Aalto University
School of Business

Centrality, influence and selective communication in an online investor network: Case Shareville

Finance
Master's thesis
Jaakko Virtanen
2015

Department of Finance
Aalto University
School of Business

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Jaakko Virtanen

Title of thesis Centrality, influence and selective communication in an online investor network: Case Shareville

Degree Master of Science in Economics and Business Administration

Degree programme Finance

Thesis advisor(s) Markku Kaustia

Year of approval 2015 **Number of pages** 52 **Language** English

Abstract

Online interactions form a growing share of the information exchange between investors, but existing research on online investor networks is still relatively scarce and focused on forex trading platforms. This thesis is the first study to observe and analyze user behavior in an online investor network focused on the stock market, specifically the Shareville community consisting of customers of the Nordic bank Nordnet. Its purpose is to investigate how activity and portfolio performance determine an individual investor's centrality in the Shareville network, whether investors influence each other's trading decisions, and whether the selective communication hypothesis, i.e. the tendency to discuss profitable trades more frequently than losses, holds in an environment where portfolio contents and returns are public information.

This thesis consists of three studies regarding 1) factors determining the centrality of a user in the Shareville network, 2) the influence of central users on trades by their followers, and 3) selective communication in the network. In Study 1, regression analysis is used to determine how past performance and communication in the Shareville network are correlated with the number of followers a user has, and how users whose own followers are well-connected within the network differ from the baseline. In Study 2, buy-side trades made by users with followers are analyzed to determine how frequently the followers copy the trade by the end of the next trading day. In Study 3, returns from sell-side trades made by Shareville users are reconstructed by analyzing two years of trading data, and trade outcomes are combined with comments users make on their own trades. The commenting frequency on gains and losses is analyzed separately for Finnish and Swedish Shareville members.

Results indicate that 1) historical portfolio performance and activity in the network are relevant but inadequate variables in explaining the number of followers a user has. Commenting on trades appears to be the most relevant form of activity in this regard, while e.g. comments in discussion groups or on individual instruments have less or no value. Results also provide 2) strong evidence of users copying trades of those users they are following. Finally, 3) the knowledge that portfolios and trades can be openly viewed by other users does not prevent users from practicing selective communication, so that trades resulting in gains are discussed more frequently than those resulting in losses. Statistically significant results on selective communication are obtained only for Swedish users, although for Finnish users the lack of statistical significance may result from a small sample size rather than complete absence of the effect.

The results and discussion presented in this thesis will hopefully aid in further development of Nordnet's Shareville platform, inform future studies on the characteristics and behavior of retail stock market investors in online investor networks, and provide retail investors with a better understanding of their own behavior and that of others in the network – for example, how phenomena such as selective communication give a biased view of investor skill.

Keywords investor network, behavioral finance, retail investor, social trading

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Jaakko Virtanen

Työn nimi Keskeisyys, vaikutusvalta ja valikoiva viestintä sijoittajien nettiverkostoitumisessa: Tapaus Shareville

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Markku Kaustia

Hyväksymisvuosi 2015 **Sivumäärä** 52 **Kieli** Englanti

Tiivistelmä

Inhimillisen kanssakäymisen saadessa yhä uusia muotoja tietoverkoissa tarvitaan tutkimustietoa sijoittajien vuorovaikutuksesta heille suunnatuissa palveluissa. Aihepiiriä on tutkittu jonkin verran valuuttasijoittajien verkostojen osalta, mutta tässä tutkielmassa perehdytään ensimmäistä kertaa osakesijoittajien käyttäytymiseen heille suunnatulla verkostoitumisalustalla. Tutkimuksen aineisto on peräisin Nordnet-pankin Shareville-palvelusta, ja sen tarkoitus on selvittää, miten aktiivisuus ja historiallinen tuotto vaikuttavat käyttäjän keskeisyyteen Sharevillen sosiaalisessa verkostossa, miten sijoittajat vaikuttavat toistensa kaupankäyntiin, sekä kommentoivatko sijoittajat todennäköisemmin voitollisia kuin tappiollisia kauppojaan (valikoivan viestinnän hypoteesi), vaikka salkku ja osakekaupat ovat palvelussa automaattisesti muiden nähtävillä.

Opinnäytetyö koostuu kolmesta osatutkimuksesta, joiden aiheet ovat 1) käyttäjän keskeisyyteen vaikuttavat tekijät Sharevillen sosiaalisessa verkostossa, 2) keskeisten käyttäjien vaikutus heidän seuraajiensa kauppoihin, sekä 3) valikoiva viestintä Sharevillessä. Tutkimuksessa 1 selvitetään regressioanalyysillä käyttäjän historiallisen tuoton ja viestinnän korrelaatiota hänen seuraajiensa lukumäärän kanssa. Lisäksi arvioidaan, miten ne käyttäjät, joiden seuraajat ovat tavallista paremmin verkostoituneita, eroavat muista käyttäjistä. Tutkimuksessa 2 selvitetään, miten todennäköisesti käyttäjän seuraajat matkivat tämän osakeostoja alkuperäisestä ostohetkestä seuraavan kaupankäyntipäivän iltaan mennessä. Tutkimuksessa 3 rekonstruoidaan käyttäjien osakemyyntien tuotot kahden vuoden ajalta ja verrataan näitä käyttäjien todennäköisyyteen kommentoida yksittäisiä osakemyyntejä. Suomalaisten ja ruotsalaisten Shareville-käyttäjien taipumusta kommentoida voittoja ja tappioita tarkastellaan erikseen.

Tulokset osoittavat, että 1) historiallinen tuotto ja aktiivinen viestintä, erityisesti osakekauppojen kommentointi, ovat merkittäviä selittäjiä käyttäjän seuraajien määrälle, mutta ne eivät yksinään selitä aineistossa esiintyvää vaihtelua. Toisaalta jotkut viestintäkanavat kuten Sharevillen keskusteluryhmät vaikuttaisivat olevan vähemmän merkityksellisiä verkostoitumiselle. Lisäksi osoittautui, että 2) käyttäjät toisinaan kopioivat seuraamiensa käyttäjien kauppoja, ja 3) tieto osakesalkun ja kauppojen julkisuudesta ei poista valikoivaa viestintää, eli käyttäjät kommentoivat voitollisia kauppoja todennäköisemmin kuin tappiollisia. Tilastollisesti merkitseviä todisteita valikoivasta viestinnästä saatiin vain ruotsalaisten käyttäjien ryhmästä, mutta koska ruotsalaisia oli aineistossa ylipäätään enemmän kuin suomalaisia, on mahdollista että ilmiö hukkui suomalaisten ryhmässä tilastolliseen kohinaan.

Tässä työssä esitetyt tulokset ja pohdinta toivottavasti tukevat Sharevillen kehitystyötä, toimivat pohjana tuleville tutkimuksille piensijoittajien käyttäytymisestä ja vuorovaikutuksesta tietoverkkojen välityksellä, sekä auttavat piensijoittajia ymmärtämään paremmin omaan ja muiden sijoittajien käyttäytymiseen liittyviä ilmiöitä sosiaalisen median palveluissa, kuten valikoivan viestinnän antamaa harhaanjohtavaa kuvaa sijoittajan taidoista.

Avainsanat sijoittajaverkosto, käyttäytymistieteellinen rahoitus, piensijoittaja, sosiaalinen kaupankäynti

Table of Contents

1 Introduction

How does the individual stock market investor interact and share knowledge with their peers? Countless books on investing and trading strategies have been written by successful investors, traders, and fund managers. Friends and co-workers are often eager to recount tales of successful trades over a mug of beer. Yet such communication can be biased in numerous ways. For example, Kaustia and Knüpfer (2012) find evidence that positive stock market returns are followed by an increase in future stock market entry rates in the local neighborhood, but the opposite effect is not present for negative returns. This suggests that investors tend to share only successful trades with their peers (the selective communication hypothesis), which can lead to e.g. selection and attribution biases when evaluating the trading behavior and strategies of peer investors. The purpose of this thesis is to determine how activity and portfolio performance determine an individual investor's influence on peers in an online environment, and whether the selective communication hypothesis holds in an environment where portfolio contents and returns are public information. This thesis represents the first effort to explore these questions within the context of an online investor network of retail stock market investors.

1.1 Online investor networks – untapped source for behavioral finance research

Developments in communication technology provide unprecedented opportunities for retail investors and traders to interact and learn from each other. Electronic trading platforms have been available to private individuals since the 1990s, but the more recent proliferation of social media has led to the creation of investor networking platforms which enable investors and traders to anonymously engage in discussions, view each other's portfolios and trades, and even automatically replicate the trades of their peers. Most of these platforms, such as eToro, Tradeo, and ZuluTrade, are focused on trading in currency markets and stock derivatives such as contracts for difference. Others, such as the Shareville platform[1] launched by the Nordic bank Nordnet in 2014, are targeted to retail stock market investors, most of whom are unable or unwilling to practice day trading and are also unfamiliar with the derivative markets. An important feature of these platforms is the opportunity to minimize selective communication: since portfolio contents, trades, and returns are constantly visible to all users, investors should in principle have a better opportunity to learn from their peers' mistakes. However, it is reasonable to assume that investor behavior in online investor networks will still be governed by the same laws and biases that apply in more traditional settings. Well-known examples include the attribution bias, base rate fallacy, and confirmation bias introduced by Tversky and Kahneman (1974).

[1] https://www.shareville.fi

Since the appearance of online investor networks is a recent phenomenon, little research has been conducted on peer influence on investor behavior in these networks. Heimer and Simon (2014) investigate how selective communication drives increased currency trading within an unnamed online investor network. Heimer (2014) suggests that rational investors who seek to form connections with skilled investors in order to acquire information may give rise to seemingly irrational loss aversion and tendency to close winning positions too early. Based on EUR/USD currency trading data from the eToro platform, Pan et al. (2012) show that the strategy of automatically mirroring (replicating) the trades of the most successful traders based on historical performance significantly outperforms the average independently placed trade and produces positive returns. Exploring an extended version of the same dataset, Liu et al. (2014) provide results in line with the prospect theory originally formulated by Kahneman and Tversky (1979), which predicts that individuals are disposed to realizing even small gains but are reluctant to realize mounting losses. This kind of behavior leads to poor performance on the market because it prevents large gains from winning positions while leading to large losses from losing positions. Liu et al. (2014) also provide prospect theory-based suggestions for behavioral metrics that would allow traders to better recognize which peers to follow based on expected future performance.

Scarcity aside, the main limitation of the existing research on investor networks is its focus on currency trading which is, to a large extent, based on exchange rate speculation and statistical arbitrage, seeking profits mainly from the aggregate behavior of other market participants. The stock market, on the other hand, provides in principle a vehicle for investing in a company's business, and the long-term value of the investment is driven by the profitability of that company's business. Nevertheless, as discussed by Thaler (1999) and Shiller (2003), research into behavioral finance has unequivocally shown that the stock market is rife with inefficiencies resulting from various mechanisms. When these are combined with the individual investor's own knowledge, experiences, and cognitive biases, the stock market forms a complex environment with ample opportunities for the individual investor to improve their long-term returns through peer learning on investor networks. The best, most granular readily available data on the influence of social networking on trading and learning behavior of stock traders and investors is constantly being created and collected on platforms such as Shareville.

1.2 Contribution of this thesis

This thesis consists of three independent studies on investor behavior in the Shareville network. The research is based on six months of panel data collected from the Shareville platform and documenting the social network links and discussions of individual investors, as well as concurrent stock trading data from the Nordnet accounts linked with the Shareville profile. All trades in equity shares are included in the analyses, but not derivatives, funds, bonds, or commodities. Data for the thesis was provided in an anonymized format by Nordnet, and the results will also aid Shareville development.

Each of the three studies explores one key research question to characterize peer influence and communication in Shareville. The research questions are chosen to not only provide descriptive information about Shareville users, but also to explore topics that the existing literature does not adequately cover. First, previous studies on online investing networks have not examined in detail what factors influence connection formation within the network. The limited research that exists on network formation and user behavior concerns primarily currency trading networks. To address this topic in the context of Shareville's stock investor network, Study 1 investigates *What attributes and behavior determine an investor's centrality in the network?*

Second, while copying other users' actions is a popular trading strategy within online currency trading networks, so that copy trading can even be automated on many popular platforms, it is not known whether stock investors exhibit similar behavior on networking platforms focused on the stock market. Therefore, Study 2 seeks to quantify *To what extent do central users influence trades by their followers?*

Third, as noted earlier, numerous biases influence both the originating and receiving party in investor communication. While online investor networks have the potential to mitigate some of these biases, their effectiveness in this regard is by no means guaranteed. Existing behavioral finance literature provides circumstantial evidence of selective communication (i.e. an investor's increased propensity to communicate after gains in the market), but the topic has not been studied at the level of individual trades and communication related to them. To both examine the utility of investor networks in mitigating biased communication and on the other hand search for direct evidence to support or refute the selective communication hypothesis, Study 3 focuses on the question *Are investors more inclined to discuss profitable trades than losses?*

Another obvious question related to the utility of investor networks is whether they support investor learning and allow investors to make better, i.e. more profitable or less risky, investment decisions. However, the Shareville data available in this thesis covers a period of only 6 months, while equity investors often have a longer time horizon. For example, Benartzi and Thaler (1995) show that empirical evidence for the prospect theory is compatible with a model where investors evaluate their asset allocation with a roughly one-year horizon. Therefore, the question of investors' ability to learn from their peers and make better investment decisions is outside the scope of this thesis.

The next chapter summarizes the existing research on peer influence and trading behavior in investor networks and introduces the three research hypotheses examined in this thesis. Chapter 3 describes the data used, Chapters 4-6 summarize the motivations, methods, and key results of the three studies, while the results from each study are interpreted and discussed in Chapter 7. Chapter 8 summarizes the conclusions and provides suggestions for further research topics.

Unless otherwise indicated, throughout this work "investor network" and similar terms refer to explicit connections between individual investors and traders on a social networking platform such as Shareville. Any real-life social networks of these individuals are outside the scope of this study. Also, the broad topic of this study, social investing, should not be confused with socially responsible investing. Finally, the term "investor" is used as a general term covering individuals with various short-term and long-term investment strategies, but in some parts of the text, "investor" and "trader" may be used for convenience to differentiate buy-and-hold approaches from short-term speculation and arbitrage. Naturally, individual users' strategies are more varied, but a detailed categorization of investment strategies is not necessary for this work.

2 Background

This chapter provides an overview of relevant research on investor peer influence and behavior, first in traditional settings and then in online investor networks, and presents the research questions and hypotheses examined in this thesis.

2.1 Literature on investor behavior and peer interactions

Although research on investor peer interaction is too extensive to present in its entirety, evidence overwhelmingly indicates that peer interactions influence market entry decisions, adoption of investment strategies, and also investing goals. For example, Hong et al. (2004) provide evidence that households are more likely to invest in the stock market when their peers participate in the market, and Kaustia and Knüpfer (2012) show that positive historical stock returns in the local neighborhood increase stock market entry rate, particularly in areas with good opportunities for social learning. Pool et al. (in press) find that real-life social interactions lead to more overlapping portfolios among professional fund managers, and information sharing can improve risk-adjusted returns. A more extensive literature review on social interaction in the stock market is presented by König (2013).

Since evidence unequivocally demonstrates the influence of social interaction on everyday investment decisions, an obvious argument for the utility of online social networks is their potential for distribution of knowledge and learning between investors. As discussed by Thaler (1999) and Shiller (2003), since rational decision making and information symmetry between market participants are important theoretical prerequisites of an efficient market, real-life market inefficiencies and investor behavioral biases enable skilled investors to earn abnormal risk-adjusted returns. On the other hand, especially inexperienced investors fall victim of fallacies identified in the academic literature decades ago. Since the focus of this thesis is investor peer interaction, the reader interested in individual investor behavior is guided to the recent, excellent summary of the relevant literature by Barber and Odean (2013).

Although outside the topic of online investor networks, the work by Ozsoylev et al. (2014) provides an interesting view into trading behavior and information diffusion between central and peripheral investors in a stock market. Using a complete 12-month dataset of all trades in the Istanbul stock exchange, they reconstruct an empirical investor network where two investors are connected if they execute similar trades (same instrument and direction) within 30 minutes of each other at least three times during the 12-month period analyzed. The authors find that central actors in the empirical investor network earn higher returns and utilize information more quickly than other individuals.

They also suggest that information propagates in the investor population through a decentralized diffusion mechanism rather than e.g. through mainstream media channels.

The network reconstruction approach chosen by Ozsoylev et al. (2014) has some shortcomings. For example, the median number of trades for individual investors in their data is eight, so that to establish a connection between two hypothetical “median investors”, three out of these eight trades would need to be very closely timed copies of each other (except for the amount traded). This seems intuitively unlikely for two retail investors even under regular information exchange. A second, related issue is the high number of first-degree connections implied: for the “median investor”, eight trades during the 12-month period would establish connections to 158 other traders. Obviously, no real trader communicates with such a high number of individuals on a few trades, so the method used is likely to generate a large number of connections that do not exist in the real world.

There may be alternative explanations to the large number of connections identified in the Ozsoylev et al. (2014) study, such as certain individuals preferring to trade in certain stocks and times of day, which would give rise to more coincidental connections than a completely random process. The authors cover some potential shortcomings of their approach in their paper, and their conclusions on information diffusion and identifying central investors in the network are certainly plausible. However, it appears intuitively unlikely that their approach to reconstructing the investor network reflects actual connections between retail investors.

A common and potentially valid counterargument to the quality and reliability of information disseminated through an investor network is that skilled investors would prefer to keep private any information or experience that might give them an advantage in the market. However, using empirical data from real-life investor network settings, Krasny (2013) and Dijk et al. (2014) demonstrate that alongside monetary gains, also factors such as the pursuit of social status influence investor portfolio composition. In both of these studies social rank is represented by relative wealth, but it is plausible that also other measures of social status, such as the number of followers within the investor network, and even altruistic reasons would motivate at least some successful investors to share their experience and learnings with others.

2.2 Research on online investor networks

An obvious benefit from studying investor interactions in online investor networks is that nearly all communication on the platform is documented, and connections between individuals are made explicit. Due to the relative novelty of social media, research on investor behavior and trading activity on investor networking platforms is limited, but there are two bodies of work that are especially relevant for this thesis. The first one comprises research on a dataset from an unnamed social networking website that allows users to anonymously share their currency trading history and interact with each other by messaging and forming bilateral friendships. The research has been presented in Heimer and Simon (2014) and Heimer (2014).

Heimer and Simon (2014) investigate how selective communication drives increased trading within the social network as proposed in a theoretical model by Han and Hirshleifer (2015). The authors show that traders who earn positive returns are more likely to send peer-to-peer messages, and extreme returns tend to be advertised by active traders due to higher portfolio volatility. This leads to the message recipients erroneously attributing the sender's success to their strategy (attribution bias). The increased social interaction propagates active trading strategies in the network and leads to increased portfolio volatilities in the long term, but the increased volatility is typically not compensated with higher returns.

The second paper by Heimer (2014) focuses on the role of social interaction in explaining the disposition effect. Key elements of the disposition effect are loss aversion (unwillingness to realize losses) and reflection effect (risk seeking for losing positions, risk aversion for winning positions) as depicted in the prospect theory formulated by Kahneman and Tversky (1979), and Kaustia (2010) provides an extensive literature review on the subject. Heimer (2014) argues that the disposition effect arises naturally from the behavior of rational investors who seek to form connections with skilled investors in order to acquire information. By closing winning positions and holding on to losing positions, the investor can demonstrate higher realized returns and thus increase their bargaining power in the market for information.

While Heimer (2014) refutes various alternative explanations to his observations, he does not consider the possibility that the heightened disposition effect could arise from reluctance to publicly admit mistakes in the social network. In other words, consciousness of the social network could plausibly intensify the regret aversion effect described by Shefrin and Statman (1985). Both interpretations would be compatible with the evidence by Krasny (2013) and Dijk et al. (2014) that pursuit of social status influences investor portfolio composition. All in all, the large body of literature suggests that different authors are capturing coexisting and complementary aspects of the disposition effect.

Whereas the work by Heimer and Simon (2014) and Heimer (2014) primarily focuses on the communication between individual traders, another body of work analyzes the aggregate behavior of

EUR/USD currency traders using data from the eToro trading and networking platform. Pan et al. (2012) analyze the performance of eToro users who can place trades independently (single trade), replicate one trade by another user (copy trade), or automatically copy all trades made by another user (mirror trade, where the traded amount is proportional to the portfolio size). One of the authors' key findings is that the mean return on investment (ROI) from single and copy trades is negative, with single trades performing more poorly, while mirror trades generate on average small but significantly positive returns.

Pan et al. (2012) also evaluate "wisdom of the crowd" by comparing the mean daily ROI of the top 5, 10, 50, and 100 users selected by 1) historical returns and 2) number of followers (mirrorers) in the social network. For the top 5 and 10 users, approaches 1) and 2) produce similar ROI, but as the number of users increases, approach 1) produces a higher ROI than 2). This implies that wisdom of the crowd is inferior to a simple rule-based method in identifying the best experts to follow (beyond the top 10). A third observation by the same authors is that except for the 50 or so most popular users, the number of followers appears to be largely independent of the user's accumulated returns. In summary, while a number of individual forex traders may be able to consistently earn positive returns on the eToro platform, historical returns appear to be only one criterion by which users pick other users to follow in the online network, and users as a group are consequently suboptimal in choosing which users to imitate.

Liu et al. (2014) also analyze the trading behavior of eToro users using at least partially the same dataset. Their data illustrates how a vast majority of the trading positions are open for a relatively short time, and very rarely longer than a day. They also demonstrate that mirror trades are on average the most profitable type of trade and generate positive ROI, similar to the previous study. The main contribution of their study is the introduction of three behavioral metrics that could be used to identify users with high future probability of placing successful trades. The metrics are based on the risk-reward ratio of trades, holding time of winning and losing positions, and win-loss ROI ratio. In essence, these metrics quantify loss aversion and reflection effect as depicted in the prospect theory formulated by Kahneman and Tversky (1979). While the authors suggest that these metrics can assist in choosing which experienced investors to follow in the network, individual users could also use such metrics to identify their own tendencies for inefficient trading behavior in order to avoid such behavior in the future.

2.3 Research questions and hypotheses

There are numerous areas in online investor networking and behavioral finance that the nascent literature does not yet cover, and the purpose of this thesis is to fill some of those gaps. First of all, results from currency trading studies cannot be applied directly to investor behavior in stock markets. The most obvious difference between the two is that currency trading seeks to benefit from short-term (typically intraday) fluctuations in market prices, while stock positions are often taken with a medium or long-term perspective (months, years, or in some cases even decades) to passively benefit from value creation by the underlying company and possibly perceived mispricing of the company's stock.

Stock market investors may also compare companies on multiple quantitative and qualitative dimensions such as size, balance sheet, profitability, market valuation, sector, strategy, etc. Dissemination of knowledge, opinions, and analyses about individual companies, market sentiment, and behavioral finance topics through an investor network may allow individual investors to improve their social status and gain followers in the network. Online commentaries become even more valuable if they can be evaluated in light of the commenter's own historical performance and activity in the market, but the existing research on online investor networks does not cover these aspects.

To shed light on the accrual and practical implications of influence in the network, this thesis explores two research questions: *What attributes and behavior determine an investor's centrality in the network?*, examined in Study 1, and *To what extent do central users influence trades by their followers?*, examined in Study 2. Another interesting topic explored by Heimer (2014) and Kaustia and Knüpfer (2012) is the influence of profits and losses on the willingness of investors to communicate with others. The authors provide evidence that investors are more inclined to discuss their profits than their losses with their peers. The data available in this thesis uniquely enables analyzing the subject at the individual investor and trade level. Therefore, the research question explored in Study 3 is *Are investors more inclined to discuss profitable trades than losses?*

Based on these research questions, three hypotheses are formed:

Hypothesis 1 (Study 1): *The centrality of an investor in the network is determined by their past returns and activity on the network, such as commenting on individual stocks and communicating with other investors.*

Hypothesis 2 (Study 2): *If an influential user trades in a particular stock, their followers are more likely to trade in that stock more actively than other Shareville users.*

Hypothesis 3 (Study 3): *An individual user is more likely to comment on trades in which they realize gains than on trades in which they realize losses.*

The next chapter describes the data available for the three studies, after which the methods and results for each study are presented separately in the following chapters.

3 Data

The dataset used to explore the research questions and hypotheses presented in the previous chapter contains Shareville user data from 26 September 2014 (date when Shareville was launched for the wider public) to 31 March 2015. Data extraction and analysis were performed using Oracle SQL and the R language and software environment[2]. The data was anonymized so that personal data such as users' names had been removed or masked.

Shareville is only available to Nordnet customers, and enabling all features requires the user to share at least one portfolio associated with their Nordnet account. However, the Shareville service itself is free of charge. Most Shareville users are Swedish, Finnish, Norwegian or Danish, with Swedish and Finnish users forming the two largest groups. Most users choose to use an anonymous username.

Shareville users can follow other users and specific instruments (i.e. configure the service to provide automatic notifications about the activities of a certain user, and about trades and comments made on a particular instrument), publish comments on completed trades, individual instruments, other users' walls, and in topical discussion groups, and send private messages to other users. Each Shareville user can also view the instruments owned by another user, and for each instrument its share of total portfolio value, but the portfolio value itself is not disclosed. Figures 1 and 2 illustrate the user's view of another user's wall and a specific instrument, respectively. All users can see other users' portfolio returns for up to three years and a 1-year Sharpe ratio relative to risk-free return. In addition, each user is assigned a star rating based on their returns and Sharpe ratio: users with positive 1-year portfolio returns receive one star, users with a Sharpe ratio in the top 50% of all shared portfolios with positive returns receive two stars, and users in the top 10% receive three stars.

The dataset in this study contains for each Shareville user the users and instruments they are following, discussion group memberships, comments they have made, and metadata of private messaging between individuals (i.e. messaging participants and timestamps of messages but not message content). It contains also the portfolio returns and Sharpe ratios and star rankings as of 31 March 2015.

To be included in the analyses, a user account had to be active on 31 March 2015 (i.e. user had not deleted their account) and share at least one Nordnet portfolio. A small number of portfolios were flagged as abnormal in the original data, e.g. because they appear to have invalid historical returns, and were therefore excluded. Also, by default, Shareville users are configured to follow four Nordnet

[2] http://www.r-project.org/

bloggers in the network. These four individuals were excluded from all analyses and calculations. The total number of Shareville accounts included in the analyses is 39,920 (as of 31 March 2015; see Figure 3).

The dataset also contains monthly account and daily trading data from the Shareville users' Nordnet accounts spanning the period from January 2012 to March 2015. Each account can contain cash deposits and financial instruments such as stocks, options, warranties, bonds, mutual funds, etc. In this thesis, only trading activity and returns related to stocks are considered. Nordnet customers trade primarily in their home countries' stock exchanges, but they also have access to other Nordic (Finnish, Swedish, Norwegian, Danish), German and North American markets.

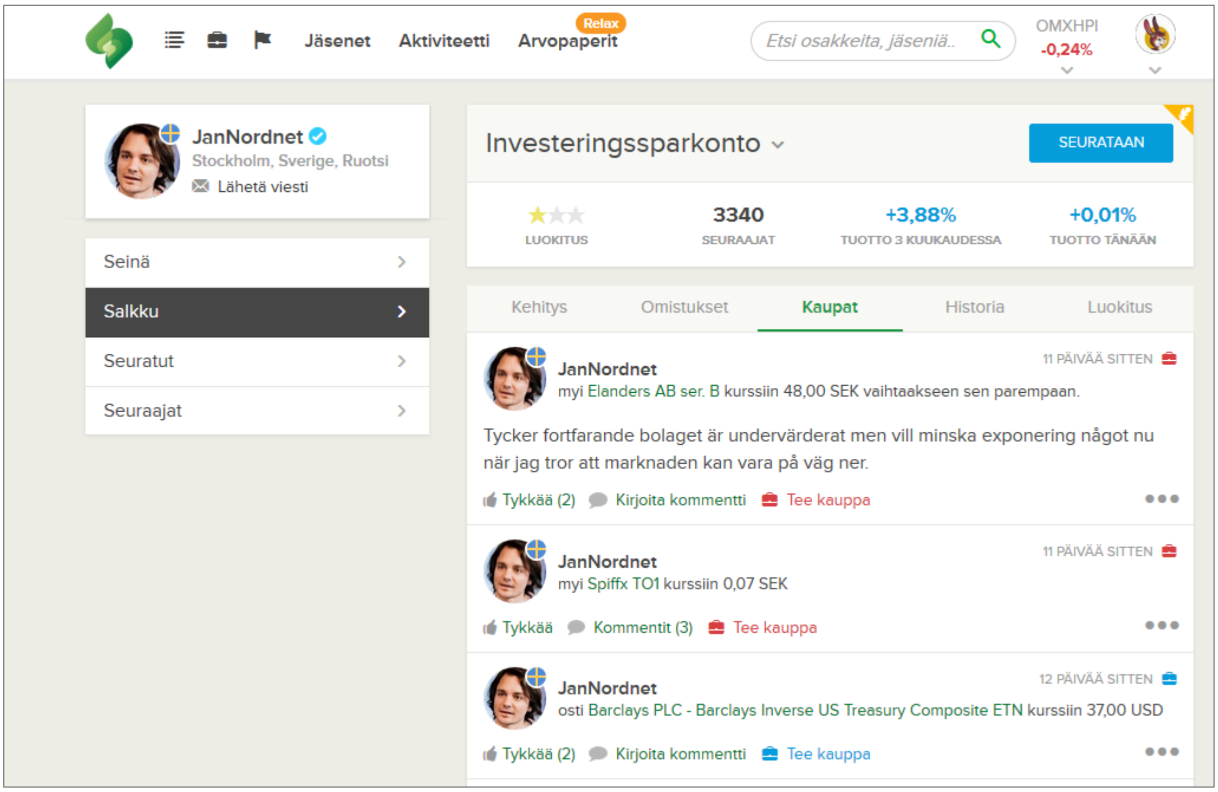

Figure 1: Screenshot illustrating the user's view of Shareville. The screenshot displays the profile of a followed user, including their star rating based on volatility adjusted returns, number of followers, past returns, and recent trades and comments.

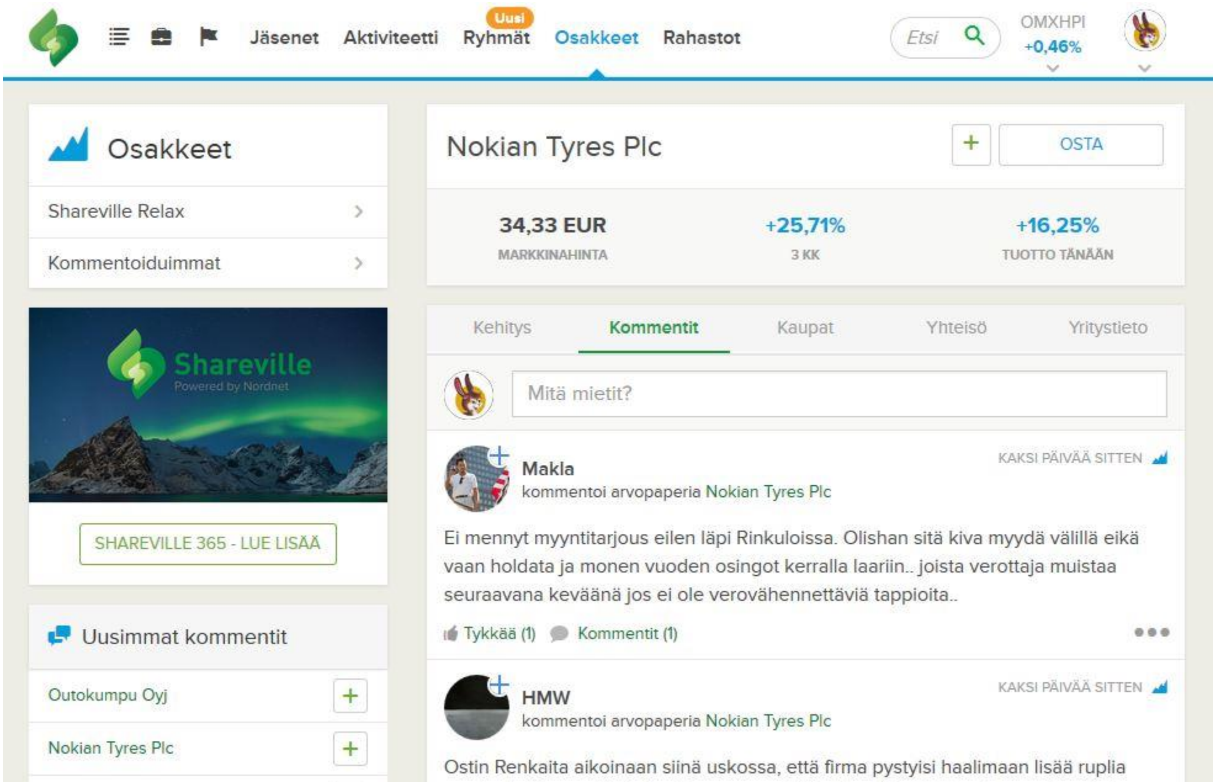

Figure 2: Shareville screenshot showing market data and discussions related to a specific share.

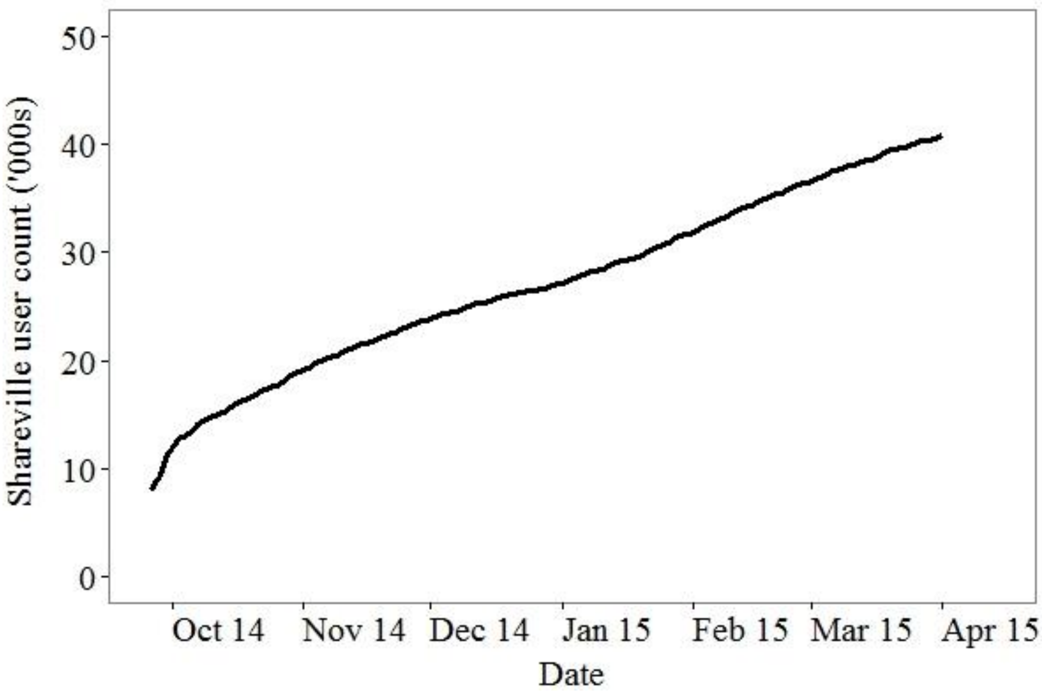

Figure 3: Growth of Shareville user base since the service was publicly launched on 26 Sep 2014.

4 Study 1: Attributes valued by Shareville users

Hypothesis 1: *The centrality of an investor in the network is determined by their past returns and activity on the network, such as commenting on individual stocks and communicating with other investors.*

4.1 Motivation

Hypothesis 1 addresses an essential question of human networking: why does a connection arise between two individuals? Previous studies on online investing networks have not examined this subject in detail. An intuitive answer in the context of Shareville is that users with high historical returns and users who are active communicators in the network will attract the most followers. However, without any quantitative foundation such statements are easily challenged, and their implications remain unclear. Does better performance automatically mean more followers? What is the relative importance of portfolio performance and activity in the network? What kind of communication activity is relevant? To what extent do these variables explain network formation? Even when considering the matter from a purely business point of view, it is difficult to justify and allocate investments in service development based on vague intuition. This study analyzes well-connected individuals in the Shareville network to provide a data-based representation of characteristics relevant to connection formation.

4.2 Methods

A review of basic network analysis concepts and methods is provided by Newman (2003). The Shareville network can be represented as a directed network, where each user corresponds to a node in the network, and following another user corresponds to a directed connection (edge) from the follower to the followed user. Centrality is a measure of the relative importance of an individual node in the network, and to examine Hypothesis 1, a quantitative definition of centrality in the Shareville investor network must be chosen.

The most basic network centrality measure is the in-degree centrality C_D, i.e. the number of incoming first-degree connections (followers) a Shareville user has in the network. However, in-degree centrality treats all connections as equally important. In the case of Shareville, it can be argued that having followers who themselves have many followers is a better indication of influence in the network. Eigenvector-based centrality measures take also into account the centrality of neighboring nodes. For example, the Pagerank centrality C_P by Page et al. (1999) is defined such that the centrality of an individual node represents the probability that a random walk through the network will terminate at that node. Thus, nodes with inbound connections from other nodes with many inbound connections will have a higher Pagerank centrality.

To examine Hypothesis 1, C_D and C_P are calculated for each Shareville user with at least one follower as of 31 March 2015. R functions used for the network analysis can be found in the `igraph` package, and the functions `degree` and `page.rank` are used to calculate the centrality measures, with the Pagerank damping factor (representing the probability of continuing the random walk after each step) set to 0.85 (default value used in many applications of the method). Figure 4 shows that the two centrality measures are strongly correlated, but may also differ considerably for some users. Analysis and comparison of C_D and C_P is expected to expose features and behavior considered valuable by Shareville users.

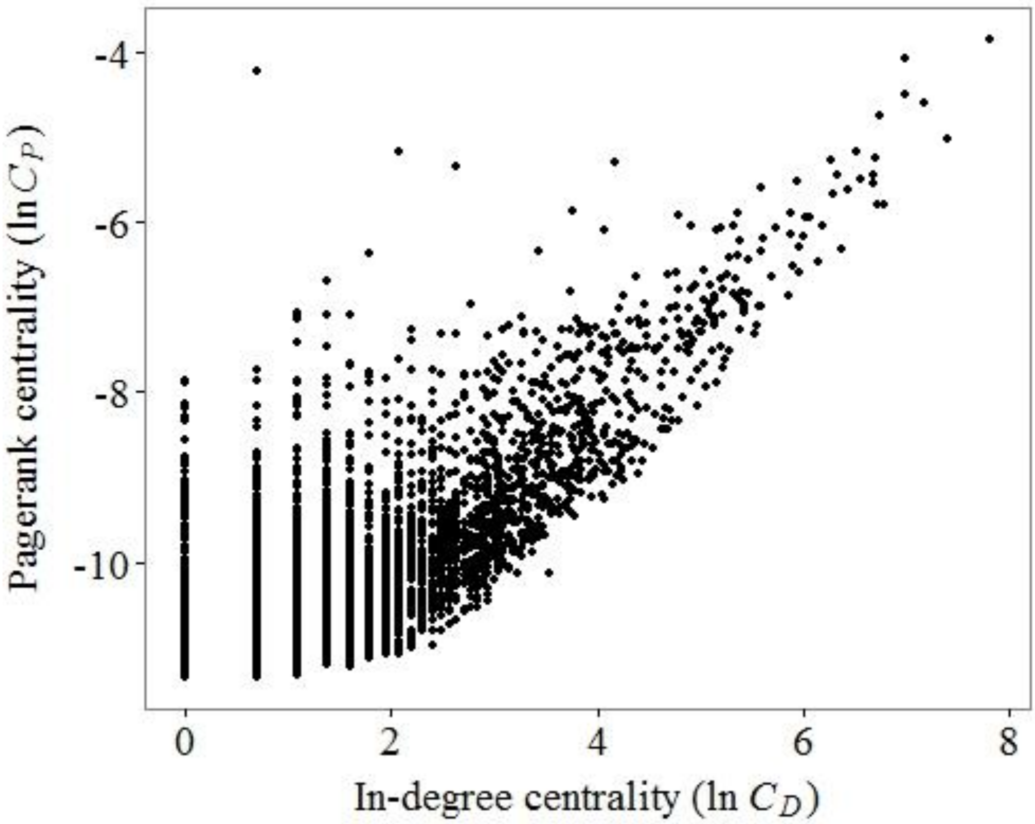

Figure 4: Natural logarithms of in-degree centrality and Pagerank centrality in the Shareville network as of 31 March 2015. Each data point corresponds to one user with at least one follower, and the centrality values are incremented by one to avoid taking logarithm of zero. Total number of data points is 8,412, and the largest number of followers for one user is 2,466.

First, treating the natural logarithm of C_D as the dependent variable, a multiple linear regression is performed on the user's historical returns and various descriptive variables in the Shareville social network. Table 1 summarizes the symbols and regression variables used for examining Hypothesis 1. All variables in the regression are calculated as of 31 March 2015, and the regression coefficients are analyzed to determine whether they differ significantly from zero (details provided below; the threshold of statistical significance is set at $p = 0.05$):

$$\begin{aligned}\ln C_{D,j} = {} & \alpha_0 + \alpha_1 \cdot Y_j + \alpha_2 \cdot H_j + \alpha_3 \cdot S_{1,j} + \alpha_4 \cdot S_{2,j} + \alpha_5 \cdot S_{3,j} \\ & + \alpha_6 \cdot \ln\left(A_{G,j} + 1\right) + \alpha_7 \cdot \ln\left(A_{T,j} + 1\right) + \alpha_8 \cdot \ln\left(A_{I,j} + 1\right) \\ & + \alpha_9 \cdot \ln\left(A_{U,j} + 1\right) + \alpha_{10} \cdot \ln\left(M_{1,j} + 1\right) + \alpha_{11} \cdot \ln\left(M_{2,j} + 1\right) \end{aligned} \tag{1}$$

where subscript j denotes the individual user, Y_j denotes the 1-year return for user j, H_j the 1-year Sharpe ratio, and $S_{k,j}$ are mutually exclusive dummy variables corresponding to the Sharpe ratio based star rating of users within Shareville ($k = 1$ indicates one star, $k = 2$ indicates two stars, and $k = 3$ indicates three stars). While the 1-year return and Sharpe ratio are also visible to other users, the star classification is more prominently displayed in the user interface, and provides a simple means for evaluating a user's past performance relative to their peers. Thus, it is interesting to see whether the absolute numeric values of Y_j and H_j provide any additional predictive power to the regression beyond the simple star rating.

The variables $A_{k,j}$ denote various activity measures in the network: the number of comments made by the user in Shareville discussion groups ($k = G$), comments on trades ($k = T$), comments on individual instruments ($k = I$), and comments on other users' walls ($k = U$). The remaining two variables are $M_{1,j}$, the number of individual Shareville users, excluding followers of user j, whom user j has contacted with a private message, and $M_{2,j}$, the number of individual Shareville users, excluding followers of user j, who have contacted user j with a private message. Users who are not followers of user j at the time of the messaging are counted towards $M_{i,j}$ even if they later become followers. These two variables are included to examine whether actively seeking contact with other users will lead to gaining followers, and whether other users tend to seek out the opinions of influential users.

Table 1: Summary of regression variables for Hypothesis 1.

Symbol	**Description**
C_D	In-degree centrality of user (i.e. number of followers)
C_P	Pagerank centrality
Y	1-year returns
H	1-year Sharpe ratio
S_1	Dummy for one star rating (positive 1-year returns)
S_2	Dummy for two stars rating (positive 1-year returns with higher Sharpe ratio than 50% of users with positive returns)
S_3	Dummy for three stars rating (positive 1-year returns with higher Sharpe ratio than 90% of users with positive returns)
A_G	Number of comments made in Shareville discussion groups
A_T	Number of comments made on individual trades
A_I	Number of comments made on individual instruments
A_U	Number of comments posted on other users' walls
M_1	Number of (non-follower) users contacted via private message
M_2	Number of (non-follower) users who have contacted the user via private message

A similar regression is formed for the natural logarithm of C_P using otherwise same regressors as in Equation 1, but adding also the logarithm of the in-degree centrality, $\ln C_{D,j}$, as a regressor:

$$\ln C_{P,j} = \beta_0 + \beta_1 \cdot \ln C_{D,j} + \beta_2 \cdot Y_j + \beta_3 \cdot H_j + \beta_4 \cdot S_{1,j} + \beta_5 \cdot S_{2,j} + \beta_6 \cdot S_{3,j} + \beta_7 \cdot \ln\left(A_{G,j} + 1\right) + \beta_8 \cdot \ln\left(A_{T,j} + 1\right) + \beta_9 \cdot \ln\left(A_{I,j} + 1\right) + \beta_{10} \cdot \ln\left(A_{U,j} + 1\right) + \beta_{11} \cdot \ln\left(M_{1,j} + 1\right) + \beta_{12} \cdot \ln\left(M_{2,j} + 1\right) \tag{2}$$

The purpose of the second regression is to identify factors that contribute to Pagerank centrality in the network beyond the number of followers the user has, i.e. factors that are relevant to influential users in the network when they decide whom to follow.

The linear regression is calculated separately for both centrality measures using the ordinary least squares (OLS) method:

$$\hat{\alpha} = (X'X)^{-1}X'c \tag{3}$$

where $\hat{\alpha}$ are the coefficient estimates, X the design matrix, and c the vector of in-degree or Pagerank centralities. Student's *t*-test is used to determine whether individual regression coefficients are significantly different from zero. The variance-covariance matrix of the regression coefficients is

$$C = \hat{\sigma}^2(X'X)^{-1} \tag{4}$$

where $\hat{\sigma}^2$ is the mean square error (MSE) of the regression fit. Then, the test statistic for regression coefficient *k* is

$$(t_0)_{\hat{\alpha}_k} = \frac{\hat{\alpha}_k}{\sqrt{C_{kk}}} \tag{5}$$

where C_{kk} is the k'th diagonal element of C. The linear regression and test statistic calculation are performed by standard R functions such as `lm`.

After performing the regression for each full model, the non-significant regressor with the test statistic value closest to zero is removed from the model and the OLS regression is repeated. Goodness of the regression fit is measured by the adjusted coefficient of determination, which takes into account the degrees of freedom in the model (p. 799 in De Veaux et al., 2008):

$$\bar{R}^2 = \frac{MS_{Regression}}{MS_{Total}} \tag{6}$$

where $MS_{Regression}$ is the mean squared error explained by the regression and MS_{Total} is the total mean squared error of the data. If $\bar{R}^2$ of the model does not decrease, this process is iterated until the model contains only statistically significant regressors to identify a minimally complex regression model explaining the centrality with accuracy similar to the full model. The proportion of variance explained by the remaining regressors (relative to unadjusted R^2) is then examined using analysis of variance (ANOVA) as implemented by function `anova` in R.

4.3 Results

Table 2 provides selected statistics on all Shareville users, top 50 users based on number of followers as of 31 March 2015, and top 50 users based on 1-year Sharpe ratio as of 31 March 2015. For 1-year returns and Sharpe ratio, the 1st and 100th percentiles of the whole dataset are excluded to remove outliers before calculating the statistics. The two top 50 groups have no common members, suggesting that Shareville users as a group are not following the most skilled investors in the community (insofar as the 1-year Sharpe ratio at a particular point in time actually reflects investor skill).

Table 2: Selected statistics on Shareville users. For 1-year returns and Sharpe ratio, the 1st and 100th percentiles of the whole dataset are excluded to remove outliers.

	Mean	Median	Min	Max
All users (*N* = 39,920)				
Followers (C_D)	2.0	0	0	2466
1-year returns	19%	21%	-73%	130%
1-year Sharpe ratio	1.13	1.15	-2.06	4.28
Top 50 by followers				
Followers (C_D)	532.8	385.5	214	2466
1-year returns	67%	64%	-8%	126%
1-year Sharpe ratio	2.57	2.82	-0.08	4.06
Top 50 by Sharpe ratio				
Followers (C_D)	14.1	2	0	154
1-year returns	60%	57%	27%	117%
1-year Sharpe ratio	4.24	4.24	4.22	4.28

Table 3 shows the results of the regression specified in Equation 1 for the full model and the final model, as well as the variance in the data explained by the regressors in the final model (ANOVA). The variables explaining most variance are commenting on trades ($\ln(A_T + 1)$) and the Sharpe ratings (S_i), so that higher risk-adjusted returns compared to other Shareville users indicate higher number of followers. There is no remaining component in the variance that would be explained by non-risk-adjusted 1-year returns (Y), 1-year Sharpe ratio (H), or comments in discussion groups ($\ln(A_G + 1)$). The residual (unexplained) variance is moderately high, 75% of total variance.

Table 4 shows the results of the regression specified in Equation 2 for the full model and the final model, as well as the variance in the data explained by the regressors in the final model (ANOVA). The predicted number of followers is, as can be expected, a statistically significant predictor of a user's Pagerank centrality, explaining 55.1% of variance in the data. Most of the other regressors are eliminated from the model or have only a marginal role in explaining variance in the data. Of the Sharpe ratio based variables, S_3 is the only statistically significant one and has a negative coefficient. Residual (unexplained) variance is 41.8%.

The evidence presented in Tables 3 and 4 is sufficient to confirm Hypothesis 1. Further implications of the results are discussed in Chapter 7.

Table 3: Regression results for C_D. The total number of users included in the regression is 8,412.

	Full model		**Final model**		
Regressor	**Coefficient (Std. error)**	**Significance**	**Coefficient (Std. error)**	**Significance**	**Variance explained**
(intercept)	0.9318 (0.0177)	+++	0.9317 (0.0177)	+++	
Y	0.0000 (0.0000)				
H	0.0000 (0.0000)				
S_1	0.1259 (0.0237)	+++	0.1261 (0.0237)	+++	
S_2	0.2774 (0.0241)	+++	0.2778 (0.0241)	+++	6.2% (S_1, S_2 and S_3 combined)
S_3	0.7169 (0.0306)	+++	0.7180 (0.0306)	+++	
$\ln(A_G + 1)$	0.0041 (0.0201)				
$\ln(A_T + 1)$	0.1758 (0.0154)	+++	0.1763 (0.0153)	+++	12.4%
$\ln(A_I + 1)$	0.0278 (0.0146)		0.0283 (0.0144)	+	0.7%
$\ln(A_U + 1)$	0.2458 (0.0189)	+++	0.2473 (0.0176)	+++	2.8%
$\ln(M_1 + 1)$	0.6594 (0.0367)	+++	0.6595 (0.0367)	+++	2.8%
$\ln(M_2 + 1)$	-0.0857 (0.0361)	+	-0.0854 (0.0360)	+	0.1%
				Residual variance:	75.0%

+ $p < 0.05$ ++ $p < 0.01$ +++ $p < 0.001$

Table 4: Regression results for C_P. The total number of users included in the regression is 8,412.

	Full model		Final model		
Regressor	**Coefficient (Std. error)**	**Significance**	**Coefficient (Std. error)**	**Significance**	**Variance explained**
(intercept)	-11.06 (0.01)	+++	-11.07 (0.01)	+++	
$\ln C_D$	0.5830 (0.0066)	+++	0.5816 (0.006)	+++	55.1%
Y	-0.0000 (0.0000)				
H	0.0000 (0.0000)				
S_1	0.0046 (0.0177)		0.0046 (0.0177)		
S_2	-0.0175 (0.0181)		-0.0124 (0.0181)		1.8% (S_1, S_2 and S_3 combined)
S_3	-0.176 (0.0235)	+++	-0.1676 (0.0235)	+++	
$\ln(A_G + 1)$	0.0107 (0.0150)				
$\ln(A_T + 1)$	0.0347 (0.0116)	++			
$\ln(A_I + 1)$	-0.0864 (0.0109)	+++			
$\ln(A_U + 1)$	0.104 (0.0142)	+++	0.0716 (0.0105)	+++	0.4%
$\ln(M_1 + 1)$	0.2610 (0.0277)	+++	0.2622 (0.0277)	+++	0.4%
$\ln(M_2 + 1)$	0.1202 (0.0268)	+++	0.1120 (0.0267)	+++	0.4%
				Residual variance:	41.8%

\+ $p < 0.05$ ++ $p < 0.01$ +++ $p < 0.001$

5 Study 2: Copy trading in Shareville

Hypothesis 2: *If an influential user trades on a particular stock, their followers will trade that stock more actively than other Shareville users.*

5.1 Motivation

Propagation of information, ideas, and behavior from one individual to another is an inseparable element of social networking. In the context of investor networking, this includes not only investment analyses and strategies, but also individual investment (and divestment) decisions. While many forex trading and networking platforms have gone so far as to allow users to automatically copy other users' trades, such copy trading makes sense only in the case of high-frequency, time-consuming day trading. It also requires a high level of trust in the other user's trading skills, at least compared to the copying user's own skills.

In contrast to forex traders, most retail stock market investors are not high-frequency traders, and their investment decision are, at least in theory, based on an evaluation of listed companies along multiple dimensions of ambiguous information. The propagation of investment strategies and individual investment decisions within online investor networks can therefore be expected to be a relatively complex phenomenon, and apart from Heimer and Simon (2014), who demonstrate the influence of increased communication on the propagation of active trading strategies within a currency trading network, there is very little research on the subject. On the other hand, online platforms offer an excellent opportunity to investigate trading propagation, since the connections between investors are made explicit, and trades can be monitored on the intraday level. Recognizing the gap in the existing literature, this study comprises the first effort to quantify trading propagation in an online investor network focused on the stock market.

5.2 Methods

While Study 1 examines what factors contribute to the decision to follow another user, this study examines the degree of influence a user has on their followers. Arguably, since most Shareville users have no followers, any user with at least one follower can be considered to be influential to at least some degree. Thus, for each Shareville user with at least one follower, all buy-side trades executed between 26 Sep 2014 and 31 March 2015 (141,248 trades) are analyzed to see how many followers copy them. Sell-side trades are omitted since copying them would require the follower to hold the instrument in the first place, limiting the sample size and incurring a greater risk of sampling bias due to e.g. preferences related to investing style.

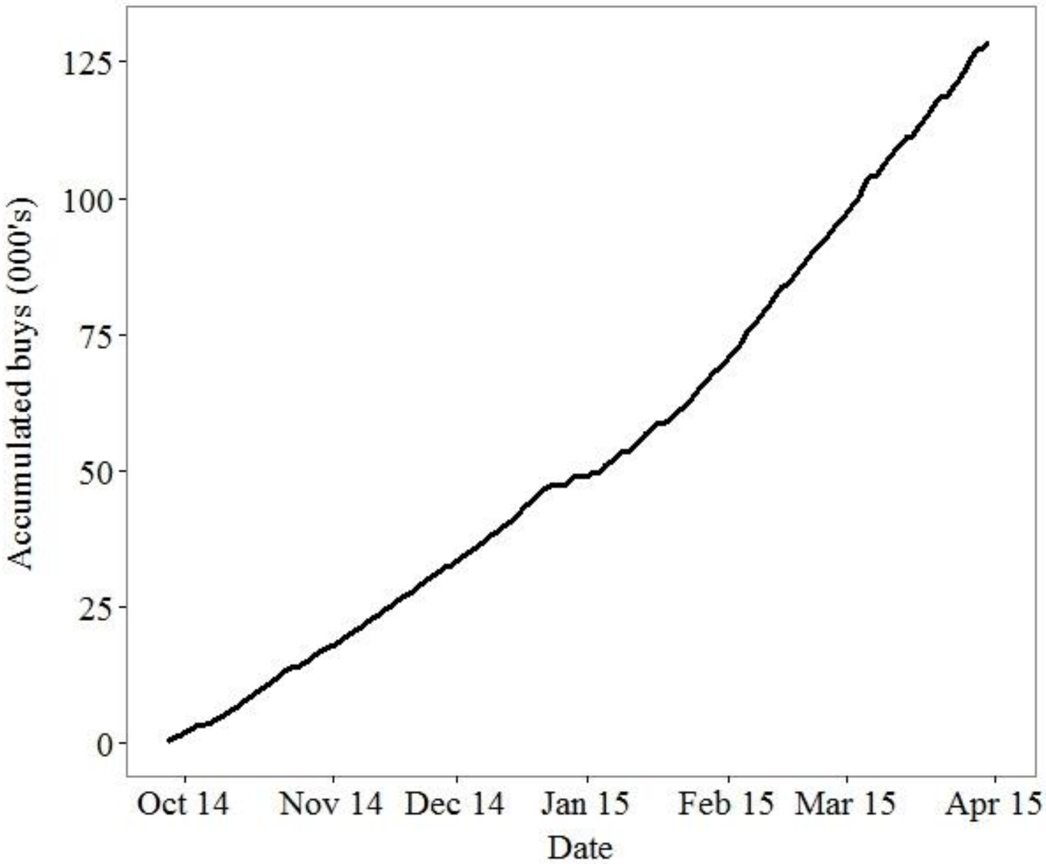

Figure 5: Accumulation of buy-side trades included in the analysis over the study period (from 26 Sep 2014 to 31 Mar 2015).

Figure 5 shows how the buy-side trades included in the analysis are accumulated over the study period. Although the Shareville user base more than quadruples over the study period, the trades are collected relatively evenly over the period (cf. Figure 3). This is because the share of users with followers is necessarily lowest among new users, so that the accumulation of trades by users with followers lags behind user base growth.

For each buy-side trade, Figure 6 shows the number of followers the user making the trade has at the time of the trade. To further examine the role of influence in the network on copy trading, the trades are divided into two groups based on the number of followers the user making the trade has at the time of the trade. Group 1 contains trades by users with less than 50 followers and Group 2 trades by users with at least 50 followers. The limit of 50 followers is chosen arbitrarily to some degree, but as shown in Figure 6, it divides the trades so that the two groups span roughly similar intervals on a logarithmic scale with regard to the number of followers at time of trade. Group 1 contains 97% of buy-side trades analyzed.

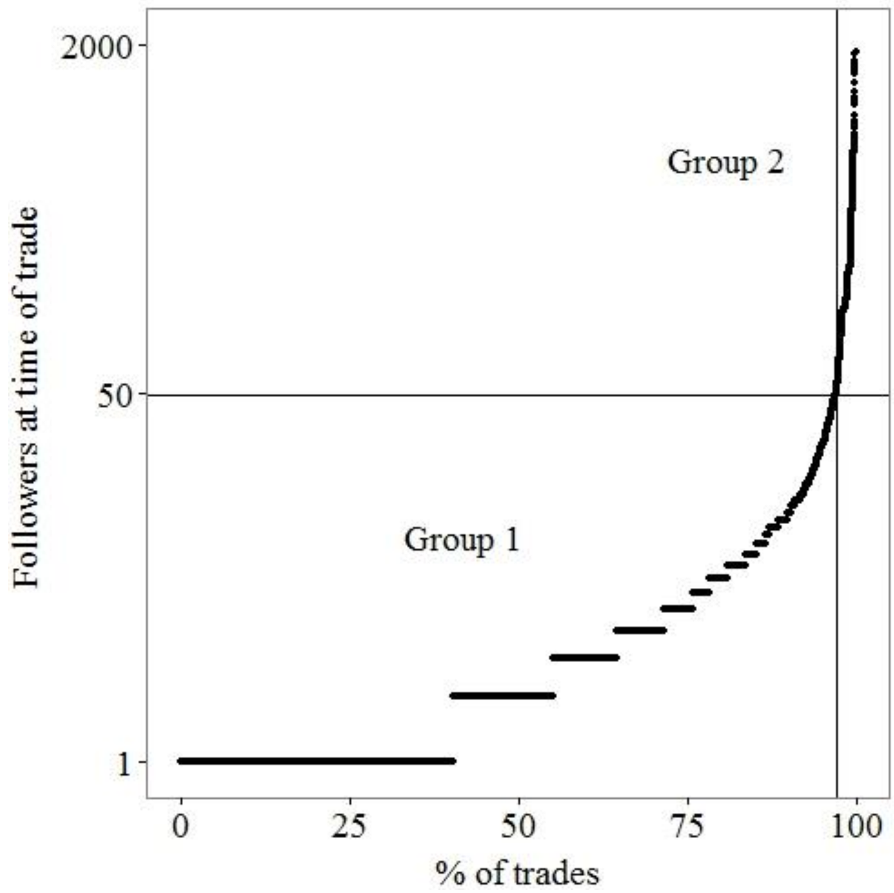

Figure 6: For each buy-side trade, the number of followers the user making the trade has at the time of the trade. Trades are arranged by number of followers of the user making the trade, so that all trades in Group 2 are to the right of the vertical line. Only trades by users with at least one follower are included.

Let now $k \in \{1,2,\dots,K_j\}$ denote the buy-side trades made by user j, $n_{j,k}$ the number of followers of user j who replicate trade k, and $N_{j,k}$ the total number of followers user j has at time of trade k. In this context, replication is defined as buying the same instrument later on the same date or during the next day (for Fridays, by end of next Monday). The copy trade rate per trade is then defined as

$$r_{j,k} = \frac{n_{j,k}}{N_{j,k}} \tag{7}$$

Since most users have just one or two followers, for most trades $r_{j,k}$ is equal to zero or a simple fraction of small integers such as $\frac{1}{1}$ or $\frac{1}{2}$ (if $n_{j,k} > 0$ and $N_{j,k}$ is small). Also, let $S_{j,k}$ denote the total number of Shareville users who are from the same country as user j at the end of the day when user j makes trade k, and $s_{j,k}$ the total number of Shareville users who replicate trade k. A reference copy trade rate for $r_{j,k}$ is then defined as

$$R_{j,k} = \frac{s_{j,k} - n_{j,k}}{S_{j,k} - N_{j,k}} \tag{8}$$

The reference rate provides an empirical estimate for the expected rate of replication arising from users coincidentally carrying out the same trade. There are several possible ways of defining the reference group, but here the use of $S_{j,k}$ reflects the intuitive assumption that Shareville users from the same country trade more homogeneously than users from different countries, and thus provide the most natural reference group. For example, Finnish users prefer to trade in companies listed in the Helsinki stock exchange. Equation 8 thus provides a relatively simple means for estimating $R_{j,k}$, and not restricting $s_{j,k}$ and $N_{j,k}$ to the user's own country in practice further inflates the reference rate slightly to avoid underestimation. A more sophisticated estimate of the reference rate would require matching users based on behavioral characteristics, e.g. trading activity, and would add considerable complexity.

Adjusting $r_{j,k}$ for coincidental trade replication by subtracting $R_{j,k}$ and averaging the adjusted rate over all users and trades yields an estimate for the copy trade rate arising from followers intentionally copying a trade by the followed user:

$$\bar{r} = \overline{r_{j,k} - R_{j,k}} = \frac{\sum_{j,k}(r_{j,k} - R_{j,k})}{\sum_j K_j} \tag{9}$$

To confirm or reject Hypothesis 2, Student's one-tailed *t*-test[3] is used to examine the null hypothesis that followers of a user are just as likely to replicate the user's trades as other Shareville users, i.e. that the mean $\overline{r_{j,k} - R_{j,k}} = 0$. Hypothesis 2 can be confirmed if this null hypothesis is rejected at the $p < 0.05$ level so that $\overline{r_{j,k} - R_{j,k}} > 0$. The R function `t.test` is used for the statistical test.

The values $r_{j,k} - R_{j,k}$ for Group 1 and Group 2 (less than and at least 50 followers at time of trade) are compared using Welch's two-tailed *t*-test for unequal variances ($p < 0.05$) to see whether the number of followers influences the copy trade rate. The R function `t.test` is used for the statistical test. Also, the cumulative number of copy trades is calculated for both groups to evaluate the prevalence of copy trading in the Shareville network. It is expected that trades in Group 2 will result in considerably more copy trades per original trade, since the number of potential copiers (followers) is higher than in Group 1.

[3] Although the distribution of $r_{j,k} - R_{j,k}$ cannot be expected to be normally distributed, the number of samples is in the thousands, so Student's *t*-test can be applied based on the central limit theorem.

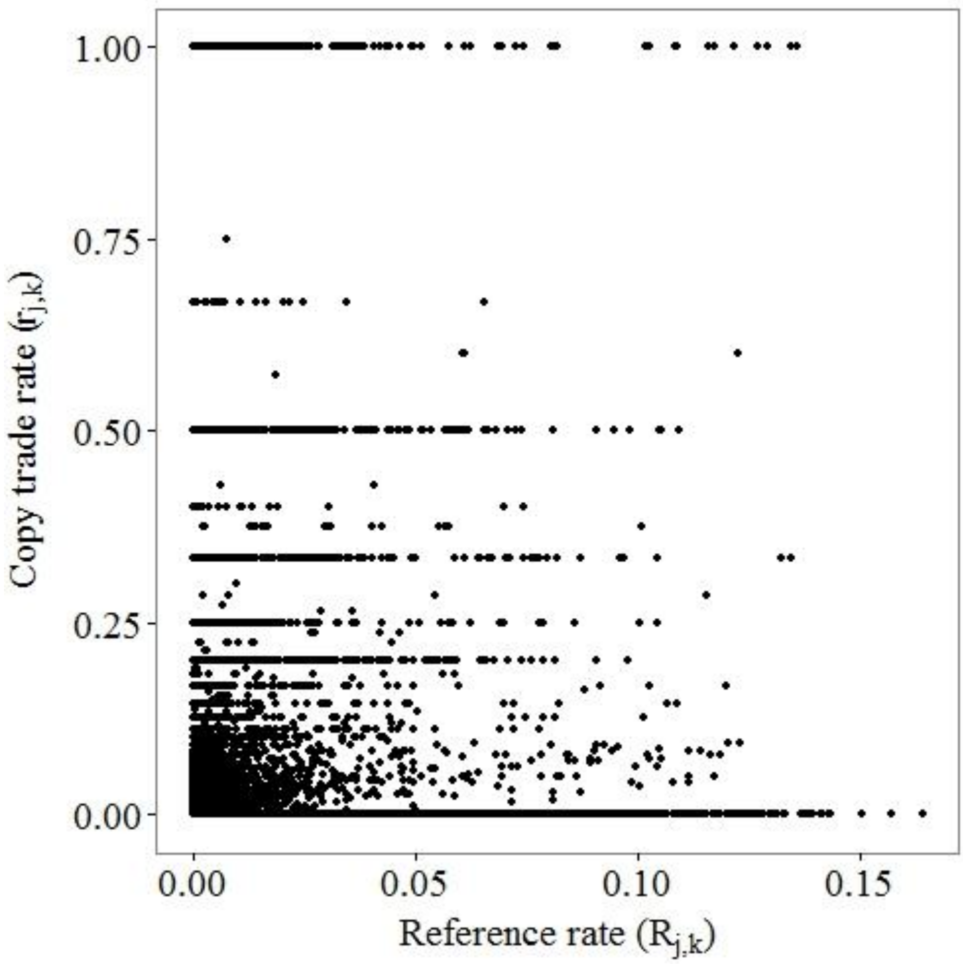

Figure 7: Joint distribution of copy trade rate and reference rate for all buy-side trades (*N*=141,248) by users with at least one follower. Each point corresponds to one trade.

5.3 Results

Figure 7 shows the joint distribution of the unadjusted copy trade rate $r_{j,k}$ and the corresponding reference rate $R_{j,k}$ for all buy-side trades made by users with at least one follower (7,459 users and 141,248 trades in total). The average value of $r_{j,k}$ is 0.012, well above the corresponding average for the reference trade rate $R_{j,k}$ which is 0.003. Student's one-tailed *t*-test for $r_{j,k} - R_{j,k}$ results in a value of t = 37.5 for the test statistic and a *p*-value below 2.2×10^{-16}, indicating that the null hypothesis that $\overline{r_{j,k} - R_{j,k}} = 0$ can be rejected. Thus, Hypothesis 2 is confirmed.

When partitioning trades into Group 1 (trader has less than 50 followers) and Group 2 (trader has at least 50 followers), the results show that $\bar{r}$ (average adjusted copy trade propensity per follower) is higher for Group 1 ($\bar{r}$ = 0.009) than for Group 2 ($\bar{r}$ = 0.002). The difference is statistically significant

(Welch's two-tailed *t*-test, $p < 2.2\times10^{-16}$). Figure 8 gives a more detailed view of how the adjusted copy trade rate decreases with the number of followers at time of trade. The adjusted copy trade rate is in practice close to zero for all trades by users with more than 50 followers. However, as shown in Figure 9, trades in Group 2 produce approximately one third of all copy trade activity in Shareville, even though the number of trades in Group 2 is only 3% of all trades (cf. also Figure 6). This is because the large number of followers more than compensates the lower copy trade rate per follower in Group 2. Table 5 shows summary statistics of the data and Groups 1 and 2.

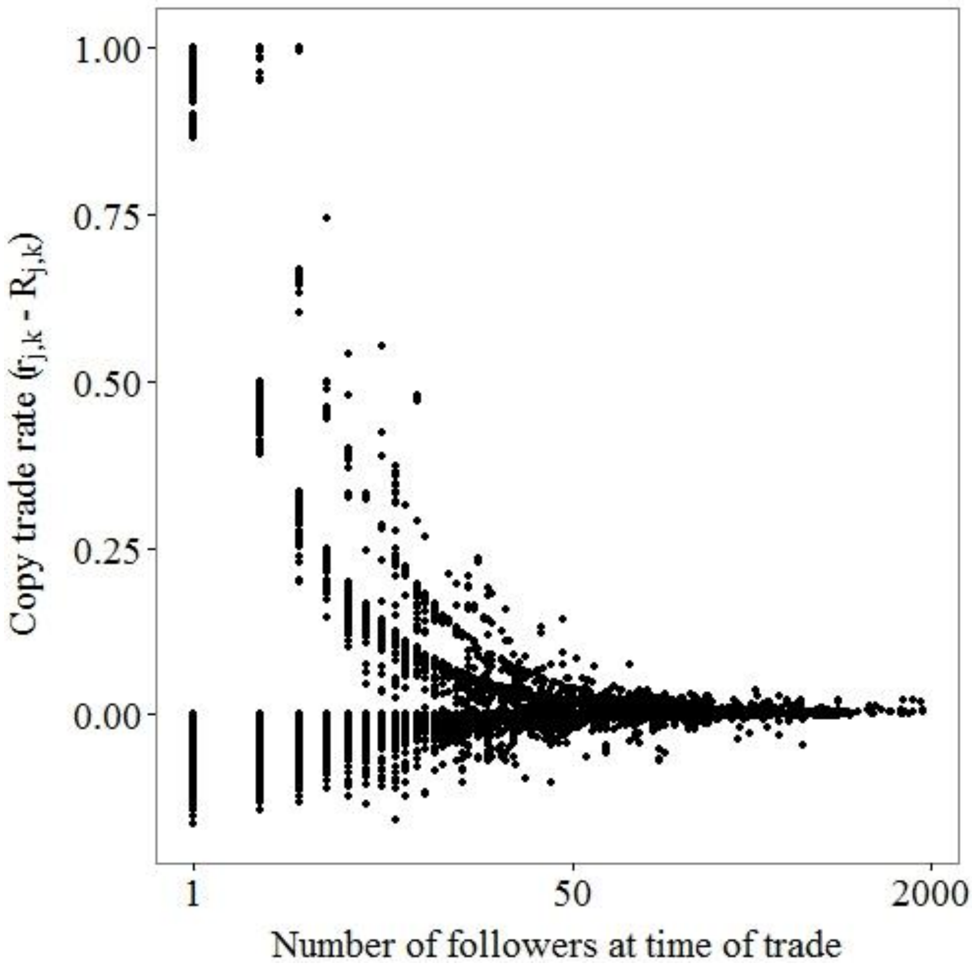

Figure 8: Copy trade rate as a function of the number of followers at time of trade. Each point corresponds to one trade. Negative values correspond to trades where the subtracted reference rate is greater than the unadjusted copy trade rate.

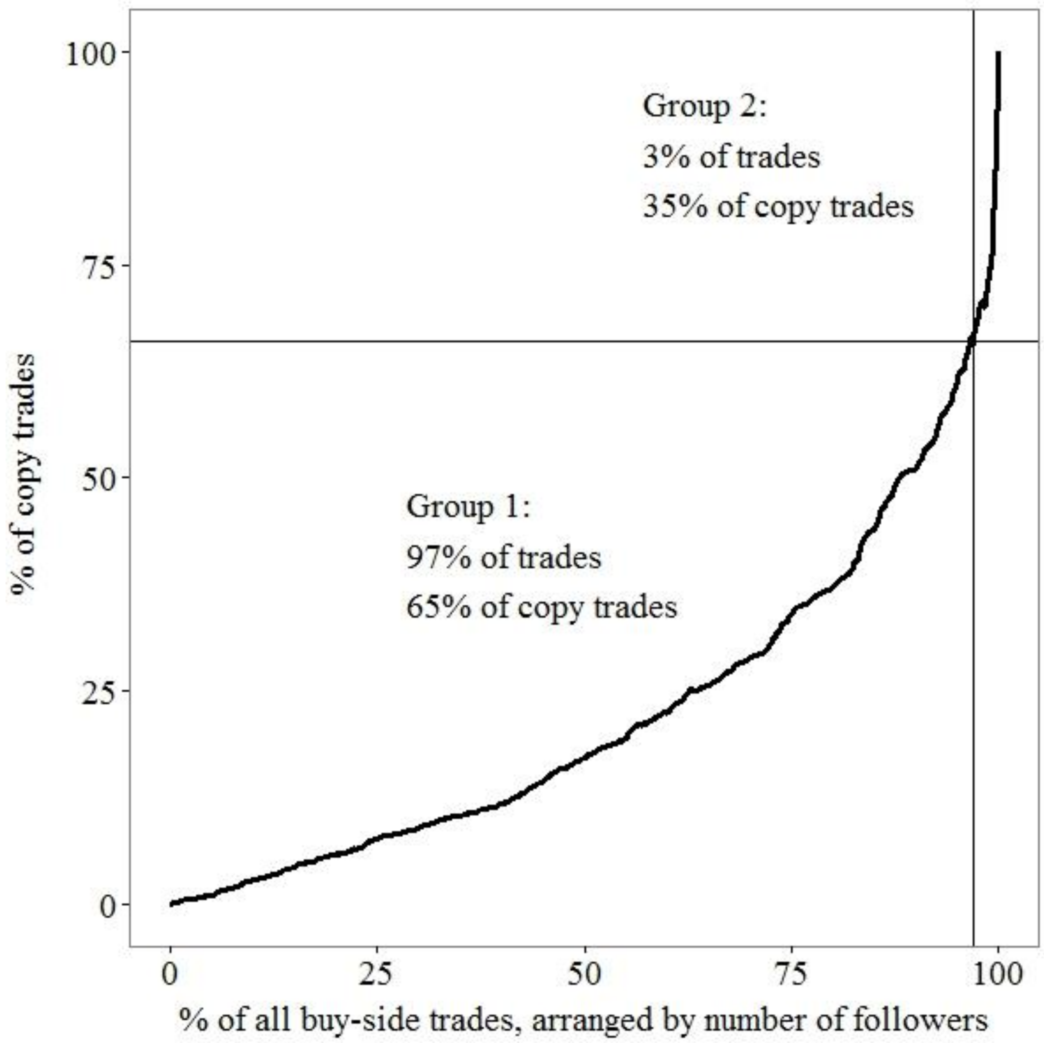

Figure 9: Cumulative shares of trades and copy trades for Group 1 and Group 2. Trades are arranged by number of followers of the user making the trade, so that all trades in Group 2 are to the right of the vertical line. Copy trades are adjusted for coincidental copying as described in the text.

Table 5: Summary statistics on Hypothesis 2 results. The total number of distinct users is greater than the total number of users with followers since the same individual's trades may be included in both groups. The third column indicates whether the difference between Group 1 and Group 2 is statistically significant (Welch's two-tailed *t*-test).

	Total	**Group 1**	**Group 2**	***p* < 0.05**
Buy-side trades	141,248	137,107	4,141	n/a
Distinct users	7,459	7,392	195	n/a
Followers at time of trade (mean)	10.9	4.8	213.6	Yes
$\overline{r_{J,k}}$	0.013	0.013	0.005	Yes
$\overline{R_{J,k}}$	0.003	0.003	0.003	Yes
$\bar{r} = \overline{r_{J,k} - R_{J,k}}$	0.009	0.009	0.002	Yes
Estimated number of copy trades	6,174	4,008	2,166	n/a

6 Study 3: Selective communication in the network

Hypothesis 3: *An individual user is more likely to comment on trades in which they realize gains than on trades in which they realize losses.*

6.1 Motivation

As discussed by e.g. Kaustia and Knüpfer (2012) and Heimer (2014), returns influence investor communication behavior according to the selective communication hypothesis: investors are less likely to communicate after losses in the market. Krasny (2013) and Dijk et al. (2014) have also shown that the investor's awareness of their social status in the investor network influences portfolio allocation. Understanding and quantifying the selective communication phenomenon is therefore an important step in aiding investors to debias themselves against a natural tendency to downplay or rationalize poor trading decisions. This study contributes to the literature by demonstrating the selective communication effect at the individual trade level and quantifying its magnitude.

6.2 Methods

When a user sells an instrument above (below) its acquisition price, they realize gains (losses). Hypothesis 3 postulates that winners are more likely to brag about their success than losers are to lament their losses in front of the community. The first step towards analyzing this matter is establishing the correct approach to calculate the effective acquisition price and thus the returns. For example, when determining capital gains tax on sale of shares, Swedish tax authorities use the volume-weighted average acquisition cost (AAC) of all shares of the same type, even if only a portion of the shares is sold. Finnish tax authorities use the first-in-first-out (FIFO) method, where the acquisition price is based on chronological order, so that the oldest shares of the same type are sold first. Consequently, a Swedish investor buying shares in two or more transactions at different prices and selling some (but not all) of them later could end up making a profit, while a Finnish investor executing the same transactions at the same prices might make a loss, at least as far as taxation is concerned.

The sell-side trade data available in this study does not carry acquisition price information, and portfolio data is available only on monthly level, so returns for sell-side trades are reconstructed from the available data for Nordnet accounts linked to a Shareville profile as follows. For each Swedish and Finnish Shareville user, trades in shares they did not hold on their Nordnet account at the end of 31 Dec 2012 are tracked from 1 Jan 2013 onwards. Shares held at the end of 2012 are excluded to ensure that all acquisition prices are calculated similarly, and to focus on trades with a short-to-mid-term horizon (shares held for approximately two years or less). For investors who joined Nordnet after 2012, tracking starts on the calendar month after joining, and shares held at the end of the month of joining are excluded (these are often transferred from another service provider without acquisition price information). The remaining trade data then consists mostly[4] of cases where the user buys shares of one type on one or more dates, and possibly sells them later on one or more dates. Then, for each sell-side trade, returns after trading costs are calculated using both the AAC and FIFO method. Trades with positive return are classified as gains and trades with zero or negative return are classified as losses.

To examine Hypothesis 3, trades made after the user joins Shareville are linked with Shareville comment data by the user. The user can publish a comment on a specific trade they make, and they can in some cases also choose a description for the trade from a small set of predefined descriptions, such as “short-term trade”, “long-term trade”, etc. In the present analysis, such descriptions are also counted as comments. The dataset does not provide a direct connection between Nordnet trade data and Shareville comments, so the trades and comments are linked based on traded instrument and trade date. Any comments made on trades by other users are ignored, as are comments on the user’s own trades that are made only after another user has commented on the trade.

If a user makes several trades on the same share and same side (buy or sell) during one day, these trades are excluded at this stage of the data processing. This ensures that each comment is linked to the right trade in the subsequent analysis of commenting behavior. Also, it may be expected that frequent (day) traders have a decreased propensity to comment on each individual trade, so their commenting behavior should in any case be analyzed separately from other users. Trades are also excluded from the analysis if the total traded value is suspiciously low (trading costs are 5% or more of traded value), which typically results from selling a very small number of shares (often just one) for some reason, or the estimated returns appear to be artificially low or high (less than -100% or over 1,000%).

To analyze the impact of loss and gain on commenting behavior, sell-side trades are divided into two categories: those that the trader chooses to comment, and those that they choose not to comment. This information is contrasted with the reconstructed gain and loss information to form a 2×2 contingency table (gain and comment, loss and comment, gain and no comment, loss and no comment). Pearson’s

[4] In a minority of cases the user may sell shares that were not acquired in the market (e.g. transfer from another account, gift, option, or short selling), resulting in a negative balance on that share in the reconstructed portfolio. Since acquisition prices are not available for these sales, only returns from shares acquired in the market are considered in the analysis.

chi-squared test is then used to determine whether the observed commenting frequencies are higher for trades resulting in gains than for trades resulting in losses. The test is performed separately for Finnish and Swedish users, and for the AAC and FIFO methods of calculating the returns. Hypothesis 3 is rejected if no statistically significant (p < 0.05) difference is observed in the commenting frequencies. The R function `prop.test` is used for the Pearson's chi-squared test.

To further analyze the relationship between returns and commenting behavior, the observed cumulative distribution functions (CDFs) of non-commented and commented sells are formed as a function of the AAC return percentage of the sell. Hypothesis 3 can then be interpreted as the statement that the difference

$$\mathrm{CDF}_{\mathrm{diff}}\,(\mathrm{AAC\ return\ \%}) = \mathrm{CDF}_{\mathrm{not.commented}} - \mathrm{CDF}_{\mathrm{commented}} \qquad (10)$$

should be greater than zero and increasing for negative values of AAC return. Furthermore, the difference should decrease for positive values of AAC return.

In addition to analyzing commenting on sell-side trades of all users, a separate analysis is carried out for the subset of users who have realized both gains and losses as determined by AAC return. For user i, the commenting rates for gains and losses are determined:

$$r_{i,gain} = \frac{\text{Number of commented gains (AAC) by user } i}{\text{Number of gains (AAC) by user } i} \qquad (11)$$

$$r_{i,loss} = \frac{\text{Number of commented losses (AAC) by user } i}{\text{Number of losses (AAC) by user } i} \qquad (12)$$

Hypothesis 3 can then be expressed as $r_{i,loss} - r_{i,gain} < 0$, while the corresponding null hypothesis is that the difference is zero or positive. Student's paired one-tailed *t*-test is used to determine whether the null hypothesis may be rejected at the p < 0.05 level, and the analysis is carried out separately for Finnish and Swedish users. If the null hypothesis can be rejected, this provides further support for the selective communication hypothesis.

The dataset in this study does not contain dividend information, so shareholder returns are slightly underestimated for those positions that are held over the ex-dividend date and then sold, leading to a potential systematic bias against observing the hypothesized commenting behavior (in cases where the dividends turn the position from losing to profitable). The impact of this bias is mitigated by the facts that it applies only to those positions that are held over the ex-dividend date, and on average dividend returns from positions held over a period of several months and closed after Shareville was launched can be expected to be small compared to share price growth (the period from Jan 2013 to March 2015 witnessed roughly 50% gains in the Helsinki and Stockholm general stock market indices). Also, a user closing a losing position might still behave as if the outcome was a loss even if their total return ends up being slightly positive due to dividend returns.

6.3 Results

Table 6 shows a summary of the sell-side trades by Finnish and Swedish Shareville users between 26 September 2014 and 31 March 2015. From a total of 111,781 sell-side trades, 86,927 were included in the analysis and 24,854 were excluded based on the criteria specified earlier (e.g. selling the same instrument multiple times during the same day). Compared to Finnish users, Swedish users have mean and median returns slightly closer to zero for both winning and losing trades, but higher standard deviation for returns from losing trades and smaller standard deviation for returns from winning trades. This indicates that especially the return distribution for winning trades is more heavy-tailed for the Finnish than Swedish trades. The differences in means between the Finnish and Swedish groups are statistically significant (Welch's *t*-test for unequal variances, $p < 0.05$).

Table 7 shows the contingency table of comment and sales data and the outcome of Pearson's chi-squared test. Overall, the propensity to comment trades is relatively low, so that only 2% of sells are commented by the trader. The commenting rate is similar for Finnish and Swedish trades. Pearson's chi-squared test indicates that losing trades by Swedish Shareville users are less likely to receive comments than winning trades, with an effect magnitude of approximately 10% (1.9% vs. 2.1%). No difference is observed for the Finnish users. The method used for calculating trade returns (AAC or FIFO) does not have any substantial impact on the results.

As further evidence to support the selective communication hypothesis, Figure 10 displays the difference in the observed cumulative distributions of non-commented trades and commented trades for the Finnish and Swedish data. The Swedish data in Figure 10, with the peak at almost exactly 0% AAC returns, provides strong support for the selective communication hypothesis. The narrow trough just left of the peak most likely represents noise in the data. The same behavior is not observed for the Finnish data, although the difference moves from zero to negative at approximately 0%, which indicates an increase in the propensity to comment trades when returns turn from negative to positive.

Table 6: Summary statistics of Study 3 data.

	Finnish			Swedish		
Number of sell-side trades	29,652			57,275		
Losing trades (AAC/FIFO)	11,489 / 11,403			23,757 / 23,578		
Users in total	4,150			4,915		
Users commenting own trades	434			602		
	Median	**Mean**	**Std. dev.**	**Median**	**Mean**	**Std. dev.**
AAC returns						
Losing trades	-6.0%	-12.0%	± 15.9%	-4.6%	-11.2%	± 17.1%
Winning trades	6.4%	16.3%	± 54.1%	5.4%	12.5%	± 24.6%
FIFO returns						
Losing trades	-5.6%	-10.9%	± 17.7%	-4.3%	-9.8%	± 20.4%
Winning trades	6.3%	15.4%	± 54.0%	5.2%	11.4%	± 23.7%

Table 7: Contingency table for Hypothesis 3. The *p*-value is for the null hypothesis that the ratio of commenting is same for losing and winning trades (Pearson's one-sided chi-squared test).

	Finnish			Swedish		
	Comment	**No comment**	**Ratio**	**Comment**	**No comment**	**Ratio**
AAC returns						
Losing trades	238	11,251	2.1%	439	23,318	1.9%
Winning trades	366	17,728	2.1%	690	32,626	2.1%
		***p*-value**	0.5975		***p*-value**	0.0317
FIFO returns						
Losing trades	238	11,165	2.1%	430	23,148	1.9%
Winning trades	366	17,809	2.1%	699	32,791	2.1%
		***p*-value**	0.6526		***p*-value**	0.0141

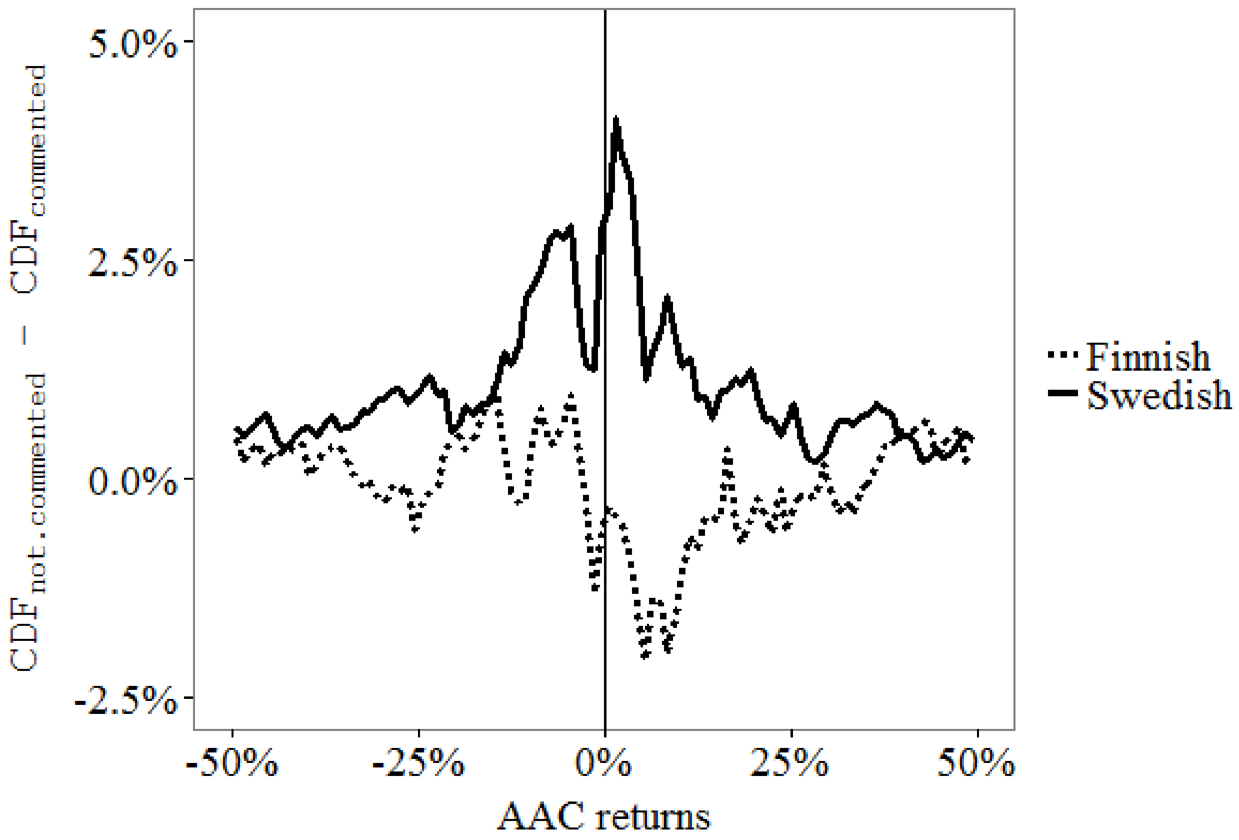

Figure 10: Difference between CDFs of non-commented and commented trades as a function of AAC returns. For example, based on data in Table 7, the value of the CDF for non-commented Swedish trades at zero returns is 41.7% and for commented trades 38.9%. The difference, 2.8%, can be read from the graph at the point where AAC returns = 0%. 86% of Finnish and 88% of Swedish trades are contained within the [-25%, 25%] return interval.

Table 8 summarizes the results from the paired t-test on $r_{i,loss} - r_{i,gain}$ among Finnish and Swedish Shareville users with both losses and gains, and Figures 11 and 12 show the distribution of $r_{i,loss} - r_{i,gain}$ for the Finnish and Swedish users, respectively. Also here the results for the Swedish users are statistically significant (p-value 0.015) and in line with the selective communication hypothesis with an effect magnitude of approximately 20% (1.7% vs. 2.1%), while no statistically significant results are obtained for the Finnish users (p-value 0.16). However, the data in both Table 8 and Figure 11 suggest that the Finnish users also exhibit selective communication, and the lack of statistical significance may be merely a question of insufficient sample size in relation to the effect magnitude. In both Finnish and Swedish users the observed effect magnitude is greater than in the previous analysis, where the commenting rates were calculated over winning and losing trades as opposed to individual investors. Repeating the analysis with FIFO returns did not affect the results appreciably.

Based on these results, Hypothesis 3 is confirmed for Swedish but not for Finnish Shareville users.

Table 8: Results from paired *t*-test on loss and gain commenting rates of Finnish and Swedish users. Data are given as mean (standard deviation) calculated over the users. The *p*-value is for the null hypothesis that $r_{i,loss} - r_{i,gain} \geq 0$.

	$r_{i,loss}$	$r_{i,gain}$	$r_{i,loss} - r_{i,gain}$	p-**value**
Finnish (N = 2046)	1.9% (9.8)	2.2% (9.3)	−0.3% (13.1)	0.16
Swedish (N = 2702)	1.7% (8.1)	2.1% (9.2)	−0.5% (11.4)	0.015

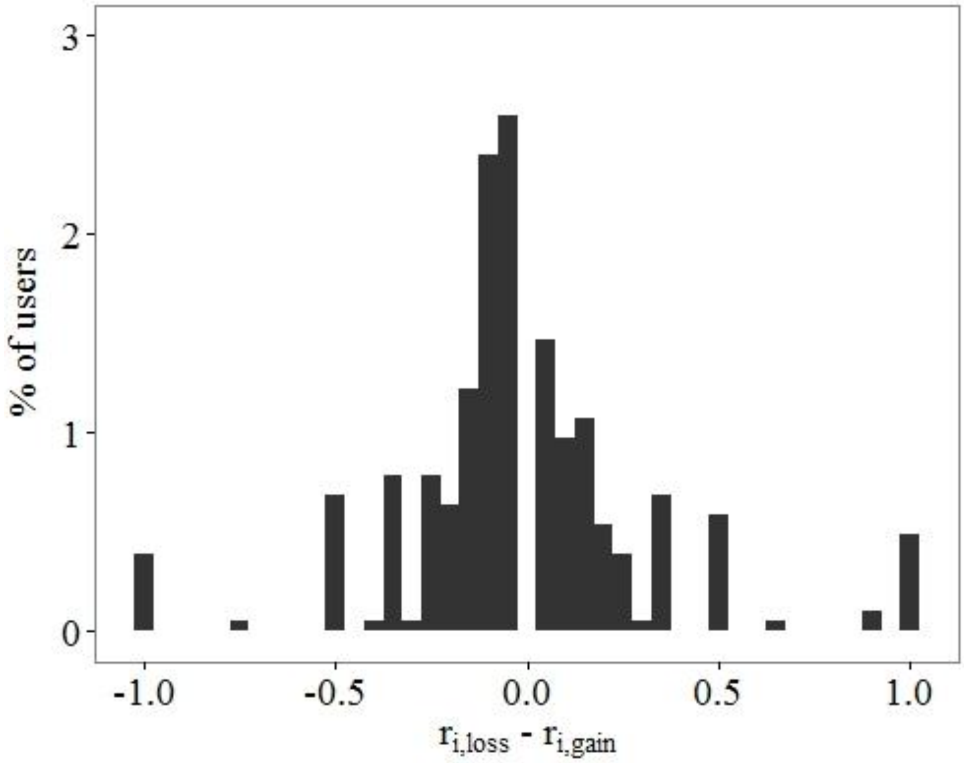

Figure 11: Distribution of the difference in commenting rates $r_{i,loss} - r_{i,gain}$ for Finnish users with both gains and losses. Bin width in the figure is 0.05. The bin at 0.0 corresponds to 84% of the users and has been removed to better show the distribution in other bins.

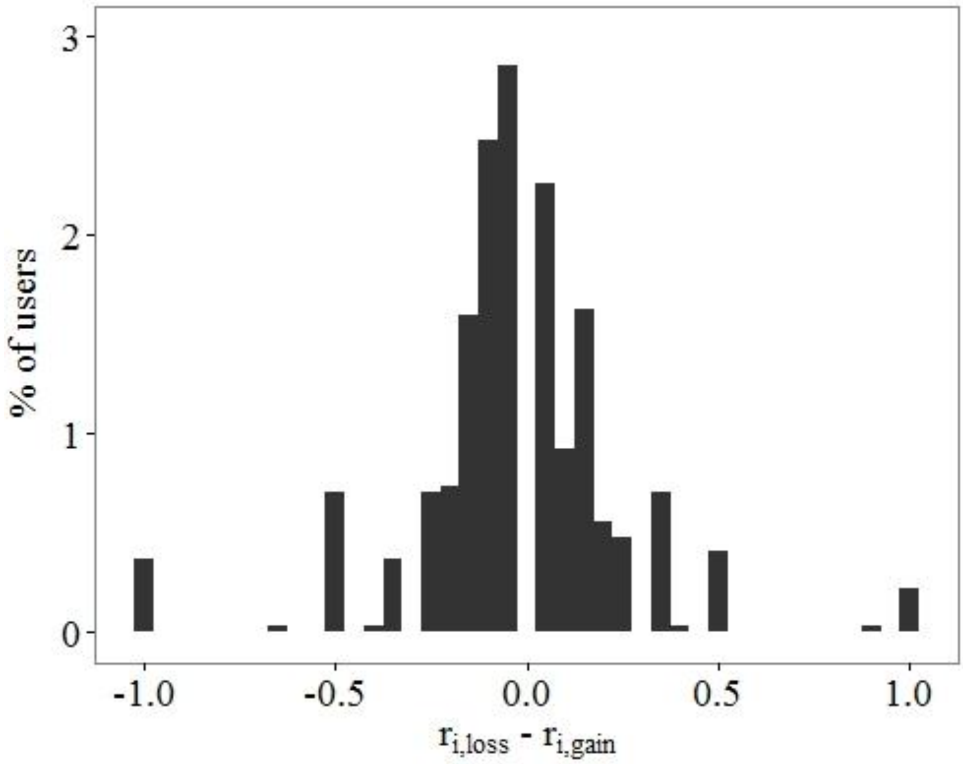

Figure 12: Distribution of the difference in commenting rates $r_{i,loss} - r_{i,gain}$ for Swedish users with both gains and losses. Bin width in the figure is 0.05. The bin at 0.0 corresponds to 83% of the users and has been removed to better show the distribution in other bins.

7 Discussion

This thesis presents three novel studies on the behavior of investors in an online investor network primarily focused on the stock market. The studies provide answers to three key research questions, namely *What attributes and behavior determine an investor's centrality in the network?*, *To what extent do central users influence trades by their followers?*, and *Are investors more inclined to discuss profitable trades than losses?* The results also provide a general characterization of the online investor network, e.g. what can be considered typical attributes and behavior of an investor in such a network. Results from each study are discussed below, including their implications and relationship to previous research.

7.1 Study 1: Attributes valued by Shareville users

Investors in the Shareville network are able to comment on trades, individual instruments, and engage in public and private conversations with other investors. They are also able to view the trades and portfolio contents and past performance of other investors. The results obtained for Hypothesis 1 indicate that when selecting which users to follow, users value particularly other users' past performance (as reflected by the Sharpe ratio based ranking of the user within the network) and active commenting of trades. Both are intuitively plausible findings, since users can be expected to seek to learn from those users who have superior past performance and who are willing to communicate with others. Also, the Shareville user interface readily displays user rankings and recent commenting activity, so high-performing active users are the easiest to find.

There is also a sizable unexplained component to the number of followers a user has. This may result from factors such as the portfolio composition of the user, the content of their communication with others, and especially the number of followers itself, so that users who already have many followers may attract more followers than other users with similar historical returns and other attributes. The latter effect could in principle be quantified by a longitudinal study, but some of the relevant data was only available in cross-sectional format at the end of the study period. In addition to these Shareville-related confounding variables, factors external to Shareville may influence the networking. For example, some popular Shareville users maintain investing related blogs and cross-link the blog with their Shareville profile. Such extra visibility will certainly add to the number of followers these users have.

Although past performance as reflected by the star rating appears to be a key factor in choosing which users to follow, none of the top 50 users by Sharpe ratio belong to the top 50 users by number of followers as indicated by Table 2. The Sharpe ratio is calculated over a 1-year period that overlaps by up to six months with the user's Shareville membership tenure, which is also the time period during which followers are attracted. To the extent that the Sharpe ratio can be used as a proxy for investor skill, this suggests that Shareville users as a group are not recognizing the most skilled investors for either past or future performance. This finding is reminiscent of the observation by Pan et al. (2012) that for most users, the number of followers is a poor predictor of future performance. On the other hand, when the simple Sharpe ratio-based star ranking is used as a regressor of the number of followers a user has, 1-year return and Sharpe ratio do not provide any additional explanatory power. This is analogous to research on Morningstar fund ratings and investment flows by Del Guercio and Tkac (2008), who observe that retail investors prefer simple summarizing information, such as star ratings, over more specific quantitative performance measures when allocating investments into mutual funds.

It is also possible that high Sharpe ratios are symptomatic of a high-volatility investment style relying on active trading in a small number of instruments and use of leverage, and the top 50 users by Sharpe ratio simply represent the "lucky few" who have recently obtained large gains with the method. In this case, a high Sharpe ratio alone might not be a very good proxy for investor skill, and should be combined with some performance measure that emphasizes consistently outperforming the relevant market index or average peer returns. However, due to the relatively short data period of this study compared to the typical stock market investment horizon, no proper analysis of factors linked to historical and future performance could be carried out.

Results from the Pagerank centrality analysis provide a view into the preferences of well-connected users, i.e. those whose followers themselves have many followers. While Pagerank centrality is naturally strongly correlated with the number of followers a user has, it is less tightly linked to belonging to the highest performing three-star category of users. The magnitude of the effect is not very large, but it is statistically significant and may reflect behavior where users with many followers put more weight on non-performance related criteria, such as investment style or portfolio composition, when choosing which users to follow. For example, it might be expected that the three-star category of users contains a disproportionate share of users trading actively in only one or two instruments, since well-balanced portfolios are less likely to produce extreme negative or positive returns. The users with most followers could be more cautious investors who are not interested in following portfolios with a small number of instruments or generating returns based on high turnover and use of leverage.

As with the number of followers, a large share of the variance in Pagerank centrality is not explained by the regression variables used. This, too, is a natural consequence of the lack of suitable variables for e.g. quantifying the quality of interactions within the Shareville network, and omitting any positive feedback effects in the level of connectedness of the user.

Previous studies on online investing networks have not examined in detail what factors influence connection formation within the network. Pan et al. (2012) find a weak relationship between the historical returns of a user and their rank by number of followers, mainly visible for the top 50 or top 100 users by number of followers (total size of user base not reported). They suggest that users tend to erroneously interpret the number of existing followers as a proxy for trading skill, even though historical returns are visible to all. Their observations are to some extent in line with the findings of this study: although the number of followers in Shareville is correlated with the 1-year Sharpe ranking, this correlation explains only a small portion of total variance in the number of followers, and e.g. active communication within the network, especially commenting on trades, appears to have a greater role. However, direct comparison of these results with Pan et al. (2012) is not possible due to differing data and methods.

Since Shareville content is primarily generated by the users within the constraints set by Shareville features, the observation that some user-generated content is poorly correlated with centrality in the network may signal that the corresponding features are underutilized, or are not perceived to add value. For example, it might be expected that sharing information such as valuation analyses on individual instruments would add to an individual’s influence and centrality in the network by attracting new followers. However, the fact that commenting on instruments has only very minor explanatory power towards the user’s centrality suggests that this feature is not very important for Shareville users. This interpretation is supported by Shareville user feedback (not published here) criticizing the low information content of many comments and lack of quality share analyses compared to other investing communities such as Seeking Alpha.

Since each additional feature tends to clutter the user interface and increases complexity of the service, the Shareville service must be continuously developed in order to either replace low value-add features with more useful ones, or identify means for increasing the value of the existing features. A guiding philosophy in Shareville development has been an emphasis on in-service recognition to encourage certain behavior, such as automatically granting a special token for users who comment on their own trades, instead of external rewards such as discounts on Nordnet trading commissions. One such improvement could be further gamification of Shareville with the help of badges or points, similar to those used in various crowdsourced online wiki projects, to recognize valuable individual contribution such as equity analysis.

As indicated above, the chosen regression analysis approach to identifying attributes valued by Shareville users when they decide which users to follow ignores many qualitative variables and path dependencies that may influence user behavior. However, even though some of these might be quantified as continuous or dummy regression variables, the interpretation of the resulting regression results would be increasingly ambiguous and would require other methods to corroborate them. The user attributes used in the regression in this study are ones that are both easily quantifiable, such as the number of various types of messages published, and represent content that is immediately visible to other users in the Shareville user interface, such as the star rating, so they provide information about the relative importance of the various features to Shareville users. A more in-depth analysis is beyond the scope of this thesis and should include multiple methods for measuring service usage and user experience, including qualitative ones such as surveys.

Shareville users form a directed network where each user decides which users to follow, and the followed user may or may not decide to reciprocate in kind. This study examines properties which attract followers in general, but it does not delve into the finer dynamics of the follower-followed relationship. In particular, potentially relevant areas for further investigation include the properties of the follower (i.e., what kind of users follow other users), the determinants of two-way connections (i.e., when is the followed user most likely to start following the follower), and predictors of connection formation between two specific users (i.e., why does a particular user choose to follow a particular user). Examining these questions would provide a richer picture of user behavior and network formation in Shareville, and would thus also aid Shareville development.

7.2 Study 2: Copy trading in Shareville

This is the first study to investigate the role of peer influence in an online investor network as a determinant of share trading behavior. The act of imitating another’s actions is a fundamental manifestation of peer influence, as witnessed by the popularity of copying other users’ trades in online currency trading networks. Although in the case of Shareville there is no way to conclusively determine which trades are made in imitation of a followed user, and individual users most likely base trading decisions on multiple sources of information, the results obtained for Hypothesis 2 provide strong evidence of copy trading behavior in Shareville and a quantitative estimate of its rate of occurrence. Adjusting for potential coincidental copying further improves the reliability of the results.

A central finding of this study is that copy trading is at least thus far a relatively minor driver of trading activity in Shareville. For one hundred buy-side trades by users with followers, an average of six copy trades are estimated to be generated. For comparison, Liu et al. (2014) report that approximately 68% of the forex trades in their data are generated by automatically or manually copying a trade by another user. The difference most likely stems from the nature of forex trading as a high-frequency, high-turnover activity largely based on intraday timing of trades according to currency price fluctuations, whereas most stock investors make relatively few trades, the investment horizon is long enough to mitigate the effect of daily market fluctuations, and trading decisions are typically based on the investor's personal opinion of the target company's future prospects. Instead of Shareville users blindly copying trades by other users, it seems more plausible that trades by well-connected users act as triggers or reminders for their followers to consider trading in the same instrument.

The small total number of copy trades also highlights the problem with reconstructing the investor network from closely timed stock exchange trades as done by Ozsoylev et al. (2014). In the Shareville data, where the actual connections between users are observed, the average share of followers who replicate a given trade within a time window of over one day is just 1.3%, and even smaller when accounting for coincidental replication. This means that reconstructing the network based on the observed replications would miss most of the actual connections, and more importantly well over 90% of trades by Shareville users are made independently, i.e. without evidence of copying another user. The network reconstruction by Ozsoylev et al. (2014) may be effective in identifying which traders are well-connected and able to take advantage of early access to information affecting stock prices, but such a network should not be interpreted to represent actual connections between traders, especially retail investors who trade infrequently.

As could be expected, a large share of the copy trades is generated by the followers of a few central users. Some investor networking platforms, such as eToro, provide monetary rewards to the most copied traders, since this encourages skilled traders to join and share their trades in the network. This makes the network more attractive to all users, and consequently boosts both the number of users and the average trading volume per user. However, monetary rewards also fundamentally alter the relationship between the user and service provider by blurring the line between spontaneously generated and sponsored content. This may result in other users questioning the motives behind individual comments and posts by central users, which in turn detracts from the value of the whole network as a platform for sharing and learning.

Users will return to Shareville only if they perceive they are getting value out of it, so any increase in an individual user's trading activity should come from making "better" trades than before and learning from past experiences as well as the actions of others. This tenet applies to both central users and their followers, and provides a guideline for developing new use cases for Shareville around the existing networking, trading, commenting, and messaging features. For example, the platform could automatically identify when two connected users make similar trades within a given timespan and offer them a chance to briefly discuss the motivation for the particular trade. When a user sells a financial instrument, Shareville could notify their followers who hold the same instrument. Such features might require an opt-out functionality for users with hundreds of followers or active traders who don't want to engage in discussions with others on a regular basis. On the other hand, most people seek confirmation of their beliefs from other like-minded individuals, so such features could at the same time be welcomed as valuable but lead to an echo chamber effect that hinders the actual learning process.

An interesting issue related to copy trading is the potential for creating market disruptions at the level of individual instruments. In theory, a central user in the network might buy a low-liquidity instrument, triggering a series of copy trades that would increase demand for the particular instrument and push the price up, at which point the central user could sell their recently purchased shares with profit. For example, daily trading volumes of some of the smaller companies in the Helsinki stock exchange are of the order of 10,000 EUR, and the number of trades is in the single or low double digits. On the other hand, some Shareville users have up to 2,000 followers, so a single trade by such a user could generate two or three copy trades with a high probability, and have a measurable impact on the daily trading volume for an illiquid share. While it is not certain that the average copy trade rate applies in the extreme case of the most connected users trading in low-volume instruments, the possibility for deliberate or unintentional price manipulation cannot be ruled out based on this study, especially as the Shareville community is still constantly growing and new connections are constantly formed. Further research is therefore needed to address this topic.

The approach taken in this study analyzes copy trading as the aggregate behavior of a user's followers. The results allow estimating macroscopic properties of copy trading, such as the number of individuals participating and trading volumes generated. This study does not analyze the factors influencing an individual user's decision to copy a particular trade, nor the probability that an individual follower will copy a particular trade. Also, since individuals are constrained in the amount of capital they can invest and by transaction costs from trading, copy trading volumes may to some extent cannibalize other trading activity, and the impact of this effect is not evaluated in this study.

From the point of view of the individual user, it would be useful to know whether copy trading generally leads to higher or lower returns than trades placed without direct influence from the network. The results from Pan et al. (2012) and Liu et al. (2014) would suggest that since individual users are generally not very good at identifying skill in others, and discretionary copy trades perform poorly compared to automatic mirroring of a successful trader's actions, copying the trades of other Shareville users may not be the best way to improve portfolio returns. However, as noted earlier, copy trades in Shareville are unlikely to represent just blind copying of another user's actions, and the user may close the position independently of the user whose actions were originally copied, so it could turn out that whether or not an individual trade is a copy trade has little discernible impact on the eventual outcome.

As Shareville user base expands and new connections are formed in the network, also the volume of social and copy trading may be expected to increase. A logical continuation of the present study would therefore be to investigate the propagation of behavior, investing styles, and investing strategies from one user to others in the network. Behavioral patterns of interest and relevance include, among others, the propensity to communicate with other users in the network, a focus on large, mid, or small cap companies, a focus on growth or value stocks, active vs. passive trading, and trading in non-domestic stock exchanges. Future studies may delve more deeply into these areas.

7.3 Study 3: Selective communication in Shareville

The selective communication hypothesis describes a basic human tendency to emphasize information that portrays oneself in a positive light when communicating with others, while withholding evidence of the opposite. It may also reflect unconscious cognitive biases, such as the confirmation and attribution biases introduced by Tversky and Kahneman (1974), so that the individual inadvertently focuses on those facts and events that support their self-image as a skilled investor, and furthermore attributes successful investment decisions to their own judgment instead of broader market movements or random chance.

Previous evidence for the selective communication hypothesis has been indirect, relying on e.g. geographical proximity and population aggregates of stock traders as presented by Kaustia and Knüpfer (2012), or on the individual trader level based on aggregate returns of forex traders, as presented by Heimer and Simon (2014). The latter study also gives a rough estimate of the event magnitude, reporting that moving from the 10^{th} to the 90^{th} return percentile increases a trader's communication propensity by approximately 25% of the average communication propensity.

This thesis presents the first effort to investigate selective communication in the stock market at the individual trade level. Statistically significant evidence of selective communication is found for Swedish but not for Finnish investors. The share of users commenting their own sales is 10-12% in the sample analyzed, depending on the country, and the magnitude of the selective communication effect in the Swedish investor population corresponds to an approximately 10-20% lower propensity for an investor to comment on their losing trades compared to winning trades. The observed magnitude of the selective communication effect is broadly comparable to that reported by Heimer and Simon (2014), especially considering the differences in study settings. Also, the effect is already seen clearly between slightly negative and slightly positive returns, as demonstrated in Figure 10. This is what would be expected if the relevant variable is the sign of the return, as suggested by the selective communication hypothesis, and not just the magnitude.

In a separate unreported robustness check, the analyses were repeated only for trades with returns between -20% and +20% to prevent extreme gains or losses from distorting the results. The results from the robustness check were nearly identical to the analysis done with the full data. In another unreported test, a regression analysis was carried out to determine whether the level of returns is significant in determining the magnitude of the selective communication effect. In the end, no statistically significant connection was identified, possibly because commenting propensity is apparently diminished considerably already for small losses (cf. Figure 10).

It is not clear why the Swedish trade data produces statistically significant results while the Finnish data does not. The returns were estimated using both Swedish and Finnish tax authority practices with virtually identical outcomes. However, there are approximately twice as many Swedish trades as there are Finnish trades in the data, and a selective communication effect of similar magnitude in the Finnish data as was seen in the Swedish data would correspond to a difference of just few tens of comments between losing and winning Finnish trades. Also, in the Finnish data of Figure 10 the observed distribution of commented and non-commented trades as a function of returns does bear some similarities to the expected distribution under the selective communication hypothesis, and when comparing loss and gain commenting propensities within the same individual, there is a non-significant difference in the same direction as in the Swedish data but of smaller magnitude. Therefore, it is possible that the influence of the selective communication effect in the Finnish data is simply masked by noise due to the smaller sample size.

Heimer and Simon (2014) suggest that selective communication propagates active trading strategies in online investor networks, since the extreme positive returns of some active traders increase communication between users, which in turn stimulates trading, while negative returns don’t have a similar effect. This hypothesis was not tested with the Shareville data, and implementing a test with sufficient statistical power might prove challenging, considering the small total number of comments in the data. However, considering that only a small fraction of Shareville trades are commented by the trader in the first place, and the difference in commenting frequency between winning and losing trades amounts to an imbalance of some tens of comments over the 6-month study period, any impact of selective communication on active trading strategy propagation in Shareville can be expected to be relatively minor so far.

It is plausible that individuals more prone to selective communication are also more susceptible to the disposition effect. Heimer (2014) presents evidence indicating that the magnitude of the disposition effect at the level of the individual user may be inversely correlated with the tendency of the user to communicate in an investor network. He explains this as a manifestation of social bargaining between investors of different experience levels, but future research could provide alternative explanations. For example, it should be investigated whether selective communication, which measures the investor’s readiness to admit their own mistakes to others, is correlated with the disposition effect, which relates to the investor’s readiness to consciously admit their own mistakes and treat gains and losses on equal terms.

If a connection between the disposition effect and selective communication were observed, various user-specific metrics could be introduced to services like Shareville to aid users in debiasing themselves and learning from their own mistakes as well as from other users. For example, useful information might include the number of trades made, how many of these incurred profits or losses, summary statistics on comments made on winning and losing trades, how long they hold instruments on average, and how they compare in these measures to other Shareville users. From the Shareville users’ point of view, better awareness of the selective communication effect might help individual investors to monitor their own trading behavior more objectively.

The methods used for investigating the selective communication hypothesis in this thesis are focused on individual trades and their comments. Further research would help in determining the wider implications of the observed results, such as the influence of the selective communication effect on copy trading and propagation of investment styles in Shareville. Also, it would be interesting to determine how the selective communication effect manifests itself on the individual user level: is the propensity to publish comments slightly smaller for losing trades for nearly all users, or is there a smaller subgroup of users who are very keen on commenting winning trades but never comment losing trades? The analysis presented here could be extended by combining the trade and comment data with user background data to reveal differences between users in their tendency to practice selective communication. This could help in identifying risk factors that predispose investors to biased messaging, and in identifying the negative consequences of selective communication for both the recipient and originator of the communication.

8 Conclusions

Online interactions form a growing share of information exchange between investors. Key features distinguishing online investor networks like Shareville from traditional interaction settings are user anonymity, extensive disclosure of portfolio contents, trades, and returns, and the possibility to contact, follow, and copy trades by any member of the community regardless of physical location or prior acquaintance. Providing access to an online investor network can allow an online brokerage to differentiate itself from competition, especially in regard to traditional banks. As the popularity of such services increases, there will undoubtedly be growing demand for features that can be shown to help users make better investment decisions.

This is the first study to observe and analyze user behavior in an online investor network focused on the stock market. Previous results from currency trading network studies are not generalizable to stock investor networks, since trading strategies and investor value creation are based on different premises in these two markets. This thesis therefore establishes a quantitative baseline of how retail stock market investors may be expected to interact and influence each other in an online network. The results will also hopefully aid in further development of Nordnet's Shareville platform, and provide retail investors a better understanding of their own behavior and that of others in the network.

Among the key findings of this thesis is the observation that historical portfolio performance and activity in the network are relevant but inadequate variables in explaining the number of followers a user has, suggesting that the "wisdom of the crowd" is also guided by other, more qualitative factors as well as chance when identifying which users are worth following in the network. The results also indicate that when evaluating other investors, users emphasize simplified and salient information in the user interface, to the extent that variables such as historical portfolio performance and Sharpe ratio may be condensed to a single four-level indicator (the Shareville star rating). Additional indicators may be therefore developed to facilitate identifying and following users with other attractive qualities beyond portfolio performance.

This thesis also provides evidence that Shareville users' investment decisions are sometimes influenced by the users whom they are following. The practice of copy trading is, however, much rarer than in forex trading networks, most likely reflecting other considerable differences between stock investing and forex trading. Since the probability to copy a trade is low for an individual user, the aggregate influence of a single trade on the rest of the investor network via copy trading is typically negligible, except in the case of the small minority of users who have hundreds or even thousands of followers. As the popularity of online investor networks grows, the influence of individual investors on the decisions of others can be expected to grow and provide opportunities to study also the propagation of stock market trading strategies.

Finally, research presented in this thesis shows that Shareville users are more willing to comment on trades resulting in gains than those resulting in losses, indicating that the knowledge that portfolios and trades can be openly viewed by other users does not preclude selective communication. The magnitude of the selective communication effect is even greater when comparing gain and loss commenting by the same user, and may amount to up to a 20% decrease in communicating propensity for trades where investors realize losses.

While statistically significant results on selective communication were obtained only for Swedish and not Finnish users, the evidence does not necessarily reflect cultural differences between the two nationalities, but rather the smaller size of the Finnish sample. Since the practice of commenting on one's own trades is not yet very widespread in Shareville, the influence of selective communication on e.g. the propagation of trading strategies is expected to be minor, but knowledge of the phenomenon may help in developing meaningful tools for all Shareville users to analyze their past trades and own performance more objectively, and become better investors in the process.

References

Barber, B. M. and Odean, T. (2013) "The Behavior of Individual Investors", in Constantnides, G. M., Harris, M., and Stulz, R. M. (eds.) *Handbook of the Economics of Finance*, Elsevier, 1533-1569.

Benartzi, S., and Thaler, R. H. (1995) "Myopic loss aversion and the equity premium puzzle", *The Quarterly Journal of Economics*, 110(1), 73-92.

Del Guercio, D., and Tkac, P. A. (2008) "Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow", *Journal of Financial and Quantitative Analysis*, 43(4), 907-936.

De Veaux, R. D., Velleman, P. F., and Bock., D.E. (2008) *Stats: data and models*, 2nd edition, Pearson Education, USA.

Dijk, O., Holmen, and M., Kirchler, M. (2014) "Rank matters – The impact of social competition on portfolio choice", *European Economic Review*, 66(1), 97-110.

Han, B., and Hirshleifer, D. A. (2015) "Self-Enhancing Transmission Bias and Active Investing", working paper.
Available online: http://ssrn.com/abstract=2032697 (Retrieved 7 Jun 2015).

Heimer, R. Z. (2014) "Peer Pressure: Does Social Interaction Explain the Disposition Effect?", working paper.
Available online: https://sites.google.com/a/brandeis.edu/rheimer/research (Retrieved 23 Feb 2015).

Heimer, R. Z., and Simon, D. (2014) "Facebook Finance: How Social Interaction Propagates Active Investing", working paper.
Available online: https://sites.google.com/a/brandeis.edu/rheimer/research (Retrieved 23 Feb 2015).

Hong, H., Kubik, J. D., and Stein, J. C. (2004) "Social Interaction and Stock-Market Participation", *The Journal of Finance*, 59(1), 137-163.

Kahneman, D., and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47(2), 263-291.

Kaustia, M., and Knüpfer, S. (2012) "Peer performance and stock market entry", *Journal of Financial Economics*, 104, 321-338.

König, F. (2013) "Soziale Interaktion auf Finanzmärkten", doctoral dissertation at Johann Wolfgang Goethe-Universität, Universätsbibliothek Frankfurt am Main, Germany.
Available online: http://publikationen.ub.uni-frankfurt.de/opus4/frontdoor/index/index/docId/29900 (Retrieved 28 Feb 2015).

Krasny, Y. (2013) "The effect of social status on decision-making and prices in financial networks", in Altshuler, Y., Elovici, Y., Cremers, A. B., Aharony, N., Pentland, A. (eds.), *Security and Privacy in Social Networks*, Springer, New York, 85–131.

Liu, Y., Nacher, J. C., Tomoshiro, O., Martino, M., and Altshuler, Y. (2014) "Prospect Theory for Online Financial Trading", *PLOS ONE*, 9(10), e109458.

Newman, M. E. J. (2003) "The structure and function of complex networks", *SIAM Review*, 45, 167-256.

Ozsoylev, H. N., Walden, J., Yavuz, M. D., and Bildik, R. (2014) "Investor networks in the stock market", *Review of Financial Studies*, 27(5), 1323-1366.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999) "The Pagerank citation ranking: Bringing order to the web", technical report, Stanford University, USA.

Pan, W., Altshuler, Y., and Pentland, A. (2012) "Decoding Social Influence and the Wisdom of the Crowd in Financial Trading Network", In 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, 203-209. Institute of Electrical and Electronics Engineers, USA.

Pool, V. K., Stoffman, N., and Yonker, S. E. (in press) "The People in Your Neighborhood: Social Interactions and Mutual Fund Portfolios", *The Journal of Finance*.

Shefrin, H., and Statman, M. (1985) "The disposition to sell winners too early and ride losers too long", *The Journal of Finance*, 40(3), 777-790.

Shiller, R. J. (2003) "From Efficient Markets Theory to Behavioral Finance", *Journal of Economic Perspectives*, 17(1), 83-104.

Thaler, R. H. (1999) "The End of Behavioral Finance", *Financial Analysts Journal*, 55(6), 12-17.

Tversky, A., and Kahneman, D. (1974) "Judgment under Uncertainty: Heuristics and Biases", *Science*, 185(4157), 1124-1131.