012) Heterogeneity and the dynamics of technology adoption. *Quantitative Marketing and Economics* 10:63–109.

Sakakibara T, Matsui T, Mutoh A, Inuzuka N (2015) Clustering mutual funds based on investment similarity. *Procedia Computer Science* 60:881–890.

Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. *Communications of the ACM* 18(11):613–620.

Sanderson E, Windmeijer F (2016) A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics* 190(2):212–221.

Satornino CB, Allen A, Shi H, Bolander W (2023) Understanding the performance effects of "dark" salesperson traits: Machiavellianism, narcissism, and psychopathy. *Journal of Marketing* 87(2):298–318.

Scharfstein DS, Stein JC (1990) Herd behavior and investment. *The American Economic Review* 465–479.

Schweinberger M, Snijders TA (2007) Markov models for digraph panel data: Monte carlo-based derivative estimation. *Computational Statistics & Data Analysis* 51(9):4465–4483.

Seru A, Shumway T, Stoffman N (2010) Learning by trading. *The Review of Financial Studies* 23(2):705–739.

Shen B, Dong C, Minner S (2021) Combating copycats in the supply chain with permissioned blockchain technology. *Production and Operations Management* .

Shen Z, Zheng E, Jiang W (2022) A machine learning approach to mitigating irrationality in copy trading. *Available at SSRN 4298755* .

Shome A (2021) Exclusive: etoro ended 2020 with $600 million in revenue. Accessed April 17, 2025, https://www.financemagnates.com/forex/brokers/exclusive-etoro-ended-2020-with-600-million-in-revenue/.

Simonsohn U, Ariely D (2008) When rational sellers face nonrational buyers: evidence from herding on ebay. *Management Science* 54(9):1624–1637.

Singh A, Hosanagar K, Gandhi A (2020) Machine learning instrument variables for causal inference. *Proceedings of the 21st ACM Conference on Economics and Computation*, 835–836.

Singh PV, Tan Y, Youn N (2011) A hidden markov model of developer learning dynamics in open source software projects. *Information Systems Research* 22(4):790–807.

Snijders T, Steglich C, Schweinberger M (2017) Modeling the coevolution of networks and behavior. *Longitudinal models in the behavioral and related sciences*, 41–71 (Routledge).

Snijders TA (2001) The statistical evaluation of social network dynamics. *Sociological Methodology* 31(1):361–395.

Snijders TA, Lomi A, Torló VJ (2013) A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks* 35(2):265–276.

Stammann A, Heiss F, McFadden D (2016) Estimating fixed effects logit models with large panel data .

Steinberg L, Monahan KC (2007) Age differences in resistance to peer influence. *Developmental Psychology* 43(6):1531.

Sun H (2013) A longitudinal study of herd behavior in the adoption and continued use of technology. *MIS Quarterly* 1013–1041.

Sun H, Fan M, Tan Y (2020) An empirical analysis of seller advertising strategies in an online marketplace. *Information Systems Research* 31(1):37–56.

Sun T, Viswanathan S, Zheleva E (2021) Creating social contagion through firm-mediated message design: Evidence from a randomized field experiment. *Management Science* 67(2):808–827.

Sweet K (2024) New funds will make investing in bitcoin easier. here's what you need to know. URL `https://apnews.com/article/bitcoin-exchange-traded-funds-etf-sec-59a5bb81ab891af57a1bd1765024144f`, [Online; accessed 3-Mar-2024].

Te'eni D, Yahav I, Zagalsky A, Schwartz D, Silverman G, Cohen D, Mann Y, Lewinsky D (2023) Reciprocal human-machine learning: A theory and an instantiation for the case of message classification. *Management Science* Forthcoming.

The European Business Review (2020) Forex social trading pros cons – an expert guide. Accessed April 17, 2025, https://www.europeanbusinessreview.com/forex-social-trading-pros-cons-an-expert-guide/.

Train KE (2009) *Discrete choice methods with simulation* (Cambridge university press).

Trogdon JG, Nonnemaker J, Pais J (2008) Peer effects in adolescent overweight. *Journal of health economics* 27(5):1388–1399.

Trueman B (1994) Analyst forecasts and herding behavior. *The Review of Financial Studies* 7(1):97–124.

Veeraraghavan S, Debo L (2009) Joining longer queues: Information externalities in queue choice. *Manufacturing & Service Operations Management* 11(4):543–562.

Veeraraghavan SK, Debo LG (2011) Herding in queues with waiting costs: Rationality and regret. *Manufacturing & Service Operations Management* 13(3):329–346.

Verbeek M, Wang Y (2013) Better than the original? the relative success of copycat funds. *Journal of Banking & Finance* 37(9):3454–3471.

Wang M, Hussainey K (2013) Voluntary forward-looking statements driven by corporate governance and their value relevance. *Journal of Accounting and Public Policy* 32(3):26–49.

Wang Q, Li B, Singh PV (2018) Copycats vs. original mobile apps: A machine learning copycat-detection method and empirical analysis. *Information Systems Research* 29(2):273–291.

Wang X, Jiang M, Han W, Qiu L (2022) Do emotions sell? the impact of emotional expressions on sales in the space-sharing economy. *Production and Operations Management* 31(1):65–82.

Wang X, Wu L, Hitt LM (2024) Social media alleviates venture capital funding inequality for women and less connected entrepreneurs. *Management Science* 70(2):1093–1112.

Wang X, Xu F (2023) The value of smart contract in trade finance. *Manufacturing & Service Operations Management* 25(6):2056–2073.

Wardman M (1988) A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy* 71–91.

Weber M, Candia B, Coibion O, Gorodnichenko Y (2023) Do you even crypto, bro? cryptocurrencies in household finance. *Chicago Booth Working Paper* 23-10.

Wei KD, Wermers R, Yao T (2015) Uncommon value: The characteristics and investment performance of contrarian funds. *Management Science* 61(10):2394–2414.

Wei X, Fan M, You W, Tan Y (2021) An empirical study of the dynamic and differential effects of prefunding. *Production and Operations Management* 30(5):1331–1349.

Wei Y, Dukes A (2021) Cryptocurrency adoption with speculative price bubbles. *Marketing Science* 40(2):241–260.

Wei Yong J, Tellis GJ (2022) Machine learning for creativity: Using similarity networks to design better crowdfunding projects. *Journal of Marketing* 86(2):87–104.

Wu CA, Jin C, Chen YJ (2021) Strategic inventory, network externality and copycats. *Working Paper* Available at SSRN: https://ssrn.com/abstract=3826955.

Wu L, Jin F, Hitt LM (2018) Are all spillovers created equal? a network perspective on information technology labor movements. *Management Science* 64(7):3168–3186.

Wu L, Kane GC (2021) Network-biased technical change: How modern digital collaboration tools overcome some biases but exacerbate others. *Organization Science* 32(2):273–292.

Xiao S, Ho YC, Che H (2021) Building the momentum: Information disclosure and herding in online crowdfunding. *Production and Operations Management* 30(9):3213–3230.

Xitong L, Lynn W (2018) Herding and social media word-of-mouth: Evidence from groupon. *MIS Quarterly* 42(4):1331 – 1351, ISSN 02767783.

Xue L, Song P, Rai A, Zhang C, Zhao X (2019) Implications of application programming interfaces for third-party new app development and copycatting. *Production and Operations Management* 28(8):1887–1902.

Yan L, Peng J, Tan Y (2015) Network dynamics: how can we find patients like us? *Information Systems Research* 26(3):496–512.

Yang M, Zheng Z, Mookerjee V (2022) How much is financial advice worth? the transparency-revenue tension in social trading. *Management Science* 68(7):5252–5268.

Yoo E, Gu B, Rabinovich E (2019) Diffusion on social media platforms: A point process model for interaction among similar content. *Journal of Management Information Systems* 36(4):1105–1141.

Young HP (2009) Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review* 99(5):1899–1924.

Zaheer M, Kottur S, Ravanbakhsh S, Poczos B, Salakhutdinov RR, Smola AJ (2017) Deep sets. *Advances in neural information processing systems* 30.

Zeng X, Wei L (2013) Social ties and user content generation: Evidence from flickr. *Information Systems Research* 24(1):71–87.

Zhang J (2010) The sound of silence: Observational learning in the us kidney market. *Marketing Science* 29(2):315–335.

Zhang J, Liu P (2012) Rational herding in microloan markets. *Management Science* 58(5):892–912.

Zhang J, Liu Y, Chen Y (2015) Social learning in networks of friends versus strangers. *Marketing Science* 34(4):573–589.

Zhang J, Savin S, Veeraraghavan S (2023) Revenue management in crowdfunding. *Manufacturing & Service Operations Management* 25(1):200–217.

Zhang Y, Godes D (2018) Learning from online social ties. *Marketing Science* 37(3):425–444.

Zheludev I, Smith R, Aste T (2014) When can social media lead financial markets? *Scientific Reports* 4(1):4213.

Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M (2020) Graph neural networks: A review of methods and applications. *AI open* 1:57–81.

Zhou Z, Liu Y, Ma C, Tan Y (2024) Effects of scarcity-induced demand on crowdfunding market. *Production and Operations Management* Forthcoming.

Zhu K (2004) Information transparency of business-to-business electronic markets: A game-theoretic analysis. *Management Science* 50(5):670–685.

Appendix A
ONLINE APPENDIX TO CHAPTER 3

A.1 A broader discussion on adopting similar trading strategies in traditional financial markets

Traders in traditional financial markets exhibit similar trading behavior, often characterized by groups of investors collectively entering or exiting a specific industry or stock over a defined period (Nofsinger and Sias 1999, Choi and Sias 2009, Koch 2017, Brown et al. 2014). This behavior is referred to as institutional herding. Interestingly, studies indicate that sharing a similar trading strategy as others is generally not a successful strategy. There are several explanations for the underperformance. First, when a large portion of traders follow a similar strategy, it may lead to market inefficiencies and bubbles due to irrational exuberance or panic selling, which could distort prices away from their fundamental values (Nofsinger and Sias 1999, Choi and Sias 2009, Park and Sabourian 2011). When the market eventually corrects these inefficiencies, these investors may experience significant losses. Second, adopting similar trading strategies results in a lack of diversification, which can increase risk. With many traders making the same investments, the impact of an adverse event affecting those investments is likely to be greater (Koch 2017). Last, traders who follow common strategies may miss out on unique opportunities that offer higher returns. By pursuing less common strategies, traders can uncover overlooked or undervalued investment options and achieve superior performance (Wei et al. 2015).

A.2 Additional data description

Our data includes a well-diversified mix of users from various countries. The total number of countries in our dataset is 108. Table A.1 presents the top 10 countries represented in our dataset. This geographic diversity supports the generalizability of our results across different countries, and our empirical findings are also applicable to the United States.

Table A.1: Distribution of Traders Across the Top 10 Countries

Country	Percentage of Traders
Indonesia	7.65%
China	6.24%
Russia	6.12%
United States	3.84%
Germany	3.21%
France	2.40%
Egypt	2.37%
United Kingdom	2.25%
India	2.07%
Italy	1.92%

Table A.2 summarizes the top 10 most frequently traded currency pairs in our dataset. We observe a diverse mix of bullish, bearish, and mixed market conditions in our dataset, which supports the generalizability of our findings across different market environments.

A.3 Justification of our proposed algorithm on the trader similarity network

To illustrate how we measure similarity, consider two traders at a specific period. Their trading activities are summarized in Table A.3.

Table A.2: Top 10 Most Frequently Traded Instruments with Market Trends

Currency Pair	Percentage of Trades	Trend
EUR/USD	21.84%	Bullish
GBP/USD	12.98%	Mixed
USD/JPY	8.66%	Mixed
GBP/JPY	7.72%	Mixed
EUR/JPY	4.60%	Bullish
USD/CAD	4.23%	Bearish
AUD/USD	3.64%	Bullish
GBP/AUD	3.49%	Bearish
EUR/GBP	2.83%	Bullish
XAU/USD	2.43%	Mixed

Notes: Bullish — The average monthly closing price consistently increased over the observation period. Bearish — The average monthly closing price consistently decreased. Mixed — No consistent trend in average monthly closing prices.

Table A.3: Illustrative Example for Similarity Measurement

	(Currency, Buy/Sell, Open/Close) (k)	Number of Trades (N_{ikt})	Average Price ($AvgPrice_{ikt}$)	Average Lot Size ($AvgVolume_{ikt}$)
Trader A	(USD/EUR, Buy, Open)	1	0.92	1
	(USD/JPY, Buy, Open)	1	130	1
Trader B	(USD/EUR, Buy, Open)	5	0.89	100
	(USD/JPY, Buy, Open)	5	135	100

Below, we detail the calculation of our similarity score step by step.

- Step 1 (Raw binary vector): Note that these two traders have the same traded currency and types, thus their vector representations are also the same. Moreover, since the zero-valued element does not affect the calculation, we only consider a shortened vector with non-zero values in the example, without loss of generalizability. Thus, both trader A and trader B have the vector of $[1, 1]$, where the first element indicates whether a trader trades (USD/EUR, Buy, Open) and the second element indicates (USD/JPY, Buy, Open). That is, $U_{it} = U_{jt} = [1, 1]$.

- Step 2 (Adjustment for trading frequency, price differences, and volume):

  – For $k =$ (USD/EUR, Buy, Open)

$$FreqSIM_{ijkt} = 1 - \frac{|N_{ikt} - N_{jkt}|}{max(N_{ikt}, N_{jkt})} = 1 - \frac{|1 - 5|}{max(1, 5)} = 0.2$$

$$PriceSIM_{ijkt} = 1 - \frac{|AvgPrice_{ikt} - AvgPrice_{jkt}|}{max(AvgPrice_{ikt}, AvgPrice_{jkt})} == 1 - \frac{|130 - 135|}{max(130, 135)} \approx 0.96$$

$$VolumeSIM_{ijkt} = 1 - \frac{|AvgVolume_{ikt} - AvgVolume_{jkt}|}{max(AvgVolume_{ikt}, AvgVolume_{jkt})} = 1 - \frac{|1 - 100|}{max(1, 100)} = 0.01$$

– For $k =$ (USD/JPY, Buy, Open)

$$FreqSIM_{ijkt} = 1 - \frac{|N_{ikt} - N_{jkt}|}{max(N_{ikt}, N_{jkt})} = 1 - \frac{|1-5|}{max(1,5)} = 0.2$$

$$PriceSIM_{ijkt} = 1 - \frac{|AvgPrice_{ikt} - AvgPrice_{jkt}|}{max(AvgPrice_{ikt}, AvgPrice_{jkt})} == 1 - \frac{|0.89 - 0.92|}{max(0.89, 0.92)} \approx 0.97$$

$$VolumeSIM_{ijkt} = 1 - \frac{|AvgVolume_{ikt} - AvgVolume_{jkt}|}{max(AvgVolume_{ikt}, AvgVolume_{jkt})} = 1 - \frac{|1-100|}{max(1,100)} = 0.01$$

- Step 3 (Calculation of similarity score):

$$\begin{aligned} S_{ijt} &= \frac{\sum_k U_{ikt} U_{jkt} FreqSIM_{ijkt} PriceSIM_{ijkt} VolumeSIM_{ijkt}}{\sqrt{\sum_k U_{ikt}^2 \sum_k U_{jkt}^2}} \\ &= \frac{1 \times 0.2 \times 0.96 \times 0.01 + 1 \times 0.2 \times 0.97 \times 0.01}{\sqrt{(1+1) \times (1+1)}} = 0.00193 \end{aligned}$$

Next, we consider a weighted-vector approach to calculate the cosine similarity. To construct the weighted vector for each trader, we assign a value to each element (each combination of Currency, Buy/Sell, and Open/Close) based on $N_{ikt} \times AvgPrice_{ikt} \times AvgVolume_{ikt}$. Thus, we derive a vector of [0.92, 1] for trader A and [485, 67500] for trader B. Though exhibiting substantial differences in the magnitude of the vector, the cosine similarity calculated by these two vectors is 0.99.

We also use the weighted-vector approach to compute the cosine similarity score for each trader pair in our dataset. For comparison, we computed similarity scores using the raw binary vector (without incorporating the trading frequency, price differences, and volume) as the benchmark. The similarity scores derived from the weighted vector approach and the benchmark show a high correlation coefficient of 0.98, providing additional evidence that the weighted vector does not meaningfully capture the influence of trading frequency, price, or volume. This finding further supports our current approach that adjusts for differences in trading frequency, price, and volume, rather than a weighted vector approach.

A.4 Human evaluation (label) form

We present the instruction and evaluation (label) form distributed to human participants — specifically, undergraduate students majoring in finance — who were recruited to assess the similarity between the trading strategies of two traders.

Instructions for Labeling Similar Trader Pairs

Background and Purpose

Social trading platforms (e.g., eToro, Zulutrade) combine trading platforms with social features. On those platforms, all trading history of a trader is publicly disclosed. This allows us to identify similar trading behaviors among traders and investigate the impact of such similarities.

To begin, we need training data indicating the ground truth of whether two traders' trades are similar, which requires human labeling.

Input

You will receive multiple CSV files, each containing the trading records of two traders. The format is as follows:

- Two traders' trading records are separated by two empty rows.
- The filename format is: `week_Traderid1_Traderid2.csv`

Each CSV file contains the following columns:

- **trader_id**: A unique identifier for a trader.
- **date**: The date and time of the transaction.
- **currency**: The financial instrument being traded.
- **type**: `Buy` or `Sell` a trade.
- **price**: The instrument's price at the transaction time.
- **action**: `Open` or `Close` a trade.
- **volume**: The lot size of a trade.

Instructions for Labeling Similar Trader Pairs

Output

For each CSV file, you should assign a label to the two traders and indicate whether you think the two traders are similar. Similarity is judged based on trading the same currency/type/action with comparable frequency within a similar time frame. After evaluation, please record your label in the 4th column of `Label_master.csv`.

An example of a similar trader pair

trader_id	date	currency	type	action	volume	price
149607	2017-06-09 15:22:49	EUR/USD	SELL	close	100000.0	1.11916
149607	2017-06-09 17:01:39	EUR/USD	SELL	close	100000.0	1.12016
149607	2017-06-09 17:19:06	XAU/USD	SELL	close	100.0	1269.90700

trader_id	date	currency	type	action	volume	price
144471	2017-06-09 15:22:49	EUR/USD	SELL	close	100000.0	1.11913
144471	2017-06-09 16:55:32	EUR/USD	SELL	close	100000.0	1.12013
144471	2017-06-09 17:22:58	XAU/USD	SELL	close	100.0	1270.11600

A.5 Sampling strategy

Our snowball sampling procedure is designed to include both isolated and connected traders (non-isolated nodes). We begin by randomly selecting initial seed nodes from the complete set of 3,332 traders. These seeds (nodes) may be either isolated or non-isolated. For each non-isolated seed, we include all of its first-tier, second-tier, and third-tier neighbors (nodes) in the sub-sample, thereby expanding the network coverage. In contrast, isolated seeds do not have any connections (links), so no additional nodes are added to the sub-sample in those cases. As a result, our sub-sample includes both isolated and non-isolated traders.

In Table A.4 and Table A.5, we report the network statistics for our sub-sample and the complete data. *Average Degree* captures the average number of links each non-isolated trader has in the network. *Number of Clusters* denotes the number of connected components in the network. *Density* provides a global perspective of the network's overall interconnection level. Specifically, it calculates the ratio of the actual number of links to the possible number of links among all non-isolated nodes. Based on these statistics, we see some variation in the similarity network structure, reflecting the dynamic nature of the trading activity.

By comparing the network statistics in Table A.4 and Table A.5, we observe that while our sub-sample is in general similar to the complete dataset, it is relatively denser than the complete dataset. This is expected due to the nature of snowball sampling, which tends to oversample more connected nodes for sample expansion.

Table A.4: Network Statistics of the Sub-sample

Period	Number of Nodes	Percentage of Isolated Nodes	Number of Links	Average Degree	Number of Clusters	Number of Nodes per Cluster	Density
1	387	89%	171	8.14	8	5.25	0.40
2	387	90%	113	5.94	10	3.80	0.32
3	386	88%	133	5.54	11	4.36	0.24

Notes: Percentage of Isolated Nodes = $\frac{\text{Number of Isolated Nodes}}{\text{Total Number of Nodes}}$;
Average Degree = $\frac{2\times\text{Number of Links}}{\text{Number of Non-isolated Nodes}}$;
Number of Nodes per Cluster = $\frac{\text{Number of Non-Isolated Nodes}}{\text{Number of Clusters}}$;
Density = $\frac{2\times\text{Number of Links}}{\text{Number of Non-isolated Nodes}\times(\text{Number of Non-isolated Nodes}-1)}$.

Table A.5: Network Statistics of the Complete Data

Period	Number of Nodes	Percentage of Isolated Nodes	Number of Links	Average Degree	Number of Clusters	Number of Nodes per Cluster	Density
1	3332	96%	321	5.06	35	3.63	0.08
2	3332	97%	152	3.42	31	2.87	0.08
3	3332	96%	316	4.75	39	3.41	0.07

Notes: Percentage of Isolated Nodes = $\frac{\text{Number of Isolated Nodes}}{\text{Total Number of Nodes}}$;
Average Degree = $\frac{2\times\text{Number of Links}}{\text{Number of Non-isolated Nodes}}$;
Number of Nodes per Cluster = $\frac{\text{Number of Non-Isolated Nodes}}{\text{Number of Clusters}}$;
Density = $\frac{2\times\text{Number of Links}}{\text{Number of Non-isolated Nodes}\times(\text{Number of Non-isolated Nodes}-1)}$.

A.6 A simple numerical illustration of network evolution

We provide a detailed numerical illustration of the network evolution model. We begin by describing a very small demo dataset, including a similarity network and the values of node-level variables at time (period) 0. Next, we demonstrate how one network link change is determined using our model. To make this process clear, we first calculate each variable (or network effect), and then compare the utilities of all possible network changes.

Consider a simple similarity network consisting of four nodes (A, B, C, and D), where only A and B are linked, while C and D are isolated, as illustrated in Figure A.1. This means that traders A and B are similar peers. The key attribute of interest is the number of followers. Without loss of generalizability, we include only one control variable, $Volume$. The values for these variables at the initial period, along with each node's SimilarPeers, are presented in Table A.6.

Figure A.1: Similarity Network Illustration

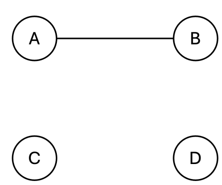


Table A.6: Summary of Node Attributes in the Initial Period

	A	B	C	D
The Number of Followers	1	2	3	4
Volume	0.25	0.75	1	0.25
SimilarPeers	1	1	0	0

Without loss of generalizability, following Section 3.5.1, we consider four effects that influence the network changes: the Degree effect, the Follower Alter effect, the Follower Homophily effect, and the Volume Homophily effect.

The network evolution process involves two steps: 1) selecting a node to initiate a network change, and 2) determining which link change (either an addition or a removal) the selected node will make. We next assume $\lambda_i = \lambda_m = 1$, which means any trader i's waiting time until the next opportunity to change a link is exponentially distributed with a rate of 1. Suppose the randomly generated waiting times at time 0 for nodes A through D are 0.1, 0.2, 0.3, 0.4, respectively. Then, at the next time $t = 0.1$, node A is selected to make a network link change.

Next, we consider all possible network changes involving node A, given the current state of the network: removing the existing link with B (denoted as AB), or adding a new link to C (denoted as AC) or D (denoted as AD). We omit the time subscript t since the example considers a single period. The utility of a link change, defined in Equation (3.8), is simplified in this example and given below,

$$f_A(x_{Aj}, z, \vec{c}) = \beta_1 Degree_A + \beta_2 FollowerAlter_A + \beta_3 FollowerHomophily_A + \beta_4 VolumeHomophily_A.$$

Note that we replace i with A as node A is selected to change the network (i.e., $i = A$). Following the definitions of each effect in Section 3.5.1, we first compute the values of these effects for each of A's potential network changes, shown in rows 1 to 4 of Table A.7. We then present the utility and resulting probability of each potential change in the last two rows of Table A.7, assuming $\beta_{1\text{-}4} = 1$.

For the degree effect, the degree value resulting from any link addition is the current degree of the node plus one, while for any link removal, it is the current degree minus one. This implies that a positive coefficient (β_1) indicates that nodes with higher degrees are more likely to form new network links, regardless of whom they connect to. Conversely, a negative coefficient suggests that nodes with higher degrees are less likely to initiate new links.

For the follower alter effect, since node D has more followers than node C, the potential

Table A.7: Illustration of Step-by-Step Calculations

	Formula	Potential network change		
		Remove AB ($x_{AB}=0$ $x_{AC}=0$ $x_{AD}=0$)	Add AC ($x_{AB}=1$ $x_{AC}=1$ $x_{AD}=0$)	Add AD ($x_{AB}=1$ $x_{AC}=0$ $x_{AD}=1$)
Degree	$\sum_j x_{Aj}$	0	2	2
Follower Alter	$\sum_j x_{Aj} Follower_j$	0	5	6
Follower Homophily	$\sum_j x_{Aj}\left(1-\frac{\lvert Follower_j - Follower_A\rvert}{Range(Follower)}\right)$	0	$\frac{2}{3}$	0
Volume Homophily	$\sum_j x_{Aj}\left(1-\frac{\lvert Volume_j - Volume_A\rvert}{Range(Volume)}\right)$	0	$\frac{2}{3}$	1
Utility	$f_A(x_{Aj}, z, \vec{c}) = \beta_1 Degree_{A,Aj} + \beta_2 FollowerAlter_{A,Aj} + \beta_3 FollowerHomophily_{A,Aj} + \beta_4 VolumeHomophily_{A,Aj}$	0	$\frac{25}{3}$	9
Probability	$p(x^{(\pm Aj)}) = \frac{exp^{f_A(x_{Aj}, z, \vec{c})}}{\sum_{h=1}^{n} exp^{f_A(x_{Ah}, z, \vec{c})}}$	0	0.34	0.66

link change "Add AD" has a higher follower alter value than "Add AC", which contributes to a higher utility for forming the link with D. Importantly, this higher utility is not related to any characteristics of node A. For example, if we compare the follower alter effect of adding link BC versus adding link BD, we observe the same pattern: "Add BD" has a higher follower alter value due to D's greater number of followers. Therefore, a positive (negative) coefficient on the follower alter effect suggests that nodes with a bigger number of followers are more (less) likely to form links — or, more precisely, are more likely to be selected as targets of new links.

The follower and volume homophily effects capture the proximity between two nodes regarding specific characteristics. For example, node A and node C have closer values in the number of followers, so the link change "Add AC" has a higher value for the follower homophily effect. In contrast, node A and node D have closer values in volume, leading to a higher volume homophily value for "Add AD". Therefore, a positive (negative) coefficient on these homophily effects suggests that nodes with more similar characteristics are more (less) likely to form a network link.

A.7 Summary of notations of parameters and variables

Table A.8, A.9, and A.10 summarize all the notations of parameters and variables in our analysis.

Table A.8: Summary of Parameters and Variables (Part 1)

Symbol	Description
Similarity Network	
U_{it}	A vector for trader i at period t (with the dimension d)
d	Number of unique combinations of financial instrument and trade type (with the value 380)
k	The index of element (combination) in the vector
U_{ikt}	k-th element of vector U_{it} for trader i at period t
N_{ikt}	Number of trades for the k-th combination by trader i at period t
$AvgPrice_{ikt}$	Average transaction price for the k-th combination by trader i at period t
$AvgVolume_{ikt}$	Average lot size for the k-th combination by trader i at period t
$FreqSIM_{ijkt}$	Frequency similarity adjustment between traders i and j for the k-th combination at period t
$PriceSIM_{ijkt}$	Price similarity adjustment between traders i and j for the k-th combination at period t
$VolumeSIM_{ijkt}$	Volume similarity adjustment between traders i and j at period t
S_{ijt}	Similarity score between traders i and j at period t

Table A.9: Summary of Parameters and Variables (Part 2)

Symbol	Description
Network Evolution	
$\mathcal{N}$	A set of node (trader) in the network, denoted by $\{1, 2, ..., n\}$.
X_t	$n \times n$ adjacency matrix representing the similarity network at time t
z_t	$\{z_{i,t}\}_{i\in\mathcal{N}}$, Key attribute value (i.e., number of followers) of all traders in the similarity network at time t
$\vec{c}_t$	$\{\vec{c}_{i,t}\}_{i\in\mathcal{N}}$, Control variables of all traders in the similarity network at time t
$x_{ij,t}$	Binary link (tie) variable between traders i and j (with a value of 1 if a link exists)
λ_i	Network change rate of trader i
λ_m	Network change rate in period m
$x^{(\pm ij,t)}$	Adjacency matrix after changing link $x_{ij,t}$ to $1 - x_{ij,t}$ at time t
$p(x^{(\pm ij,t)})$	Probability of trader i selecting to change link with trader j at time t
$f_{i,t}(x_{ij,t}, z_t, \vec{c}_t)$	Utility of trader i for changing link $x_{ij,t}$ to $1 - x_{ij,t}$ at time t
$s_{il,t}$	The l-th network effect in the utility of trader i at time t
β_l	Coefficient for the l-th network effect
$Degree_{i,t}$	Number of links of trader i in the similarity network at time t
$FollowerAlter_{i,t}$	Total number of followers of traders who have a link with trader i at time t
$FollowerHomophily_{i,t}$	Homophily in number of followers of trader i at time t
$WinRatioHomophily_{i,t}$	Homophily in trading performance (the percentage of profitable trades out of the total trades) of trader i at time t
$TradesHomophily_{i,t}$	Homophily in trading frequency (number of trades) of trader i at time t
$ViewsHomophily_{i,t}$	Homophily in social attention (number of profile clicks) of trader i at time t
$VolumeHomophily_{i,t}$	Homophily in trading volume (lot size) of trader i at time t
$ExperienceHomophily_{i,t}$	Homophily in trading experience (number of months) on the platform of trader i at time t
$RankHomophily_{i,t}$	Homophily in the rank of trader i at time t

Table A.10: Summary of Parameters and Variables (Part 3)

Symbol	Description
Attribute Evolution	
$Z_{i,t}$	Number of followers of trader i at time t
$dZ_{i,t}$	Incremental change in $Z_{i,t}$ from time t to $t + dt$
$dW_{i,t}$	Increment of Wiener process from time t to $t + dt$
τ_m	Period-specific attribute evolution rate
a, $b_{0\text{-}2}$, $c_{1\text{-}6}$	Coefficients in attribute evolution equation
$SimilarPeers_{i,t}$	Number of links (degree) of trader i at time t
$WinRatio_{i,t}$	Percentage of profitable trades out of the total trades made by trader i at time t
$Trades_{i,t}$	Number of trades made by trader i at time t
$Views_{i,t}$	Social attention (number of profile clicks) of trader i at time t
$Volume_{i,t}$	Trading volume (lot size) of trader i at time t
$Experience_{i,t}$	Trading experience (number of months) on the platform of trader i at time t
$Rank_{i,t}$	The rank of trader i at time t by the platform

A.8 Estimation of the coevolution model

We present how to estimate the parameters in the coevolution model. The estimation framework and algorithms were proposed by Snijders et al. (2017) and Niezink et al. (2019).

The coevolution model is presented in Equations (A.1) and (A.2).

$$p(x^{(\pm ij,t)}) = \frac{exp^{f_{i,t}(x_{ij,t},z_t,\vec{c}_t)}}{\sum_{h=1}^{n} exp^{f_{i,t}(x_{ih,t},z_t,\vec{c}_t)}}, \tag{A.1}$$

where $f_{i,t}(x_{ij,t}, z_t, \vec{c}_t) = \sum_l \beta_l s_{il}$.

$$\begin{aligned} dZ_{i,t} =& \tau_m(aZ_{i,t} + b_0 + b_1 SimilarPeers_{i,t} + c_1 WinRatio_{i,t} \\ & + c_2 Trades_{i,t} + c_3 Views_{i,t} + c_4 Volume_{i,t} + c_5 Experience_{i,t} + c_6 Rank_{i,t})dt + \sqrt{\tau_m} dW_{i,t} \end{aligned} \tag{A.2}$$

From Equations (A.1) and (A.2), we shall estimate 1) period-specific parameters, λ_m (the speed of network evolution) and τ_m (the speed of attribute evolution); 2) β_l (parameters in the network dynamics); 3) a, b_0, b_1, c_1, c_2, c_3, c_4 and c_5 (parameters in the attribute dynamics).

A.8.1 Method of moments

The likelihood function for the coevolution model cannot be computed explicitly, which makes maximum likelihood or Bayesian estimation difficult, and hence, the method of moments (MoM) is adopted to estimate the coevolution model (Snijders 2001, Snijders et al. 2017). According to MoM (Bowman and Shenton 1985), parameter estimates are the values for which the expected data given the parameters and the observed data are most similar.

For a general statistical model with data Y and parameters θ, the MoM estimator based on the statistic $u(Y)$ is defined as the parameter value $\hat{\theta}$ for which the expected and observed values of $u(Y)$ are the same,

$$\mathbb{E}_{\hat{\theta}}(u(Y)) = u(y) \tag{A.3}$$

where $u(y)$ is the observed value, and θ and $u(Y)$ are vectors with the same dimensions. Equation (A.3) is called the moment equation.

In panel data, observations on the stochastic process are available at several discrete time moments. Together with the Markov property, the adapted MoM estimator for the coevolution model relies on the conditional expectation. Therefore, the moment equation for the period-specific parameters is

$$\mathbb{E}_\theta[u_m\left(Y(t_{m+1}), Y(t_m)\right)|Y(t_m) = y(t_m)] = u_m\left(y(t_{m+1}), y(t_m)\right). \tag{A.4}$$

For parameters in the network dynamics and attribute dynamics that are constant across all the time periods, the moment equation is

$$\sum_{m=1}^{M-1} \mathbb{E}_\theta[u_m\left(Y(t_{m+1}), Y(t_m)\right)|Y(t_m) = y(t_m)] = \sum_{m=1}^{M-1} u_m\left(y(t_{m+1}), y(t_m)\right), \tag{A.5}$$

where M is the total number of time periods. In our data, $M = 3$.

A.8.2 Statistics for moment estimation

To establish moment equations, we next find the statistics $u(Y)$. Since a closed-form solution does not exist, Snijders et al. (2017) propose a way to find $\hat{\theta}$. The intuition is to find such $u(Y)$ that partially increases monotonically with respect to one-dimensional parameter θ_h holding other parameters constant. Therefore, we can find a unique solution of θ_h by simulating different values.

Statistics of the rate parameters Here, we elaborate on how to estimate the rate of network evolution (λ_m) and the speed of attribute evolution (τ_m). Intuitively, to estimate λ_m, we use the number of changes occurring in the network during a period as in Equation (A.6). The higher the value of λ_m is, the higher the number of changes.

$$\sum_{i,j} |x_{ij,t}(t_m) - x_{ij,t}(t_{m-1})| \tag{A.6}$$

The estimation of τ_m follows the same logic. We use the total absolute change in the attribute as in Equation (A.7). The higher value of τ_m is, the higher the statistic

$$\sum_{i} |Z_i(t_m) - z_i(t_{m-1})|. \tag{A.7}$$

Statistics of parameters in the network evolution In Equation (3.7), β_l interacts with only s_{il}. Therefore, the statistic for estimating β_l is

$$\sum_i s_{il}(X(t_m), z(t_{m-1})). \tag{A.8}$$

Statistics of parameters in the attribute evolution The statistics to estimate the parameters in Equation (A.2) are derived from an autoregression model (Niezink et al. 2019). To estimate the feedback parameter a, we use

$$\sum_i Z_i(t_m) z_i(t_{m-1}). \tag{A.9}$$

To estimate the parameters b_0, b_1, c_1, c_2, c_3, c_4, c_5 and c_6, we use

$$\sum_i Z_i(t_m) q_i(t_{m-1}), \tag{A.10}$$

where q_i denotes the corresponding variables with respect to b_0, b_1, and $c_{\text{1-6}}$, for example, $Degree_{i,t}$, $WinRatio_{i,t}$, etc.

A.8.3 Standard error

Based on the delta method and the implicit function theorem, it can be proven that the asymptotic covariance matrix of the moment estimator is

$$cov_\theta(\hat{\theta}) = D_\theta^{-1} \Sigma_\theta D_\theta^{\prime -1} \tag{A.11}$$

where D_θ is the matrix of partial derivatives,

$$D_\theta = \frac{\partial \mathbb{E}_\theta(u(Y))}{\partial \theta}, \tag{A.12}$$

and Σ_θ is the covariance matrix of $u(Y)$,

$$\Sigma_\theta = cov_\theta(u(Y)). \tag{A.13}$$

Thus, the standard errors can be obtained. Both the partial derivatives D_θ and the covariance matrix Σ_θ are estimated using the Monte Carlo method. Details can be found in Schweinberger and Snijders (2007).

A.8.4 Stochastic approximation

Although the conditional expectation in the moment equations (A.4) and (A.5) cannot be calculated explicitly, the stochastic process can be easily simulated. Therefore, the stochastic approximation method (stochastic iterative algorithms) can be used to solve the moment equations (Robbins and Monro 1951). The basic iteration step in such algorithms is

$$\hat{\theta}_{N+1} = \hat{\theta}_N - a_N D_0^{-1}(U_N - u(y)), \tag{A.14}$$

where a_N is a sequence of positive numbers converging to 0. The optimal choice of D_0 depends on the distribution of U_N, and U_N is generated according to the probability distribution defined by the parameter value $\hat{\theta}_N$. If all eigenvalues of the matrix of partial derivatives (see Equation (A.12)) have positive real parts and certain regularity conditions are satisfied, then convergence at an optimal rate can be achieved when D_0 is fixed, for example, the identity matrix, with α_N a sequence of positive numbers converging to 0 at the rate $N^{-\delta}$, where $0.5 < \delta < 1$ (Snijders et al. 2017). U_N is defined as

$$U_N = \sum_{m=2}^{M} u_m \left(y(t_{m-1}), Y^{sim}(t_m) \right), \tag{A.15}$$

where $Y^{sim}(t_m)$ is the simulated value for time t_m starting at time t_{m-1} with the observed value $Y(t_{m-1}) = y(t_{m-1})$ for parameter value θ_N.

The observed outcome $u(y)$ is

$$u(y) = \sum_{m=2}^{M} u_m(y(t_{m-1}), y(t_m)). \tag{A.16}$$

A.9 Model choice

To illustrate why considering the number of similar peers by modelling the network evolution process is important, we adopt the explained variance, a metric for model comparison proposed by Niezink et al. (2019). This metric serves a similar role as R^2 in traditional regression, capturing how much additional structure in the model helps explain variation in the outcome (that is, a trader's number of followers). Specifically,

$$R_m^2 = 1 - \frac{\sum_{i=1}^{n} \left(z_{i,t} - \hat{z}_{i,t}^{m} \right)^2}{\sum_{i=1}^{n} \left(z_{i,t} - \hat{z}_{i,t}^{m_0} \right)^2}, \tag{A.17}$$

where $z_{i,t}$ is the observed number of followers following trader i at period t, $\hat{z}_{i,t}^{m}$ is the predicted number of followers following trader i at period t using our coevolution model, $\hat{z}_{i,t}^{m_0}$ is the predicted number of followers following trader i at period t using the baseline model. We define the baseline model as a linear OLS regression without including $SimilarPeers_{i,t}$ as an independent variable. This metric captures the relative improvement in predictive accuracy when we model and endogenize $SimilarPeers_{i,t}$ in our coevolution model. A higher explained variance indicates that modeling the evolution of the trader similarity network provides a meaningful improvement in explaining the follower dynamics. The explained variance of our model is 92.36%, indicating that incorporating the trader similarity network accounts for a substantial portion of the variation in the number of followers.

A.10 Interpretation of the network evolution model

Since we use the multinomial logit to model the network change, the magnitude of the coefficient estimate should be interpreted by the difference in utility, not the absolute values directly (Train 2009). Therefore, regarding the coefficient of Follower Alter effect, we focus on the interpretation of log-odds of two possible network changes of trader i: ($+ij$: adding a link between i and j; $+iw$: adding a link between i and w), and everything is the same between j and w except for the number of followers. The log-odds can be expressed as:

$$log(\frac{P(x^{+ij}|Z_{j,t})}{P(x^{+iw}|Z_{w,t})}) = \beta_2(Z_{j,t} - Z_{w,t}). \tag{A.18}$$

Suppose trader j with 10 followers and trader w with 5 followers, trader j will be much more likely to attract trader i to be her similar peer than trader w (with an odds ratio of $e^{1.1795\cdot(10-5)} = 364.13$, given $\beta_2 = 1.1795$). We see that the meaningful interpretation depends on the number of followers and the network structure, together with the estimated coefficient value. More generally, the positive and significant coefficient of the Follower Alter effect suggests that traders with more followers are more likely to have more similar peers.

We next interpret the coefficient of the Follower Homophily effect. Suppose everything is the same between trader j and trader w except for their number of followers. The log-odds can be expressed as:

$$log(\frac{P(x^{+ij}|Z_{j,t})}{P(x^{+iw}|Z_{w,t})}) = \beta_3 \left(\frac{|Z_{w,t} - Z_{i,t}|}{Range(Z)} - \frac{|Z_{j,t} - Z_{i,t}|}{Range(Z)} \right) \tag{A.19}$$

Suppose trader i has 15 followers, trader j has 5 followers, and trader w has 10 followers. Therefore, we get $|Z_{j,t} - Z_{i,t}| = |10 - 15| = 5$, $|Z_{w,t} - Z_{i,t}| = |5 - 15| = 10$, and

$$\frac{10-5}{Range(Z)} = \frac{5}{Range(Z)} > 0.$$

Given that $\beta_3 = 4.6951$ (a positive value), this results in a positive log-odds in favor of trader j over trader w. Therefore, trader j is more likely to be trader i's similar peer than trader w due to being more similar in terms of the number of followers. This demonstrates the presence of homophily: traders are more likely to connect with others who are similar to themselves in terms of the number of followers. A similar interpretation also applies to other significant homophily effects.

Appendix B
ONLINE APPENDIX TO CHAPTER 4

B.1 Network statistics

Table B.1 shows the weekly summary statistics of three networks: the network extracted from the social interactions ($SocNet_t$), the network extracted from the leader-copier relationship ($CopyNet_t$), and the combined network of the two (Net_t). The coverage of users denotes the fraction of users connected with each other relative to the total number of users. The number of receivers denotes the number of users who receive information (identified as a copier from $CopyNet_t$ or a user who comments or likes another user's post from $SocNet_t$). The number of senders denotes the number of users who send information (identified as a leader from $CopyNet_t$ or a user who makes a post from $SocNet_t$). The percentage of reciprocal links denotes the percentage of reciprocal links relative to all the links. The density of a network is calculated as the ratio of the number of existing directed links to the total number of possible links in the network. Outdegree denotes the number of out-going links; indegree denotes the number of in-coming links. Adoption rate denotes the percentage of Bitcoin adoption rate among senders and receivers, respectively. We observe that the network is sparse. To ensure that the sparsity of the network does not alter our conclusions, we obtain a denser network by only including more active users who have been active on the platform for over one month and have executed at least five trades during our observation period. We re-run our analysis based on the denser network as a robustness check in Section 4.5.

Table B.1: Summary statistics of the network data

	Net_t	$CopyNet_t$	$SocNet_t$
Overall statistics			
Number of users	21,619	11,988	12,568
Coverage of users	9%	5%	5%
Number of receivers	20,181	10,872	11,368
Number of senders	5,719	1,377	4,820
Number of directed links	75,826	25,369	50,456
Percentage of reciprocal links	15.1%	0.1%	19.3%
Density	0.0002	0.0002	0.0003
Outdegree			
Min	1	1	1
Median	2	2	2
Mean	13	18	10
Max	5,095	3,209	2,349
Indegree			
Min	1	1	1
Median	2	1	1
Mean	4	2	4
Max	848	132	842
Adoption rate			
Senders	12.7%	12.5%	12.7%
Receivers	7.0%	10.9%	4.6%

B.2 Definition of variables and distribution of categorical variables

Table B.2 summarizes all defined variables in our empirical analysis. Table B.3 provides the probability distribution of all categorical variables.

Table B.2: Definition of variables

Variable	Description
1. Neighbor-related variables	
$\text{NeighborAdopt}_{i,t}$	Fraction of Bitcoin adopters over user i's neighbors in period t
$\text{AdoptBTCPositive}_{i,t}$	Fraction of adopters who post positively about Bitcoin over all Bitcoin adopters in user i's neighbors in period t
$\text{NonAdoptBTCPositive}_{i,t}$	Fraction of non-adopters who post positively about Bitcoin over all Bitcoin non-adopters from user i's neighbors in period t
$\text{Neighbor\_N}_{i,t}$	The total number of user i's neighbors in period t
$\text{Post\_N}_{i,t}$	Average number of posts made by user i's neighbors in period t
$\text{AdoptReturn}_{i,t}$	Average profit of adopters in user i's neighborhood in period t
$\text{NonAdoptReturn}_{i,t}$	Average profit of non-adopters in user i's neighborhood in period t
2. Individuals' trading behavior	
$\text{AccReturn}_{i,t}$	The average profit of user i by period t
$\text{AccSD}_{i,t}$	The standard deviations of user i's profit by period t
$\text{AccTrade}_{i,t}$	The total number of trades of user i by period t
$\text{NTrade}_{i,t}$	The number of trades traded by user i in period t
$\text{Diversification}_{i,t}$	The total number of unique instruments traded by user i by period t
$\text{Tenure}_{i,t}$	The number of weeks user i has traded on the platform by period t since January 2016
3. Individual time-invariant characteristics	
Male_i	1, if male, otherwise, 0
AgeRange_i	Age range of user i
Income_i	Annual income of user i (in U.S. dollars)
Experience_i	Trading experience of user i before joining the platform (in years)
RiskLevel_i	Self-reported risk level of user i

Table B.3: Probability distribution of categorical variables

Panel A: Gender and Age								
	Gender		Age range					
	Female	Male	18-24	25-34	35-44	45-54	55-64	≥65
Total	11,678	98,165	41,733	36,274	16,774	7,541	4,060	3,075
Panel B: Experience								
	< 1 year	1 year	1-3 years	≥3 years	Missing			
Percent	4.8%	23.4%	16.1%	11.7%	44.1%			
Panel C: Risk level								
	Very Low	Low	Medium	High	Very High	Missing		
Percent	5.9%	6.6%	15.2%	14.8%	13.8%	43.6%		
Panel D: Income								
	10K	25K	50K	100K	200K	≥500K	Missing	
Percent	15.0%	1.7%	25.2%	11.6%	11.7%	3.1%	40.3%	

Note: Panel A reports the gender and age distributions of the users in our dataset. Panel B, C, and D report users' self-reported trading experience, risk level, and income upon registering at the platform.

B.3 Multicollinearity test

To mitigate the concern of multicollinearity of independent variables, we show the correlation matrix in Table B.4. From Table B.4, we do not observe variable pairs that exhibit extremely high correlation coefficients. We further test Variance Inflation Factor (VIF), reported in Table B.5. We find that no variable has a VIF score exceeding 10, indicating that multicollinearity does not significantly distort the outcomes of our empirical analysis.

Table B.4: Correlation matrix

	Neighbor_N	Post_N	AccTrade	Tenure	NTrade	Diversification	AccReturn	AccSD	NeighborAdopt	AdoptBTCPositive	NonAdoptBTCPositive
Neighbor_N	1	.1001	.0959	-.1574	.3268	.2418	.0577	-.0015	.2565	.1226	.0587
Post_N	.1001	1	-.0897	-.0902	-.0334	-.0593	.0087	-.0442	.0155	.0136	.0333
AccTrade	.0959	-.0897	1	.5881	.6423	.7973	.0046	.3841	-.068	-.028	.0108
Tenure	-.1574	-.0902	.5881	1	.0902	.1547	.0121	.0403	-.1111	-.0398	-.0342
NTrade	.3268	-.0334	.6423	.0902	1	.7281	.0237	.3531	.0159	.0088	.0445
Diversification	.2418	-.0593	.7973	.1547	.7281	1	-.0073	.4393	-.0093	-.0073	.0456
AccReturn	.0577	.0087	.0046	.0121	.0237	-.0073	1	-.2861	.0137	.0045	.0007
AccSD	-.0015	-.0442	.3841	.0403	.3531	.4393	-.2861	1	-.0378	-.013	-.0004
NeighborAdopt	.2565	.0155	-.068	-.1111	.0159	-.0093	.0137	-.0378	1	.2192	.0041
AdoptBTCPositive	.1226	.0136	-.028	-.0398	.0088	-.0073	.0045	-.013	.2192	1	.0081
NonAdoptBTCPositive	.0587	.0333	.0108	-.0342	.0445	.0456	.0007	-.0004	.0041	.0081	1

Table B.5: Results of VIF

Variable	VIF score
AccTrade	7.40
Diversification	5.24
Tenure	2.49
NTrade	2.39
AccSD	1.52
Neighbor_N	1.35
AccReturn	1.16
NeighborAdopt	1.13
AdoptBTCPositive	1.06
Post_N	1.02
NonAdoptBTCPositive	1.01
Mean VIF score	2.34

B.4 The effect of diversity of opinions on cryptocurrency adoption

To further shed light on how the size of the neighborhood affects the focal user's Bitcoin adoption, we add another control variable that directly captures the diversity of opinions in one's neighbor. $NeighborDiversity$ is defined by the standard deviation of the sentiment of neighbors' opinions. We label positive sentiment with a value of 1, and 0 otherwise, and take the standard deviation of sentiments of all neighbors. We set $NeighborDiversity$ to be zero when no neighbors post about Bitcoin. A greater standard deviation denotes a higher degree of diversity in neighbor opinions towards Bitcoin. The correlation coefficient between the $NeighborDiversity$ and $Neighbor\_N$ is 0.290, suggesting a significant positive relationship between the diversity of opinion and number of neighbors. After adding the new control variable, we re-run our model. Table B.6 shows the estimation results. Our findings indicate that the diversity of opinions within the neighborhood has a negative impact on a user's decision to adopt Bitcoin.

Table B.6: Diversity in neighboring opinions

Depedent Variable	Adopt
NeighborAdopt	0.098
	(0.060)
AdoptPosBTC	10.732***
	(2.192)
NonAdoptPosBTC	-5.423***
	(1.106)
Neighbor_N	-0.068***
	(0.017)
NeighborDiversity	-3.762***
	(0.807)
Post_N	0.000***
	(0.000)
AccTrade	-0.018***
	(0.003)
Tenure	0.034***
	(0.006)
NTrade	0.004***
	(0.001)
Diversification	0.156***
	(0.009)
AccReturn	-0.000***
	(0.000)
AccSD	-0.001***
	(0.000)
Observations	2,316,559

Notes: Standard errors clustered at the user level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix C

ONLINE APPENDIX TO CHAPTER 5

C.1 Proof of Theorem 1

Proof. *Sufficiency.* Take any measurable triplet (ρ, ϕ_1, ϕ_2) and define $\tilde{f}_{mj} = \rho\Big(\phi_1(X_{mj}, E_{mj}) + \sum_{k\neq j} \phi_2(X_{mk}, E_{mk})\Big)$. Because the inner summation is a multiset operation, $\tilde{f}_{mj}$ is automatically permutation–invariant for the non-focal nodes and does not depend on their ordering; identity independence follows by construction. Hence any function that can be written in this form meets Assumptions 2–3.

Necessity. Let $\mathbb{C}$ be the countable universe of possible pairs (X, E) obtained after discretising X to rational vectors. Partition $\mathbb{C}$ into the set of even indices $\mathbb{E} = \{2n : n \in \mathbb{N}\}$ and odd indices $\mathbb{O} = \{2n{+}1 : n \in \mathbb{N}\}$. Because $\mathbb{C}$ is countable, there exist mappings $c^{\mathrm{e}} : \mathbb{C} \to \mathbb{E}$ and $c^{\mathrm{o}} : \mathbb{C} \to \mathbb{O}$. Define

$$\phi_1(X_{mj}, E_{mj}) = 4^{-c^{\mathrm{e}}(X_{mj}, E_{mj})}, \qquad \phi_2(X_{mk}, E_{mk}) = 4^{-c^{\mathrm{o}}(X_{mk}, E_{mk})}.$$

Because even and odd codes never collide, the quantity $T_{mj} = \phi_1(X_{mj}, E_{mj}) + \sum_{k\neq j} \phi_2(X_{mk}, E_{mk})$ provides a *unique* representation of the focal node and the non-focal nodes in the network. Hence a function $\rho : \mathbb{R} \to \mathbb{R}$ can always be found such that

$$f_{mj} = \rho\Big(\phi_1(X_{mj}, E_{mj}) + \sum_{k\neq j} \phi_2(X_{mk}, E_{mk})\Big)$$

completing the proof.

VITA

Ye Liu is a researcher specializing in Information Systems. Her research interests lie at the intersection of FinTech, digital platforms, and societal issues in digital economic contexts. She employs a range of methodologies, including causal inference, structural modeling, and machine learning, to study how technological and social forces shape behavior and market outcomes. She received her Ph.D. from the Michael G. Foster School of Business at the University of Washington. Prior to that, she earned her bachelor's degree in Finance from Shanghai Jiao Tong University. She will join the Information Technology Management group at the Scheller College of Business, Georgia Institute of Technology as an Assistant Professor in July 2025.