CENTER FOR TAX AND ACCOUNTING RESEARCH

Center for Risk Management

TAXATION, ACCOUNTING, AND FINANCE

TAF WORKING PAPER

No. 74 / July 2022

Social trading, communication, and networks

Deng, Jiaying / Yang, Mingwen / Pelster, Matthias / Tan, Yong

Social trading, communication, and networks*

Jiaying Deng[1], Mingwen Yang[2], Matthias Pelster[3], Yong Tan[4]
Michael G. Foster School of Business, University of Washington, Seattle, WA, United States[1,2,4]
Paderborn University, Center for Risk Management, Department for Taxation, Accounting and Finance, Paderborn, Germany[3]

Social trading is an emerging market in the sharing economy, allowing inexperienced investors (followers) to automatically follow the trades of experts (leaders) in real time. We use a separable temporal exponential random graph model (STERGM) to analyze the formation and dissolution of links in a large social trading network. In contrast to traditional social networks, social trading networks are characterized by a rapid dissolution of links, thereby increasing the importance of studying network dissolution. We investigate how social communication, along with financial performance and demographics, affects dynamic network evolution and address the existing dependence in leader-follower links. The determinants of link formation and dissolution are asymmetric. Different types of social communication, such as posts and comments, have different implications for link formation and dissolution. In addition, our results show that financial performance and demographic characteristics are also important determinants of link formation. However, once a link is formed, followers mainly focus on financial performance in addition to social communication but not on demographic characteristics. Our findings have important implications for both investors and social trading platforms.

Key words: Social Trading, Copy Trading, Social Communication, STERGM

1. Introduction

Online engagement among individual investors has grown significantly in recent years. The recent Reddit hype (Pedersen 2022) vividly shows that social media plays an important role in financial markets and in transmitting relevant information to potential investors. Social trading platforms incorporate elements from the worlds of social media *and* online trading and have recently garnered tremendous attention in both research and practice (e.g., Ammann and Schaub 2021, Apesteguia et al. 2020, Yang et al. 2021). Social trading is a novel form of investing that allows retail investors to observe the trading behavior of other investors and to automatically follow their investment strategies through so-called "copy trading" or "mirror trading" (Apesteguia et al. 2020). An autocopy service (mirror

* Data were obtained from a large social trading platform. We thank our data provider for use and explanations of their data. Matthias Pelster gratefully acknowledges financial support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) – Individual Research Grants project number 434732045. Any errors, misrepresentations, and omissions are our own.

trading) enables novice investors (followers) to link their trading accounts to those of expert investors (leaders) and thereby delegate their trading activities (Doering et al. 2015). Experienced investors are able to earn additional income by sharing their trading knowledge with a large group of followers. Large social trading platforms, such as eToro, Zulutrade, and FX Junction, have gained popularity, as evidenced by the growing pool of investors on such platforms. For example, in 2019, eToro was operating in 140 countries with over 10 million users.[1]

Social trading platforms offer several unique features. First, social trading platforms offer a very transparent information flow, as (potential) followers are able to see the details of the transactions completed by other investors and track their gains and losses in real time. Second, these platforms allow for straightforward and transparent communication among investors. Investors can share their opinions, publish posts, and leave comments in a news feed that is publicly available to all users. Third, different from mutual fund managers, most participants in social trading platforms are individual traders who lack institutional endorsement.

Social trading platforms also require a new perspective on considerations of the evolution of social networks, as their network structure follows a different dynamic from that of traditional social networks such as Facebook and Twitter, which have been studied extensively (Li et al. 2017, Kim et al. 2018). In the social networks on social trading platforms, the information flow among users is directly tied to cash flows because of the copy trading feature. Individual investors may become leaders who share their trading strategies or may become followers who copy the trading strategies of leaders. Platforms typically share some of their revenue with leaders. As a result of this monetary incentive, in contrast with other traditional social networks such as Facebook or Twitter, link dissolutions in social trading networks are more frequent. A link connecting two individuals on a traditional social network is commonly characterized by stability and longevity. The link connecting two individuals on a social trading network is short-lived and volatile (Pelster 2017). Thus, not only the process of link formation but also that of link dissolution is crucial. Considering the increasing spread of these networks and their economic implications, an extensive understanding of the evolution of these networks is important. However, the evolution of

[1] `https://www.coindesk.com/company/etoro`, last accessed Jan. 8, 2021.

networks with a frequent dissolution of links has not yet been studied in detail. A large number of studies have focused on the preformation process, i.e., how a social network is formed, but none have analyzed the postformation process. Our study fills this void.

We study how the directed leader-follower networks on the largest social trading platform evolve over time. Building on the theory of soft information and hard information (Liberti and Petersen 2019), we investigate the determinants of link formation and link dissolution. Social trading platforms provide a transparent environment that releases two types of information with which potential followers can evaluate leaders: their *trading activities* (financial performance, i.e., hard information), and their *social activities* (social communication, i.e., soft information). The combination of these data is typically difficult to obtain. For example, in traditional mutual funds, researchers can observe the financial performance of a mutual fund manager but typically lack soft communication information. While some mutual fund managers may have social media channels on, for example, YouTube (e.g., Cathie Wood), this is not the case for all managers. In addition, the mutual fund industry allows for private communication between managers and investors that is unobservable to other investors, which may affect investment decisions. In contrast, social trading platforms do not allow for private communication between leaders and followers.[2] The access to various information on social trading platforms allows us to examine the role of soft and hard information in this innovative form of delegated investment management.

We place a particular emphasis on social network features and study the impact of social communication on link formation and dissolution. While prior studies document that investors chase past financial performance, the role of social communication is not clear. Financial performance signals a trader's trading ability. The platform summarizes such information in a highly transparent manner, and it does not allow users to modify or manipulate the data, making them trustworthy. Social communication provides an additional channel for leaders to convince potential followers of their superior investment skill and thus to follow their investment strategies. Given that the primary goal of investors is to make money, followers may focus mainly on financial performance, which provides an objective measure of investment skill, instead of non-monetary soft information—in particular, since, in contrast to the objective features of financial performance, the textual

[2] Except in rare cases where leaders and followers may know each other in real life.

information in communication is more complex to interpret and evaluate. It requires more time for followers to read through text messages and filter out the irrelevant information. The limited attention of followers may make social communication less effective. In addition, it is also questionable how reliable social communication is, given that individually disclosed information is not screened by the platform. Hence, social communication may not be as trustworthy as financial performance, and the role of social communication is unclear.

In this study, we build a separable temporal exponential random graph model (STERGM) to disentangle the reasons why a follower follows *and* the reasons why she or he unfollows a leader in the social trading context. To capture unobserved heterogeneity and address potential endogeneity concerns, we incorporate Chamberlain correlated random effects (Chamberlain 1980) into our STERGM. We find that financial performance, social communication, and demographic characteristics are important determinants of link formation. However, once a link is formed, followers mainly focus on financial performance and social communication (instead of demographic characteristics) to decide whom to unfollow. We also find that the impact of these factors is asymmetric in the link formation and dissolution processes. Different types of social communication, such as posts and comments, have different implications for link formation and dissolution. Both the quality and the quantity of a leader's posts increase the follower's probability of forming a new link and of maintaining an existing link. Followers rely on "peer reviews": Leaders who receive more positive comments are more likely to attract new followers and keep existing ones. Followers are less likely to form new links or sustain existing links if the leaders receive more negative comments. Moreover, the impacts of negative and positive comments are asymmetric; negative comments have a larger impact than positive comments in both link formation and link dissolution. Overall, we find that social communication plays an important role in leaders' ability to convince potential followers to follow their trading strategies and existing followers to sustain their links.

Our work makes several contributions to the extant literature. First, this study is the first to model leader-follower network evolution on social trading platforms. Different from those in traditional social networks, relations between investors on social trading platforms involve a monetary dimension, and therefore, social trading features frequent link formation *and* dissolution as investors adjust their investment strategies. Our findings enrich the literature on the determinants of social networks by providing empirical evidence of the evolution process of an innovative network structure. Second, our study contributes to the literature on financial advice. Recent developments in fintech have made it easier and more convenient for investors to share their trading knowledge and turn to other investors for advice. With the increasing importance of social interactions on financial markets, our results contribute to this stream of literature by showing that social communication, especially from leaders, can generate economic impacts (i.e., leaders can attract or maintain more followers to earn higher compensation). Third, our study contributes to the literature on individual investor behavior. While financial performance is an important signal of traders' trading skills, we find that followers also rely on communication when evaluating peer traders. Fourth, we contribute to the literature on hard and soft information. We find evidence that both hard information (i.e., financial performance) and soft information (i.e., social communication) play important roles in the link formation and dissolution processes on social trading platforms. Finally, from a methodological perspective, we incorporate Chamberlain random effects into the STERGM to alleviate concerns about confounding effects from individual-level unobserved heterogeneity in the network analysis.

Our paper has important managerial implications. While social trading has some features that are comparable to mutual funds in the sense of "delegated portfolio management" (Doering et al. 2015)[3], the extreme flexibility of followers in dissolving links and thereby terminating their relationship instantaneously brings about large income uncertainty for the leader. Thus, for a leader, a thorough understanding of network evolution and its determinants is crucial. In this context, social communication can mitigate information asymmetries and help to build trust. Thus, our results on the impact of social communication can provide important guidance for leaders on when and how to communicate with followers. Second, our findings provide implications for the providers of social trading platforms. For their business model to work, platforms need to ensure that both leaders and followers are satisfied with the services provided and thus need to provide a positive investment experience. Third, as most recently demonstrated by the GameStop frenzy,

[3]In particular, investors who invest in mutual funds entrust their money to a third party who then makes specific investment decisions for them. This is the same in copy trading on social trading platforms: Investors entrust their money to a leader, and the leader makes specific investment decisions for them. Due to these similarities, the Markets in Financial Instruments Directive (MiFID) characterizes social trading as portfolio management (see, `https://financefeeds.com/mifid-ii-entering-age-completely-self-directed-traders-final-nail-goes-copy-trading-coffin/`).

vocal leaders on social media may exert a significant influence on financial markets (see, e.g., Pedersen 2022). Thus, a better understanding of the evolution of social networks with an investment focus is important for regulators.

The remainder of this paper is structured as follows. Section 2 summarizes the relevant literature and presents the theoretical background. Our model is introduced in Section 3. Section 4 describes the data. Section 5 presents the results, and Section 6 presents several robustness checks. We discuss the implications and conclude the study in Section 7.

2. Theoretical background

While the literature has studied social networks in detail (Kane et al. 2014), it is unclear how insights from other social networks can be applied to social trading platforms. The social networks on social trading platforms differ from traditional social networks. Platforms typically do not allow for social relations akin to friendships on Facebook or follower relations on Twitter. Instead, the relations between users in such networks are directly tied to cash flows. The platform provides a service that allows "copy trading" for its customers—duplicating investment strategies from other investors with one's own money without approving each individual transaction. Investors (i.e., followers) who make use of the social features of the platform and form a copy trading link with other investors can delegate their trading; at the same time, leaders (or signal providers) can earn additional income, receiving compensation from the platform for contributing to its business model. As a result, compared to other traditional social networks such as Facebook or Twitter, link dissolution occurs more frequently. While prior literature has mainly focused on link formation, link dissolution has received little attention in the literature—partially because link dissolution is a relatively rare event in many traditional social networks.

2.1. Social trading

Our study contributes to the fast-growing literature on social trading, which, generally, can be divided into three groups. The first group addresses the institutional aspects of social trading (see, e.g., Doering et al. 2015). A key feature of social trading platforms is a high level of information transparency. Investors can observe the trading behavior of their peers at the trade level and in real time. Considering that some investor trades may contain valuable information, making these trades available in real time potentially undercuts the platforms' payoff potential. To resolve this issue, Yang et al. (2021) propose a personalized trade-level information release policy that allows platforms to optimize their revenues.

A large second group of studies examines how the information transparency that allows investors to observe other investors in real time may affect their trading behavior and the performance implications of social trading. Gemayel and Preda (2017, 2018a) label the state of permanent reciprocal observation and scrutiny that are typical of social trading platforms a “scopic regime”. Trading in a scopic regime alters investors’ behavioral biases such as the disposition effect (Heimer 2016, Pelster and Hofmann 2018).[4] Focusing on the copy trading feature, the experimental study by Apesteguia et al. (2020) suggests that having the option to directly follow other investors significantly increases the risk-taking behavior of investors. This increased risk taking does not, however, yield superior investment returns (e.g., Pan et al. 2012). Several studies have found that on average, traders on social trading platforms do not outperform the market in the long term (Dorfleitner et al. 2018, Oehler et al. 2016). Only a few investors are able to earn significant short-term excess returns (Dorfleitner et al. 2018, Oehler et al. 2016).

Most closely related to our paper, the last group of papers studies network relationships on social trading platforms. Ammann and Schaub (2021) find that social traders are more likely to have their investment strategies duplicated within three weeks of making (positive) posts on trading platforms. These posts do not, however, seem to contain valuable information, as they do not have any predictive power over future performance. While Ammann and Schaub (2021) focus on fund flows in their analysis, we, in contrast, focus on individual activity-based relationships between traders, i.e., the social ties within the network. Additionally, focusing on fund flows, Röder and Walter (2019) document a positive flow-performance relationship in social trading portfolios, which is limited to the top past performers (i.e., investment flows chase past performance). To optimally exploit the copy trading function, Lee and Ma (2015) develop a system called “W2F (whom to copy)” that enables users of social trading platforms to “discover expert traders” who consistently realize high risk-adjusted performance.

2.2. Link formation and dissolution in (social) networks

The evolution of leader-follower networks, or in other words, when and why an investor (follower) follows or unfollows another investor (leader) on a social trading platform, is one of the fundamental questions in social trading. Prior research has mostly focused on

[4] The disposition effect is an anomaly discovered in behavioral finance. It is related to the tendency of investors to sell assets that have increased in value while holding assets that have dropped in value.

the trading behaviors of leaders in this context but has largely ignored followers' decisions about whom to follow or unfollow. In this study, we aim to shed light on this question by focusing on the factors that drive the following and unfollowing decisions of followers on social trading platforms. In this subsection, we review the related literature that discusses the determinants of tie formation and dissolution in other (related) settings and discuss social trading-specific features that may influence tie formation and dissolution.

Trust. It is widely acknowledged that a good relationship between managers and investors is beneficial. Trust is important for managing relationships and regulating their quality (Kaiser and Berger 2021). Trust is relevant when a person (the trustor) has specific expectations of another person (the trustee) and is vulnerable to whether the trustee fulfills those expectations, regardless of control (Mayer et al. 1995). Various factors can influence the existence of trust in a relationship. As noted by Mayer et al. (1995), trustworthiness must be established before any factor can lead to trust. Trustworthiness refers to the trustee and relates to specific characteristics of the trustee, for example, benevolence, integrity, or ability. The most common factor that can establish trustworthiness and ultimate trust is communication (Kaiser and Berger 2021). Communication and timely feedback increase trust; other factors that establish trust are reputation, quality, and partner fit (Kaiser and Berger 2021).

Recent developments in fintech have led to new challenges in managing relationships, particularly because investors cannot establish a personal one-on-one trust relationship but must instead seek to build a relationship with a more or less anonymous mass (Kaiser and Berger 2021). As noted by Wohlgemuth et al. (2016), trust among online community members plays an even more important role in the online trading context, where investors can automatically, without further evaluation, duplicate the investment strategies of their peers.

Social communication. Social communication can help build trust and reduce information asymmetries (Xu and Chau 2018). Duarte et al. (2012) show that borrowers who appear to be more trustworthy based on their pictures have a higher probability of having their loans funded. Xu and Chau (2018) find that both credit grades and lender-borrower communication affect funding outcomes on peer-to-peer lending platforms. In particular, social communication can be a tool that particularly allows listers with poor hard information (e.g., credit grades) to improve their chances of being funded (Xu and Chau 2018).

Social communication can be regarded as soft information. Soft information refers to qualitative information such as media press, communication texts, or market commentary, whereas hard information refers to quantitative information such as stock returns or credit ratings. Even though soft information is usually qualitative, it can be "hardened" using information technologies such as text mining and converted into a quantitative measure. Thus, the main difference between hard and soft information is that the former can be objectively verified and is independent of context, while the quantification of the latter makes use of various degrees of freedom. An immediate consequence of the nature of soft information is that its assessment (e.g., how trustworthy another agent is or other informational cues are) depends on each agent's personal standards. When evaluating information from others, the quality of this information is important for building trust (Xu and Chau 2018). We hypothesize that social communication is important for investors to establish trust on social trading platforms. Thus, we incorporate social communication variables, including a proxy for the quality of social communication, and explore how they affect link formation and dissolution.

Financial performance. As noted above, hard information is also an important determinant of whether investors decide to fund a project, sell a stock, or make a loan (Liberti and Petersen 2019). Social trading platforms are comparable to mutual and hedge funds, as they allow for some form of delegated portfolio management (Doering et al. 2015). Both mutual and hedge funds have received considerable attention in the financial literature, with a particular focus on the determinants of their performance (e.g., Agarwal et al. 2009, Grinblatt et al. 2020) and the relationship between fund performance and (net) fund flows (Sirri and Tufano 1998, Goetzmann et al. 2003). As investors can infer the skills, at least to some degree, of mutual and hedge fund managers from their past performance, (net) fund flows should be explained by past performance (Barber et al. 2016). This stream of the literature also documents not only that mean performance is important but also that the volatility of performance is negatively related to fund flows (Sirri and Tufano 1998, Huang et al. 2007). Importantly, investors seem to determine their inflows and outflows differently (Ivković and Weisbenner 2009). The importance of returns has also been documented in nonprofessional settings, such as online crowdfunding markets (Lin and Viswanathan 2016). Based on this stream of the literature, we examine how the formation and dissolution of leader-follower links are affected by financial performance.

Demographic factors and homophily. A large stream of the social networks literature documents that similarity (homophily) breeds connections (McPherson et al. 2001). Homophily describes the preference of people to favor others who are similar to them rather than those who are dissimilar to them. In financial markets, so-called home bias is a prominent example of a preference for similarity (Coval and Moskowitz 1999, Lin and Viswanathan 2016). In the social trading context, potential cultural differences or language barriers in online communications may also contribute to homophily. In addition, the profile picture (image) and biography on an investor's public profile page can affect how he/she is perceived by others (Duarte et al. 2012).

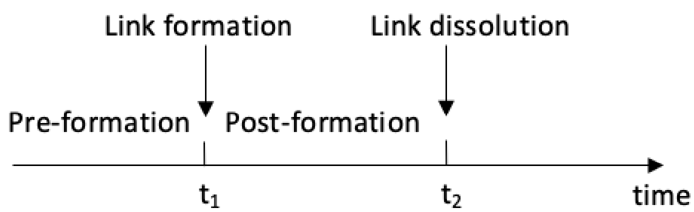
Dissolution of ties. Even though link dissolution happens less frequently in traditional social networks such as Facebook or Twitter, the dissolution of ties in interfirm networks or in financial markets is a rather frequent event. Studying interfirm networks, Greve et al. (2010) argue that the dissolution of ties may happen in particular when embeddedness is low. Polidoro et al. (2011) further study the importance of embeddedness for tie dissolution and argue that network centrality, i.e., positional embeddedness, does not promote stability, but having common partners, i.e., structural embeddedness, does. Translated to social trading, where common partners do not play a meaningful role and where there are almost no costs to dissolving a tie and engaging in a new relationship (aside from transaction costs that are charged via the spread), these findings suggest that ties could be dissolved rather quickly. Baker et al. (1998) argue that unsatisfactory performance is a main driver of tie dissolution—an argument that can be translated to social trading at face value.

Shafi et al. (2020) study the dissolution of ties in early entrepreneurial finance and argue that a tie discontinuation can have important ripple effects on other ties. In particular, once well-established investors cut their ties with a start-up, smaller investors may follow suit. Thus, discontinuation may have important repercussions for start-ups (Shafi et al. 2020). In social trading, despite the high level of transparency, investors face some degree of uncertainty when deciding to follow other investors. As commonly stated on delegated investment opportunities, "past performance is no guarantee of future results". Consequently, investors may not be satisfied with the outcomes of a given tie and decide to discontinue the relationship, or in the words of Shafi et al. (2020): "The decision to withdraw financial support may be primarily related to a venture's underperformance." Based on this stream of literature, we hypothesize that financial performance, or hard information, is a key factor that determines link dissolution. In addition, the role of soft information in link dissolution is not clear. We thus investigate how hard information and soft information affect link dissolution.

3. Model

Due to the nature of social trading networks, particularly the frequent link formation and dissolution in such networks, it is important to study both link formation and dissolution. Figure 1 illustrates a typical link formation and dissolution process on social trading platforms. A link between a follower and a leader is formed at t_1 and dissolved at t_2. Following a leader is equivalent to (automatically) copying the trading strategy of the leader. A link is formed when a follower follows a leader; the link is dissolved when the follower stops following the leader. The network is constructed through leader-follower links. We use the word "follower" for consistency with prior literature (Ammann and Schaub 2021, Yang et al. 2021).

Figure 1 **An illustration of link formation and dissolution**



3.1. Separable temporal exponential random graph model

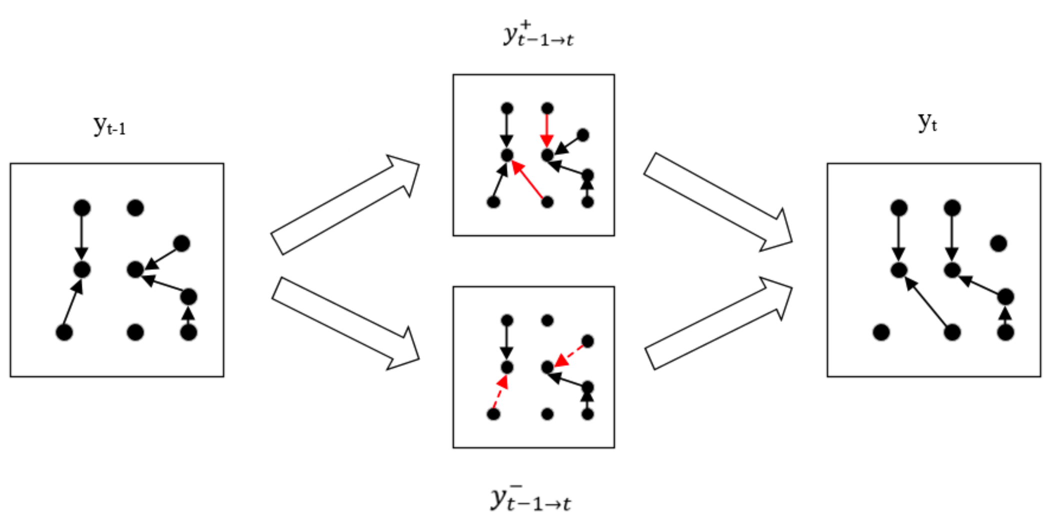
We use extensions of the exponential random graph model (ERGM) (Snijders et al. 2006, Robins et al. 2007) to model the network evolution. ERGMs represent a general class of models based on exponential family theory that can be used to specify the probability distribution underlying a set of random graphs or networks (Robins et al. 2007, Snijders et al. 2006) and are widely used for network analyses in the field of information systems (Yan et al. 2015, Hwang et al. 2022). The aim of the ERGM is to identify the factors that affect link formation in a network by comparing the probability of the realized network structure with all alternative network configurations. However, the conventional ERGM neither accounts for the intertemporal dependence in longitudinally observed networks nor models the link dissolution process. In this study, we adopt the STERGM (Krivitsky and

Handcock 2014), an extension of the ERGM, to model the network dynamics that allows us to capture both intertemporal dependence and the link dissolution process.

We consider dynamic leader-follower networks with a total of T time periods. At time t, suppose that there are N_t nodes, and let Y_t be an $N_t \times N_t$ adjacency matrix for a random network. $y_{ijt} = 1$ indicates a link between nodes i and j at time t, and $y_{ijt} = 0$ indicates that there is no link between these nodes at time t. We define $\mathcal{Y}_t$ as the set of all possible networks among the nodes and y_t as a realized network for $y_t \in \mathcal{Y}_t$ at time t.

Figure 2 provides an illustration that visualizes directed network changes from time $t-1$ to t. We show the realized network at times $t-1$ and t, denoted as y_{t-1} and y_t, respectively. We define two networks to track the evolution of the network from time $t-1$ to t: the *formation network* y^+ and the *dissolution network* y^-. $y^+_{t-1\to t}$ is defined as network y_{t-1} plus the links established from time $t-1$ to t. Similarly, $y^-_{t-1\to t}$ is defined as network y_{t-1} minus the links dissolved from time $t-1$ to t. In our illustration, two new links are added (denoted by red solid arrows), and two existing links are removed (denoted by red dashed arrows). Thus, we are able to track the network evolution in terms of links from time $t-1$ to t. Although we observe only networks y_{t-1} and y_t, we can recover $y^+_{t-1\to t}$ and $y^-_{t-1\to t}$ since $y^+_{t-1\to t} = y_{t-1} \cup y_t$ and $y^-_{t-1\to t} = y_{t-1} \cap y_t$. Appendix A gives a detailed description of how we track the network evolution and construct y^+ and y^- in each period.

Figure 2 **A visualization of network changes from time $t-1$ to t**



Mathematically, the formation process is modeled as

$$\mathrm{P}(Y^{+}_{t-1\to t} = y^{+}_{t-1\to t} | Y_{t-1} = y_{t-1}; \theta^{+}) = \frac{e^{(\theta^{+\prime} g^{+}(y^{+}_{t-1\to t}, X_{t-1}))}}{\kappa(\theta^{+}, X_{t-1}, \mathcal{Y}^{+}(y_{t-1}))}, \tag{1}$$

and the dissolution process is modeled as

$$\mathrm{P}(Y^{-}_{t-1\to t} = y^{-}_{t-1\to t} | Y_{t-1} = y_{t-1}; \theta^{-}) = \frac{e^{(\theta^{-\prime} g^{-}(y^{-}_{t-1\to t}, X_{t-1}))}}{\kappa(\theta^{-}, X_{t-1}, \mathcal{Y}^{-}(y_{t-1}))}, \tag{2}$$

where $g^{+}(y^{+}_{t-1\to t}, X_{t-1})$ $(g^{-}(y^{-}_{t-1\to t}, X_{t-1}))$ is the vector of model covariates for formation network $y^{+}_{t-1\to t}$ (dissolution network $y^{-}_{t-1\to t}$), and θ^{+} (θ^{-}) is the vector of coefficients for network $y^{+}_{t-1\to t}$ $(y^{-}_{t-1\to t})$. The denominators in Equations (1) and (2) are normalizing factors that represent the sum of the numerator over all possible networks to ensure that the probability of observing the realized formation (dissolution) network is between 0 and 1. Mathematically, the factor is defined as follows:

$$\kappa(\theta^{+}, X_{t-1}, \mathcal{Y}^{+}(y_{t-1})) = \sum_{z^{+} \in \mathcal{Y}^{+}(y_{t-1})} e^{(\theta^{+\prime} g^{+}(z^{+}, X_{t-1}))} \tag{3}$$

and

$$\kappa(\theta^{-}, X_{t-1}, \mathcal{Y}^{-}(y_{t-1})) = \sum_{z^{-} \in \mathcal{Y}^{-}(y_{t-1})} e^{(\theta^{-\prime} g^{-}(z^{-}, X_{t-1}))}, \tag{4}$$

where z^{+} (z^{-}) denotes a possible formation (dissolution) network from time $t-1$ to t.

3.2. Identification

A dynamic network analysis of thousands of nodes requires significant computing resources and is computationally intractable (Yan et al. 2015). Thus, we adopt a degenerate statistical model to estimate the coefficients in the link formation and dissolution processes, similar to maximum pseudolikelihood estimation (Strauss and Ikeda 1990).

A common issue in network analysis is endogeneity. First, we use lagged independent variables (i.e., from period $t-1$) to mitigate potential reverse causality. Second, in our context, the information provided on the platform is rich and highly transparent. We observe the information that is observed by followers on the platform, which may affect followers' link formation and dissolution decisions. We have access to the complete transactions history and social communications of each trader, and rich demographic information; the platform does not allow for a private chat channel. We construct various covariates, including follower characteristics, leader characteristics, homophily, and network structure, as elaborated in a later section. However, some determinants that explain link formation and dissolution may still be unobserved, at least to researchers. For example, when followers make their decisions, their investment goals on the platform and their intrinsic trust in others might affect their link formation and dissolution. Hence, to mitigate the concern of omitted variables, we control for follower-specific unobservables (η_i) in the link formation and dissolution model. Specifically, for the link formation process[5], we define

$$y_{ijt} = \begin{cases} 1 & y^*_{ijt} > 0, \text{and} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

y_{ijt} is a binary variable that is equal to 1 if follower i forms a link with leader j from period $t-1$ to t, and y^*_{ijt} is the corresponding latent utility. The utility of follower i from forming a link to leader j at time t is defined as follows:

$$y^*_{ijt} = \boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij} + \boldsymbol{\tau} C_i + \eta_i + \epsilon_{ijt}, \tag{6}$$

where X_{it-1} is a vector of follower i's time-variant covariates at period $t-1$, W_{jt-1} is a vector of leader j's time-variant covariates at period $t-1$, V_{ij} is a set of dummies indicating whether follower i and leader j share the same demographics (homophily), C_i is a set of follower-specific time-invariant observable controls, and η_i is the follower-specific fixed effect. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\tau}$ are the corresponding vectors of coefficients to be estimated.

A conventional approach to estimating fixed effects is to treat η_i as parameters and use maximum likelihood estimation. However, such estimation is inconsistent when the number of nodes is large and the number of time periods is finite, which is the incidental parameter problem (Neyman and Scott 1948). To correct for the incidental parameter problem, Chamberlain (1980) proposes a correlated random effects model (Wooldridge 2010). We control for follower-specific Chamberlain correlated random effects in our model.[6] Chamberlain (1980) allows for follower-specific unobservables to be correlated with the independent variables. Specifically, we implement Chamberlain correlated random effects following Mundlak (1978). η_i is defined as follows:

$$\eta_i = \psi + \boldsymbol{\xi} \bar{X}_i + a_i, \tag{7}$$

[5]The link dissolution process is defined in the same fashion. In the dissolution process, y_{ijt} is equal to 1 if follower i dissolves the link from leader j in period t.

[6]We also estimate link formation and dissolution using an alternative estimation approach, the conditional logit estimator. The estimation results remain consistent with the results from the main model. Please refer to Appendix G for more details.

where a_i follows a normal distribution with mean zero and variance σ_a^2, ψ is a constant, and $\bar{X}_i$ is the time average of the follower's time-variant observables. $\bar{X}_i = (\Gamma_i)^{-1} \sum_{t=1}^{\Gamma_i} X_{it}$, where Γ_i equals the number of periods that follower i exists on the platform times the number of leaders that follower i can potentially follow in each period.

In Equation (7), a_i is not assumed to depend on X_i, and the model allows for dependence between η_i and X_i by adding $\bar{X}_i$ to the equation. From Equation (7), we see that η_i follows a conditional normal distribution, that is, $\eta_i|X_i \sim Normal(\psi + \bar{X}_i\xi, \sigma_a^2)$. Thus, unlike the conventional fixed effects model, the coefficients on the follower-specific time-invariant controls C_i in Equation (6) can be identified.

Although no regulatory changes were made to social trading in general during our sample period and the regulation remains intact, link formation and dissolution might be affected by some other events that are time dependent—for example, some policy change on the platform. To mitigate this concern, we include time fixed effects, which allows us to control for time-specific peculiarities. The estimation results remain consistent with the results from the main model. Please refer to Appendix B for more details.

Finally, despite the various leader characteristics included in the model, there might still be some unobservables that affect link formation and dissolution. For example, it is possible for leaders on the platform to advertise themselves via other social media platforms. To mitigate omitted variables on the leader's side, we also include leader-specific Chamberlain correlated random effects in the link formation model. The estimation results remain consistent with the results from the main model. Please refer to Appendix C for more details.

4. Data and variables

In this section, we first introduce the data and then describe how we construct the variables used in the analysis.

4.1. Data

We obtained our data from eToro, the largest social trading platform. Similar to other online trading brokerage services, the platform allows its customers to trade stocks and contracts-for-differences (CFDs) on indices, commodities, currency pairs, and crypto assets. For each trade, the broker charges transaction fees as a portion of the bid-ask spread. In addition, the platform incorporates various features that are typical of social media.

Specifically, the platform contains a news feed in which investors can disclose their trading activities (*open book trading*) and publish posts. Here, investors can conveniently discuss their trading strategies, like and comment on others' trading activities, and automatically copy the trades of other investors. Investors who have their trades copied receive monetary compensation from the brokerage service in relation to their number of followers, their assets under management, and their investment performance, similar to professional fund managers. Each investor has a public profile page, which shows detailed and transparent information on his/her past trading activities, including financial performance, social activities (e.g., posts, comments, likes, and replies), and number of followers.

Our data cover the complete social activity and trading activity histories of all investors in 2016 and 2017. Social activity histories include all posts, comments, replies, and likes together with the exact timestamp of each activity. Trading activity histories include detailed information on each trade. In addition, the data include the dynamics in the leader-follower networks (i.e., when a follower follows or unfollows a leader). In other words, for each link in the network, we know the exact timestamp of the formation and dissolution of the link between the follower and the leader. Finally, the data include each investor's nationality, age, gender, use of a profile image, publication of a biography, trading experience before joining the platform, wealth, income, and desired risk level upon registration.

In our analysis, we consider each investor to be a node in the network. If an investor (follower) follows or autocopies another investor (leader), then this relation is modeled as a directed link between the follower and leader. We set $y_{ijt} = 1$ if a link exists between nodes i and j in period t. We use the data from 2016 to proxy for historical trading performance (e.g., average profit and standard deviation of profit) to guarantee a long-term horizon for followers to evaluate leaders. We examine link formation and dissolution using the leader-follower network in 2017. We define each period at the monthly level. We first illustrate how we sample investors in period 1. We select all leaders who have at least 5 followers (to alleviate the sparsity of the network and exclude some casual investors) in period 1 and stay on the platform in two successive months (i.e., period 1 and period 2), ending with 462 leaders.[7] We then obtain information on all the followers of these leaders, ending with 13,533 unique followers who exist during these two successive periods. As the large

[7]Existence over two successive months is the minimum requirement because the formation network (dissolution network) is constructed by tracking the links added (removed) between two successive periods.

number of nodes can cause computational intractability issues in network analyses, we randomly sample 600 followers out of 13,533, resulting in 1,057 unique investors (because some investors may be both followers and leaders, the total number of investors is less than the summation of leaders and followers). In period 2, as some investors sampled from period 1 may exit the platform and some new investors may join the platform, we first keep those investors who remain on the platform over two successive months (i.e., period 2 and period 3). Then, using the number of investors in period 1 as an anchor, we add investors by randomly sampling from the new investors who join the platform in period 2 and stay on the platform during both period 2 and period 3. We repeat this procedure across all the periods. Table 1 summarizes the network statistics: the number of nodes, the number of links, and the density of the network over time.

Table 1 Network dynamics

Period	Nodes	Links	Density
1	1057	1595	0.0014
2	1053	1588	0.0014
3	1053	1658	0.0015
4	1053	1703	0.0015
5	1053	2025	0.0018
6	1054	2012	0.0018
7	1055	1923	0.0017
8	1055	1857	0.0017
9	1056	1832	0.0016
10	1056	1809	0.0016
11	1057	1742	0.0016
12	1057	1250	0.0011

From Table 1, we see that the number of unique nodes varies slightly across different periods because the number of traders who are both a follower and a leader is different in each period. Although the number of nodes does not vary much across periods, the node set does change since new nodes can join and existing nodes can exit. 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 were conducted on three city-level subsamples consisting of 336, 129, and 146 users. Yan et al. (2015) examine the driving forces behind patients' social network formation and evolution using a subsample consisting of 1,322 individuals. Although the number of nodes is approximately 1,057 in our network analysis, the leader-follower relations between the nodes are described by a 1,057×1,057-dimensional matrix in each period, and there are 12 periods in total. The total number of observations in the network formation model is 11,000,219. However, we acknowledge that this current sample size is still small compared to the large user base on the platform. To further address concerns about the external validity of our findings, we take another random sample and re-estimate the model; our main results are generally consistent. Please refer to Appendix D for more details.

4.2. Variables

Based on the theoretical background presented in Section 2, we consider four groups of variables that may affect the dynamics of follower-leader networks. We next describe in detail how we construct these variables.

4.2.1. Social communication An important feature of social trading platforms is that investors are able to conveniently interact with other investors. For example, eToro allows its users to publish posts, comment on posts, and distribute likes. Investors can, for example, publish posts to broadcast their recent achievements, explain their trading strategies, share their financial advice, or simply communicate with others about recent events. Other investors can comment on these posts to voice their opinions, to request additional information, or to ask for clarification regarding comments.[8] As investors sometimes choose to reply to a comment that is made on their original post by leaving another comment, we label these types of comments "replies" to distinguish them from the original comments. All social interactions are shown on the platform news feed and in the investor's public profile, similar to the widely adopted news feeds used on typical social media platforms such as Facebook or Twitter. eToro does not provide its users with the ability to chat privately. Consequently, the news feed is the only way that users can communicate with each other, and all social activities on the platform are public. Examples of different types of social communication (posts, comments, and replies) are presented in Appendix E.

[8]We do not observe automated comments in our data and are confident that investors manually post these comments. In addition, we do not work under the assumption that all users read all comments. In fact, given the substantial number of comments and the limited attention of investors (Hirshleifer and Teoh 2003), we believe this to be very unlikely. However, if comments are not being read, then the impact on relationships should be zero and statistically insignificant, which would be reflected in our estimation results.

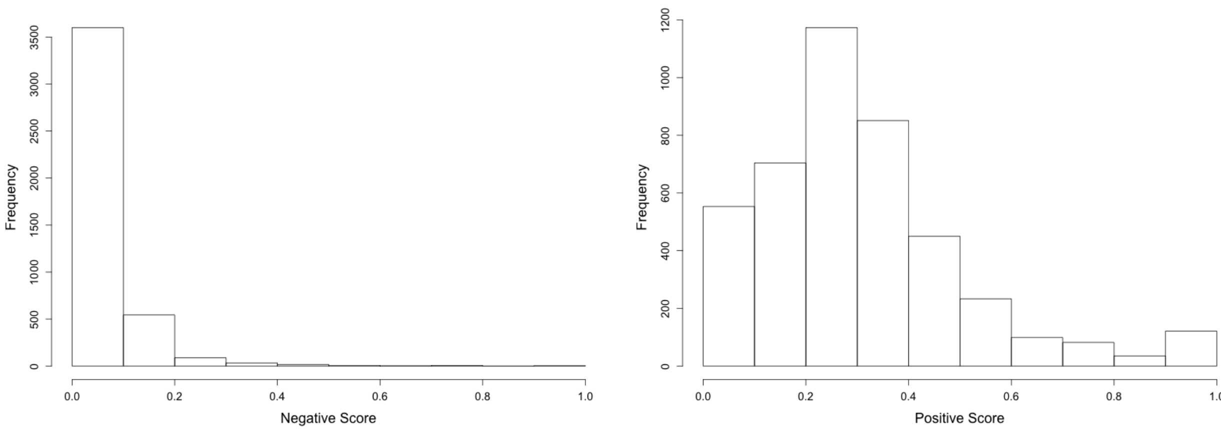
We measure investors' social activities using the following variables. For each investor, we use the total number of posts over period t to measure his/her post writing intensity (*post quantity*). Following the literature (Khern-am-nuai et al. 2018, Cao et al. 2011), we use the number of likes that a post receives to measure its quality. Then, we take the average over all posts in period t as a proxy for an investor's *post quality*. Considering prior evidence that sentiments in user-generated content play an important role in agents' decision-making processes (Xu and Chau 2018), we do not simply examine the effect of the number of comments but instead focus on the sentiments expressed in those comments. Due to the international customer base of the platform, comments are posted in different languages. Thus, to conduct a sentiment analysis, we first use a Google cloud translation application programming interface (API) to translate all comments into English.[9] We then remove stop words, perform word stemming, and use a lexicon-based content analysis to perform our sentiment analysis. We implement the Valence Aware Dictionary and sEntiment Reasoner (VADER), specifically attuned to sentiments expressed on social media (Hutto and Gilbert 2014). VADER has been recently applied in finance and trading and performs as well as individual human raters at matching ground truth (Hutto and Gilbert 2014). The package enables us to label each comment with positive and negative sentiment scores by calculating the percentage of the text that falls in each category. Then, we average the positive and negative sentiment scores for all comments on an investor's post in period t and obtain the variables *comment positive* and *comment negative*, respectively.[10] Figures 3a and 3b show the distribution of negative and positive comment scores, where a higher negative score indicates that a comment contains a larger percentage of negative words. Finally, we count the replies provided by investors to their received comments in period t using the variable *reply*.

4.2.2. Financial performance To measure an investor's financial performance, we first calculate their average daily profit over period t. However, statistics from only one period (month) may not reflect the investor's overall performance. Thus, we use their historical average profit until period t to measure their *average profit*. Similarly, we first calculate

[9]This procedure is consistent with the practices of the platform, which provides a "translate" icon for each post and comment that allows users to view all posts and comments in English.

[10]As Tirunillai and Tellis (2012) found that the effect of negative and positive user-generated content is asymmetric, we include both positive and negative comments in the model.

Figure 3 **Distribution of comment sentiment scores**



(a) **(b)**

the standard deviation of daily returns over period t and then take the average of the historical returns until period t as a proxy for investors' volatility (in line with, e.g., Sirri and Tufano 1998, Huang et al. 2007).

We construct the following variables to proxy for investors' trading strategies. *Holding time* measures the duration from opening to closing a particular position, reflecting the extent to which a trader prefers "day trading" versus a buy-and-hold strategy. We account for investors' portfolio features using the Herfindahl-Hirschman index (*HHI*), a measure of diversification based on the sum of squared portfolio weights (Dorn et al. 2008). A smaller HHI indicates a better diversified portfolio. We include a measure for investors' preferences for investing in lottery-type stocks following Kumar (2009), based on the observation that retail investors are attracted to lottery stocks (i.e., stocks with positively skewed returns) and that this attraction to lottery stocks can increase as a result of social interactions even if investors do not have inherent preferences for skewness (Han et al. 2022). In this vein, Bali et al. (2021) show that social interactions aggravate the lottery anomaly. We define the variable *lottery preference* as the fraction of trades that a given investor executes in lottery-type stocks to all trades by the investor.

4.2.3. Demographics We include some variables to control for homophily based on demographic characteristics. First, we use a dummy variable that takes a value of one if investors come from the same country and zero otherwise (*Nationality*). In a similar fashion, we control for homophily along the investor gender dimension (*Gender*). We also construct a dummy variable to indicate whether investors are in the same age range (*Age*). Social investors may also generate trust by having a detailed profile page that includes their image and/or a biography (Wohlgemuth et al. 2016). Consequently, we include two dummy variables for leaders, *Image* and *Bio*, to denote whether a profile picture[11] or a biography are provided on an investor's profile page. Finally, we incorporate investor characteristics upon registration, including trading experience in years before joining the platform (*Experience*), total wealth in dollars (*Wealth*), annual income in dollars (*Income*), and the reported desired risk level (*Risk*), to capture observed heterogeneity.

4.2.4. Network structure In addition to node characteristics and dyadic covariates, the network structure may affect network evolution through reciprocity and transitivity (Wasserman et al. 1994, Holland and Leinhardt 1971). In our data, the number of mutual links (i.e., $i \to j$ and $j \to i$) is zero, and thus, there is no reciprocity effect, which is intuitive, as it is unlikely that leaders will follow their followers' trading strategies in a social trading network. Therefore, we do not consider reciprocity in our study. However, we incorporate a triadic term to capture potential transitivity. When links $i \to j$ and $j \to k$ exist, the likelihood that a new link $i \to k$ will be formed increases. While triadic effects represent the local hierarchy within the network, we also incorporate the global hierarchy among all nodes within the network—node-level in-degree-related popularity and out-degree-related activity (Hunter et al. 2008). It is possible that the effects of popularity and activity are different for a leader and a follower. Therefore, we distinguish between leader and follower nodes. For a leader node, we measure how in-degree-related popularity and out-degree-related activity affect the propensity of others forming a link with that leader. A positive in-degree-related popularity effect indicates that others find it attractive to follow a leader with more in-ties (in-links). Similarly, a positive effect from out-degree-related activity means that others find it attractive to follow a leader with more out-ties (out-links). For follower nodes, we measure the propensity of the follower to follow someone else (form a link). Specifically, a positive estimate for in-degree-related popularity (out-degree-related activity) implies that followers with more in-ties (out-ties) are more likely to follow leaders. For all the above variables, we provide brief definitions and summary statistics in Table 2.

[11]Our dataset does not contain more detailed information on the picture, for example, whether it is a symbolic image or a photo showing a real person. Due to the anonymous nature of the data, we are not able to collect this information and merge it with our dataset. We study a random sample of the profile pictures of eToro users to analyze how many fantasy pictures, on average, are used. Our analysis of slightly over 500 randomly selected profile pages shows that approximately 80% of them contain photos showing a real person (that is not a well-known celebrity).

Table 2 Data description and statistics

Variable	Description	Mean	Std. Dev.	Min.	Max.
Social communication					
Post quantity	The number of posts made by investors	23.77	146.39	0	5406
Post quality	The number of likes that an investor's posts receive	4.28	9.59	0	236
Comment positive	The percentage of positive words in the comments that an investor receives	0.31	0.21	0	1
Comment negative	The percentage of negative words in the comments that an investor receives	0.06	0.08	0	1
Reply	The number of replies to comments	20.57	62.49	1	1551
Financial performance					
Average profit	The average profit	0.0008	0.04	-0.55	0.97
Std. dev. profit	The standard deviation of profit	0.05	0.11	0	3.11
Holding time	The duration of a particular position from opening to closing	14.00	38.23	0	749.21
HHI	Herfindahl-Hirschman index of portfolio diversification	0.19	0.31	0	1
Lottery preference	Fraction of trades in lottery-type stocks	0.02	0.06	0	1
Demographics					
Gender	Dummy =1 if both investors are females or both are males and 0 otherwise	0.83	0.38	0	1
Age	Dummy =1 if both investors are in the same age range and 0 otherwise	0.29	0.45	0	1
Nationality	Dummy =1 if both investors are from the same country and 0 otherwise	0.07	0.26	0	1
Image	Dummy =1 if the investor uploads a profile picture and 0 otherwise	0.48	0.50	0	1
Bio	Dummy =1 if the investor uploads a biography to his/her profile and 0 otherwise	0.22	0.41	0	1
Experience	Trading experience before joining the platform at the time of registration	1.53	1.07	0	3
Wealth	Reported wealth at the time of registration	105,541	251,475.2	10,000	2,000,000
Income	Reported annual income at the time of registration	113,644	209,495.7	10,000	2,000,000
Risk	Reported risk at the time of registration	24.35	15.41	3	48
Network structure					
In-degree popularity	The number of incoming ties in an investor's social network	1.70	10.68	0	371
Out-degree activity	The number of outgoing ties initiated by an investor in his/her social network	1.70	2.18	0	43
Transitivity	The number of triadic closures for each node	0.21	1.42	0	42

5. Results

We apply the STERGM with Chamberlain correlated random effects to investigate link formation and dissolution in social trading. We study the determinants presented in Section 4.2 and estimate the coefficients that best fit our model using a maximum likelihood procedure. We provide several additional robustness checks in Section 6 and in the Appendix.

Table 3 summarizes our main estimation results separately for link formation and dissolution. A positive coefficient in the formation model indicates an increased probability of forming a new link, whereas a positive coefficient in the dissolution model indicates a positive effect on link duration (i.e., the link is less likely to dissolve). We distinguish between the variables for a leader's account and those for a follower's account since these actors play different roles in the network evolution. Model 1 is the baseline model without social communication. Model 2 is the full model that includes all explanatory variables. Hereafter, we discuss our main results based on Model 2, which includes social communication, financial performance, demographic characteristics, and network structure. Overall, we find evidence that hard information (financial performance) and soft information (social communication and demographic characteristics) play different roles in the link formation and dissolution processes. We discuss the findings for each group of variables in detail.

5.1. Social communication

Table 3 shows positive coefficients for a leader's post quantity for both link formation (coefficient of 0.3162, $p < 0.01$) and link dissolution (coefficient of 0.1163, $p < 0.01$). Similarly, the coefficients for a leader's post quality during link formation (coefficient of 0.5097,

Table 3 Estimation results

	Model 1		Model 2	
Variable	Formation	Dissolution	Formation	Dissolution
Social communication				
Leader's post quantity			0.3162*** (0.0268)	0.1163*** (0.0283)
Leader's post quality			0.5097*** (0.0250)	0.0439** (0.0206)
Leader's number of replies			0.0238 (0.0277)	−0.0105 (0.0310)
Leader's comment received positive score			1.1122*** (0.1397)	0.5754*** (0.1458)
Leader's comment received negative score			−4.7688*** (0.7525)	−2.1181*** (0.4715)
Follower's post quantity			0.0655 (0.0474)	−0.5311*** (0.0536)
Follower's post quality			−0.3348*** (0.0797)	−0.0044 (0.0749)
Financial performance				
Leader's average profit	0.0536*** (0.0114)	0.0383** (0.0175)	0.0903*** (0.0136)	0.0707*** (0.0183)
Leader's std. dev. profit	−0.3113 (0.4760)	−1.3945* (0.7306)	−2.1610*** (0.5567)	−2.6485*** (0.7615)
Follower's average profit	0.0031 (0.0116)	0.0416*** (0.0124)	0.0012 (0.0114)	0.0440*** (0.0120)
Follower's std. dev. profit	−0.8744* (0.5144)	−8.8636*** (0.5827)	−0.6665 (0.5086)	−8.1553*** (0.5582)
Leader's average holding time	0.1740*** (0.0416)	−0.0106 (0.0644)	0.3465*** (0.0446)	0.0652 (0.0661)
Leader's lottery preference	0.5191 (0.3220)	1.1677*** (0.3937)	0.2019 (0.3619)	1.0743*** (0.3948)
Leader's HHI	−0.8337*** (0.0913)	−0.0956 (0.0917)	−0.6094*** (0.0905)	−0.1435 (0.0929)
Demographics				
Nationality	0.8856*** (0.0691)	0.3435*** (0.0790)	0.7456*** (0.0704)	0.3408*** (0.0797)
Age	0.1028** (0.0508)	0.0456 (0.0539)	0.1221** (0.0505)	0.0519 (0.0543)
Homophily (male)	0.9723*** (0.1358)	−0.1302 (0.1152)	0.9862*** (0.1359)	−0.1510 (0.1164)
Homophily (female)	−0.8507** (0.4164)	0.2365 (0.3145)	−0.8602** (0.4166)	0.2906 (0.3177)
Image	2.9703*** (0.4171)	0.6320 (0.4759)	2.4784*** (0.4238)	0.7280 (0.4652)
Bio	3.6225*** (0.1778)	−0.1228 (0.1870)	2.9204*** (0.1869)	−0.1499 (0.1886)
Experience	−0.0511 (0.0372)	0.1489*** (0.0422)	−0.0444 (0.0371)	0.1460*** (0.0404)
Wealth	0.0217 (0.0328)	0.0795** (0.0364)	0.0243 (0.0324)	0.0618* (0.0347)
Income	−0.0027 (0.0396)	0.0205 (0.0429)	−0.0090 (0.0393)	0.0239 (0.0409)
Risk	−0.0922* (0.0492)	−0.0624 (0.0553)	−0.0839* (0.0488)	−0.0397 (0.0528)
Network structure				
Leader's popularity	0.0157*** (0.0003)	0.0021*** (0.0003)	0.0082*** (0.0004)	0.0014*** (0.0004)
Leader's activity	−0.1376*** (0.0172)	0.0312*** (0.0117)	−0.0353*** (0.0123)	0.0380*** (0.0119)
Follower's popularity	−0.0726*** (0.0130)	−0.0201*** (0.0069)	−0.0546*** (0.0127)	−0.0089 (0.0065)
Follower's activity	0.0009 (0.0070)	−0.0838*** (0.0088)	0.0026 (0.0071)	−0.0769*** (0.0088)
Transitivity	0.0938*** (0.0267)	−0.0057 (0.0666)	0.0820*** (0.0256)	−0.0218 (0.0656)
Constant	−19.2557*** (5.2126)	0.8493 (0.6597)	−15.8578*** (5.6635)	2.0938*** (0.6446)
Log Likelihood	-14,418.19	-9,345.29	-13,620.36	-9,152.26
Observations	11,000,219	19,744	11,000,219	19,744

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of the posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

$p < 0.01$) and link dissolution (coefficient of 0.0439, $p < 0.05$) are also positive, indicating that the propensity to form a link and maintaining existing links increases as leaders publish a larger number of posts of higher quality.[12] This observation is consistent with the notion that high-quality posts provide useful information and increase the transparency of a leader's investment strategy, which in turn increases trust in the leader, attracts more incoming links, and helps maintain existing links. Overall, this notion is consistent with the

[12] Also of particular interest is the interplay between post quality and post quantity. A robustness check including a quantity-quality interaction term indicates that post quality has a positive moderating effect on post quantity in the link formation process. The interaction term is not significant in the link dissolution process. Please refer to Appendix F for more details.

literature that argues that communication can increase trustworthiness and trust (Kaiser and Berger 2021).

The coefficients on positive comments for link formation and dissolution are significantly positive (coefficients of 1.1122 with $p < 0.01$ and 0.5754 with $p < 0.01$, respectively), indicating a higher probability that a follower will form or maintain a link if the leader receives more positive comments. Similarly, negative and significant estimates for negative comments indicate that negative comments are associated with a lower probability of link formation or a shorter link duration (coefficients of -4.7688 with $p < 0.01$ and -2.1181 with $p < 0.01$, respectively). The different magnitudes of the coefficients further indicate that the impacts of negative and positive comments are asymmetric and that negative comments are particularly relevant in online contexts, consistent with previous evidence from social media (Xu and Chau 2018).[13]

Comparing the estimates for the effect of social communication in link formation and dissolution, we find that social communication plays an important role in both processes. Thus, communication on social trading platforms seems to have an economic impact on leaders given that the links in the leader-follower network are tied to cash flow and directly affect the compensation that a leader receives from the platform. Communication, a type of soft information, not only helps leaders attract new links but also helps to maintain existing links. Our coefficients indicate that the effect of communication is stronger for link formation than for link dissolution, which may be explained by the fact that posts can affect link formation through an additional channel (i.e., the attention channel) that is less relevant for dissolution (see, e.g., Barber and Odean 2008, for a similar argument in the financial markets context).

Regarding the social activities of followers, we find that the probability of establishing new links decreases as the quality of posts increases (coefficient of -0.3348, $p < 0.1$). Intuitively, followers with higher-quality posts may have greater financial knowledge and expertise and consequently may be more likely to trade by themselves instead of following others.

[13]The t-test of the difference in coefficients between negative comments and positive comments is 7.85 in the link formation model and 5.57 in the link dissolution model, indicating that they are significantly different.

5.2. Financial performance

Turning to financial performance, we find that a leader's average profit tends to attract followers for both link formation and link duration (coefficients of 0.0903 with $p < 0.01$ and 0.0707 with $p < 0.01$, respectively). In addition, greater volatility in the leader's financial performance is negatively associated with link formation and link duration (coefficients of -2.1610 with $p < 0.01$ and -2.6485 with $p < 0.01$, respectively). Overall, these observations are in line with prominent findings from the mutual fund flow literature that investors chase past performance (Barber et al. 2016) and with previous evidence on social trading (Doering et al. 2015).

Next, we focus on the performance of followers. We find that they are insensitive to their past profit on the platform when choosing a new link, as the coefficients on a follower's average profit and the standard deviation of profit are insignificant in the link formation model (coefficients of 0.0012, $p = 0.91$ and -0.6665, $p = 0.19$, respectively). However, once a link is formed, followers tend to be more likely to maintain the link if they have a higher average profit (coefficient of 0.0440, $p < 0.01$) and a lower volatility level (coefficient of -8.1553, $p < 0.01$).

Finally, we turn to the trading strategies of leaders and find that followers prefer leaders who tend to follow diversified buy-and-hold strategies and are more likely to establish links with those leaders. Once links are established, investing in lottery-like stocks is associated with a longer link duration.

5.3. Demographics and network structure

With respect to demographic characteristics, we find—most notably—that followers tend to establish and maintain links with leaders who have the same nationality (coefficients of 0.7456, $p < 0.01$ and 0.3408, $p < 0.01$), which is consistent with studies on peer-to-peer credit markets (Lin and Viswanathan 2016). These effects may be driven by language barriers or cultural differences. We also find evidence in support of age homophily (coefficient of 0.1221, $p < 0.05$) and gender homophily among male investors (coefficient of 0.9862, $p < 0.01$) during link formation. However, female followers are more likely to form a link with male leaders (coefficient of -0.8602, $p < 0.05$). In line with previous findings from the literature (Wohlgemuth et al. 2016), we observe that the presence of a picture on a leader's profile page and of a biographical description significantly increases the likelihood that followers will form a new link (coefficients of 2.4784, $p < 0.01$ and 2.9204, $p < 0.01$, respectively). The disclosure of a profile picture or a biography may increase the perceived trustworthiness of the leader and therefore the likelihood of new links.

Next, we consider a follower's experience, wealth, and risk preference. Our results show that followers with a higher risk score are less likely to form a new link (coefficient of -0.0839, $p < 0.1$), whereas followers with a higher experience level or wealth level tend to maintain their existing links (coefficients of 0.1460, $p < 0.01$ and 0.0618, $p < 0.1$, respectively).

Comparing the link formation and dissolution processes, we find strong differences in the impact of the demographic characteristics of leaders. This observation is intuitive in the sense that once followers have considered the demographic characteristics and established a link, there is no need to consider them again, as demographics remain stable over time. Thus, other factors, such as financial performance and social communication, become more relevant.

Finally, we briefly discuss the variables that capture the network structure. A leader's popularity (in-ties) increases his/her propensity to attract followers who form new links and to maintain existing links (coefficients of 0.0082, $p < 0.01$ and 0.0014, $p < 0.01$, respectively), indicating preferential attachment. In contrast, the coefficient on a leader's activity (out-ties) is significantly negative (coefficient of -0.0353, $p < 0.01$), indicating that leaders who follow other investors are less attractive to potential followers. Followers with higher popularity are less likely to form new links (coefficient of -0.0546, $p < 0.01$), whereas those with higher levels of activity are less likely to maintain existing links (coefficient of -0.0769, $p < 0.01$). We also find a significant transitivity effect (coefficient of 0.0820, $p < 0.01$), indicating that the presence of a link from i to j and from j to k increases the likelihood of a direct link between i and k being formed.

5.4. Heterogeneous effects across follower age

Prior studies have found that individuals who are younger in age are more likely to blog, visit social network sites, and rely on social media in their decision-making than those who are older in age (Chou et al. 2009). Thus, in this subsection, we further scrutinize the implications related to age. Considering that social trading is a novel way to participate in financial markets that may particularly attract younger individuals, it is natural to ask whether our findings hold across all age groups. We thus examine whether the impact of the social communication and financial performance variables varies across investors by age group.

We split the dataset into two subsamples based on followers' age ranges. The first group includes followers between 18 and 44 years of age, and the second group includes followers who are older than 44 years. We again apply the STERGM with Chamberlain correlated random effects and summarize the estimation results in Table 4. Younger followers are rather sensitive to positive and negative comments in the link dissolution process (coefficients of 0.6684, $p < 0.01$ and -2.3591, $p < 0.01$, respectively), whereas the effects of comments are not significant for older followers (coefficients of 0.3637, $p = 0.19$ and -1.5742, $p = 0.11$, respectively). Interestingly, we observe that a leader's post quantity increases the probability of younger followers maintaining existing links (coefficient of 0.1302, $p < 0.01$), while the post quality becomes insignificant. At the same time, however, for older followers, it is a leader's post quality (coefficient of 0.1076, $p < 0.01$) rather than post quantity that increases the probability of maintaining existing links.

Table 4 Estimation results: Heterogeneous effects by follower age

	Group 1		Group 2	
Variable	Formation	Dissolution	Formation	Dissolution
Social communication				
Leader's post quantity	0.2975*** (0.0317)	0.1302*** (0.0334)	0.3624*** (0.0501)	0.0770 (0.0537)
Leader's post quality	0.5126*** (0.0296)	0.0217 (0.0241)	0.5041*** (0.0471)	0.1076*** (0.0407)
Leader's number of replies	0.0517 (0.0326)	−0.0220 (0.0364)	−0.0400 (0.0531)	0.0222 (0.0596)
Leader's comment received positive score	1.1623*** (0.1662)	0.6684*** (0.1721)	0.9892*** (0.2590)	0.3637 (0.2780)
Leader's comment received negative score	−4.2946*** (0.8679)	−2.3591*** (0.5416)	−6.1927*** (1.5113)	−1.5742 (0.9766)
Follower's post quantity	0.1476*** (0.0528)	−0.5766*** (0.0599)	−0.1739 (0.1121)	−0.3404*** (0.1270)
Follower's post quality	−0.4948*** (0.0893)	0.0136 (0.0801)	0.2838* (0.1720)	−0.1798 (0.2439)
Financial performance				
Leader's average profit	0.0914*** (0.0166)	0.0741*** (0.0222)	0.0885*** (0.0238)	0.0603* (0.0338)
Leader's std. dev. profit	−2.4100*** (0.6752)	−2.9491*** (0.9208)	−1.6267* (0.9773)	−1.7755 (1.4176)
Follower's average profit	−0.0021 (0.0124)	0.0348*** (0.0132)	0.0662** (0.0275)	0.0880*** (0.0301)
Follower's std. dev. profit	−0.1756 (0.5337)	−7.8365*** (0.6058)	−5.0396** (1.9718)	−10.4074*** (1.4808)
Leader's average holding time	0.3505*** (0.0532)	0.0917 (0.0762)	0.3371*** (0.0828)	−0.0311 (0.1390)
Leader's lottery preference	0.1726 (0.4300)	1.4006*** (0.4611)	0.3151 (0.6711)	0.1110 (0.7660)
Leader's HHI	−0.6326*** (0.1070)	−0.0990 (0.1082)	−0.5813*** (0.1696)	−0.2657 (0.1834)
Demographics	Yes		Yes	
Network structure	Yes		Yes	
Log Likelihood	-9,795.51	-6,779.52	-3,762.49	-2,327.82
Observations	8,015,758	14,476	1,658,708	5,268

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

6. Robustness tests

In this section, we present a series of robustness checks.

6.1. Two-stage selection model

Since the total number of leaders on the platform is large and followers have limited attention, it is possible that some leaders are more visible than others, which might affect whether potential followers follow any particular leader. Thus, in this robustness check, we develop a two-stage selection model that attempts to accurately model the link formation process. In the first stage, to account for different exposure to leaders' profiles among followers, we model the probability of followers being aware of leader j as

$$Pr(A_{jt}=1)=\frac{e^{\gamma z_{jt}}}{1+e^{\gamma z_{jt}}}, \tag{8}$$

where z_{jt} denotes the number of followers of leader j in period t. We choose the number of followers in the first step because, first, the finance literature provides substantial evidence that herding is a relevant behavioral trait in financial markets (Devenow and Welch 1996). In the context of social trading, Gemayel and Preda (2018b) show that the scopic regime can increase the extent of herding. Given that the number of followers already reflects financial performance to some degree (i.e., investors with poor performance are less likely to have many followers), we argue that investors will start their filtering decision based on fewer criteria in order to simplify their decision process as much as possible. Second, we refer to the concept of preferential attachment in social networks from the IS literature, and leaders with large follower bases should be more likely to attract additional followers (Neyman and Scott 1948).

In the second stage, followers decide whether to form a link with the leader based on the hard information and soft information that they see on the leader's profile page. The second stage is identical to the main model; the probability of follower i following leader j during period t is modeled as

$$Pr(y_{ijt}=1)=\frac{e^{(\boldsymbol{\alpha} X_{it-1}+\boldsymbol{\beta} W_{jt-1}+\boldsymbol{\lambda} V_{ij}+\boldsymbol{\tau} C_i+\eta_i)}}{1+e^{(\boldsymbol{\alpha} X_{it-1}+\boldsymbol{\beta} W_{jt-1}+\boldsymbol{\lambda} V_{ij}+\boldsymbol{\tau} C_i+\eta_i)}}. \tag{9}$$

Taking step 1 and step 2 together, we derive the overall likelihood as follows:

$$\begin{aligned} L_{ijt} &= y_{ijt}\times(Pr(A_{jt}=1)\times Pr(y_{ijt}=1|A_{jt}=1)) \\ &+(1-y_{ijt})\times(Pr(A_{jt}=1)\times Pr(y_{ijt}=0|A_{jt}=1)+Pr(A_{jt}=0)) \end{aligned} \tag{10}$$

The overall log-likelihood value is further written as follows:

$$TLL(\gamma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\tau})=\sum_{i=1}^{I_t}\sum_{t=1}^{T_i}\sum_{j=1}^{J_{it}}\ln(L_{ijt}) \tag{11}$$

where I_t is the number of followers in period t, T_i is the number of periods that follower i exists on the platform, and J_{it} is the number of leaders that follower i can potentially follow in period t. We estimate our proposed two-stage selection model by maximizing the overall log-likelihood value. For computational tractability, following Heckman and Singer (1984), we apply a non-parametric approach to estimate the follower's random effects after controlling for the time average of the follower's time-variant observables. The estimation results are reported in Table 5. We find that potential followers are more likely to be aware of a leader if that leader has a larger number of followers of his/her account (coefficient of 0.3483, $p < 0.01$). The results from the second stage are generally consistent with the findings from the main model.

6.2. Alternative sentiment dictionary

In the main analysis, we use the widely applied VADER sentiment dictionary to calculate the sentiment scores for comments. It is possible that different sentiment dictionaries generate different sentiment scores, which might affect the estimated effects of positive and negative scores for a leader's comments on link formation and dissolution. Therefore, in this robustness check, we first apply an alternative sentiment dictionary to calculate the sentiment scores for comments. In particular, we adopt the Harvard General Inquirer[14] dictionary, another widely adopted dictionary for extracting sentiment from social media, to perform the sentiment analysis (Ammann and Schaub 2021). After that, we re-estimate the STERGM with Chamberlain correlated random effects. The estimation results are presented in Table 6. We find that the results are generally consistent with the findings in the main model.

6.3. Alternative measure of financial risk

The volatility of performance is a symmetric measure of risk that takes into account both positive and negative deviations from the mean. Investors may, however, be mostly concerned with extreme negative profit outcomes, i.e., with large losses. Consequently, we consider the maximum drawdown (MDD) as an alternative risk measure that accounts for large losses. The MDD measures the monthly maximum observed loss in a leader's daily profit. Similar to the standard deviation, the MDD is a widely used risk measure (Cvitanić and Karatzas 1999, de Melo Mendes and Lavrado 2017). We re-estimate the STERGM

[14] `http://www.wjh.harvard.edu/~inquirer/`, last accessed Jan. 8, 2021.

Table 5 Estimation results of the two-stage selection model

Variable	Formation	
First stage		
Num. of followers	0.3483***	(0.0430)
Second stage		
Social communication		
Leader's post quantity	0.2297***	(0.0407)
Leader's post quality	0.2674***	(0.0392)
Leader's number of replies	0.0927**	(0.0431)
Leader's comment received positive score	1.1130***	(0.2025)
Leader's comment received negative score	−3.4844***	(0.9125)
Follower's post quantity	−0.0424	(0.0644)
Follower's post quality	−0.1392	(0.1166)
Financial performance		
Leader's average profit	0.9184***	(0.1867)
Leader's std. dev. profit	0.5326	(0.6810)
Follower's average profit	−0.0414	(0.1270)
Follower's std. dev. profit	0.3549	(0.5457)
Leader's average holding time	0.0875	(0.0748)
Leader's lottery preference	0.8159	(0.4981)
Leader's HHI	−0.9456***	(0.1388)
Demographics		
Nationality	1.1929***	(0.1285)
Age	0.1773**	(0.0759)
Homophily (male)	0.9895***	(0.1729)
Homophily (female)	−1.1487**	(0.5223)
Image	−1.4969***	(0.5574)
Bio	1.4299***	(0.2641)
Experience	−0.5724	(0.3572)
Wealth	−0.4306	(0.3099)
Income	0.0968	(0.3723)
Risk	−1.2409***	(0.4730)
Network structure		
Leader's popularity	4.8850***	(0.4789)
Leader's activity	−3.0419**	(1.5117)
Follower's popularity	−17.6188***	(2.0538)
Follower's activity	30.8006***	(2.4906)
Transitivity	0.9729***	(0.0922)
Constant	17.6665***	(4.6990)
Log Likelihood	-13,426.99	
Observations	11,000,219	

Notes: The number of posts, the quality of posts (the number of likes) and the number of replies are log-transformed. Average profit and std. dev. profit are scaled by a factor of 10. Average holding time is scaled by a factor of 1/100. Popularity and activity are scaled by a factor of 1/100. Wealth, income and risk are log-transformed and then scaled by a factor of 1/10. Experience is scaled by a factor of 1/10.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

with Chamberlain correlated random effects; the estimation results are presented in Table 7. The coefficients on key determinants are generally consistent with our main results, and the MDD shows the expected negative effect.

7. Conclusions

Social trading is a novel form of trading that combines online brokerage with traditional social media features. It has attracted a large number of investors and increasing attention from practitioners and academia. Social trading allows investors to seek financial advice from their peers, to observe their peers' trading strategies, and to directly follow other

Table 6 Estimation results using an alternative sentiment dictionary

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3177***	(0.0262)	0.1213***	(0.0282)
Leader's post quality	0.4695***	(0.0243)	0.0356*	(0.0208)
Leader's number of replies	−0.0151	(0.0273)	−0.0227	(0.0311)
Leader's comment received positive score	1.4295***	(0.1617)	0.7911***	(0.1834)
Leader's comment received negative score	−0.7144	(0.4436)	−1.0227**	(0.4143)
Follower's post quantity	0.0648	(0.0474)	−0.5332***	(0.0535)
Follower's post quality	−0.3354***	(0.0797)	−0.0036	(0.0748)
Financial performance				
Leader's average profit	0.0913***	(0.0135)	0.0706***	(0.0183)
Leader's std. dev. profit	−2.0776***	(0.5485)	−2.6433***	(0.7622)
Follower's average profit	0.0002	(0.0115)	0.0425***	(0.0119)
Follower's std. dev. profit	−0.6806	(0.5085)	−8.0848***	(0.5558)
Leader's average holding time	0.3416***	(0.0441)	0.0737	(0.0661)
Leader's lottery preference	0.0585	(0.3629)	1.0407***	(0.3940)
Leader's HHI	−0.5866***	(0.0904)	−0.1197	(0.0929)
Demographics				
Nationality	0.7461***	(0.0703)	0.3395***	(0.0797)
Age	0.1165**	(0.0506)	0.0494	(0.0543)
Homophily (male)	0.9602***	(0.1359)	−0.1622	(0.1164)
Homophily (female)	−0.8334**	(0.4166)	0.3173	(0.3180)
Image	2.4534***	(0.4237)	0.7248	(0.4656)
Bio	2.9068***	(0.1873)	−0.1510	(0.1885)
Experience	−0.0450	(0.0371)	0.1445***	(0.0402)
Wealth	0.0220	(0.0324)	0.0603*	(0.0346)
Income	−0.0046	(0.0393)	0.0291	(0.0408)
Risk	−0.0893*	(0.0487)	−0.0395	(0.0526)
Network structure				
Leader's popularity	0.0086***	(0.0004)	0.0014***	(0.0183)
Leader's activity	−0.0347***	(0.0123)	0.0382***	(0.7622)
Follower's popularity	−0.0534***	(0.0126)	−0.0085	(0.0119)
Follower's activity	0.0024	(0.0070)	−0.0763***	(0.5558)
Transitivity	0.0798***	(0.0257)	−0.0203	(0.0661)
Constant	−14.2699***	(5.4388)	2.1800***	(0.6442)
Log Likelihood	-13,643.37		-9,158.25	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

investors in real time. Thus, inexperienced retail investors may benefit from their peers, while experienced investors are able to provide signals and earn additional income. Due to the monetary aspects involved in these leader-follower relationships, network evolution follows a distinct pattern that differs from that of traditional social media platforms. In particular, link dissolution is an important part of social trading.

We study a dynamic social trading network using a STERGM and examine how various factors, such as social communication, financial performance, and demographics, affect the link formation and dissolution processes. We show that social communication, financial performance, and demographics have different implications for the link formation and dissolution processes. Followers consider financial performance, social communication, and demographics when deciding whom to follow (link formation process). However, once a link

Table 7 Estimation results with the MDD as the measure of financial risk

Variable	Formation		Dissolution[a]	
Social communication				
Leader’s post quantity	0.3093***	(0.0267)	0.1001***	(0.0279)
Leader’s post quality	0.5150***	(0.0251)	0.0450**	(0.0207)
Leader’s number of replies	0.0242	(0.0279)	−0.0071	(0.0310)
Leader’s comment received positive score	1.1092***	(0.1396)	0.5300***	(0.1459)
Leader’s comment received negative score	−4.7028***	(0.7531)	−2.0796***	(0.4720)
Follower’s post quantity	0.0652	(0.0475)	−0.5491***	(0.0548)
Follower’s post quality	−0.3370***	(0.0798)	−0.0298	(0.0755)
Financial performance				
Leader’s average profit	0.0369***	(0.0052)	0.0199***	(0.0068)
Leader’s MDD	0.0221	(0.0219)	−0.0625**	(0.0247)
Follower’s average profit	−0.0019	(0.0106)	0.0574***	(0.0114)
Follower’s MDD	−0.1071***	(0.0254)	−0.2378***	(0.0262)
Leader’s average holding time	0.3761***	(0.0434)	0.0397	(0.0666)
Leader’s lottery preference	0.1687	(0.3651)	0.9932**	(0.3947)
Leader’s HHI	−0.5842***	(0.0906)	−0.1560*	(0.0927)
Demographics				
Nationality	0.7488***	(0.0704)	0.3282***	(0.0797)
Age	0.1173**	(0.0506)	0.0470	(0.0543)
Homophily (male)	0.9847***	(0.1361)	−0.1252	(0.1167)
Homophily (female)	−0.8613**	(0.4167)	0.2514	(0.3172)
Image	2.5500***	(0.4236)	0.7448	(0.4688)
Bio	2.8402***	(0.1833)	−0.1102	(0.1883)
Experience	−0.0453	(0.0374)	0.1181***	(0.0426)
Wealth	0.0333	(0.0328)	0.0915**	(0.0368)
Income	−0.0144	(0.0397)	0.0176	(0.0433)
Risk	−0.0899*	(0.0492)	−0.1038*	(0.0558)
Network structure				
Leader’s popularity	0.0083***	(0.0004)	0.0016***	(0.0004)
Leader’s activity	−0.0369***	(0.0125)	0.0400***	(0.0119)
Follower’s popularity	−0.0563***	(0.0129)	−0.0112	(0.0069)
Follower’s activity	0.0048	(0.0070)	−0.0714***	(0.0090)
Transitivity	0.0840***	(0.0257)	−0.0316	(0.0660)
Constant	1.7241	(9.2812)	2.9661***	(0.6828)
Log Likelihood	-13,617.23		-9,231.18	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, MDD, wealth, income, and risk are log-transformed. Average profit is scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

is formed, demographic characteristics become less important, as followers mainly focus on leaders’ financial performance as well as social communication to decide whether to sustain the link or not (link dissolution process). Focusing on the different types of social communication, we find that the quality and quantity of a leader’s posts increase the likelihood of followers forming a new link and sustaining existing links. Followers are less likely to form new links or sustain existing links with leaders who receive more negative comments. Leaders who receive more positive comments are more likely to attract new followers and to keep existing followers. In addition, the impacts of negative and positive comments are asymmetric. Negative comments have a larger impact than positive comments on the link formation and dissolution processes.

Our study contributes to the growing literature on social trading by first modeling the dynamics of leader-follower networks. Our granular data allow us to thoroughly examine the implications of various factors on the link formation and dissolution processes. Our study also contributes to a better understanding of how hard information (e.g., financial performance) and, in particular, soft information (e.g., social communication) affect leader-follower network evolution in the social trading context.

Our study has practical managerial implications. We document link formation and link dissolution processes and thereby broaden and deepen our understanding of leader-follower network evolution in social trading. Social trading platforms were established in the aftermath of the global financial crisis to provide retail investors with an alternative to traditional wealth management in response to the eroding trust in financial markets following the crisis (Doering et al. 2015). Although social trading platforms provide a high level of informational transparency, investors face new challenges in building trust, in particular because most investors on the platform are individuals who lack institutional endorsements and the relationship is online with a more and less anonymous mass. In our study, we find that social communication plays an important role in leaders' ability to convince potential followers to follow their trading strategies and existing followers to sustain their links. Social communication is effective in building trust among investors on social trading platforms. Therefore, leaders have to use caution when making posts. If not used properly, those posts receiving more negative comments can backfire, which can reduce link formation and increase link dissolution. Negative comments have a larger impact on link formation and dissolution than positive comments. In addition, leaders should make high-quality posts, as high-quality posts both increase link formation and reduce link dissolution. Thus, by communicating in a balanced manner, leaders can attract new followers to follow their trading strategies and encourage existing followers to sustain their links. Our results can guide leaders on when and how to communicate with followers on social trading platforms. Given the importance of social communication in the evolution of leader-follower networks with real money flow, social trading platforms should carefully regulate social communication to sustain a healthy ecosystem.

Our analysis has some caveats. First, we abstract away from the potential dependence between a follower's following decisions and a leader's subsequent trading strategies. Thus, we assume that a leader's trading strategy does not change regardless of any given follower's decision. However, such dependencies do exist (Pelster and Hofmann 2018) and may affect network evolution. Future research may aim to study these coevolution effects in more detail. Second, we focus only on a follower's decision regarding whether to follow a leader, without considering the monetary amount assigned to the link. Future research can consider the amount of money allocated to a link in a weighted directed (leader-follower) network. Such adjustments may reveal additional insights regarding the role of social communication in network evolution. Finally, we incorporate Chamberlain correlated random effects in our model to address the omitted variable issue; for example, some leaders might advertise themselves on other social media platforms, which might affect link formation. Such behavior indicates a leader's general propensity to engage in advertising their trading on social media, and such general propensities are rather stable over time. We acknowledge that the Chamberlain correlated random effects capture only the time-invariant unobservables and, for instance, do not account for the possibility that the advertising activities of investors on other social platforms change over time.

References

Agarwal, V., N. D. Daniel, N. Y. Naik. 2009. Role of managerial incentives and discretion in hedge fund performance. *The Journal of Finance* **64**(5) 2221–2256.

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

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

Baker, W. E., R. R. Faulkner, G. A. Fisher. 1998. Hazards of the market: The continuity and dissolution of interorganizational market relationships. *American sociological review* **63**(2) 147–177.

Bali, T. G., D. Hirshleifer, L. Peng, Y. Tang. 2021. Attention, social interaction, and investor attraction to lottery stocks. Tech. rep., National Bureau of Economic Research.

Barber, B., X. Huang, T. Odean. 2016. Which factors matter to investors? evidence from mutual fund flows. *The Review of Financial Studies* **29**(10) 2600–2642.

Barber, B. M., T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies* **21**(2) 785–818.

Cao, Q., W. Duan, Q. Gan. 2011. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems* **50**(2) 511–521.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *The Review of Economic Studies* **47**(1) 225–238.

Chou, W.Y.S., Y.M. Hunt, E.B. Beckjord, R.P. Moser, B.W. Hesse. 2009. Social media use in the united states: implications for health communication. *Journal of medical Internet research* **11**(4) e1249.

Coval, J. D., T. J. Moskowitz. 1999. Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance* **54**(6) 2045–2073.

Cvitanić, J., I. Karatzas. 1999. On dynamic measures of risk. *Finance and Stochastics* **3**(4) 451–482.

de Melo Mendes, B. V., R. C. Lavrado. 2017. Implementing and testing the maximum drawdown at risk. *Finance Research Letters* **22** 95–100.

Devenow, A., I. Welch. 1996. Rational herding in financial economics. *European economic review* **40**(3-5) 603–615.

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

Dorfleitner, G., L. Fischer, C. Lung, P. Willmertinger, N. Stang, N. Dietrich. 2018. To follow or not to follow–an empirical analysis of the returns of actors on social trading platforms. *The Quarterly Review of Economics and Finance* **70** 160–171.

Dorn, D., G. Huberman, P. Sengmueller. 2008. Correlated trading and returns. *The Journal of Finance* **63**(2) 885–920.

Duarte, J., S. Siegel, L. Young. 2012. Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies* **25**(8) 2455–2484.

Gemayel, R., A. Preda. 2017. 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., A. Preda. 2018a. Does a scopic regime erode the disposition effect? evidence from a social trading platform. *Journal of Economic Behavior & Organization* **154** 175–190.

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

Goetzmann, W. N., J. E. Ingersoll Jr., S. A. Ross. 2003. High-water marks and hedge fund management contracts. *The Journal of Finance* **58**(4) 1685–1718.

Greve, H. R., J. A. C. Baum, H. Mitsuhashi, T. J. Rowley. 2010. Built to last but falling apart: Cohesion, friction, and withdrawal from interfirm alliances. *Academy of Management Journal* **53**(2) 302–322.

Grinblatt, M., G. Jostova, L. Petrasek, A. Philipov. 2020. Style and skill: Hedge funds, mutual funds, and momentum. *Management Science* **66**(12) 5505–5531.

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

Heckman, J., B. Singer. 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society* 271–320.

Heimer, R. Z. 2016. Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies* **29**(11) 3177–3209.

Hirshleifer, D., S. H. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* **36** 337–386.

Holland, P. W., S. Leinhardt. 1971. Transitivity in structural models of small groups. *Comparative group studies* **2**(2) 107–124.

Huang, J., K. D. Wei, H. Yan. 2007. Participation costs and the sensitivity of fund flows to past performance. *The Journal of Finance* **62**(3) 1273–1311.

Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, M. Morris. 2008. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software* **24**(3) nihpa54860.

Hutto, C., E. Gilbert. 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*, vol. 8. 216–225.

Hwang, E. H., X. Guo, Y. Tan, Y. Dang. 2022. Delivering healthcare through teleconsultations: Implications for offline healthcare disparity. *Information Systems Research* **forthcoming**.

Ivković, Z., S. Weisbenner. 2009. Individual investor mutual fund flows. *Journal of Financial Economics* **92**(2) 223 – 237.

Kaiser, M., E. S. C. Berger. 2021. Trust in the investor relationship marketing of startups: a systematic literature review and research agenda. *Management Review Quarterly* **71**(2) 491–517.

Kane, G. C., G. J. Labianca, S. P. Borgatti. 2014. What's different about social media networks? a framework and research agenda. *MIS Quarterly* **38**(1) 275–304.

Khern-am-nuai, W., K. Kannan, H. Ghasemkhani. 2018. Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research* **29**(4) 871–892.

Kim, H., A. Kankanhalli, S. Lee. 2018. Examining gifting through social network services: A social exchange theory perspective. *Information Systems Research* **29**(4) 805–828.

Krivitsky, P. N., M. S. Handcock. 2014. A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **76**(1) 29.

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

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

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

Li, Z., X. Fang, X. Bai, O.R.L. Sheng. 2017. Utility-based link recommendation for online social networks. *Management Science* **63**(6) 1938–1952.

Liberti, J. M., M. A. Petersen. 2019. Information: Hard and soft. *Review of Corporate Finance Studies* **8**(1) 1–41.

Lin, M., S. Viswanathan. 2016. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* **62**(5) 1393–1414.

Mayer, R. C., J. H. Davis, F. D. Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* **20**(3) 709–734.

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

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society* 69–85.

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

Oehler, A., M. Horn, S. Wendt. 2016. Benefits from social trading? empirical evidence for certificates on wikifolios. *International Review of Financial Analysis* **46** 202–210.

Pan, W., Y. Altschuler, A. S. Pentland. 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*. Washington DC, 203–209.

Pedersen, L. H. 2022. Game on: Social networks and markets. *Journal of Financial Economics* **forthcoming**.

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

Polidoro, F., G. Ahuja, W. Mitchell. 2011. When the social structure overshadows competitive incentives: The effects of network embeddedness on joint venture dissolution. *Academy of Management Journal* **54**(1) 203–223.

Robins, G., P. Pattison, Y. Kalish, D. Lusher. 2007. An introduction to exponential random graph (p*) models for social networks. *Social networks* **29**(2) 173–191.

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

Shafi, K., A. Mohammadi, S. A. Johan. 2020. Investment ties gone awry. *Academy of Management Journal* **63**(1) 295–327.

Sirri, E. R., P. Tufano. 1998. Costly search and mutual fund flows. *The journal of finance* **53**(5) 1589–1622.

Snijders, T. A., P. E. Pattison, G. L. Robins, M. S. Handcock. 2006. New specifications for exponential random graph models. *Sociological methodology* **36**(1) 99–153.

Strauss, D., M. Ikeda. 1990. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association* **85**(409) 204–212.

Tirunillai, S., G. J. Tellis. 2012. Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science* **31**(2) 198–215.

Wasserman, S., K. Faust, et al. 1994. *Social network analysis: Methods and applications*. Cambridge University Press.

Wohlgemuth, V., E. S. Berger, M. Wenzel. 2016. More than just financial performance: Trusting investors in social trading. *Journal of Business Research* **69**(11) 4970–4974.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Xu, J., M. Chau. 2018. Cheap talk? the impact of lender-borrower communication on peer-to-peer lending outcomes. *Journal of Management Information Systems* **35**(1) 53–85.

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

Yang, M., Z. Zheng, V. Mookerjee. 2021. How much is financial advice worth? the transparency-revenue tension in social trading. *Management Science* **forthcoming**.

Appendix

A. Network evolution

To model network evolution over time, we track link changes by defining two networks: the formation network (y^+) and the dissolution network (y^-). $y^+_{t-1\to t}$ consists of network y_{t-1} plus the links formed from time $t-1$ to t, and $y^-_{t-1\to t}$ consists of network y_{t-1} minus the links removed from time $t-1$ to t. In the data, we observe y_{t-1} and y_t. Given the observed data, we are able to recover formation network $y^+_{t-1\to t}$ and dissolution network $y^-_{t-1\to t}$.

Table A-1 shows 4 possible transitions between nodes i and j. If there is no link between nodes i and j at both times $t-1$ and t (the first row in Table A-1), then the value of $y^+_{t-1\to t}$ is set to 0 in the formation network, indicating that there is no link formation from $t-1$ to t. However, we are not able to come to any conclusion regarding link dissolution because there is no link that can be dissolved between nodes i and j from time $t-1$ to t (denoted by - in Table A-1). If the link between nodes i and j exists at time $t-1$ but no longer exists at time t (the third row in Table A-1), then the value of $y^-_{t-1\to t}$ is set to 0 in the dissolution network, indicating that the link is dissolved from $t-1$ to t. If the link exists both at time $t-1$ and at time t (the last row in Table A-1), then the value of $y^-_{t-1\to t}$ is set to 1 in the dissolution network, meaning that the link is sustained from time $t-1$ to t. However, we cannot infer anything about formation (denoted by - in Table A-1) when $y_{t-1} = 1$, as the link between nodes i and j already exists at time $t-1$. Note that a value of 1 in $y^-_{t-1\to t}$ means that the link is sustained and that a value of 0 means that the link is dissolved. In contrast, a value of 1 in $y^+_{t-1\to t}$ means that the link is formed.

Table A-1 Network evolution

y_{t-1}	y_t	$y^-_{t-1\to t}$	$y^+_{t-1\to t}$
0	0	-	0
0	1	-	1
1	0	0	-
1	1	1	-

B. Estimation with time fixed effects

Link formation and dissolution may be affected by some events that are time dependent, for example, some policy change on the platform, the availability of crypto currencies on the platform, or a regulatory change in the social trading industry. During our sample period (2016 and 2017), no changes were made to social trading regulations in general, and all existing regulation remained intact.

The relevant regulators classify social trading as portfolio management per the Markets in Financial Instruments Directive (MiFID). In particular, the European Securities and Markets Authority (ESMA), the EU's securities market regulator, announced in 2008 that financial operators active on a social trading network could exercise "investment discretion by automatically executing the trade signals of third parties", which implied that brokers and market-makers active in that field were assimilated into the group of other financial intermediaries that need ad hoc authorization for portfolio management per the Markets in Financial Instruments Directive (MiFID). Subsequently, the same authorization requirements were confirmed by ESMA in 2012. Whenever a service provider makes an investment through an automated algorithm in view of trade signals coming from third parties—in relation to MiFID financial instruments—this implies that the provider has to perform some consequent duties related to a suitability assessment, the completion of business obligations, and information standards for both clients and authorities.

The corresponding directive, Directive 2004/39/EC, was first introduced in 2004. The main objective of MiFID is to create a European financial market that encourages honest competition between the participating companies and, at the same time, increasing consumer protection. MiFID has been in force since January 31, 2007 and was superseded by MiFID II on January 3, 2018. MiFID II did not bring meaningful changes to the regulation of social trading.

Similarly, the Financial Conduct Authority (FCA) specified that the service of social trading falls within Article 4(1)(9) of MiFID. This article defines "portfolio management" as "managing portfolios in accordance with mandates given by clients on a discretionary client-by-client basis where such portfolios include one or more financial instruments." In copy trading and mirror trading, investment decisions are implemented with no intervention by the client other than an agreement ("mandate") between the service provider and the client on the discretionary service provided.[15] This interpretation has not changed over time.

Despite the lack of regulatory changes during our sample period, it is possible that link formation and dissolution are affected by other events that are time dependent. To mitigate this concern, we include time fixed effects, which allows us to control for time-specific peculiarities. We estimate our extended model and present the results in Table B-1. The estimation results are generally consistent with the results from the main model.

Table B-1 Estimation results with time fixed effects

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.2653***	(0.0271)	0.1090***	(0.0290)
Leader's post quality	0.5212***	(0.0250)	0.0510**	(0.0213)
Leader's number of replies	0.0971***	(0.0282)	0.0140	(0.0317)
Leader's comment received positive score	1.2130***	(0.1420)	0.4852***	(0.1489)
Leader's comment received negative score	−5.2520***	(0.7781)	−1.9831***	(0.4792)
Follower's post quantity	0.0511	(0.0493)	−0.5559***	(0.0556)
Follower's post quality	−0.3151***	(0.0813)	0.0066	(0.0774)
Financial performance				
Leader's average profit	0.1069***	(0.0112)	0.0505***	(0.0187)
Leader's std. dev. profit	−0.3588	(0.4516)	−1.5600**	(0.7802)
Follower's average profit	−0.0028	(0.0106)	0.0551***	(0.0126)
Follower's std. dev. profit	−0.2900	(0.5032)	−8.7381***	(0.5961)
Leader's average holding time	0.2492***	(0.0454)	0.0587	(0.0682)
Leader's lottery preference	0.6185*	(0.3555)	1.0449***	(0.4019)
Leader's HHI	−1.0280***	(0.0929)	−0.3822***	(0.0967)
Demographics				
Nationality	0.7866***	(0.0706)	0.3577***	(0.0810)
Age	0.1132**	(0.0509)	0.0582	(0.0552)
Homophily (male)	0.9514***	(0.1362)	−0.1272	(0.1189)
Homophily (female)	−0.8487**	(0.4171)	0.2573	(0.3209)
Image	2.3251***	(0.4258)	0.9411*	(0.4832)
Bio	2.8937***	(0.1931)	−0.1689	(0.1937)
Experience	−0.0724*	(0.0380)	0.1277***	(0.0443)
Wealth	0.0120	(0.0332)	0.0666*	(0.0381)
Income	0.0038	(0.0401)	0.0276	(0.0449)
Risk	−0.0809	(0.0498)	−0.0252	(0.0579)
Network structure				
Leader's popularity	0.0110***	(0.0004)	0.0017***	(0.0004)
Leader's activity	−0.0401***	(0.0125)	0.0312***	(0.0119)
Follower's popularity	−0.0570***	(0.0130)	−0.0084	(0.0072)
Follower's activity	−0.0047	(0.0078)	−0.1032***	(0.0097)
Transitivity	0.0856***	(0.0255)	−0.0256	(0.0665)
Constant	−1.5666	(6.0335)	2.9661***	(0.6968)
Time fixed effects	Yes		Yes	
Log Likelihood	-13,039.84		-9,000.62	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

[15]`https://www.fca.org.uk/firms/copy-trading`, last accessed Jul. 8, 2022

C. Estimation with leader fixed effects

As mentioned in Section 3.2, in our context, the information provided on the platform is rich and highly transparent. We observe the information that is observed by followers on the platform, which may affect followers' link formation and dissolution decisions. We have access to the complete transaction and social communication histories of each trader, and the platform does not allow for a private chat channel. In addition, given the size and international reach of the platform, we believe that personal relationships between investors are very unlikely and affect at most very few investors. Our data include 106,007 unique traders of more than 140 nationalities. In addition, only approximately 14.6 percent of the total links are between members of the same nationality. Considering this large size and international reach, it is unlikely that (many) investors know each other outside of the platform.

However, some investors may decide to advertise their trading via other social media channels such as YouTube and include links to their eToro profile page on YouTube. Despite the various control variables included in our network analysis, some determinants of link formation and dissolution may be unobserved, at least to researchers. For example, it is possible that leaders' advertising activities could be found on other social media sites, which may affect followers' following decision. However, due to the anonymity of our data, we are unable to match data from other social media sites to our trading data. To mitigate this concern, we include leader fixed effects to capture a leader's general propensity to advertise their trading on social media. We believe that such general propensities are rather stable over time. Some investors are, in general, willing to advertise their investment strategies on alternative social media, while others are not.

We model only the link formation process because leader advertisements on other channels are more likely to affect the link formation process. We estimate the model using the same Chamberlain approach described in Section 3.2. To guarantee computational tractability with two random effects (follower-specific and leader-specific), we keep all leaders and randomly sample 250 followers from the sample used for the main model. The results are reported in Table C-1, and our main findings hold qualitatively.

Table C-1 Estimation results with leader fixed effect

Variable	Formation	
Social communication		
Leader's post quantity	0.0936	(0.0819)
Leader's post quality	0.1600**	(0.0692)
Leader's number of replies	0.1521*	(0.0907)
Leader's comment received positive score	1.1432***	(0.3365)
Leader's comment received negative score	−7.1147***	(1.9477)
Follower's post quantity	0.1544***	(0.0536)
Follower's post quality	−0.6185***	(0.1068)
Financial performance		
Leader's average profit	0.1196***	(0.0393)
Leader's std. dev. profit	−4.2705**	(1.8278)
Follower's average profit	−0.0615***	(0.0195)
Follower's std. dev. profit	−0.9975	(1.1685)
Leader's average holding time	0.0713	(0.1680)
Leader's lottery preference	−0.7884	(0.9843)
Leader's HHI	−0.1140	(0.2744)
Demographics		
Nationality	1.5814***	(0.1619)
Age	0.1193	(0.1234)
Homophily (male)	0.3063	(0.3620)
Homophily (female)	−0.3198	(0.8204)
Image	1.6256***	(0.6251)
Bio	0.9437***	(0.3473)
Experience	−0.0871	(0.0634)
Wealth	0.1410**	(0.0567)
Income	0.0263	(0.0654)
Risk	−0.1986**	(0.0851)
Network structure		
Leader's popularity	0.0001	(0.0015)
Leader's activity	−0.0371	(0.0320)
Follower's popularity	−0.0722***	(0.0148)
Follower's activity	0.1165***	(0.0123)
Transitivity	0.1580***	(0.0312)
Constant	27.8232**	(10.9007)
Log Likelihood	-3,121.92	
Observations	1,592,411	

Notes: The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

D. Alternative data sample

We estimate the main model using a random subsample of all users on the platform. In this appendix, we take another random sample and estimate the main model to address a potential concern about the external validity of the findings. The results are reported in Table D-1, and our findings are generally consistent with those from the main model.

Table D-1 Estimation results with an alternative sample

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3015***	(0.0267)	0.1025***	(0.0276)
Leader's post quality	0.5340***	(0.0246)	0.0637***	(0.0204)
Leader's number of replies	−0.0024	(0.0271)	−0.0090	(0.0297)
Leader's comment received positive score	1.0002***	(0.1404)	0.4649***	(0.1456)
Leader's comment received negative score	−3.8560***	(0.6862)	−2.8281***	(0.4554)
Copier's post quantity	−0.0794*	(0.0447)	−0.3592***	(0.0489)
Copier's post quality	−0.2615***	(0.0818)	−0.1304*	(0.0735)
Financial performance				
Leader's average profit	0.0898***	(0.0118)	0.0429**	(0.0167)
Leader's std. dev. profit	−2.5028***	(0.4889)	−1.2995*	(0.6940)
Copier's average profit	−0.0252	(0.0155)	0.0938***	(0.0149)
Copier's std. dev. profit	−1.3589**	(0.6327)	−8.5795***	(0.5875)
Leader's average holding time	0.3524***	(0.0423)	−0.0783	(0.0619)
Leader's lottery preference	0.1861	(0.3453)	0.7816**	(0.3876)
Leader's HHI	−0.5040***	(0.0863)	−0.3184***	(0.0871)
Demographics				
Nationality	0.6705***	(0.0669)	0.1022	(0.0783)
Age	0.0134	(0.0502)	−0.0408	(0.0539)
Homophily (male)	1.2767***	(0.1488)	−0.0717	(0.1152)
Homophily (female)	−1.5970**	(0.7136)	0.4126	(0.4167)
Image	2.6399***	(0.4239)	1.0131**	(0.5009)
Bio	2.6410***	(0.1573)	0.0152	(0.1797)
Experience	0.0277	(0.0385)	0.0780*	(0.0399)
Wealth	−0.0128	(0.0333)	0.1109***	(0.0344)
Income	0.0056	(0.0390)	−0.0138	(0.0405)
Risk	−0.0513	(0.0487)	−0.1273**	(0.0513)
Network structure				
Leader's popularity	0.0089***	(0.0004)	0.0016***	(0.0004)
Leader's activity	−0.0479***	(0.0125)	0.0283**	(0.0117)
Copier's popularity	−0.0489***	(0.0119)	−0.0033	(0.0057)
Copier's activity	0.0036	(0.0059)	−0.0147**	(0.0062)
Transitivity	0.0825***	(0.0231)	0.0439	(0.0609)
Constant	−18.7072***	(4.9729)	1.3530**	(0.6745)
Log Likelihood	-14,202.79		-9,541.34	
Observations	11,000,237		20,796	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

E. Examples of social communication texts

We provide some examples of posts, comments, and replies in Table E-1. As shown in the table, the example posts were written by leaders. Leaders may want to advertise their performance, share their trading strategies or simply welcome new followers. Followers can ask for clarification in the comments (see comment 1), provide positive feedback to the leader (see comment 2), or complain about the leader's performance by leaving a negative comment (see comment 3). Finally, a leader can reply to a follower's question (see reply 1) or share his/her insights about the market by replying to the comment (see reply 2).

Table E-1 Examples of social communication

Post	**Text**
1	After being very long out I made some big mistakes. Now I trusted my instinct and made every trade without the knowledge from someone else. With this I turned back in the green and closed 2016 in green! awesome
2	The $EURUSD is a buy for me for the next few months... been at its lowest in years lately & I will definitely be looking for buy positions
3	@adelaya Hi, thank you for following. I wish you happy successful trading :)
Comment	**Text**
1	Why does it seem to fall in the after hours charts, though? Shouldn't it surge? I'm a newbie.
2	Your current investments are looking awesome!
3	I has lost so much money since i copied u and i never taste the earning sweet
Reply	**Text**
1	@Seregaomsk Additional funds will be used only when I open new deals, to distribute them to open positions while the only option is to stop copying and then copy with a new amount.
2	The beginning of the fall was undoubtedly connected with the general correction in the market, after which they are still only recovering. The current fall, in my opinion, is largely speculative, since there have been no negative indicators, news or decisions regarding YNDX lately. In this regard, I plan to keep them for now, as I look forward to recovery in the coming weeks.

Notes: Some social communications were not originally in English, and thus, we present the translated versions.

F. Estimation with post quality-quantity interaction term

In this appendix, we add an interaction term between the post quantity and post quality of a leader, keeping everything else the same as in the main model, to investigate the moderating effect of post quality. We estimate the extended model and show the results in Table F-1. We find that post quantity has a positive moderating effect on post quality in the link formation process, whereas the interaction term is statistically not significant in the link dissolution process.

Table F-1 Estimation results with post quality-quantity interaction term

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.2942***	(0.0279)	0.1226***	(0.0288)
Leader's post quality	0.4140***	(0.0356)	0.0678**	(0.0283)
Leader's number of replies	−0.0263	(0.0310)	0.0037	(0.0331)
Leader's comment received positive score	1.2167***	(0.1411)	0.5555***	(0.1466)
Leader's comment received negative score	−4.6982***	(0.7611)	−2.1387***	(0.4719)
Follower's post quantity	0.0657	(0.0474)	−0.5301***	(0.0536)
Follower's post quality	−0.3339***	(0.0797)	−0.0049	(0.0749)
Leader's post quality * leader's post quantity	0.0636***	(0.0166)	−0.0184	(0.0149)
Financial performance				
Leader's average profit	0.0891***	(0.0135)	0.0729***	(0.0184)
Leader's std. dev. profit	−1.8736***	(0.5551)	−2.7737***	(0.7687)
Follower's average profit	0.0013	(0.0114)	0.0441***	(0.0120)
Follower's std. dev. profit	−0.6637	(0.5090)	−8.1489***	(0.5580)
Leader's average holding time	0.3451***	(0.0441)	0.0615	(0.0661)
Leader's lottery preference	0.1675	(0.3630)	1.0977***	(0.3957)
Leader's HHI	−0.6133***	(0.0906)	−0.1403	(0.0929)
Demographics				
Nationality	0.7466***	(0.0704)	0.3440***	(0.0798)
Age	0.1235**	(0.0505)	0.0514	(0.0543)
Homophily (male)	0.9849***	(0.1360)	−0.1562	(0.1165)
Homophily (female)	−0.8621**	(0.4167)	0.2938	(0.3178)
Image	2.4808***	(0.4236)	0.7285	(0.4652)
Bio	2.9513***	(0.1873)	−0.1537	(0.1886)
Experience	−0.0445	(0.0371)	0.1459***	(0.0404)
Wealth	0.0244	(0.0324)	0.0617*	(0.0347)
Income	−0.0091	(0.0393)	0.0240	(0.0409)
Risk	−0.0838*	(0.0488)	−0.0397	(0.0527)
Network structure				
Leader's popularity	0.0078***	(0.0004)	0.0015***	(0.0004)
Leader's activity	−0.0354***	(0.0124)	0.0379***	(0.0119)
Follower's popularity	−0.0544***	(0.0127)	−0.0091	(0.0065)
Follower's activity	0.0026	(0.0071)	−0.0771***	(0.0088)
Transitivity	0.0810***	(0.0257)	−0.0206	(0.0656)
Constant	−16.0395***	(5.6622)	2.0885***	(0.6444)
Log Likelihood	-13,613.02		-9,151.50	
Observations	11,000,219		19,744	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

G. Alternative estimation using conditional logit estimator

The Chamberlain correlated random effects applied in our main model require the assumption that η_i follows a conditional normal distribution depending on X_i with a constant variance, which is equivalent to a conventional random effects model that controls for the correlation function. Thus, the coefficients on the time-invariant observables (C_i) can be estimated. To correct for the incidental problem, another way to estimate the follower-specific unobservables η_i is based on the conditional logit estimator (Wooldridge 2010), which allows η_i to be arbitrarily correlated with X_i. However, under the conditional logit estimator, η_i and time-invariant covariates C_i cannot be identified simultaneously. Thus, C_i should be excluded.

We use the link formation process to illustrate the model implementation. The utility of follower i from forming a link with leader j between period $t-1$ and period t is defined as follows:

$$y^*_{ijt} = \boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij} + \eta_i + \epsilon_{ijt}, \tag{G-1}$$

where the definitions of notations are the same as those in the main model, Equation (6).

For the link formation process, we define

$$y_{ijt} = \begin{cases} 1 & y^*_{ijt} > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{G-2}$$

where y_{ijt} is a binary variable that is equal to 1 if follower i forms a link with leader j between period $t-1$ and period t. The link dissolution process is defined in the same fashion. In the dissolution process, y_{ijt} is equal to 1 if follower i dissolves the link with leader j in period t.

We denote as n_i the sum of all binary outcomes for follower i's following status over all the periods. That is, $n_i = \sum_{t=1}^{T_i} \sum_{j=1}^{J_{it}} y_{ijt}$, where T_i is the number of periods during which follower i exists on the platform and J_{it} is the number of leaders that follower i can potentially follow in period t. Each follower i has a corresponding vector with length $T_i \times J_{it}$. B_i is the set of all possible vectors in which n_i elements are equal to 1 and $(T_i \times J_{it} - n_i)$ elements are equal to 0. In other words, B_i represents all the possible scenarios in which follower i forms n_i links with the potential J_{it} leaders over T_i periods. Mathematically,

$$B_i = \{b \in \{0,1\}^{\{T_i \times J_{it}\}} | \sum_{t=1}^{T_i} \sum_{j=1}^{J_{it}} b_{jt} = n_i\}, \tag{G-3}$$

where b is one realization or scenario among all the possible scenarios and b_{jt} denotes an element in the vector b.

The conditional probability of y_i given n_i is defined as follows:

$$Pr(y_i|X_{it-1}, W_{jt-1}, n_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{e^{(y_i \times (\boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij}))}}{\sum_{b \in B_i} e^{(b \times (\boldsymbol{\alpha} X_{it-1} + \boldsymbol{\beta} W_{jt-1} + \boldsymbol{\lambda} V_{ij}))}}. \tag{G-4}$$

From Equation (G-4), we observe that the conditional probability does not depend on η_i. Thus, the conditional log likelihood is also independent of η_i and can be written as follows:

$$CLL(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{i=1}^{I_t} \sum_{t=1}^{T_i} \sum_{j=1}^{J_{it}} \ln[Pr(y_i|X_{it-1}, W_{jt-1}, n_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})], \tag{G-5}$$

where I_t is the total number of followers in period t, T_i is the number of periods that follower i exists on the platform, and J_{it} is the number of leaders that follower i can potentially follow in period t. We estimate the model by maximizing its overall log-likelihood value. The estimation results are reported in Table G-1. We find that the results are consistent with the findings in the main model.

Table G-1 Estimation results using conditional logit estimator

Variable	Formation		Dissolution[a]	
Social communication				
Leader's post quantity	0.3184***	(0.0268)	0.1167***	(0.0284)
Leader's post quality	0.5167***	(0.0250)	0.0438**	(0.0208)
Leader's number of replies	0.0211	(0.0277)	−0.0114	(0.0312)
Leader's comment received positive score	1.1088***	(0.1395)	0.6159***	(0.1485)
Leader's comment received negative score	−4.8155***	(0.7544)	−2.0508***	(0.4720)
follower's post quantity	0.0405	(0.0624)	−0.4604***	(0.0671)
follower's post quality	−0.2816***	(0.0874)	0.0585	(0.0800)
Financial performance				
Leader's average profit	0.0906***	(0.0136)	0.0734***	(0.0185)
Leader's std. dev. profit	−2.1217***	(0.5543)	−2.7319***	(0.7651)
follower's average profit	0.0298	(0.0197)	0.0085	(0.0196)
follower's std. dev. profit	−3.9422***	(1.2738)	−8.2172***	(1.2796)
Leader's average holding time	0.3466***	(0.0446)	0.0774	(0.0673)
Leader's lottery preference	0.2098	(0.3613)	1.1073***	(0.3989)
Leader's HHI	−0.5972***	(0.0904)	−0.1268	(0.0935)
Demographics				
Nationality	0.7418***	(0.0707)	0.3390***	(0.0794)
Age	0.1198**	(0.0507)	0.0548	(0.0544)
Homophily (male)	0.9848***	(0.1359)	−0.1385	(0.1170)
Homophily (female)	−0.8769**	(0.4166)	0.2625	(0.3121)
Image	2.4737***	(0.4239)	0.4781	(0.5424)
Bio	2.9224***	(0.1869)	−0.2559	(0.2008)
Network structure				
Leader's popularity	0.0077***	(0.0004)	0.0014***	(0.0004)
Leader's activity	−0.0359***	(0.0124)	0.0392***	(0.0121)
follower's popularity	−0.0364**	(0.0147)	−0.0060	(0.0087)
follower's activity	−0.0282***	(0.0074)	−0.0849***	(0.0104)
Transitivity	0.0927***	(0.0290)	−0.0218	(0.0635)
Log Likelihood	-11,339.00		-5,522.04	
Observations	4,074,702		15,083	

Notes: A positive coefficient in the dissolution model indicates increased link duration. The number of posts, the quality of posts (the number of likes), the number of replies, wealth, income and risk are log-transformed. Average profit and std. dev. profit are scaled by a factor of 100. Average holding time is scaled by a factor of 1/100.
Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.