The Impact of Finfluencers on Retail Investment*

Isaiah Hull[a,c,d], Yingjie Qi[b]

[a] *Rethinc.Labs UNC*
[b] *Copenhagen Business School*
[c] *DeepLearning.AI*
[d] *CogniFrame Inc.*

Abstract

We examine the impact of financial influencers ("finfluencers") on retail investment using real equity and derivative investments across four Nordic countries. Using an instrument that randomly assigns influencers to followers, we find the following: (1) Investors tend to follow influencers with high Sharpe ratios, frequent trades, a shared country of residence or language, and male gender. (2) Influencers affect followers' portfolios and trading behavior, particularly when they have a large following, a central network position, or participate in group discussions. This effect is strongest for investors who follow fewer influencers, female investors, and when trading passive funds.

Keywords: Social network, retail trading, financial influencers
JEL: G40, G41, G50

*We are grateful for helpful comments and suggestions from Steffen Andersen, Laurent Bach, Claire Celerier, Tony Cookson, Francesco D'Acunto, Magnus Dahlquist, Rawley Heimer, Harrison Hong, George Jiang (discussant), Runjing Lu (discussant), Julie Marx, Kasper Meisner Nielsen, Michaela Pagel, Gianpaolo Parise, Christine Parlour, Lasse Heje Pedersen, Kim Peijnenburg, Akash Raja, Alberto Rossi, Karin Thorburn, Erkki Vihriälä (discussant), Baolian Wang, Tingxuan Zhao (discussant), the seminar and conference participants at Copenhagen Business School, Lund University, AFA AFFECT Workshop, Asian Bureau of Finance and Economics Research Conference, Future of Financial Information Conference, Midwest Finance Association Annual Conference, Northeastern University Finance Conference, NFN Young Scholars Workshop, and CICF. Qi acknowledges support from the Danish Finance Institute and Center for Big Data in Finance (grant no. DNRF167). E-mail: yq.fi@cbs.dk (Yingjie Qi), isaiah@cogniframe.com (Isaiah Hull).

1. Introduction

The growth of investor-focused social media has rejuvenated interest in understanding the social transmission of ideas (Cookson, Mullins and Niessner, 2024; Cookson, Engelberg and Mullins, 2023; Bailey et al., 2022, 2018*b*; Heimer, 2016; Heimer and Simon, 2015; Hong, Kubik and Stein, 2004; Manski, 2000). As both the importance of retail investors and their exposure to financial influencers grow (FINRA 2023), it is increasingly important to determine what impact these influencers have on retail investors and to examine the incentives that underpin their suggestions and trading decisions.

In this paper, we leverage trading and network data from a social trading platform operating in four Nordic countries, combined with exogenous variation in influencer assignment, to measure the *causal* impact of financial influencers on their followers and to characterize the nature of that influence. Our dataset, which contains time-stamped daily transaction records for both influencers and their followers, spans nearly a decade from the platform's inception at the end of 2014 to early 2023.[1] This platform is affiliated with a leading Northern European brokerage firm, managing approximately €11 billion in assets under custody. As such, trades are real transactions and not investors' opinions.

We first describe the characteristics of the network of traders, including the relationships among all users on the platform. We find that an influencer's follower count positively correlates with past performance and number of trades. For the subsample of investors whose gender could be identified, we find that male users are more likely to be followed. Exploiting bilateral relations among follower-influencer pairs and using the schema introduced in Pedersen (2022*a*), we find that the most popular influencers are those that can best be described as "long-term rational" investors, followed by "fanatics." Investors who are "short-term rational" are measurably less popular. Lastly, we also find evidence that coming from the

[1] The "social trading" platform allows users to observe and interact with the trades of other investors. In order to interact and receive updates from others, one needs to choose to "follow" another person. Investors automatically receive timely updates on the activities of the individuals they follow, delivered through their platform news feed and email notifications.

same country or speaking the same language increases the likelihood of an influencer being followed.

We label the top 100 most-followed individuals in the dataset "influencers." However, having high follower counts does not necessarily translate into having high impact. Followers could follow other investors for purposes unrelated to their real investment decisions. Our main analysis attempts to measure an influencer's causal impact on their followers. This requires us to address two potential issues with endogeneity. First, the choice to follow an influencer is not random. Although we indeed observe that investment decisions overlap more within follower-influencer pairs, this association could arise from similarities in investment preferences or exposure to common information sources (omitted variable bias). Second, influencers may hold or trade the same securities as their followers to maintain popularity and relevance, reversing the direction of causality.

To overcome these challenges and causally identify whether and how much influencers impact their followers, we adopt an IV strategy in the main empirical exercise that instruments influencers who are employees of the platform (platform-made influencers). Upon joining, each user is automatically assigned to follow the employees who work for the platform at the time. These follower-influencer pairs—formed as a consequence of the platform's design—provide a source of exogenous variation. This is a mechanical feature of the website and is not determined endogenously by unobserved common interests. Furthermore, since the platform forms these connections in the network, causality cannot run in the opposite direction. First stage regressions confirm that the relevance condition for IV is indeed met.

Our IV estimates show that following an influencer is associated with substantial changes in a follower's portfolio and trading decisions. More specifically, in a regression with investor-by-year-month fixed effects, we find that, on average, a follower's overlap with an influencer increases by 3.8 percentage points (109%) for holdings, 0.3 percentage points (18%) for purchases, and 4.6 percentage points (192%) for sales.

Beyond the average effects of financial influencers on their followers, we find significant heterogeneity in the effects across influencer and follower types. The effect is more pronounced for influencers who have more followers, are central to the network, and participate actively in group discussions. The effect is also more pronounced for followers who follow fewer people. For the subsample of followers whose gender or/and age could be identified, we found that female investors are more influenced, but age does not appear to predict susceptibility to influence.

In terms of security type, we find that ETFs and passive index funds are passed through from influencer to follower; whereas risky ones, such as levered products, are not. This suggests that investors are selective when copying financial influencers.

We also examine the difference between influencers and traditional financial advisers, and discuss the implications for potential conflicts of interest. In particular, we manually identify influencers who disclose their relationship (ex-post) with the company that manages the platform. We find that they both held and traded more products issued by the platform and adopted trading styles that generated higher trading volume. This, in turn, generated higher commissions for the platform, but did not translate into a higher Sharpe ratio.

Our work sheds light on a recent driver of stock market activity: the rise and impact of social media influencers on retail investors. From a policy perspective, it is important to understand the implications of this phenomenon. On the one hand, financial influencers could promote stock market participation at a lower cost, having a broadly positive impact on markets and investors' wealth accumulation. On the other hand, they could increase noise and herding in markets, especially if their incentives do not align well with their followers'.

The granularity, long duration, and broad coverage of our data, combined with the unique instrument, enable a deeper understanding of social trading and clean identification of the role of financial influencers. In doing so, we contribute to the two following literature. The first is the behavioral finance literature and, in particular, work that studies drivers of retail investor behavior (Barber et al. 2022; Jiang et al. 2022; Barber and Odean 2008; Odean 1999). This literature has seen a rapid expansion in the new age of increasing social media influence, which facilitates broader and faster information diffusion among people who do not necessarily know each other in real life. Social trading not only impacts the decision-making of individuals, but also financial markets in aggregate (Aridor et al., 2024; Cookson et al., 2024; Bailey et al., 2022; Cookson, Niessner and Schiller, 2022; Ammann and Schaub, 2021; Cookson and Niessner, 2020; Bailey et al., 2018*c*).[2]

The second literature of relevance studies social networks and financial influencers.[3] In contrast to conventional influences in the financial markets—such as financial advisors, friends, and family members—the growing presence of financial influencers raises fresh concerns about information asymmetry and conflicts of interest. Specifically, finfluencers may trade or advise in a way that does not focus on absolute performance, since this may not be conducive to gaining popularity. This unique characteristic of financial influencers, which is not shared by conventional financial advisors, underscores the importance of understanding the impact and incentives influencers have in giving advice. This paper contributes to the discussion by quantifying the causal impact financial influencers have on their followers, establishing the heterogeneity of this impact across finfluencer and follower types, and uncovering evidence of conflicts of interests in these relationships.

The remainder of this article is organized as follows. Section 2 describes the background of the platform, the network, and the trading data. Section 3 investigates which factors are associated with influencer popularity. Section 4 empirically tests and quantifies the impact of influencers on their followers. Section 5 explores potential mechanisms of how influencers generate impact by exploring the heterogeneity of the baseline effects across influencer, follower, and security types. Section 6 provides a discussion of further findings

[2]See important earlier work on the social transmission of ideas in, e.g., Hirshleifer (2020); Daniel and Hirshleifer (2015); Hirshleifer (2015); Frydman et al. (2014); Barber and Odean (2013); Barber et al. (2009); Barberis and Xiong (2009); Barberis, Huang and Thaler (2006); Barber and Odean (2002); Hirshleifer (2001); Barber and Odean (2001*b*); Barberis and Huang (2001); Barberis, Shleifer and Vishny (1998); Shleifer and Summers (1990).

[3]See, e.g., Benetton et al. 2024; Kakhbod et al. 2023; Dim 2025; Sui and Wang 2022; Barber et al. 2022; Han, Hirshleifer and Walden 2022; Bikhchandani et al. 2021; Hirshleifer 2020; Cookson and Niessner 2020; Bailey et al. 2018*a*; Heimer 2016; Heimer and Simon 2015; Ozsoylev et al. 2014; Heimer 2014; Bikhchandani, Hirshleifer and Welch 1998.

and Section 7 concludes.

2. Data

This section describes the trading platform, the network and trading data, the data collection process, and the data sources. It also provides summary statistics for all variables included in the analyses.

2.1. The online platform

The online platform studied in this paper was launched in September 2014 in four Nordic countries: Denmark, Finland, Norway, and Sweden. In the first 18 months, it attracted 80,000 users and 22 billion SEK in assets under custody.[4] The platform is affiliated with a brokerage firm that has 300,000 users and manages assets under custody worth approximately 113 billion SEK (Swedish Krona), which is roughly equivalent to 11 billion USD. It is free and easy to register as a user. To do so, each investor must identify herself with an electronic ID issued by the tax authorities. This means that each investor can only have one profile on the platform.

After an investor has opened an account with the brokerage firm, she can choose to join the platform and show her trading history to other investors. We show an example of a user profile in Figure 1.[5] At the top of each user's profile page, registered users can observe the user's location, number of followers, and short bio (if any). The "View Portfolio" button allows users to view a snapshot of the investor's portfolio. Trading history is also available via the profile page. This includes detailed transaction times, prices, and security names.

[4] 1 SEK $\approx$ 0.15 USD in 2014.

[5] Users can choose to have a public or private profile. If they started as a private profile, but then switched to public, the private trades are not revealed. Alternatively, if a user starts with a public profile, but switches to private, the trades done while public remain. Furthermore, if a user leaves the platform and deactivates their account, their trades remain on the platform. This feature of the platform ensures that there is no survivorship bias in the data. According to conversations with platform employees, most users have a public profile; and, thus, our sample is unlikely to have selection issues.

Investors can choose to follow another trader on the platform by clicking "Follow" on the person's profile page. Subsequently, the investor will receive real-time notifications of the other user's trades both in email and on the platform. A user may unsubscribe from email notifications, but they will still appear in their news feed on the platform. Similarly, if one decides not to continue following another trader, she can click the "Unfollow" button on the same profile page, which will remove that person's activity from her news feed.

2.2. Network and trading data

In this subsection, we describe the network and trading data, which was collected from the platform referenced in Section 2.1. Our sample consists of 32,104 users who were either influencers or their followers over the period between 2014 and March 2023. We first identified the superset of users that contains all influencers by collecting the usernames of discussion group members. Influencers want to gain popularity and discussions groups are the best place on the platform to achieve this objective. We then collected each potential influencer's list of followers from their profile page.

After identifying the users, we collected each individual's trading history on the platform, extending back to the platform's launch in 2014. This yielded a total of 5,735,004 trades for 51,180 securities, distributed over 2,458 days. For each trade, we collected the action taken (buy or sell), the security traded, the execution price, and the currency used.[6] With the complete trading history, we were able to infer each investor's portfolio holdings on a daily basis. Since we do not have information about the amount of shares traded, we construct the portfolio with equal weights. It is important to point out that other users on the platform also do not observe this information; and all the information that would be available to a follower is collected.

In addition to trading history data, we also collected each investor's performance rating and average return since they joined the platform. We also compiled a list of users that each

[6] All scraping was conducted in accordance with the restrictions listed in the robots.txt file and with built in delays to ensure that no strain was placed on the website's servers.

investor followed, allowing us to identify the network structure of platform participants. This yielded 160,158 distinct and directed user-follower pairs.

Figures 2 and 3 visualize the social trading network. Each node in the figures represents an investor and the edges between the nodes indicate influencer-follower relationships. The node colors indicate the influencer's performance ratings based on the Sharpe ratio calculated by the platform. In order from lowest to highest rating, the groups are as follows: A node is colored orange if the investor's rating is zero, which means the investor's return is non-positive. A node is colored yellow if the investor's portfolio yielded a positive return since joining. A node is colored green if the investor's portfolio return is ranked among the top 50% in the entire platform and is blue if the investor's portfolio return is ranked among the top 10% in the entire platform. Follower nodes take the color of the influencer with the highest return rating that they follow. In the network shown in Appendix A1, we adopt an alternate coloring scheme based on investors' raw past returns.

While Figure 2 shows the full network of all investors and their influencer-follower relationships, Figure 3 shows two different influencer subnetworks, which are indicated by gray edges between the influencer and each of its followers. Although we have historical data on user transactions since the release of the platform, we do not have information about the dynamics of the network over time. We elaborate in Section 4 how our unique instrumental variable approach can help to overcome this data challenge.

As shown in the top panel of Figure 4, trading activity has grown rapidly over time on the platform. In the beginning of 2015, shortly after the platform launched, there were approximately 5000 trades per month in our sample.[7] However, by 2023, this number had increased dramatically to roughly 150,000. There was a particularly pronounced surge in trading during the COVID-19 pandemic, as shown in the sharp spike in early 2020.

Since we collect metadata for all users in our sample, we are also able to split the sample

[7] Note that we—like all users on the platform—don't observe the actual quantity of trades, but rather each trading event. This means the actual shares of securities traded are likely above the numbers shown.

by country of residence. From this, we can see that the rise in trading occurred both in aggregate and in each individual country in the sample—Sweden, Denmark, Norway, and Finland—as indicated in the top panel of Figure 4.

In the bottom panel of Figure 4, we show the daily number of trading events on the platform, which was around 10,000 in 2023. Users in Finland are the most active, followed by users in Sweden, Denmark, and then Norway. The pattern is stronger after adjusting for the total number of residents in the country.[8]

While we do not have detailed demographic information about individual investors, statistics provided by the platform suggest that the increase in trading volume seems to be partially driven by a rise in the participation of younger investors. This is precisely the group that regulators and the media have suggested may be most responsive to the advice of influencers. Figure 5 illustrates this pattern using data released by the platform on customer cohort and age.

2.3. Text data

In addition to network and trading data, we also collected text data from the platform in the form of 1) usernames; and 2) biographies when they were available. Usernames sometimes contain information that can be parsed to infer a name, birth year, or both. And influencer biographies typically contain information about the investor and his or her trading strategies and preferences.

In this section, we discuss how these two sources of textual information were processed to classify investors into groups. We then used these sample partitions in some empirical exercises in Section 3 and Section 5.

[8] Finland, Norway, and Denmark each have approximately 5 million residents during the sample period; whereas Sweden has 10 million.

2.3.1. Investor type

We use methods and models from natural language processing to classify influencers according to the behavioral scheme introduced in Pedersen (2022*b*). For influencers who had biographies, we used the `DeBERTaV3` model (He, Gao and Chen, 2021) and zero shot classification (Pushp and Srivastava, 2017) to identify the influencer's class.[9] Each influencer was assigned a probability distribution over four general classes of investor style: fanatic, naive, long-term rational, and short-term rational.[10]

The `DeBERTaV3` model is an 88 million parameter transformer model.[11] Relative to competing open source models, it achieves high scores on natural language evaluation benchmarks (He, Gao and Chen, 2021) and is computationally efficient for our task of interest. We used zero shot classification because it permits us to apply the `DeBERTaV3` model directly and without the need to train a model on our specific classification task for which we do not have labelled data.

In Appendix B, we include examples of text taken from influencer biographies and the corresponding probability distributions from applying zero shot classification with `DeBERTaV3`.

2.3.2. Gender

Demographic characteristics, such as gender, may be an important predictor of investor behavior, including their proclivity for seeking advice from a influencer (Barber and Odean, 2001*a*); however, we do not observe gender directly in our dataset because it is not included in user profiles. We do, however, have a username for each individual in our sample, which sometimes contains the user's actual name and can be used to infer gender. To identify gender, we manually screen all usernames and label those that contain common gendered

[9]Zero shot classification involves performing classification without training the model for the specific task of interest.

[10]For the purpose of the ZSC exercise, we describe the investors as "long-term rational investor," "short-term rational investor," "fanatic investor with stubborn views," and "naive investor who wants to learn."

[11]In contrast to the GPT-3.5 and GPT-4 models produced by OpenAI, the `DeBERTaV3` model is smaller, faster, and discriminative, rather than generative. The version we use is fine-tuned to perform zero shot classification accurately, consistently, and efficiently. It is also an open source model, which enables us to inspect its architecture and evaluate its training data.

names. For example, `MatildaEriksson1998` would be identified as a female given that `Matilda` is a female name, while `JensFredriksen` would be identified as a male as `Jens` is a male name across all four Scandinavian countries.

Names may appear in many languages, including non-Nordic languages, and could be associated with different genders in different countries. For this reason, we verify the quality of our manual labeling of the usernames by fine-tuning a character-level, multi-lingual large language model (LLM) to produce a classification of gender. We discuss the LLM training process and performance of the model in Appendix C. The main results make use of manually classified names, but are robust to the use of LLM-classified names.

2.3.3. Age

Similar to our identification of gender, we identify age by checking for the presence of a plausible birth year in each username. To do this, we use regular expressions to check whether each username contains a substring of the form 19** or 20**, where * is a wildcard for integers between 0 and 9. In addition to this, we require that 1) there are no additional numbers either before or after 19** or 20**; and 2) that the birth year identified implies that the user is at least 18 years old.

As an example, in the username `jan1981`, 1981 would be identified as a plausible birth year, but the 1981 in `jan198105` would not. Since 05 could refer to a month, we exclude some usernames that could plausibly contain birth years with the intention to reduce the incidence of false positives.

Even after imposing conditions (1) and (2), the identified birth year (and, therefore, age) may still be incorrect in some cases, since some users may use the year of a different event in their username; however, given the prevalence of constructing user names out of names and birth years and the conservative selection criteria, the noisiness of the measure should be relatively low.

2.4. Securities data

We obtain the name of the securities used in each trade in our sample and the category of investment. We define a trade to be stock-related if it is made directly on a stock, regardless of whether it involved buying or selling. If the underlying asset is a index fund or derivative, we classify it as non-stock trade. Figure 6 shows the composition of investment type for each country. Users from all countries trade directly on stocks more than half of the time, and this preference for stocks is most pronounced for users from Finland, followed by Norway, Denmark, and then Sweden. We also classify separately whether a trade is in a passive fund or ETF, a direct investment or derivative on a crypto asset such as the Bitcoin, or a derivative on an underlying asset that normally expires within a day with high leverage.

2.5. Final sample and summary statistics

The top 100 users by follower count are classified as influencers; whereas those who follow at least one influencer are defined as followers. We show the summary statistics for both groups and all users on the platform in Table 1, where influencer statistics are shown in panel (a) and (b), and follower statistics are shown in panel (c) and (d). In (e) and (f) we describe all users, including those that do not follow anyone.

In panel (a) of Table 1, we present summary statistics at the influencer-month level. The average (median) number of trades that an influencer conducts in a given month is 5.993 (3). The average (median) number of purchases is 2.203 (1) and number of sales is 3.79 (2). For the securities that we could identify as index products, the mean (median) number of trade is 0.39 (0). For derivatives such as options, the mean (median) number of trades by influencers at a monthly frequency is 0.224 (0).

In panel (b), we examine the cross-sectional variation among influencers. We find that the average (median) influencer has 27,436 (1,876) followers. Since the platform ranks users based on the their portfolio's Sharpe ratio (since joining the platform) using a scale of 0 and 3, we also examine influencer ratings. A rating of 3 indicates that the influencer has a Sharpe ratio that is in the top 10% of all users on the platform. A rating of 2 indicates that user's

Sharpe ratio is above the platform's median. A rating of 1 indicates that the user has one portfolio that has had positive return. And a rating of 0 indicates that the user's past return is non-positive. The average (median) influencer has a max rating of 0.94 (1) and standard deviation of 1.023, which suggests that influencers do not necessarily perform better than the average user on the platform (Barber et al. 2009; Odean 1998*a*; Odean 1998*b*). This is consistent with the pattern shown in Kakhbod et al. (2023) that financial influencers' advice is characterized by high dispersion.

Influencers trade a large number of securities and conduct a large number of trades during their time on the platform. The average (median) number of influencer trades is 211 (168) unique securities and, on average, 63% of all trades are sales and 37% are purchases. Since we track their activities since the inception of the platform, this pattern suggests that the quantity of purchases per time unit is higher than it is for sales. Their average length of active years on the platform is 4.18. For those whose gender could be identified, the majority (85.7%) are male.

Panel (c) and (d) of Table 1 report characteristics of followers. The average (median) follower follows 37.852 (16) influencers. On average they trade 106 unique securities on the platform. Sales account for 29% of the trades and purchases for 71%. The pattern that sales are in smaller quantity per time unit than purchases is similar across both influencers and followers. Followers trade both more index funds and derivatives than influencers do during their time on the platform. The average follower trades 40.016 index funds and 16.443 derivatives, compared to 13.77 and 7.92 for the average influencer. The average number of active trading years for followers is 3.593, which is slightly lower than that of the influencers. Followers are also predominantly male, with a share of 73.3%. Among those followers whose birth year could be identified, the median year is 1985.

Panels (e) and (f) describe the rest of the platform, including those users who do not follow any influencer–that is, a user that is among the top-100 most followed. Their trading behavior and demographics are similar to the followers in (c) and (d). One noticeable difference is that they trade less than those who follow influencers. This might arise from their being a different type of investor or having reduced exposure to influence from other investors in the network.

3. What correlates with popularity?

In this section, we explore factors that correlate with popularity and follower-influencer connections. Although network formation is an endogenous process and we do not observe the dynamics of the follower-influencer network, the static network can still be used to investigate the correlates of influencer popularity. We do this using two tests. First, we use each influencer's follower count as a measure of popularity and relate it to potential determinants of popularity, including past investing performance, intensity of trading and communication activities on the platform, and influencer gender. We then exploit bilateral relations among follower-influencer pairs and test how sharing a common country of origin or a common language influences the probability of pair formation. With a subsample of investors whose biographies were collected, we also examine the role of influencer trading style on popularity.

3.1. Cross-sectional tests

Using the static network, we run the following regression at the influencer-level:

$$y_i = \beta X_i + \texttt{FE} + \epsilon_i, \tag{1}$$

where the unit of observation is influencer i. The outcome variable, y_i, is the log of the total number of followers for influencer i. To account for unobserved characteristics that are shared by influencers within the same cohort (based on platform age measured in years), we include cohort fixed effects. X_i is a collection of explanatory variables that includes 1) the influencer's platform-assigned past return rating; 2) their trading intensity on the platform;

and 3) their gender. Their association with the outcome variable y_i is captured by the coefficient vector β.

The estimation results are shown in the first four columns of Table 3. Column 1 shows that—when comparing the influencers that started trading on the platform at the same time—a higher rating level is associated with more followers. Compared to unrated influencers, those with ratings of 1, 2, and 3 have follower counts that are 10%, 15.2%, and 26.2% higher on average, respectively. Given that the average number of followers for unrated influencers is 144, these numbers translate to an increase in follower counts of 14, 22, and 38. This suggests that past performance positively correlates with follower counts.

In column 2, we test whether trading intensity increases popularity and find that one additional trade is associated with an increase in the follower count by 0.2% (corresponding to 27,436.84 x 0.002 $\approx$ 55). This effect remains the same quantitatively when combined with the ratings in column 3. Last, controlling for performance rating and the intensity of trading in column 4, we find that being male is associated with a 15.3% higher follower count.

3.2. Pair-wise tests

We first demonstrated a positive cross-sectional correlation between influencer popularity and past performance, trading intensity, and gender. We now examine factors that vary within investors and across following relations. Specifically, we construct pseudo-following relations for comparisons with the real following relations. In doing so, we test what factors are associated with an investor's decision to follow or not follow an influencer, controlling for individual fixed effects.

We illustrate how the *Follow* variable is constructed in Figure 7. In the left panel, we use blue arrows to indicate the observed follower-influencer relations (*Follow*=1). In the right panel, we visualize the follower-influencer pseudo relations in orange that are absent, but could exist (*Follow*=0).

The first driver of influencer popularity we examine is homophily, which is often invoked as a means of explaining network structure in sociology. Homophily refers to the tendency of individuals with similar characteristics to group together. In our context, it seems plausible that a user is more likely to follow those who speak the same language.

To test for the importance of homophily in network formation, we use a unique feature of our setting. While trade and economic relations are similar, language and cultural barriers vary across different country pairs in our sample. Specifically, the Swedish, Danish, and Norwegian languages are all North Germanic languages and are written similarly. However, the Finnish language is a Uralic language, which is more similar to languages used in Eastern Europe and Russia. Consequently, even though investors can see each other's profile and decide to follow anyone freely, traders who are residents of Finland will have a higher language barrier in understanding and communicating with influencers from the rest of the countries in the sample than a trader from outside of Finland. And the same holds for investors from the three non-Finnish countries when it comes to decisions to engage with an investor based in Finland. We therefore define a dummy named "same language" to be equal to 1 if the follower is in the same language group as the influencer, and 0 otherwise. In addition to homophily, we also investigate whether certain types of investing styles are more popular than others.

The regression specification is a pair-wise analysis as shown below:

$$\text{Follow}_{i,f} = \beta_1 X_{i,f} + \beta_2 X_i + \texttt{FE} + \epsilon_{i,f}, \tag{2}$$

where i represents the influencer and f represents the follower. The dependent variable *Follow* is a dummy variable that takes the value of 1 if follower f follows influencer i, and 0 otherwise. $X_{i,f}$ is an indicator variable that takes the value of 1 if the follower-influencer pair shares the same country or language, and 0 otherwise. The categorical variable X_i represents an influencer's trading style. We extracted and classified trading styles based on influencers' self-disclosed biography. In total, we define four styles following the literature: naive, short-term rational, long-term rational, and fanatic.

The regression coefficients are reported in the last four columns of Table 3. We find that, controlling for the same investor, the probability of following an influencer increases both when the pairs comes from the same country and when they write in the same language. When putting these two factors in the same regression as shown in column 7, the probability of following an influencer increases by 6.3 percentage points (ppts) if the pair live in the same country. Common language increases the probability of following by another 2.1 ppts. These effects are robust to including influencer fixed effects as well.

For a subsample with investors whose trading styles could be extracted from their biographies, we include dummies for different styles in the regression as shown in column 8. The estimated coefficients show that compared to influeners with naive trading styles, the average investor is 8.3 ppts more likely to follow those with long-term rational strategies, and 0.6 ppts more likely to follow those with fanatic views. Short-term rational influencers are less popular, since the average investor is 0.9 ppts less likely to follow them.

Overall, our findings suggest that past performance, trading intensity, and similar language or shared country of residence predict influencer popularity on the platform. For the subsample of influencers whose trading styles can be identified, we find followers are more likely to follow those whose short self-description is classified as long-term rational, followed by those classified as fanatics, naive, and short-term rational, respectively.

4. Identifying influencer impact

In the previous section, we identified factors that correlate with the decision to follow another investor. However, simply following someone does not necessarily translate into being influenced by that person. It is possible that users follow a certain influencer for reasons unrelated to their investment decisions. Similarly, high follower counts and high popularity do not necessarily translate into high impact. In this section, we further investigate whether influencers generate impact on their followers' investing behavior, both in terms of portfolio choice and trading decisions.

4.1. Measuring impact

We measure an influencer's impact on a follower by calculating what fraction of the follower's decisions in each period of time is identical to the influencer's. Specifically, we look at the overlap within the same influencer-follower pair in terms of both holding and trading decisions. We aggregate the daily trades to the monthly level. In unreported results, we also aggregated the trades to the quarterly level and the findings remain qualitatively the same.

4.1.1. Measuring holdings overlap

To quantify portfolio overlap, we start with the measure used in Pool, Stoffman and Yonker (2015), but modify it slightly to capture the overlap in securities between a follower's holdings and an influencer's at a point in time, as shown in Equation 3:

$$\text{PortOverlapRatio}_{f,i,t} = \frac{\sum_{k \in \mathcal{H}_t} \min\{l_{f,k,t}, l_{i,k,t}\}}{\sum_{k \in \mathcal{H}_t} l_{f,k,t}}, \tag{3}$$

where f indexes follower, i influencer, k security, and t time; and $\mathcal{H}_t$ is the set of all the securities person f holds at time t. By construction, this variable varies between 0 and 1. It is equal to 0 if none of follower f's holdings in month t overlap with influencer i's in the same month, and it is equal to 1 if all of follower f's holdings in month t are also in influencer i's portfolio. For example, if follower f holds securities A, B, and C, and influencer i holds B, C, D, and E at time t, the overlap would be 2/3. The numerator is 2 because securities B and C are the held by both the follower f and influencer i, and the denominator is 3 because the follower f holds 3 unique securities in this month.[12]

4.1.2. Measuring trade overlap

Similar to the definition of portfolio overlap, we measure the overlap in trading behavior between a follower f and influencer i for buying and selling separately, as specified in Equations

[12] Portfolios can also be compared using cosine similarity, as in Girardi et al. (2021). Similar to our approach, this produces a bounded measure that captures the closeness of two portfolios.

(4) and (5):

$$\text{BuyOverlapRatio}_{f,i,t} = \frac{\sum_{k \in \mathcal{T}_t} \min\left\{l^+_{f,k,t}, l^+_{i,k,t}\right\}}{\sum_{k \in \mathcal{T}_t} l^+_{f,k,t}} \tag{4}$$

$$\text{SaleOverlapRatio}_{f,i,t} = \frac{\sum_{k \in \mathcal{T}_t} \min\left\{l^-_{f,k,t}, l^-_{i,k,t}\right\}}{\sum_{k \in \mathcal{T}_t} l^-_{f,k,t}}, \tag{5}$$

where f indexes follower, i influencer, k security, and t time; and $\mathcal{T}_t$ is the set of all the securities person f trades at time t. The plus sign indexes purchases and the minus sign indexes sales. Similar to the holding overlap ratio, these two variables also vary between 0 and 1. It is equal to 0 if none of follower f's trades (either a purchase or sale) in month t overlap with influencer i's trades in the same direction in the same month, and it is equal to 1 if all of follower f's trades in month t are identical to influencer i's. For example, if follower f bought securities A, B, and C, and influencer i bought B, C, D, and E at time t, the overlap would be 2/3. The same holds for sales, except we would only count the unique securities sold.

4.2. IV identification strategy

In this subsection, we first present OLS results capturing the correlation between influencers' and followers' portfolios and trades, and then employ an IV-based identification strategy to obtain causal estimates of the impact of influencers on their followers.

4.2.1. Correlational results

After defining the main dependent variables in Section 4.1, we report the characteristics of both the real and pseudo influencer-follower pairs in Table 2. The summary statistics provide suggestive evidence that portfolio overlap is higher in the real pairs (3.5% per month) than the pseudo pairs (1.6% per month). The same patterns are also present for both buying and selling decisions. To formally test this relationship between investor decisions and influencer actions of interest, we run an OLS regression based on the following specification:

$$y_{f,i,t} = \beta \text{Follow}_{f,i} + \Gamma \text{Nr of unique securities}_{f,t} + \delta \text{Nr of unique securities}_{i,t} + \texttt{FE} + \epsilon_{f,i,t}, \tag{6}$$

where the dependent variable y measures either portfolio or trade overlap between influencer i and follower f at year-month t—including both purchases and sales—as defined in Equations (3)-(5).

The main independent variable *Follow* is a dummy variable that equals 1 if follower f follows influencer i, and 0 otherwise. We further control for the total number of securities that both the follower f and influencer i trade in the same month t, and follower fixed effects or follower-time fixed effects. Standard errors are clustered at the investor-level, since we expect the main variables of interest to vary across time within the same investor.

We report the regression estimates when *PortOverlapRatio* is the dependent variable in columns 1 and 3 of Table 5. Including follower-time fixed effects—meaning that we compare variation within the same follower and month whether the overlap ratio differs depending on whether the following relation is real or pseudo—does not change the magnitude and shows that following an influencer increases portfolio overlap by 2 ppts. Given that the average overlap is 3.5% (Table 2), this represents a both economically and statistically significant 57% increase over the mean.

We report the OLS coefficients in column 1 and 3 of Table 6 for purchases. The pattern is similar to what we found for portfolio overlap. The coefficient of 0.011 in column 3 indicates that following an influencer is associated with an increase in the pair's monthly purchase overlap by 1.1 percentage points, which corresponds to an increase of 65% (0.011/0.017) over the mean.

For sales, we find similar associations in column 1 and 3 in Table 7. Specifically, when controlling for follower-time (year-month) fixed effects in column 3, following an influencer is associated with an increase in a pair's monthly sales overlap by 1.2 ppts, which corresponds to an increase of 50% (0.012/0.024) over the mean.

While these findings document a relationship between the behavior of influencers and followers, they are not sufficient to establish causality. In particular, two concerns might make us question whether the estimated relationships are causal.

The first endogeneity concern is related to omitted variable bias. Since network formation is endogenous, observing that a follower and an influencer hold or trade the same security should not be interpreted as direct evidence of the influencer's impact. Rather, the behavior of follower might be explained by other factors, such as similarities in investment preferences or shared sources of information. For example, both the follower and influencer might prefer holding and trading in the same companies, sectors, or countries. They could also react to the same type of news in a similar way. Consequently, although the follower might frequently trade similar securities as the influencer, his or her decisions could be driven by a third factor that is also driving the influencer's decisions.

The second endogeneity concern is reverse causality: in order to gain and maintain popularity, an influencer might strategically hold or trade popular securities. As such, causality could be reversed: the influence could be exerted by the average follower on the influencer. To overcome these endogeneity concerns, we implement an IV strategy, where we exploit the fact that users are set to automatically follow certain platform employees at account creation.

The exact setting is as follows: new users register a profile on the platform with their digital identification. Upon successful profile creation, each user is set to automatically follow employees selected by the platform. The link between followers and influencers is effectively random, since it is not driven by common interests or sources of information but rather the design of the website.

In total we were able to identify 7 platform-made financial influencers that have trading history on the platform. Panel (g) and (h) in Table 1 reports their summary statistics: both their trading patterns and profile look similar to self-made influencers. Figure 2 illustrates as selection of popular influencers within the social trading graph. Visual inspection confirms that their positions in the network are central. Any activity (trades or posts) by a platform-made finfluencer will appear in their followers' newsfeeds.

Although users can subsequently decide to unfollow them, most users do not, as shown in the first-stage test below. The fact that some users opted out from this setting is also not a concern for the IV strategy, as shown in Angrist and Imbens (1994). Our 2SLS estimates the average effect of the treatment among compliers, or the local average treatment effect (LATE).

4.2.2. First stage

To implement the IV estimator, we run the following first stage regression:

$$\text{Follow}_{f,i} = \beta \text{platform-made influencers}_i + \texttt{FE} + \epsilon_{f,i}, \tag{7}$$

where the dependent variable *Follow* is 1 if follower f follows influencer i and 0 if not, as previously defined and visually illustrated in Figure 7. The main independent variable, *platform-made influencers*, is equal to 1 if the user is a platform employee that investors are assigned to automatically and a 0 otherwise. We report the coefficients in Table 4.

Controlling for investor fixed effects, the probability that a user would follow a platform-made influencers is 15 ppts higher than for other influencers. The unconditional probability that a follower follows an influencer is 5 ppts,[13] which is one third of the former, supporting the relevance of the instrument. Moreover, the F-statistic of 4180 suggests that the instrument is not weak. As a result, the IV estimates are unlikely to be biased toward the OLS estimates.

[13] We compute this as 53079/(53079+925679).

4.2.3. IV main results

In the the second-stage, we estimate the impact of influencers on followers' investment decisions, following the specification in Equation (8):

$$\mathrm{y}_{f,i,t} = \beta\widehat{\mathrm{Follow}}_{f,i} + \Gamma\mathrm{Nr\ of\ unique\ securities}_{f,t} + \delta\mathrm{Nr\ of\ unique\ securities}_{i,t} + \mathtt{FE} + \epsilon_{f,i,t}, \tag{8}$$

where we exploit a user's status of being employed by the platform to instrument for the follower's choice to follow. The dependent variable y measures either portfolio or trade overlap between influencer i and follower f at year-month t—including both purchases and sales—as defined in Equations (3)-(5).

The main independent variable *Follow* is a dummy variable that equals 1 if follower f follows influencer i, and 0 otherwise. We further control for the total number of securities that both the follower f and influencer i trade in the same month t, and follower fixed effects or follower-time fixed effects. Standard errors are clustered at the investor-level, since we expect the main variables of interest to vary across time within the same investor. Regression estimates are reported in columns 2 and 4 of Table 5 for portfolio overlap, Table 6 for purchases, and Table 7 for sales.

The IV estimate magnitudes in Table 5 exceed those of the OLS regressions. In column 4, where we control for follower-time fixed effects, the coefficients indicate an increase of portfolio overlap ratio of 3.8 ppts, corresponding to a 109% (0.038/0.035) increase relative to the mean.

In Table 6, the estimated coefficient magnitudes for the IV specifications are smaller than for OLS. When controlling for follower-time (year-month) fixed effects, the act of following an influencer is associated with an increase in the pair's purchase overlap by 0.3 percentage points, which corresponds to an 18% (0.003/0.017) increase over the mean.

Finally, in Table 7, the estimated effects are larger in the IV specifications relative to OLS. When controlling for follower-time (year-month) fixed effects in column 4, the act of following an influencer is associated with an increase in the pair's sales overlap by 4.6 percentage points, which corresponds to a 192% (0.046/0.024) increase over the mean. The stark difference between purchases and sales seems to suggest that followers are more sensitive to influencers' pessimistic views on securities than the optimistic ones.

4.3. Robustness

4.3.1. Reverse causality

In the previous sub-section, we demonstrated that both portfolio and trading decisions overlap more for real follower-influencer pairs than pseudo follower-influencer pairs at a monthly frequency. However, measuring the impact using the overlap ratio does not take into consideration the order of trades: namely, do influencers or their followers tend to trade first? To investigate this, we test the following specification

$$\mathrm{y}_{f,i,s,t} = \beta \mathrm{Follow}_{f,i} + \mathtt{FE} + \epsilon_{f,i,s,t}, \tag{9}$$

where the dependent variable y measures the distance in time between influencer i's trade and follower f's trade of the same security in year-month t. The main independent variable *Follow* is a dummy variable that equals 1 if follower f follows influencer i and 0 otherwise. We further control for security fixed effects and follower-time fixed effects. Standard errors are clustered at the investor-level, since we expect time lag to vary across time within the same investor.

We report the regression estimates in Table 8 where the first two columns are for purchases and the latter two columns are for sales. Including follower-time fixed effects does not change the magnitude and shows that followers trade on average 1.183 days after their influencer trades. For sales, this number is shorter: column 4 shows that followers tend to trade on the same day that their influencer sells.

The delay is also supported by price regressions where we replace the dependent variable y with price differences in the same currency. We report the OLS estimates in Table 9. In column 2, we see that, on average, followers buy the same security with a higher price in the same currency as their influencer and sell at 4.555 units lower (column 4). These patterns further support the conjecture that followers mimic their influencers' trades.

4.3.2. The validity of the IV strategy

Angrist and Imbens (1994) show that in cases where the population only contains compliers and noncompliers, 2SLS estimates the average effect of the treatment among compliers, or the local average treatment effect (LATE). For the IV to be valid and LATE not to be biased, four assumptions need to be met.

First, the instrument should be randomized or conditionally randomized with respect to the outcome and treatment variables (the ignorability of the instrument). We believe this assumption is met given that the setting is driven by a mechanical feature of the website. There is no official disclosure about the website's decision on this matter, but according to informal discussions with employees, the purpose of assigning platform-made influencers automatically to all users is to provide technical support and to communicate company- and platform-related news. The platform-made influencers who are allocated to followers are employed by the brokerage firm—and combining with the fact that it is unlikely that investors time their entrance into the platform based on which platform-made influencer is hired by the firm at the time—these decisions are arguably random from a user's perspective.

Second, the instrument must have an effect on the treatment—that is, a nonzero association between the IV and treatment variable (the relevance condition). We believe this assumption is also met in our setting as the assignment of platform-made influencers to followers is a default feature of the profile creation process. The strong first stage results shown in Table 5 also confirm this.

Third, there must not be users who would follow influencers if this were not the default option, but will not follow if it is (no defiers). Since the main purpose of investors joining the platform is for investment and all platform-made influencers work for the platform for an extended period of time, we do not believe the IV strategy would be challenged by the defiers in our setting.

Fourth, the instrument must have no direct effect on the outcome other than indirectly through the treatment (exclusion restriction). That is to say, the employees can not influence the platform investors investment decisions other than through being included in their following list. We manually inspected the backgrounds of platform-made influencers and did not find them to be influential in any other way. If not for their inclusion on users' lists, it is unlikely that their investment decisions would independently reach as many platform users. In summary, we believe our IV is valid and satisfies all four conditions.

5. Discussion of mechanism: How do influencers impact their followers?

Having established the causal effects of financial influencers on followers' portfolio composition and trading decisions, we next provide evidence on possible mechanisms behind these findings by exploring the heterogeneity in the impact across influencers and over followers. Specifically, at the influencer level, we test whether network effects and active engagement are the factors driving their influence. At the follower level, we test whether the number of influencers followed and certain investor demographics, such as age and gender, affect the degree of susceptibility to influencer impact. We modify the main regression as specified in Equation 8 by adding an extra dummy variable D, which is interacted with $\widehat{\text{Follow}}$, and captures heterogeneity across both followers and influencers. This dummy will capture different dimensions of influencer heterogeneity in different specifications, including the extent of their popularity and participation in group discussions. The regression specification is shown as below.

$$\begin{aligned} \mathrm{y}_{f,i,t} = {} & \beta_1 \widehat{\text{Follow}}_{f,i} + \beta_2 \widehat{\text{Follow}}_{f,i} * D + \beta_3 D + \Gamma \text{Nr of unique securities}_{f,t} \\ & + \delta \text{Nr of unique securities}_{i,t} + \texttt{FE} + \epsilon_{f,i,t}, \end{aligned} \tag{10}$$

5.1. Network effects

As an influencer's popularity increases, their influence tends to rise more than proportionally. The main reason is that network effects cause the influence of these individuals to extend beyond their direct followers. Their trades can be spread by their followers to their own networks, exponentially increasing the potential impact. With a larger follower base, their trades are able to reach a wider audience. This is the main reason social trading is generating more volumes than the traditional channels through which investors interact with each other. In addition, it is also plausible that followers interpret a substantial follower count as skill and credibility, assuming they have expertise, knowledge, or private information, which leads to the already popular influencers gaining even more followers.

We test whether network effects constitute a mechanism through which influencers' trades spread with the following two tests. First, we test whether having more followers is associated with greater influence. At the influencer level, we split influencers into two groups based on their numbers of followers: influencers with follower counts in the top quartile (equal to or more than 4,395 followers) are defined as popular, while those with follower counts in the bottom quartile (less than 1,184 followers) are defined as not popular. We then interact the dummy *Follow* with an indicator *HighPopularity* for whether an influencer is of high popularity in the main regression. We show the regression results in columns 1 and 2 of Table 10. Indeed, we find that the effect is mainly driven by the popular influencers. Economically, influencers in the top quartile in terms of follower counts increase portfolio overlap by 50.8 ppts, which corresponds to an increase of 14.5% from the mean.

Second, although impactful influencers generally have a large follower base, they are not necessarily central in the network. The impact of an individual with fewer followers, each of significant influence, is plausibly greater than that of someone with a larger number of followers who have minimal impact. We therefore test the effects of the network by investigating whether more central influencers—those who are able to reach more followers given the network structure—have higher impact. To test this hypothesis, we partition influencers into three groups based on their degree centrality and define those with in the top tercile as central influencers. The central influencers will have a dummy *Central* equal to 1. The rest of the influencers are classified as non-central and assigned the value of *Central* equal to 0. We show the results in column 3 and 4 of Table 10 where we find that the effect is driven by the central influencers. Compared to a non-central influencer, the central ones are able to increase portfolio overlap by 103.3 ppts. Economically, this is a sizable increase corresponding to more than 20 times from the mean.

5.2. Influencer activeness

Lastly, we investigate influencer activeness as a factor contributing to their impact. Influencers who actively participate in content creation and group discussions could have more impact. To test this hypothesis, we split influencers based on how many group discussions they participate in. We believe this measure is an appropriate measure for influencer activeness for two reasons. First, users have to seek approval to join a group, making it an active choice. Second, in contrast to their profile—where all the past trading records are automatically published—group participation requires them to actively post.

Since the median number of groups an influencer participates in is 233, we classify those influencers who are present in more than 233 groups as being more active and having a dummy *ManyGroups* equal to 1 and 0 otherwise. We interact this dummy with *Follow* and find (in Table 10) that, compared to influencers who are less active, more active influencers generate a higher portfolio overlap ratio of 28.3 ppts. Economically, this corresponds to an increase of 8 times relative to the mean.

5.3. Investor following number

Since information each follower is able to digest likely decreases with increasing number of influencers they follow, we anticipate that the effect should be more pronounced for those that follower fewer people. All else equal, followers who follow less people—and therefore, on average, notice more each influencer followed—are more affected within a follower-influencer pair.

To test this hypothesis, we divide investors into groups based on how many users they follow. Since the median number of users an average investor in our data follows is 37, we define *LowFollowingNumber* to be a dummy equal to 1 if the investor follows less or equal to 37 people (median number of influencers followed), and 0 otherwise. We then interact *LowFollowingNumber* with the dummy *Follow* as guided by Equation 10. The regression coefficients are reported in column 1 of Table 11. We find that it is indeed the case that people who follow fewer influencers are impacted more.

For investors who follow more than the median number of total users, the 2SLS estimate indicates that following influencers increases their portfolio overlap within an influencer-follower pair by 2 ppts. However for those who follow less than 37 people and therefore likely to allocate more attention to the followers, the increase is 4.6 (=2+2.6) ppts, which is more than twice the impact than the former.

5.4. Investor demographics

For the subsample of followers whose gender and age could be identified, we also test whether these factors play a role in increasing the susceptibility to influencer impact. Ex-ante, it is unclear whether and how these factors affect the susceptibility of investors to influencers' impact. We interact the main independent variable *Follow* with a dummy variable *Male* which is equal to 1 if the follower is identified as male and 0 if identified as female. The results are summarized in column 3 and 4 of Table 11. We find that, on average, male followers are 1.5 ppts less influenced than female followers; however, the difference is only statistically significant at the 10% level when controlling for follower-time fixed effects.

In column 5 and 6, we interact *Follow* with a dummy variable indicating whether the investor is born after the median identified birth year 1985. In this case, even though the estimated coefficient indicates that young investors are more impacted, we do not find a statistically significant difference between the older and younger group. This could, however, be a consequence of the small sample size.

Overall, these findings seem to suggest that follower attention is an important channel through which influencers impact retail investment. In addition, amplified by network effects and active involvement in group discussions and idea spreading, influencers are able to generate sizable impacts on their followers' real investment decisions.

6. Extensions and discussions

In this section, we discuss additional findings and explore issues of interest for regulators. Specifically, we investigate which types of trades are more impactful and what incentives motivate influencers' actions; and examine the generalizability of our results.

6.1. What types of trades are more influential?

From a policy perspective, it is important to know what types of trades are passed through from influencers to followers. To investigate this question, we re-examine the relationship between influencer and follower trades while separately calculating the overlap ratio of trades as broadly diversified products (Exchange Traded Funds (ETF) or passive index funds) and risky ones which are derivatives with high leverage or investments in crypto assets such as Bitcoin.[14] We then repeat the main analysis as in Equation 7 (results in Table 5), but replace the dependent variables with the portfolio overlap ratio in ETF or risky security shares.

We present the findings on portfolio overlap ratio in Table 12. We scale up the dependent variable by a factor of 100 to simplify the interpretation, since the mean overlap ratio is small. In columns 1 and 2, the IV estimates show that, similar to our main findings, influencers' trades increase the index product portfolio overlap by 4.7 percentage points. However, the effect is the opposite for risky shares. In unreported results, similar patterns are present for the security purchases and sales. Overall, this suggests that the effect of influencers' investment decisions on their followers is more pronounced for the passive index funds but

[14] Index products are not necessarily of low risk but are less risky compared with the crypto-related products.

not for risky investments. In other words, followers are selective in what type of trades they mimic of the influencers in their social network.

6.2. What are influencers' incentives?

We then investigate what factors could be driving influencers' decisions to engage in activities on the platform and with their followers. In contrast to trading in traditional financial markets, financial performance might not be the only driver of financial influencers' decisions. Most financial influencers appear independent and do not receive compensation from their followers but still share their investment suggestions with them. It takes effort to become a finfluencer—since one has to be active in trading and engaging in group discussions. Therefore, it is important to understand the incentives influencers have in influencing their followers' trades—especially regarding monetary incentives.

To better understand influencers' incentives, we compare those who are affiliated with the platform (platform-made influencers) to those who are not. Platform-made influencers work or are associated with the platform in some way and, thus, may have an incentive to promote the financial interests of the platform. In addition to the platform-made influencers, we also manually identified other influencers who are affiliated with the platform through coverage in the news or the influencers' own social media accounts (Twitter, LinkedIn, blogs, and websites). Many of the relationships between these investors and the platform were only available online after they stopped working for the platform. We examine this using the regression specification shown in Equation (11):

$$y_{i,t} = \beta \text{Platform-made}_i + \texttt{FE} + \epsilon_{i,t}, \tag{11}$$

where $y_{i,t}$ is the outcome variable measuring trading characteristics of the influencer, and the main independent variable *Platform-made* is a dummy variable equal to 1 if the influencer is platform-made and 0 otherwise. We investigate whether an influencer who is affiliated with—and therefore either directly or indirectly financially compensated by—the platform trades differently than an independent influencer.

The regression estimates are reported in Table 13. In column 1, the dependent variable is an indicator for whether the product is issued by the platform. The coefficient indicates that platform-made influencers are 19.9% more likely to trade products (usually index funds or derivatives) issued by the platform. Since the average probability of trading any platform-issued product in the transaction data is 7%, this translates into a 284% increase over the mean.

In column 2, the dependent variable is the number of trades per month. We find that platform-affiliated influencers make on average 1.86 more trades than non-platform-affiliated influencers per month. This number may appear small in isolation, but after factoring in the network effects and long tenure each follower has on the platform, the volume generated by their trades could yield sizable commissions and revenue for the platform.[15]

Finally, it is possible that these platform-made influencers receive discounts on transactions fee when they trade products issued by the platform, which could explain the pattern in column 1. However, if this is indeed the case, we should observe them having a higher Sharpe ratio than the other influencers. We formally test this hypothesis and find that their performance rating is not significantly higher than the rest of the users as shown by a cross-sectional test in column 3.[16] Depending on whether the platform-made influencers are aware of this, this finding lends support to the notion that platform-made influencers tend to trade more often in order to increase total trading volume on the platform. Since the platform profits from executing trades on behalf of investors, this finding raises the concern that platform-made influencers may have incentives that do not align with those they influence.[17]

[15]On average, the platform charges between 0.069 to 0.25% in brokerage fees depending on the security type. In addition, conversations with anonymous employees confirm that they receive sizable kick-backs from other financial intermediaries when selling their products.

[16]This result is also consistent with the finding in (Barber and Odean, 2000) that active trading hurts individual investors' performance.

[17]The platform offers access for free, but profits from the other services customers buy (see e.g., Gabaix and Laibson 2006 and Qi 2024).

In summary, we find suggestive evidence that financial influencers affiliated with the platform trade in a fashion that financially benefits the platform, raising the concern over conflicts of interest.

6.3. External validity discussion

Recent developments in IV literature, summarized in Mogstad and Torgovitsky (2024), highlight the importance of accounting for unobserved heterogeneity in treatment effects across individuals. Since our estimated effects are for the compliers (LATE), we might be concerned that they are individuals who opt for treatment, limiting the generalizability of the results.

To address this concern, we conduct robustness checks where examine investors' first month of trading on the platform. The reason for this is that it takes time for investors to get familiar with the platform and notice the presence of influencers in their following list. As such, the majority of the platform users are compliers in their first month on the platform.

We repeat the tests of Equation 6 and Equation 8. We report the estimates for portfolio overlap in Table 14, the estimates for purchases in Table 15, and finally the estimates for sales in Table 16. The coefficients remain statistically significant for portfolio overlap and purchase overlap at the 1% level and decrease to the 10% level for sales, likely due to the smaller sample size. The difference in sample size for sales and purchases is reasonable: sales happen less often, since investors hold fewer securities when they are new to the platform. Compared to the main results, the economic significance is lower for portfolio overlap and sales, but increases for purchases. Overall, the main effects remain similar for non-compilers.

Given the combination of both completeness of the network and the nearly decade-long transaction history and unique IV strategy, our results have important implications for markets where financial influencers have a presence. Our findings are particularly relevant for contexts where influencers' actual holdings and trades are observable, as many policy makers have recommended they should be.

Ex-ante, it is unclear how the truthful disclosure of trading activities affects the impact influencers have on their followers. On the one hand, truthful reporting and transparency could increase their influence, as followers see the influencers' views on the underlying securities as more credible. For example, GameStop shares rose 21% in a single day when a finfluencer posted his holdings of the stock.[18] On the other hand, certain investors could take an opposing positions to profit from this information. The net effects depend on the characteristics of the influencers, types of trades, and the types of audience that they influence.

7. Conclusion

This paper constructs a new data set and uses it to quantify the impact that financial influencers have on their followers' real investment decisions. Leveraging data from an investment trading platform operating in four Nordic countries, we observe both a snapshot of the network of relationships between influencers and their followers, and a decade's history of actual daily transactions executed by traders.

Our analyses show that factors such as a shared language, a common country of residence, better past performance, and a longer trading history appear to correlate with influencer popularity. Additionally, by classifying financial influencers' trading styles into categories—following the scheme introduced in Pedersen (2022*b*)—we show that long-term rational influencers tend to be the most popular, followed by fanatics.

Employing a unique IV strategy, we find that influencers have a sizable causal impact on followers' portfolios and trading decisions. Specifically, the impact on sales is more pronounced than purchases. We also document considerable heterogeneity in the effects. For example, the measured impact is more pronounced for influencers who are popular, central, and active in group discussions. It is also stronger for followers who follow relatively fewer influencers and for female followers. In addition, we find suggestive evidence that the diver-

[18] See the full news reporting at `https://www.cnbc.com/2024/06/02/gamestop-jumps-as-roaring-kitty-trader-posts-giant-116-million-stock-position.html`.

sified products, such as passive index funds, are more impactful than risky trades in levered products.

Finally, we investigate the incentives influencers face and how it influences their behavior. We find suggestive evidence that influencers at least partially seek to monetize their influence through their activities and influence, rather than exclusively focusing on returns. As such, there may be gains to increasing transparency. These findings advance our understanding of an increasingly important phenomenon—namely, the emergence of financial influencers—and provide insights that address questions of growing urgency in the financial industry and among regulators.

References

Ammann, Manuel, and Nic Schaub. 2021. "Do Individual Investors Trade on Investment-Related Internet Postings?" *Management Science*, 67(9): 5679–5702.

Angrist, Joshua, and Guido Imbens. 1994. "Identification and Estimation of Local Average Treatment Effects." *Quarterly Journal of Economics*, 62(2).

Aridor, Guy, Rafael Jiménez Durán, Ro'ee Levy, and Lena Song. 2024. "The Economics of Social Media." Available at SSRN: `https://ssrn.com/abstract=4708840` or `http://dx.doi.org/10.2139/ssrn.4708840`.

Bailey, Michael, Drew Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2022. "Peer Effects in Product Adoption." *American Economic Journal: Applied Economics*, 14(3): 488–526.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*a*. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives*, 32(3): 259–280.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018*b*. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy*, 126(6): 2224–2276.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018*c*. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy*, 126(6): 2224–2276.

Barber, Brad M, and Terrance Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance*, 55(2): 773–806.

Barber, Brad M, and Terrance Odean. 2001*a*. “Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment.” *Quarterly Journal of Economics*.

Barber, Brad M, and Terrance Odean. 2001*b*. “The Internet and the Investor.” *Journal of Economic Perspectives*, 15(1): 41–54.

Barber, Brad M, and Terrance Odean. 2002. “Online Investors: Do the Slow Die First?” *The Review of Financial Studies*, 15(2).

Barber, Brad M, and Terrance Odean. 2008. “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors.” *Review of Financial Studies*, 21(2): 785–818.

Barber, Brad M, and Terrance Odean. 2013. “The Behavior of Individual Investors.” In *Handbook of the Economics of Finance*. Vol. 2, 1533–1570. Elsevier.

Barber, Brad M, Xing Huang, Terrance Odean, and Christopher Schwarz. 2022. “Attention-Induced Trading and Returns: Evidence from Robinhood Users.” *The Journal of Finance*, 77(6): 3141–3190.

Barber, Brad M, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean. 2009. “Just How Much Do Individual Investors Lose by Trading?” *Review of Financial Studies*, 22(2): 609–632.

Barberis, Nicholas, and Ming Huang. 2001. “Mental Accounting, Loss Aversion, and Individual Stock Returns.” *The Journal of Finance*, LVI(4).

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. “A Model of Investor Sentiment.” *Journal of Financial Economics*.

Barberis, Nicholas, and Wei Xiong. 2009. “What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation.” LXIV(2).

Barberis, Nicholas, Ming Huang, and Richard H Thaler. 2006. "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing." *The American Economic Review*, 96(4).

Benetton, Matteo, William Mullins, Marina Niessner, and Jan Toczynski. 2024. "Celebrity Tweets: This Is Not Financial Advice."

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1998. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." *Journal of Economic Perspectives*, 12(3): 151–170.

Bikhchandani, Sushil, David Hirshleifer, Omer Tamuz, and Ivo Welch. 2021. "Information Cascades and Social Learning."

Cookson, J Anthony, and Marina Niessner. 2020. "Why Don't We Agree? Evidence from a Social Network of Investors." *The Journal of Finance*, 75(1): 173–228.

Cookson, J Anthony, Joseph E Engelberg, and William Mullins. 2023. "Echo Chambers." *The Review of Financial Studies*, 36(2): 450–500.

Cookson, J. Anthony, Marina Niessner, and Christoph M. Schiller. 2022. "Can Social Media Inform Corporate Decisions? Evidence from Merger Withdrawals." Center for Open Science SocArXiv 56yrj.

Cookson, J. Anthony, Runjing Lu, William Mullins, and Marina Niessner. 2024. "The social signal." *Journal of Financial Economics*, 158: 103870.

Cookson, J. Anthony, William Mullins, and Marina Niessner. 2024. "Social Media and Finance." *SSRN Electronic Journal*.

Daniel, Kent, and David Hirshleifer. 2015. "Overconfident Investors, Predictable Returns, and Excessive Trading." *Journal of Economic Perspectives*, 29(4): 61–88.

Dim, Chukwuma. 2025. "Social Media Analysts' Skill: Evidence from Text-implied Beliefs." *Journal of Financial and Quantitative Analysis*, 1–69.

FINRA. 2023. "Gen Z and Investing: Social Media, Crypto, FOMO, and Family." CFA Institute.

Frydman, Cary, Nicholas Barberis, Colin Camerer, Peter Bossaerts, and Antonio Rangel. 2014. "Using Neural Data to Test a Theory of Investor Behavior: An Application to Realization Utility." *The Journal of Finance*, 69(2): 907–946.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics*.

Girardi, Giulio, Kathleen W. Hanley, Stanislava Nikolova, Loriana Pelizzon, and Mila Getmansky Sherman. 2021. "Portfolio similarity and asset liquidation in the insurance industry." *Journal of Financial Economics*, 142(1): 69–96.

Han, Bing, David Hirshleifer, and Johan Walden. 2022. "Social Transmission Bias and Investor Behavior." *Journal of Financial and Quantitative Analysis*, 57(1): 390–412.

Heimer, Rawley. 2016. "Peer Pressure: Social Interaction and the Disposition Effect." *Review of Financial Studies*.

Heimer, Rawley Z. 2014. "Friends Do Let Friends Buy Stocks Actively." *Journal of Economic Behavior & Organization*, 107: 527–540.

Heimer, Rawley Z., and David Simon. 2015. "Facebook Finance: How Social Interaction Propagates Active Investing." Federal Reserve Bank of Cleveland Working Paper (Federal Reserve Bank of Cleveland) 15-22.

He, Pengcheng, Jianfeng Gao, and Weizhu Chen. 2021. "DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing."

Hirshleifer, David. 2001. "Investor Psychology and Asset Pricing." *The Journal of Finance*, 56(4): 1533–1597.

Hirshleifer, David. 2015. "Behavioral Finance."

Hirshleifer, David. 2020. "Presidential Address: Social Transmission Bias in Economics and Finance." *The Journal of Finance*, 75(4): 1779–1831.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. 2004. "Social Interaction and Stock-Market Participation." *The Journal of Finance*, 59(1): 137–163.

Jiang, Lei, Jinyu Liu, Lin Peng, and Baolian Wang. 2022. "Investor Attention and Asset Pricing Anomalies." *SSRN Electronic Journal*.

Kakhbod, Ali, Seyed Kazempour, Dmitry Livdan, and Norman Schurhoff. 2023. "Finfluencers."

Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives*, 14(3): 115–136.

Mogstad, Magne, and Alexander Torgovitsky. 2024. "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects." In *Handbook of Labor Economics*. Vol. 5, 1–114. Elsevier.

Odean, Terrance. 1998*a*. "Are Investors Reluctant to Realize Their Losses?" *The Journal of Finance*, 53(5): 1775–1798.

Odean, Terrance. 1998*b*. "Volume, Volatility, Price, and Profit When All Traders Are Above Average." *The Journal of Finance*, 53(6): 1887–1934.

Odean, Terrance. 1999. "Do Investors Trade Too Much?" *The American Economic Review*, 89(5).

Ozsoylev, Han N., Johan Walden, M. Deniz Yavuz, and Recep Bildik. 2014. "Investor Networks in the Stock Market." *Review of Financial Studies*, 27(5): 1323–1366.

Pedersen, Lasse Heje. 2022*a*. "Game on: Social networks and markets." *Journal of Financial Economics*, 146(3): 1097–1119.

Pedersen, Lasse Heje. 2022*b*. "Game on: Social Networks and Markets." *Journal of Financial Economics*, 146(3): 1097–1119.

Pool, Veronika K., Noah Stoffman, and Scott E. Yonker. 2015. "The People in Your Neighborhood: Social Interactions and Mutual Fund Portfolios." *The Journal of Finance*, 70(6): 2679–2732.

Pushp, Pushpankar Kumar, and Muktabh Mayank Srivastava. 2017. "Train Once, Test Anywhere: Zero-Shot Learning for Text Classification." *CoRR*, abs/1712.05972.

Qi, Yingjie. 2024. "Big Broad Banks: How Does Cross-Selling Affect Lending?" *Review of Finance*, 28(2): 551–592.

Raffo, Julio, and Gema Lax-Martinez. 2018. "WGND 1.0."

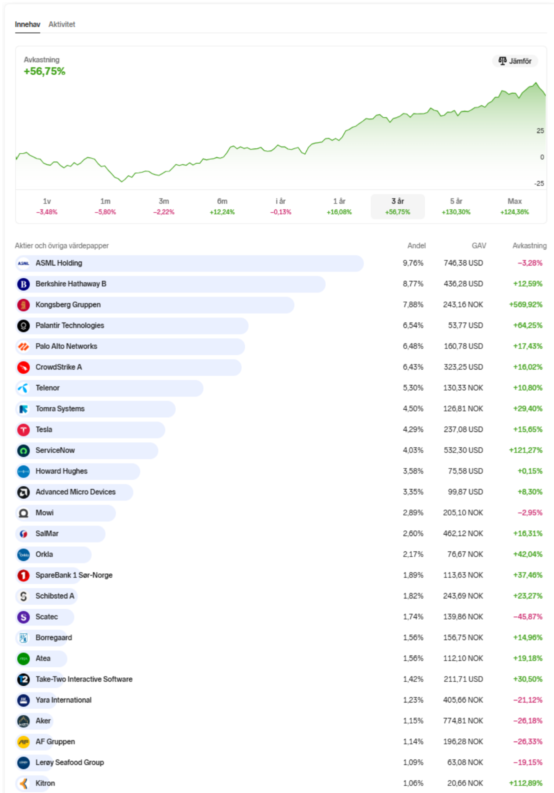
Shleifer, Andrei, and Lawrence H Summers. 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives*, 4(2): 19–33.

Sui, Pengfei, and Baolian Wang. 2022. "Social Transmission Bias: Evidence from an Online Investor Platform." *SSRN Electronic Journal*.

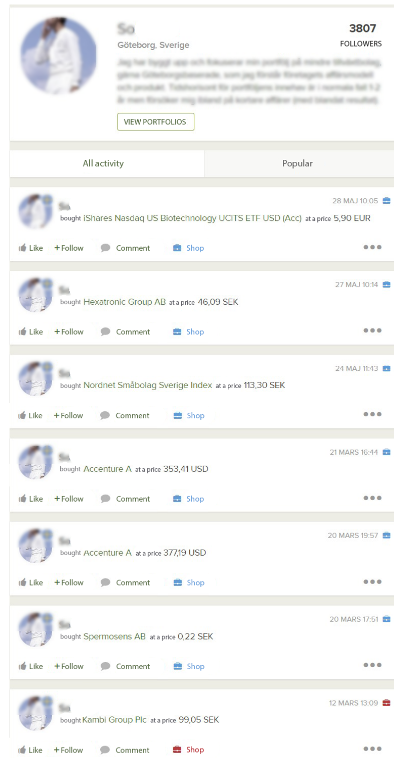
8. Figures and tables

Figure 1: Example of an investor profile on the platform

(a) Portfolio overview

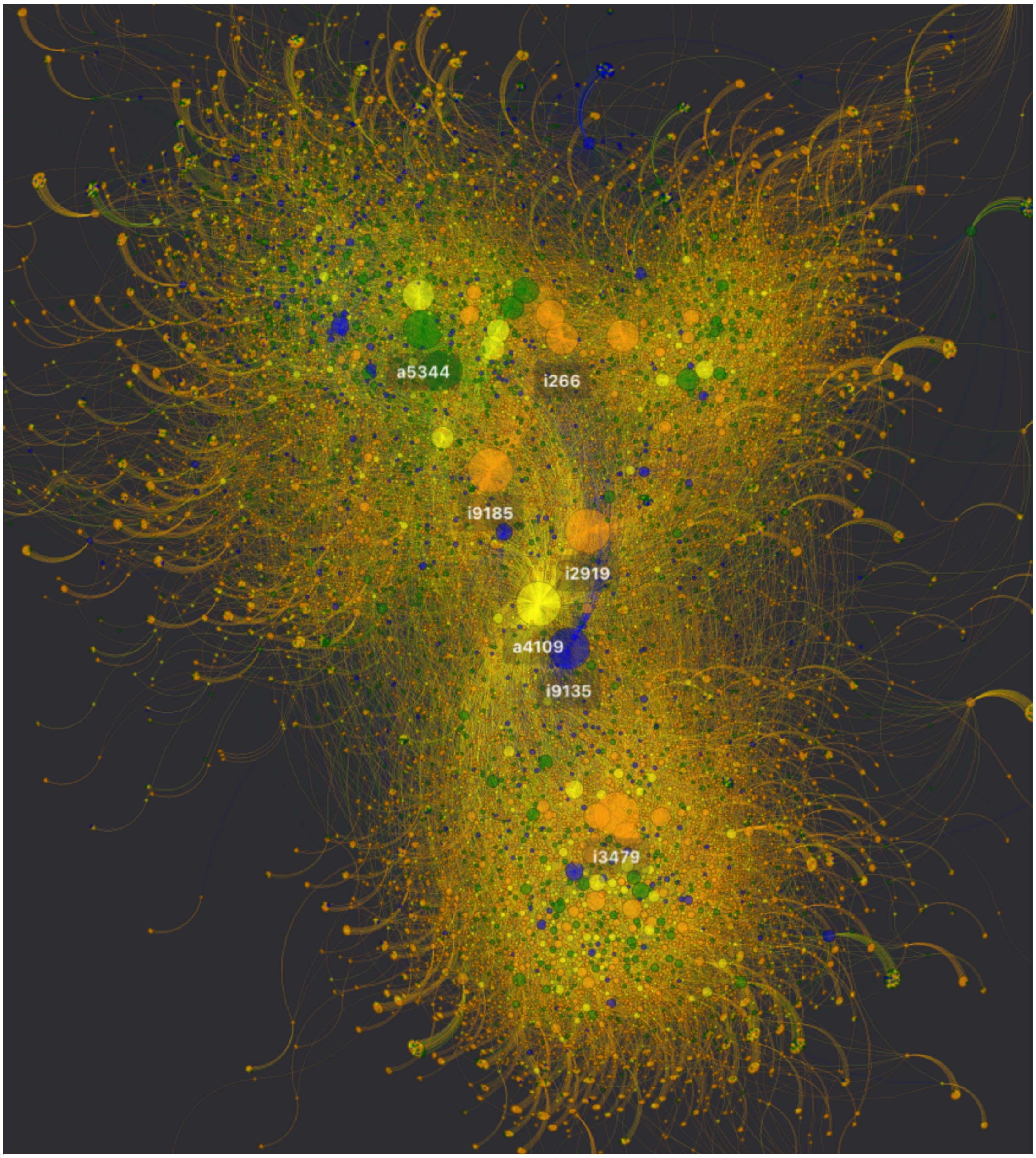


(b) Profile and activity overview



Notes: The figure above shows an example portfolio overview in the left panel and profile and activity page in the right panel. Information are extracted from the platform but personally identifiable information has been blurred. The original information on the profile page was translated from Swedish to English.

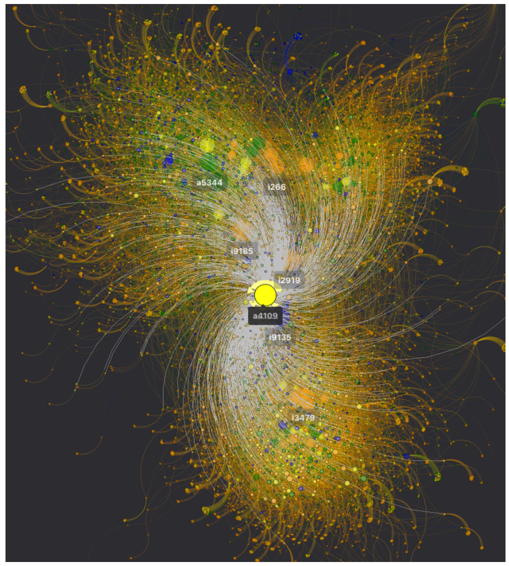
Figure 2: Full Network



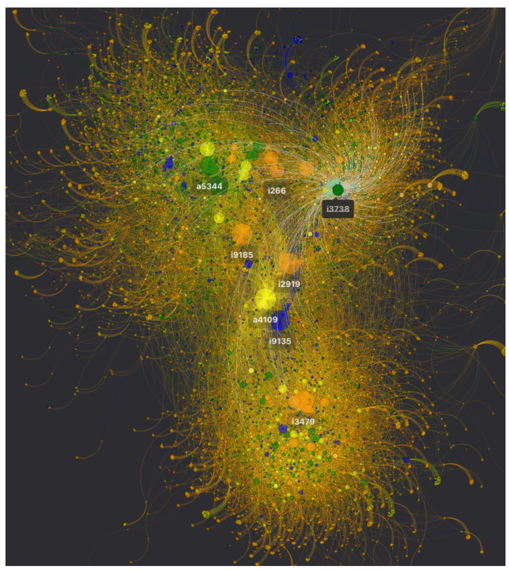
Notes: The figure above shows the full network of relationships in our sample. Colors correspond to performance ratings assigned to influencers. In order from lowest to highest return, the groups are as follows: orange (rating is zero), yellow (if the investor's portfolio has a positive return since joining), green (the investor's portfolio return is ranked among the top 50% in the entire platform), and blue (the investor's portfolio return is ranked among the top 10% in the entire platform). Follower nodes take the color of the influencer with the highest return rating that they follow. Influencers with the most followers are labelled. Influencers who are platform-made have an 'a' prefix at the start of their ID; whereas influencers who are not affiliated with the platform have an 'i'.

Figure 3: Influencer Subnetworks

(a) Influencer 1



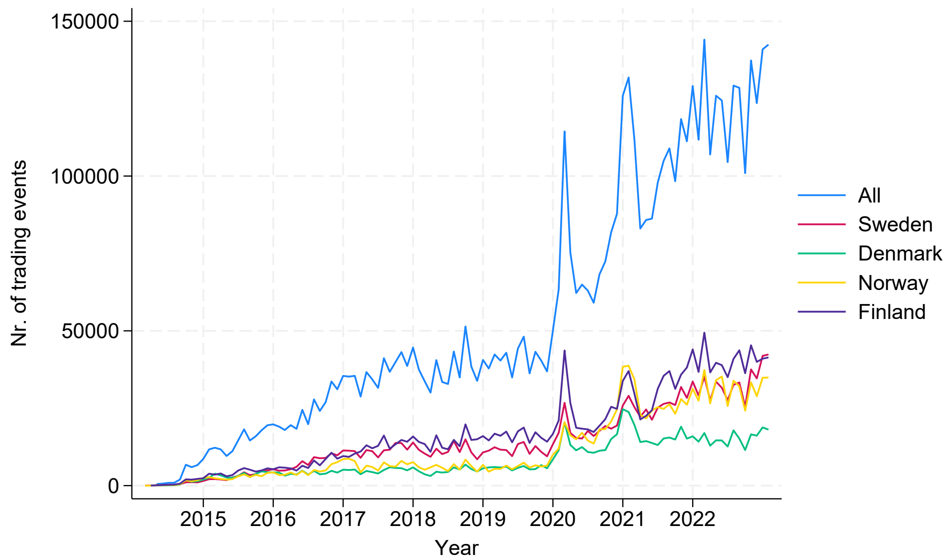
(b) Influencer 2



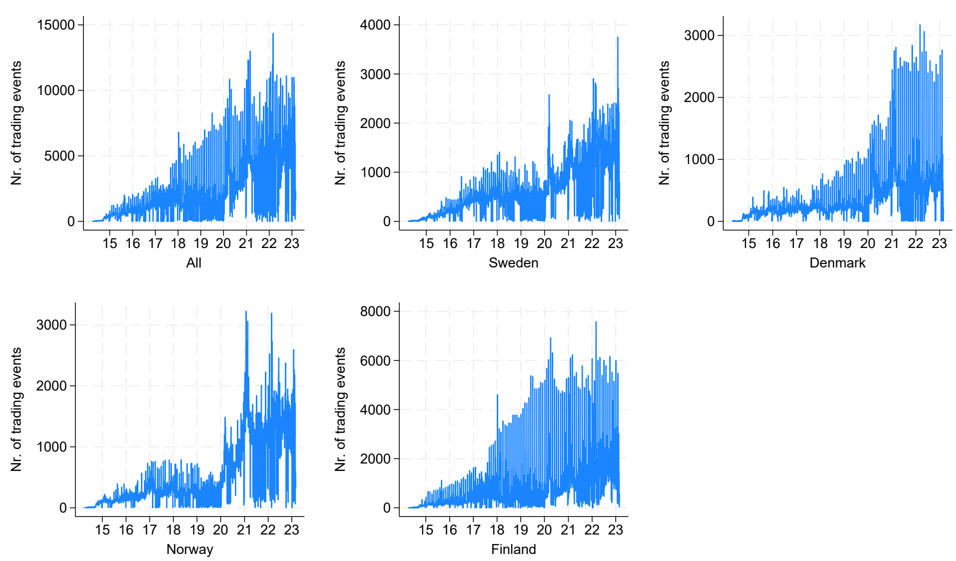
Notes: The figure above shows two example influencer subnetworks. Each node represents an investor and each edge represents a influencer-follower relationship. Colors correspond to investment returns ratings assigned to influencers. In order from lowest to highest return, the groups are as follows: orange (rating is zero), yellow (if the investor's portfolio has a positive return since joining), green (the investor's portfolio return is ranked among the top 50% in the entire platform), and blue (the investor's portfolio return is ranked among the top 10% in the entire platform). Follower nodes take the color of the influencer with the highest return rating that they follow. Gray edges indicate an influencer-follower relationship between the selected influencer and an investor in the network. Influencers with the most followers are labeled. Influencers who are platform-made have an 'a' prefix at the start of their ID; whereas influencers who are not affiliated with the platform have an 'i'.

Figure 4: Number of daily and monthly trading events

(a) Monthly

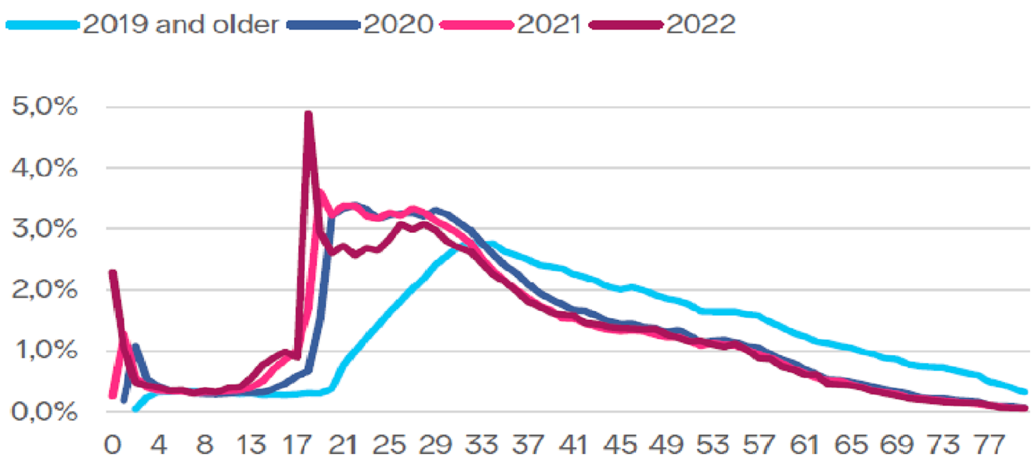


(b) Daily



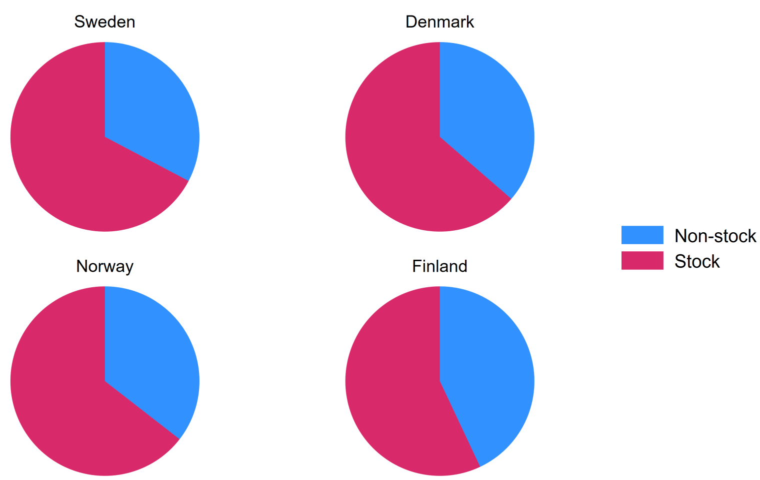
Notes: The figure in panel (a) shows the total number of trading events on the platform at a monthly frequency from September 2014 to March 2023 for trades made by users whose country of residence can be identified. The figure in panel (b) shows separate daily plots of trading activity for Sweden, Denmark, Norway, and Finland.

Figure 5: Share of customers by cohort and age



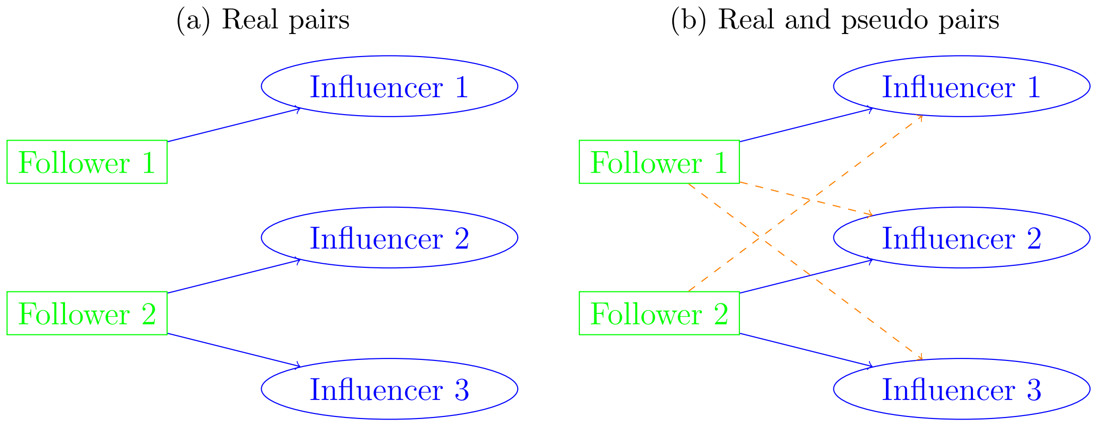
Notes: The figure above shows the investors' age distributions for the 2019 and older cohort, the 2020 cohort, the 2021 cohort, and the 2022 cohort. The horizontal axis measures age while the vertical axis measures density. Source: the anonymous platform studied in this paper.

Figure 6: Trades by type and country



Notes: The figure above shows the composition of types of investments made on the platform from September 2014 to March 2023 for trades made by users whose nation of residence can be identified. Stock (in red) represents investments made directly into a stock, while non-stock (in blue) represents investments made into non-equity products such as index funds or derivatives. We plot separately for Sweden, Denmark, Norway, and Finland.

Figure 7: Construction of pseudo pairs



Notes: The figure above shows how we construct the counterfactual of following influencers. Followers are shown in green hollow squares and influencers are shown in blue solid oval shapes. Panel a illustrates the real follower-influencer relations as shown in Figure 2 with blue directed arrows. Panel b illustrates the pseudo relations that connect followers and all influencers that they do not follow with dashed orange directed arrows.

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	P1	P50	P99
Panel a: Influencer-time level						
# of trades (month)	3530	5.993	9.095	1	3	41
# of purchases (month)	3530	2.203	4.46	0	1	19
# of sales (month)	3530	3.79	5.791	0	2	28
# of index trades (month)	3530	.39	1.816	0	0	7
# of derivatives (month)	3530	.224	1.281	0	0	6
Panel b: Influencer level						
# of trades (all)	100	211.55	153.084	2.5	168	620.5
# of purchases (all)	100	77.76	64.238	.5	67	234
# of sales (all)	100	133.79	108.325	1.5	118	559.5
# of index trades (all)	100	13.77	40.821	0	0	224
# of derivatives (all)	100	7.92	21.593	0	0	114.5
Max number of followers	100	27436.84	64956.61	853	1876	266553
Min rating	100	.16	.526	0	0	2.5
Max rating	100	.94	1.023	0	1	3
Years on the platform	100	4.18	2.83	0	4	9
Male	21	.857	.359	0	1	1
Panel c: Follower-time level						
# of trades (month)	347114	6.132	10.471	1	3	48
# of purchases (month)	347114	1.767	4.768	0	0	20
# of sales (month)	347114	4.364	6.554	0	3	30
# of index trades (month)	347114	1.209	2.025	0	0	8
# of derivatives (month)	347114	.497	5.293	0	0	11
Panel d: Follower level						
# of trades (all)	10485	202.988	146.833	10	175	420
# of purchases (all)	10485	58.501	66.281	0	33	213
# of sales (all)	10485	144.488	99.417	7	130	390
# of index trades (all)	10485	40.016	59.577	0	13	263
# of derivatives (all)	10485	16.443	68.433	0	0	354
Max number of following	10485	37.852	65.088	2	16	345
Years on the platform	10485	3.593	2.28	0	3	8
Male	1417	.733	.443	0	1	1
Birth year	87	1983.816	12.635	1955	1985	2006
Panel e: All users-time level						
# of trades (month)	981388	5.775	9.848	1	3	44
# of purchases (month)	981388	1.866	4.587	0	0	20
# of sales (month)	981388	3.907	6.123	0	2	28
# of index trades (month)	981388	.828	1.727	0	0	7
# of derivatives (month)	981388	.465	4.926	0	0	10

Panel f: All users level						
# of trades (all)	33662	168.355	144.117	1	125	420
# of purchases (all)	33662	54.408	61.786	0	29	212
# of sales (all)	33662	113.906	97.155	0	88	368
# of index trades (all)	33662	24.152	48.316	0	1	232
# of derivatives (all)	33662	13.558	56.523	0	0	337
Max number of following	12843	35.083	63.488	1	14	338
Years on the platform	32269	3.882	2.372	0	4	8
Male	5293	.788	.409	0	1	1
Birth year	104	1983.135	12.156	1956	1984	2001
Panel g: All platform influencer-time level						
Nr of trades (month)	169	8.627	15.026	1	4	62
Nr of purchases (month)	169	1.142	5.446	0	0	13
Nr of sales (month)	169	7.485	11.836	0	3	62
Nr of index trades (month)	169	2.627	5.803	0	0	28
Nr of derivatives (month)	169	.03	.202	0	0	1
Panel h: All platform influencer level						
Nr of trades (all)	7	208.286	291.601	1	127	821
Nr of purchases (all)	7	27.571	29.177	0	31	81
Nr of sales (all)	7	180.714	264.21	0	96	740
Nr of index trades (all)	7	63.429	115.678	0	18	322
Nr of derivatives (all)	7	.714	1.89	0	0	5
Max number of followers	7	117745.7	106038	1137	181250	227913
Min rating	7	0	0	0	0	0
Max rating	7	1.286	.488	1	1	2
Years on the platform	7	2.571	1.988	0	2	6
Male	5	1	0	1	1	1

Notes: The table above shows summary statistics for influencers (investors with top-100 ranked number of followers) in Panel (a) and (b), their followers in Panel (c) and (d), all investors on the platform in Panel (e) and (f), and finally all the platform-made influencers in Panel (g) and (h).

Table 2: Summary statistics: main regression

Variable	Obs	Mean	Std. Dev.	P1	P50	P99
Panel a: Pseudo pairs						
monthly level						
# of unique securities	29048853	8.315	7.618	1	6	36
Influencer # of unique securities	29048853	10.353	9.881	1	8	46
# of common securities	29048853	.127	.488	0	0	2
Portfolio Overlap Ratio	29048853	.016	.073	0	0	.333
# of unique purchase	3831718	3.281	4.011	1	2	19
Influencer # of unique purchase	3831718	3.271	3.844	1	2	18
# of common purchase	3831718	.015	.139	0	0	1
Buy Overlap Ratio	3831718	.006	.061	0	0	.167
# of unique sale	11175294	3.699	3.698	1	3	18
Influencer # of unique sale	11175294	3.555	3.641	1	2	18
# of common sale	11175294	.033	.226	0	0	1
Sell Overlap Ratio	11175294	.009	.074	0	0	.333
pair level						
Follow	925679	0	0	0	0	0
Same language	915427	.504	.5	0	1	1
Same country	915427	.244	.429	0	0	1
Panel b: Real pairs						
monthly level						
# of unique securities	1307312	9.298	8.381	1	7	39
Influencer # of unique securities	1307312	10.859	13.713	1	7	79
# of common securities	1307312	.345	.931	0	0	4
Portfolio Overlap Ratio	1307312	.035	.104	0	0	.5
# of unique purchase	170257	3.466	4.216	1	2	20
Influencer # of unique purchase	170257	3.236	3.76	1	2	17
# of common purchase	170257	.047	.257	0	0	1
Buy Overlap Ratio	170257	.017	.105	0	0	.5
# of unique sale	505048	4.071	4.099	1	3	20
Influencer # of unique sale	505048	4.496	4.537	1	3	19
# of common sale	505048	.087	.381	0	0	2
Sell Overlap Ratio	505048	.024	.115	0	0	.667
pair level						
Follow	53049	1	0	1	1	1
Same language	52722	.765	.424	0	1	1
Same country	52722	.487	.5	0	0	1

Notes: The table above shows summary statistics for the variables used in the main analyses. Statistics for the pseudo pairs are in Panel (a) and real pairs are given in Panel (b).

Table 3: Influencer popularity

	ln(1+number of followers)				Follow			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Max rating=1	0.100***		0.087***	0.115***				
	(0.001)		(0.002)	(0.006)				
Max rating=2	0.152***		0.125***	0.206***				
	(0.003)		(0.002)	(0.009)				
Max rating=3	0.262***		0.274***	0.229***				
	(0.008)		(0.008)	(0.022)				
# of trades		0.002***	0.002***	0.002***				
		(0.000)	(0.000)	(0.000)				
Male				0.153**				
				(0.042)				
Same language					0.063***		0.021***	0.033***
					(0.001)		(0.001)	(0.001)
Same country						0.080***	0.063***	0.061***
						(0.001)	(0.001)	(0.002)
Long term rational								0.083***
								(0.002)
Short term rational								-0.009***
								(0.002)
Fanatic								0.006***
								(0.002)
Influencer cohort FE	Yes	Yes	Yes	Yes	No	No	No	No
Influencer FE	No	No	No	No	Yes	Yes	Yes	No
Investor FE	No	No	No	No	Yes	Yes	Yes	Yes
Adj. R2	0.042	0.083	0.086	0.092	0.176	0.181	0.182	0.084
No of obs	22,868	22,868	22,868	3,740	968,149	968,149	968,149	124,789

Notes: The table above shows two sets of regression estimates. Columns 1-4 show cross-sectional regressions at the influencer level where the dependent variable is log(1+number of followers). Column 5-8 is a panel regression at the follower-influencer pair level where the dependent variable is a dummy *Follow* which is equal to 1 if the follower f follows influencer i and 0 otherwise. Standard errors clustered at the influencer rating level in columns 1-4 and at the follower level in column 5-8 are shown in the parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: IV first stage results

	Follow
	(1)
Instrument	0.150***
	(0.002)
Investor FE	Yes
F	4180
Adj. R2	0.060
No of obs	978,728

Notes: The table above shows the estimation results for Equation 7. The dependent variable is a dummy *Follow* which equals 1 if follower f follows influencer i and 0 otherwise. The independent variable is a dummy equal to 1 if the influencer is a platform-made influencer and 0 if not. Standard errors are clustered at the follower level and included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The impact of finfluencers' portfolio composition on followers'

	OLS	Second stage	OLS	Second stage
	Portfolio Overlap Ratio			
	(1)	(2)	(3)	(4)
Follow	0.020***	0.036***	0.020***	0.038***
	(0.000)	(0.002)	(0.000)	(0.002)
Influencer # of unique securities	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
# of unique securities	-0.000***	-0.000***		
	(0.000)	(0.000)		
Investor FE	Yes	Yes	No	No
Investor x Time FE	No	No	Yes	Yes
Adj. R2	0.039	0.019	0.046	0.020
No of obs	30,356,165	30,356,165	30,356,165	30,356,165

Notes: The table above shows OLS and 2SLS regression estimates following Equation (6) and (8). The dependent variable is *Portfolio Overlap Ratio*, which is the fraction of total number of common securities held in month t by follower f and influencer i divided by the total number of unique securities held by follower f. Standard errors clustered at the follower level are included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: The impact of finfluencers' purchases on followers'

	OLS	Second stage	OLS	Second stage
	Buy Overlap Ratio			
	(1)	(2)	(3)	(4)
Follow	0.011***	0.004***	0.011***	0.003***
	(0.000)	(0.001)	(0.000)	(0.001)
Influencer # of unique securities	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
# of unique securities	-0.000***	-0.000***		
	(0.000)	(0.000)		
Investor FE	Yes	Yes	No	No
Investor x Time FE	No	No	Yes	Yes
Adj. R2	0.012	0.004	0.024	0.004
No of obs	4,001,975	4,001,975	4,001,953	4,001,953

Notes: The table above shows OLS and 2SLS regression estimates for Equation (6) and (8). The dependent variable is *Buy Overlap Ratio*, which is the fraction of the total number of common securities bought in month t by follower f and influencer i, divided by the total number of unique securities bought by follower f. Standard errors are clustered at the follower level and shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: The impact of finfluencers' sales on followers'

	OLS	Second stage	OLS	Second stage
	Sell Overlap Ratio			
	(1)	(2)	(3)	(4)
Follow	0.013***	0.053***	0.012***	0.046***
	(0.000)	(0.002)	(0.000)	(0.002)
Influencer # of unique securities	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
# of unique securities	0.000***	0.000***		
	(0.000)	(0.000)		
Investor FE	Yes	Yes	No	No
Investor x Time FE	No	No	Yes	Yes
Adj. R2	0.033	0.008	0.042	0.011
No of obs	11,680,342	11,680,342	11,680,342	11,680,342

Notes: The table above shows OLS and 2SLS regression estimates for Equation (6) and (8). The dependent variable is *Sell Overlap Ratio*, which is the fraction of total number of common securities sold in month t by follower f and influencer i divided by the total number of unique securities sold by follower f. Standard errors are clustered at the follower level and included in parentheses. Standard errors clustered at the follower level are included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Time lag

	Purchases		Sales	
	Time lag (day)			
	(1)	(2)	(3)	(4)
Follow	0.823***	1.183***	0.276*	0.764***
	(0.100)	(0.122)	(0.146)	(0.183)
Security FE	Yes	Yes	Yes	Yes
Investor FE	Yes	No	Yes	No
Time FE	Yes	No	Yes	No
Investor x Time FE	No	Yes	No	Yes
Adj. R2	0.213	0.402	0.165	0.426
No of obs	369,184	369,184	49,514	49,514

Notes: The table above shows OLS regression estimates for Equation (9). The dependent variable is time lag, which is the difference between the timestamps when trading security s in month t by follower f and influencer i. It is calculated as follower's trading day minus influencer's trading day. In scenarios when the investor traded the same security multiple times in the same month, the first trading day is kept. Column 1 and 2 test time lag for purchases while column 3 and 4 test time lag for sales. Standard errors clustered at the follower level are included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Price difference

	Purchases		Sales	
	Price difference (unit)			
	(1)	(2)	(3)	(4)
Follow	1.934***	1.290***	-4.323***	-4.555***
	(0.363)	(0.333)	(0.988)	(1.226)
Security x Currency FE	Yes	Yes	Yes	Yes
Investor FE	Yes	No	Yes	No
Time FE	Yes	No	Yes	No
Investor x Time FE	No	Yes	No	Yes
Adj. R2	0.084	0.262	0.046	0.290
No of obs	366,630	366,630	49,005	49,005

Notes: The table above shows OLS regression estimates for Equation (9). The dependent variable is price difference, which is the difference between the execution price when trading security s in month t by follower f and influencer i. It is calculated as follower's trading price minus influencer's trading price. In scenarios when the investor traded the same security multiple times in the same month, the first trading day is kept. Column 1 and 2 test time lag for purchases while column 3 and 4 test time lag for sales. Standard errors clustered at the follower level are included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Heterogeneity at the influencer level

	Portfolio Overlap Ratio					
	(1)	(2)	(3)	(4)	(5)	(6)
Follow	-0.335***	-0.335***	-0.999***	-0.999***	-0.036***	-0.036***
	(0.042)	(0.042)	(0.231)	(0.231)	(0.009)	(0.009)
Follow x HighPopularity	0.508***	0.508***				
	(0.042)	(0.042)				
HighPopularity	-0.027***	-0.027***				
	(0.001)	(0.001)				
Follow x Central			1.033***	1.033***		
			(0.232)	(0.232)		
Central			-0.009***	-0.009***		
			(0.003)	(0.003)		
Follow x ManyGroups					0.283***	0.283***
					(0.014)	(0.014)
ManyGroups					-0.051***	-0.051***
					(0.002)	(0.002)
Time FE	Yes	No	Yes	No	Yes	No
Investor x Time FE	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
No of obs	14,351,945	14,351,945	30,356,165	30,356,165	30,356,165	30,356,165

Notes: The table above shows estimation results for the heterogeneity test where we interact the main independent variable *Follow* with the following dummy variables. *HighPopularity* is a dummy equal to 1 if the influencer is followed by more than 4395 people (top quartile) and 0 if followed by less than 1184 people (bottom quartile). *ManyGroups* is a dummy equal to 1 if the influencer i is participating in discussions in more than 233 groups (the median number of groups that influencers participate in) and 0 otherwise. *Central* is a dummy equal to 1 if the follower is identified as a central person (with a degree centrality belong to the top tercile) in the entire investor network and 0 if not. *Influencer # of unique securities* and *follower # of unique securities* are controlled for in the regression as in the main tables. Standard errors are clustered at the follower level and shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Heterogeneity at the follower level

	Portfolio Overlap Ratio					
	(1)	(2)	(3)	(4)	(5)	(6)
Follow	0.020***	0.020***	0.051***	0.047***	0.019	0.014
	(0.002)	(0.002)	(0.008)	(0.007)	(0.037)	(0.036)
Follow x LowFollowingNumber	0.027***	0.026***				
	(0.003)	(0.003)				
LowFollowingNumber	-0.001***					
	(0.000)					
Follow x Male			-0.021**	-0.015*		
			(0.009)	(0.009)		
Male			-0.002**			
			(0.001)			
Follow x Young					0.031	0.038
					(0.042)	(0.041)
Young					-0.005	
					(0.004)	
Time FE	Yes	No	Yes	No	Yes	No
Investor x Time FE	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
No of obs	30,356,165	30,356,165	4,068,163	4,068,163	236,922	236,922

Notes: The table above shows estimation results for the heterogeneity test where we interact the main independent variable *Follow* with the following dummy variables. *HighAttention* is a dummy equal to 1 if the follower f follows less than 16 people (median number of people followed by all users on the platform) and 0 otherwise. *Male* is a dummy equal to 1 if the follower is identified as a male person and 0 if female. *Young* is a dummy equal to 1 if the follower is born after 1986, the median birth year among users whose birth year was identified, and 0 otherwise. *Influencer # of unique securities* and *follower # of unique securities* are controlled for in the regression as in the main tables. Standard errors are clustered at the follower level and shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Impact by securities type

	ETF ratio		Risky ratio	
	(1)	(2)	(3)	(4)
Follow	0.0549***	0.0471***	-0.0002***	-0.0001***
	(0.0014)	(0.0012)	(0.0000)	(0.0000)
Investor FE	Yes	No	Yes	No
Investor x Time FE	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes
No of obs	30,356,165	30,356,165	30,356,165	30,356,165

Notes: The table above shows 2SLS regression estimates for a specification that is the same as Table 5, but with different dependent variables. In columns 1 and 2, the dependent variable is the ratio of ETF products; whereas, it is risky products (derivatives with extreme volatility and Bitcoin-related securities) in column 3 and 4. Standard errors clustered at the follower level and shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Influencer incentives

	Own product	Number of trade per month	Max rating
	(1)	(2)	(3)
Platform-made=1	0.199***	1.863***	0.099
	(0.037)	(0.688)	(0.194)
Fixed effects	Year-month	Year-month	Trading years
Adj. R2	0.139	0.040	0.001
No of obs	21,155	3,530	100

Notes: The table above shows the OLS regression results where we compare platform-made influencers with non-platform-related influencers trading behavior. The independent variable of interest is a dummy *Platform-made*, which is equal to 1 if the user is related to the platform at some point and 0 if not. In column 1, the dependent variable is a dummy *Own product*, which is equal to 1 if the underlying traded asset is issued directly by the platform and 0 otherwise. In column 2, the dependent variable is the total number of trades that each influencer makes at year-month level. In column 3, the dependent variable is the maximum of platform assigned rating of each influencer. Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Portfolio composition evidence from the first month

	OLS	Second stage	OLS	Second stage
	Portfolio Overlap Ratio			
	(1)	(2)	(3)	(4)
Follow	0.011***	0.024***	0.011***	0.024***
	(0.001)	(0.003)	(0.001)	(0.003)
Investor FE	Yes	Yes	No	No
Investor x Time FE	No	No	Yes	Yes
Adj. R2	0.034	0.006	0.034	0.006
No of obs	554,769	554,769	554,769	554,769

Notes: The table above shows OLS and 2SLS regression estimates following Equation (6) and (8). The dependent variable is *Portfolio Overlap Ratio*, which is the fraction of total number of common securities held in month t by follower f and influencer i divided by the total number of unique securities held by follower f. Standard errors clustered at the follower level are included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Purchase evidence from the first month

	OLS	Second stage	OLS	Second stage
	Sell Overlap Ratio			
	(1)	(2)	(3)	(4)
Follow	0.005***	0.011***	0.005***	0.011***
	(0.001)	(0.003)	(0.001)	(0.003)
Investor FE	Yes	Yes	No	No
Investor x Time FE	No	No	Yes	Yes
Adj. R2	0.020	0.003	0.020	0.003
No of obs	291,074	291,074	291,074	291,074

Notes: The table above shows OLS and 2SLS regression estimates for Equation (6) and (8). The dependent variable is *Buy Overlap Ratio*, which is the fraction of the total number of common securities bought in month t by follower f and influencer i, divided by the total number of unique securities bought by follower f. Standard errors are clustered at the follower level and shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Sale evidence from the first month

	OLS	Second stage	OLS	Second stage
	Buy Overlap Ratio			
	(1)	(2)	(3)	(4)
Follow	0.006***	0.010*	0.006***	0.010*
	(0.001)	(0.005)	(0.001)	(0.005)
Investor FE	Yes	Yes	No	No
Investor x Time FE	No	No	Yes	Yes
Adj. R2	0.023	0.004	0.022	0.003
No of obs	106,510	106,510	106,510	106,510

Notes: The table above shows OLS and 2SLS regression estimates for Equation (6) and (8). The dependent variable is *Sell Overlap Ratio*, which is the fraction of total number of common securities sold in month t by follower f and influencer i divided by the total number of unique securities sold by follower f. Standard errors are clustered at the follower level and included in parentheses. Standard errors clustered at the follower level are included in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix for "The Impact of Finfluencers on Retail Investment"

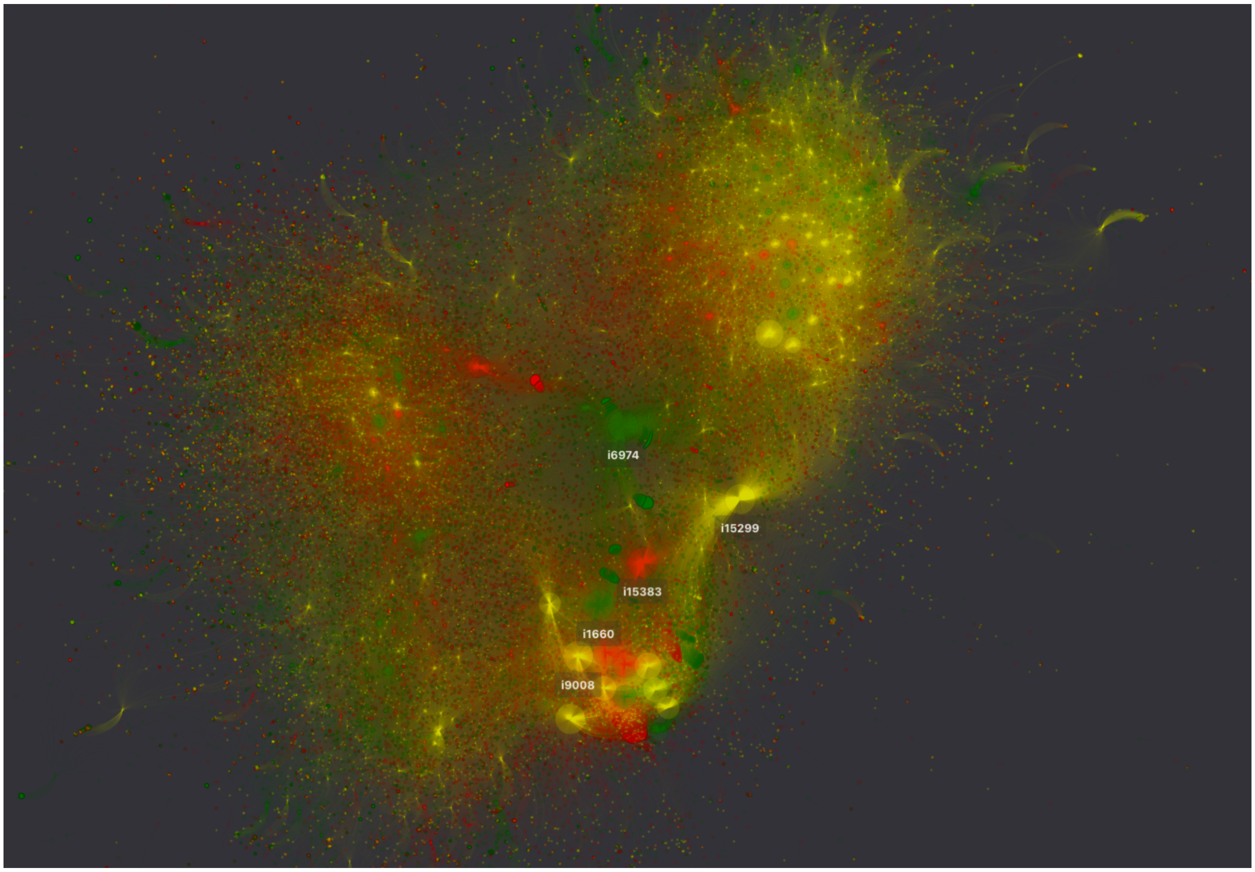
Isaiah Hull and Yingjie Qi

A. Additional tables and figures

Table A1: Variable Definitions

Variable	Description
Influencer	Takes the value of 1 for investors whose follower counts ranks among the top 100, and 0 otherwise.
Follow	Takes the value of 1 if a follower subscribes to another investor, and 0 otherwise.
Instrument	Takes the value of 1 for users who are employees of the platform, and 0 otherwise.
Same language	Takes the value of 1 for pairs where the follower understands the influencer's language, and 0 otherwise.
Same country	Takes the value of 1 for investors for pairs where the follower lives in the same country as the influencer, and 0 otherwise.
Rating	Platform assigned performance rating based on Sharpe ratio. A rating of 0 means the investor's return is non-positive. A rating of 1 means the investor's portfolio yielded a positive return since joining. A rating of 2 means the investor's portfolio return is ranked among the top 50% in the entire platform. A rating of 3 means the investor's portfolio return is ranked among the top 10% in the entire platform.
# of trades (month)	Number of trading transactions made at a monthly basis.
# of purchases (month)	Number of purchasing transactions made at a monthly basis.
# of sales (month)	Number of selling transactions made at a monthly basis.
# of index trades (month)	Number of transactions made in index products at a monthly basis.
# of derivatives (month)	Number of transactions made in derivatives at a monthly basis.
Max number of following	Maximum number of investors followed by a given investor.
Max number of follower	Maximum number of followers an influencer has.
Influencer # of unique securities	Unique number of securities held or traded (depending on which test) by an influencer in a given month
# of unique securities	Unique number of securities held or traded (depending on which test) by a follower in a given month

Figure A1: Full Network



Notes: The figure above shows the full network of relationships in our sample. Colors correspond to accumulated absolute returns of influencers. In order from lowest to highest return, the groups are as follows: red, yellow, and green. Follower nodes take the color of the influencer with the highest return level that they follow.

B. Investor classification examples

Below, we include examples of text taken from influencer biographies and the corresponding probability distributions from applying zero shot classification with `DeBERTaV3`. We have paraphrased the original text and changed details to avoid including information that personally identifies users.

> **Bio:** *"I am committed to a mix of index investing and selective stock picking, with a horizon stretching beyond a decade. My portfolio is built around firms known for their robust performance and consistent dividend payouts. And I reinvest those dividends back into more shares. Occasionally, a few emerging tech startups find their way into my collection. The strategy is all about incremental increases in my investment contributions and adhering to my long-term plan."*
>
> **Class Scores:** [fanatic: 0.003517, long-term rational: 0.872214, naive: 0.006504, short-term rational: 0.117765]

In the first example, the influencer emphasizes a gradual, incremental strategy, and the model identifies this as "long-term rational."

> **Bio:** *"I move fast. Overly concentrated in Danish tech firms. I buy when stocks are rising and sell when they fall. I try not to focus on losses, but instead think about the future."*
>
> **Class Scores:** [fanatic: 0.264301, long-term rational: 0.024658, naive: 0.12586, short-term rational: 0.585181]

The second example talks about moving fast and making mistakes. It also indicates that the investor is concentrated in Danish stock, but has a reasonable tone. The model classifies it as short-term rational with high probabilities also being assigned to fanatic and naive.

> **Bio:** *"My investment strategy is extreme and no one should follow it. I want to save the planet and accelerate automation, which is why I have invested heavily*

> *in RIVN."*
>
> **<u>Class Scores</u>:** [fanatic: 0.772131, long-term rational: 0.019979, naive: 0.170592, short-term rational: 0.037298]

The third example takes a self-deprecating tone, followed by a description of a narrow and extreme focus. The model classifies the biography as describing an investor who fits the "fanatic" category.

> **<u>Bio</u>:** *"Trading started as a fun experiment for me, and I'm still figuring things out. I began with stocks, switched to funds, but might go back since stocks seemed to work better. Not sure if I'll focus on short-term trades or long-term investments—probably a mix. There's a lot to learn, but I'm taking it step by step. Wishing everyone good luck on their trading journey!"*
>
> **<u>Class Scores</u>:** [fanatic: 0.104956, long-term rational: 0.1797789, naive: 0.519656, short-term rational: 0.1956109]

The fourth example captures the characteristics of an investor who would be classified as naive under the scheme in Pedersen (2022*b*).

C. Alternative gender identification method

We classify usernames into three categories: male, female, and neutral. In the first round of classification, we search for an exact substring match with names listed in the World Gender Name Dictionary (Raffo and Lax-Martinez, 2018). This database contains 6.2 million names for 182 countries, along with the gender classification for the name. To limit the scope of the matching exercise, we use names that are listed for the Nordic countries: Sweden, Norway, Denmark, Iceland, and Finland. We also exclude names with fewer than four characters in this step, since they will tend to generate many false matches. We also repeat this process for names of increasing minimum character length (i.e. 5 characters minimum and 6 characters minimum). This provides us with multiple variables that we can use to filter names, where longer names are less likely to be falsely matched.

In addition to exact matching, we also use the `CANINE` model, which is a 121 million parameter large language model (LLM) that is pretrained to process text at the character level in multiple languages. We fine-tune the `CANINE` model to classify names by gender. We do this by taking the names in the World Gender Name Dictionary for 104 countries (including the Nordics) and then modify them by using data augmentation. Specifically, we combine the names with randomly drawn words, symbols, and numbers. We then pair these artificially-generated usernames with the base name's classification in the database. In addition to this, we generate usernames that contain no actual name by drawing random dictionary words, numbers, and symbols.

The approach we take does not try to infer gender based on words or gender stereotypes embodied in words. Instead, we treat all words as neutral and attempt to classify usernames based on the presence of names and their associated gender in the database. It is, however, possible that the model's pretraining could embed information about gender that is not eliminated in the fine-tuning process. It is also possible that the names in the database itself, which are sometimes words that are associated with a gender, could cause certain words to tend to be classified as male or female.

We fine-tuned the model to perform gender classification of usernames with an A100 GPU. After three epochs of training, the model attains a classification accuracy of 0.8978 in the test set. Figure C2 shows the counts of true and predicted labels for each gender classification in our test set.

Since the evaluation is performed on artificially-generated names, we also checked random subsamples of the classifications for real usernames to verify that the performance generalized. Furthermore, we used both our fine-tuned version of the `CANINE` model and the exact substring matches to identify lists of usernames in our sample that were very likely to contain a male name or a female name. Our approach was conservative in the sense that the

Table C1: Username classification examples

Username	Actual	EM4+	EM5+	EM6+	Model
nelli5	F	F	F	N	F
Jeylani	M	M	M	M	M
27awatchai	M	F	N	N	M
Solgte	N	F	N	N	N
-Alkan60	M	M	M	N	M
kambiz-	M	M	M	M	M
Aghyad	M	M	M	M	M
-Jukka pekka	M	M	M	M	M
Boman	M	M	M	N	M
Palads-	N	F	N	N	N
abena	F	F	F	N	F
selvija	F	F	F	F	F
haakon47	M	M	M	M	M
Harena	F	F	F	F	F
Overstimulate100	N	F	N	N	N
eldina	F	F	F	F	F
Iobhlkin	M	F	N	N	M
aysheh-	F	F	F	F	F
Stine	F	F	F	N	F
kadar54	M	M	M	N	M

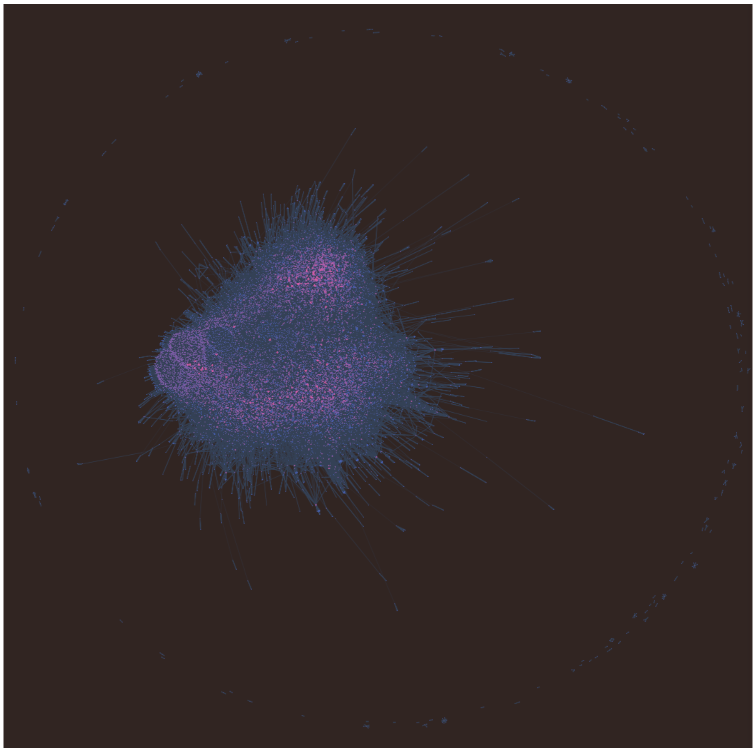
Notes: The table above provides a sample of randomly generated usernames and their classifications according to the four schemes we consider. `EMX+` indicates whether a substring in the username exactly matches a name in the World Gender Name Dictionary (Raffo and Lax-Martinez, 2018) that is `X` characters or longer. `Model` indicates which label was assigned to the username by our fine-tuned version of the `CANINE` model. `Actual` indicates the correct classification, which is based on the underlying name or word that was used to construct the username. The labels are male (M), female (F), and neutral (N).

model training and the substring matching were used to filter out usernames that were either gender neutral or did not contain a name.

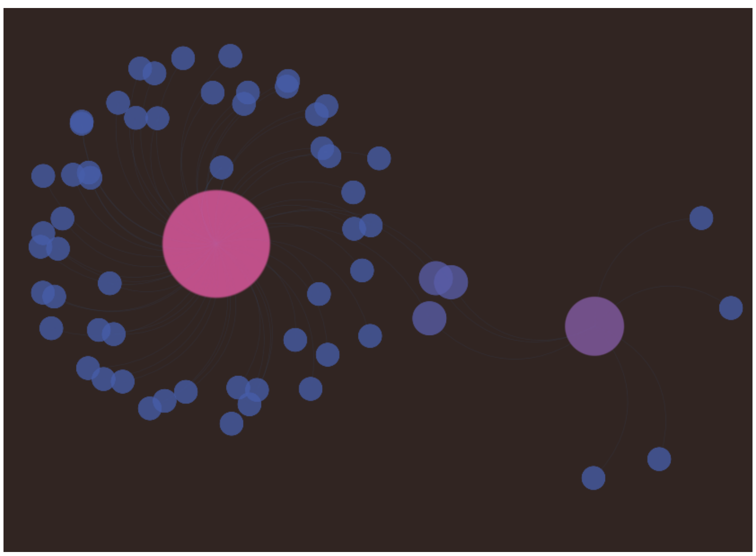
Example classifications are given in Table C1 for 20 randomly selected artificially-generated usernames. We show the actual labels for the usernames, which were based on an underlying base name or word. Names may be male, female, or neutral. We choose to treat all words as neutral. We provide three classifications produced by searching for an exact match in a substring of the username, as well as classifications produced by the model. Using an exact substring match for names in the database with six or more characters yields accurate, but conservative predictions. That is, misclassifying males names as female names or vice versa is rare; however, many usernames will be classified as neutral.

Figure C1: Degree centrality of influencers

(a) Full Network

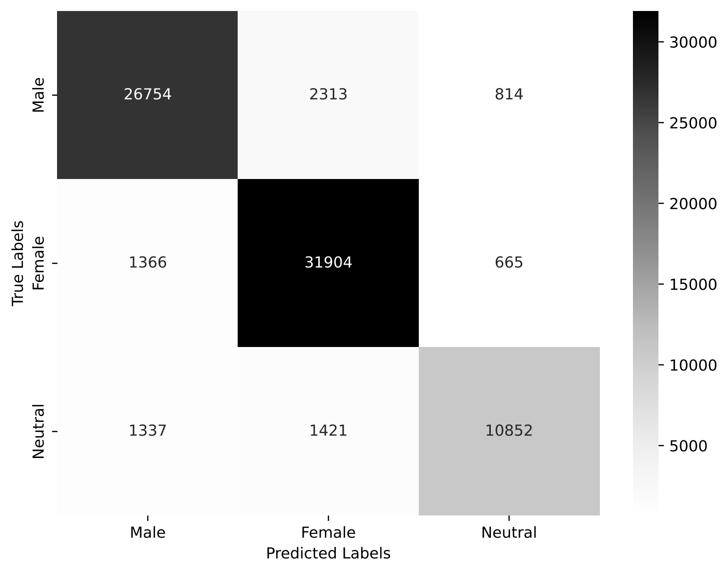


(b) Subnetwork



Notes: The figure above illustrates the degree centrality of influencers. The subfigure in the top panel shows the degree centrality of each node in the full network. The subfigure in the bottom panel focuses on two overlapping subnetworks. Note that only the influencers' connections are shown in the bottom panel. Centrality is illustrated by both the size and color of nodes. Low degree centrality influencers are illustrated as small, blue nodes. As centrality increases, nodes become increasingly large and pink.

Figure C2: Gender classification confusion matrix



Notes: The figure above shows the confusion matrix, which is produced by classifying the artificially-generated usernames in our test set.