



# Trade synchronization and social ties in stock markets

Margarita Baltakienė<sup>1\*</sup> , Kęstutis Baltakys<sup>1</sup>  and Juho Kanninen<sup>1</sup> 

\*Correspondence:

[margarita.baltakiene@tuni.fi](mailto:margarita.baltakiene@tuni.fi)

<sup>1</sup>Unit of Computational Sciences,  
Tampere University,  
Korkeakoulunkatu 1, 33720  
Tampere, Finland

## Abstract

Previous studies suggest that individuals sharing similar characteristics establish stronger social relationships. This motivates us to examine what combinations of socioeconomic investor attributes are more likely to be associated with joint trading behavior. We use a unique data set on actual social ties between investors and find that similarities in investors' age, geographical location, or length of the co-employment can affect trade synchronization under certain circumstances. Our findings have implications for the analysis of investor networks.

**Keywords:** Investor network; Trade synchronization; Social network; Insider network; Network inference

## 1 Introduction

Information is a key resource in the stock markets. Informed investors trade securities based on their expectations of the dividend streams and discount rates [1–3]. When new information arrives, investors revise their expectations, and consequently, the stock price changes due to shocks in supply and demand. While public information is an important factor that shapes the market dynamics and is widely used in asset pricing with various factors [4, 5], private information also plays a vital role in stock markets [6]. The problem, however, is that, by definition, private information is not publicly observable, which makes its research challenging. Nevertheless, the diffusion of private information and its consequences have been investigated from different points of view with direct and indirect observations. There is evidence of a strong association between day traders' instant messaging and trade synchronization [7], inside information transfer and illegal trading [6], and investor's centrality in an investor network and its relation to profit-making [8, 9]. The strength of the investor network links and, at the same time, information channels are affected by investors' socioeconomic attributes, such as age, language, and geographical distance (see, e.g., [6, 10, 11]). Overall, the extant literature provides strong evidence on the existence and use of private information in stock markets (see also [12, 13]).

While there is research about investor trade synchronization (see, e.g., [14–18]), and, in turn, trade synchronization can signal information exchange [8], less is known about what socioeconomic attributes increase the likelihood of trade synchronization between investors. To this date, a significant limitation of the research on investor networks is

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

caused by the lack of ground truth observations about the actual social connections. We use a unique combination of an observable social network between company insiders and their complete trading history. The proxy for the social network is constructed from simultaneous co-employments. Because there are insiders who sit on the boards of multiple companies, we obtain a connected network of Finnish company insiders. Hereafter, we refer to investors with ties to multiple companies as *Connectors*. Importantly, even if we extract the social network links from the data on insiders' co-employment, we do not analyze insider trading per se.

In this paper, we analyze whether trade synchronization is associated with specific attributes of social ties. We hypothesize that attributes associated with stronger social connections are more likely to be observed for those *pairs of socially connected investors who are found to synchronize their transactions than for those who trade differently*. Particularly, we test whether socially connected investors who synchronize their trades (i) are more similar in age, (ii) are located more closely to each other geographically, and (iii) have worked longer together than the pairs who do not trade synchronously.

Similarly, to understand the effect of socioeconomic attributes for a pair of disconnected investors who synchronize their trades, we test whether socially connected investor pairs on the shortest paths between them ( $i^*$ ) are more similar in age, ( $ii^*$ ) are located more closely to each other geographically, and ( $iii^*$ ) have worked together longer than other socially connected investors.<sup>1</sup>

Our research question is related to the concept of homophily, which is defined as a tendency for individuals to establish and prefer connections to others similar to them [19, 20]. Social network studies observe, among other characteristics, stronger social ties for people with similar age, profession, and geographical proximity. For example, Fischer [21], Marsden [22], and McPherson et al. [19] find that interactions in relationships with a stronger age homophily tend to be more personal, longer lived, and frequent. Geographical proximity is a strong predictor on how often friends get together to socialize [23]. Proximity induces social interactions as it is easier to connect, socialize and maintain relationship with the local contacts [19, 23–25]. Spatial proximity enables spreading of peer influence [26]. Furthermore, ties that are involved into regular business interactions survive much longer than ties of individuals not sharing the same professional role title [27]. In addition, prior co-employment facilitates social tie formation [28], persistence and decay [29]. Longer lasting relationships in work environment induce friendships and information exchange, because people with the same education background and occupation are more likely to confide in each other [19, 27]. Moreover, Barone and Coscia [30] find that businesses tend to connect with similar businesses, and interestingly, business partnerships exhibit tax fraud homophily.

We focus our analysis on the insiders of 150 companies listed in the Helsinki Stock Exchange (HSE). We combine (a) historical data about company insiderships, (b) insiders' mandatory notification of trades, and (c) an investor-level transaction data with complete trading history of all investors who traded in HSE. Typically, in the existing empirical research, real investor social relationships are neither observable nor can be determined

---

<sup>1</sup>We investigate the same set of hypotheses for different subsets of relationships: 1) between investors who synchronize their transactions and are directly connected in our social network proxy, 2) investors on the shortest paths between synchronizing individuals who are not directly connected, and two of the three hypotheses for 3) investor pairs who synchronize their transactions and differentiate the test and reference sets by whether they are socially connected. We use none, one, or two asterisks to differentiate the same hypotheses tested on different subsets of relationships.

from the trading data. Similarly to Wong et al. [31] and McEvily et al. [32], we use information about simultaneous insiderships to derive our proxy for the investor social network. By combining data sets (b) and (c), we gain access to insiders' full trading history on *all* the securities they traded in the HSE. Our research relies on a two layer network:

- An observable proxy for a social network between 6318 insiders who share insider positions in the same company. Insiders of different companies are linked through *Connectors*, who hold insider positions in multiple companies.
- An investor network with 1756 nodes, where a link between a pair of investors is based on the level of trade timing similarity, as introduced in Tumminello et al. [33].

Our research contributes to understanding the effect of socioeconomic attributes on the financial decision-making in investor networks. Differently from existing studies on information transfer in insider networks (see, e.g., [6]), our study broadens the investigation of private information in financial markets, not limiting the analysis to insider trading. We do this at a cost by making some strong assumptions. First, we assume that the network derived from the co-employment records is a valid proxy for the investor social network. Second, when interpreting the results from the point of view of private information diffusion, we consider the trade synchronization network as a proxy for the information network [8]. To the best of our knowledge, this paper is the first to analyze a trade synchronization network inferred from shareholder registration data together with an observable proxy of a social network.

## 2 Methods

With an increasing interest in network science applications in the financial domain (see, e.g., [34–36]), several techniques have been introduced to identify trade and portfolio similarities between investors [8, 9, 16, 33, 37]. Among  $O(N^2)$  investor pairs, one would want to focus only on meaningful similarities. To filter out spurious trade co-occurrences, the links are usually statistically validated against a chosen null model with a selected statistical significance threshold.

In this paper, we use Statistically Validated Network (SVN) method [14] that identifies non-random trade co-occurrences using hypergeometric distribution. Applying SVN, we project the bipartite system of investors connected to their trading days into a monopartite investor network. A link in this network represents investor trade co-occurrence. The statistical significance of trade co-occurrence between investors is estimated by leveraging the hypergeometric test. The set of statistically significant trade co-occurrences yields the trade synchronization network. Though this method does not allow to determine the direction of influence between two investors, it is sufficient for our analysis purposes. The method is presented below in more detail.

First, for each investor  $i$  and her/his traded security  $k$  on the trading day  $t$ , we calculate the scaled net-volume as

$$v_{ikt} = \frac{V_{ikt}^b - V_{ikt}^s}{V_{ikt}^b + V_{ikt}^s}, \quad (1)$$

where  $V_{ikt}^b$  and  $V_{ikt}^s$  are the total daily buy and sell volumes observed from the shareholder registration data. Next, investors' buy and sell trading states are defined using the scaled

net-volume as follows:

$$\begin{cases} b - \text{buying state, when } v_{ikt} > \theta, \\ s - \text{selling state, when } v_{ikt} < -\theta, \end{cases} \quad (2)$$

where  $\theta > 0$ . In this paper, we use  $\theta = 0.10$ . Net scaled volumes that fall between  $-\theta$  and  $\theta$  indicate a day-trading pattern, for which we do not assign a trading state. Using the hypergeometric test we assume that trading days are homogeneous in terms of investor trading activity. Similarly to [14], we find that the distribution of trading state occurrences in different trading days has a bell shape on the log-scale with a fluctuation of approximately one decade around the mean. We conclude that the limited heterogeneity should not undermine the use of the hypergeometric null model.

The null hypothesis is that the trading state co-occurrences for two investors are random. We define  $T_{ijk}$  as the length of the joint trading period of security  $k$  for investors  $i$  and  $j$ . The probability of observing  $X$  co-occurrences in  $T_{ijk}$  observations is estimated by the hypergeometric distribution, i.e.,

$$H(X | T_{ijk}, N_{ik}^P, N_{jk}^P) = \frac{\binom{N_{ik}^P}{X} \binom{T_{ijk} - N_{ik}^P}{N_{jk}^P - X}}{\binom{T_{ijk}}{N_{jk}^P}}. \quad (3)$$

Here,  $P$  is one of the trading states, i.e.,  $P \in \{b, s\}$ , and  $N_{ik}^P$  is the number of days investor  $i$  was in the trading state  $P$  for the security  $k$ . In turn, the probability of having at least  $N_{ijk}^P$  trading state co-occurrences by chance is calculated as follows:

$$p(N_{ijk}^P) = \text{Prob}(Y \geq N_{ijk}^P) = 1 - \sum_{X=0}^{N_{ijk}^P-1} H(X | T_{ijk}, N_{ik}^P, N_{jk}^P). \quad (4)$$

Using Eq. (4), we estimate  $p$ -values for our one-sided tests,  $p_{ijk}^P := p(N_{ijk}^P)$ , for all pairs of investors who traded at least once on the same day. We separately calculate and retain the minimum  $p$ -value for each unique pair of investors across multiple trading states (buy and sell sides) and securities, i.e.

$$p_{ij} = \min(p_{ijk_1}^b, p_{ijk_1}^s, p_{ijk_2}^b, p_{ijk_2}^s, \dots), \quad (5)$$

where security  $k_{(\cdot)}$  does not belong to a company where investor  $i$  or investor  $j$  is an insider.<sup>2</sup> The link is validated for investors  $i$  and  $j$ , i.e., investors have synchronized trading, if  $p_{ij} < \alpha$ , where  $\alpha$  is a chosen statistical significance threshold. To investigate the sensitivity of our results, we vary  $\alpha$  thresholds. Optionally, this procedure is combined with the multi-test correction (MTC) to reduce the type I errors. We present some results with Bonferroni multiple test correction [38].

<sup>2</sup>Insiders' transactions in their own company shares can be synchronized because of company policies or regulatory requirements on insider trading. For example, the insiders of a company must ensure that they execute trades outside of a closed period of 30 days before the announcement of an interim financial report or a year-end report, and consequently, insiders' transactions can be clustered. Even if we find synchronization in own company shares between insiders we could not exclude the possibility of the investment being synchronized due to technical reasons. Moreover, insiders know that they are closely monitored by financial supervisory agencies when they trade the securities of their own companies. For these reasons, we ignore the data about trades in the companies where both investors are insiders.

### 3 Data

#### 3.1 Data sets

This paper introduces a new worldwide-unique data set of an insider network representing insiders and related parties from 153<sup>3</sup> publicly listed companies in the HSE coupled with complete histories of their trading behavior for 22 years.<sup>4</sup> The list of insiders<sup>5</sup> contains all persons who have access to inside information<sup>6</sup> and who are working for the issuer under a contract of employment or otherwise performing tasks which give them access to inside information. Among other, the insider list includes members of the board of directors, senior company executives, and other employees with access to insider information. Related parties include spouses, other relatives, minors under guardianship, and controlled or influenced companies. Data set is comprised of three information sources:<sup>7</sup>

- (i) Shareholder registration data (SRD) from all the individual Finnish investors and companies. This includes data for more than 1.5 million investors, provided by Euroclear Finland. The data contains information about transaction date, price, volume, security ID, the postal code, year of birth, and gender of the investor. These data have been extensively used in the literature over two decades [9, 16, 40, 41].
- (ii) Insiders' mandatory notifications of trade in the HSE from 2005 to 2010 and from 2013 to 2018 for 153 companies. The data include information about family members and trading companies associated with the insiders. This data set is provided by Euroclear Finland's insider register service (SIRE).
- (iii) Insiders' assignments and positions in Finnish companies obtained from Finnish Patent and Registration Office (Virre<sup>8</sup>). This data is used for reconstructing the insider network.

More details about the three data sources can be found in Appendix A. The time spans for the collected data sets and the empirical analysis period of this paper are shown in Fig. 1. By matching data sets (i) and (ii), we have been able to track *all transactions* over *all the securities* made by a part of the insiders in the HSE. The matching procedure is described in more detail in Sect. 3.2.

#### 3.2 Data preprocessing

First, we construct the observable insider network by combining both Virre and SIRE data sets. The social insider network is a dynamic network, constructed on a daily basis. A pair of insiders have a link on the day  $t$  if they both were insiders in the same company on this day. Over the union of all daily network snapshots the network contains 13,932 nodes that represent insiders (see Fig. 2 and Table 1).

The history of co-employment establishes a potential social relationship that can be used to share information. Because not all co-employment relationships are equivalent,

---

<sup>3</sup>151 unique companies, two of them changed their names.

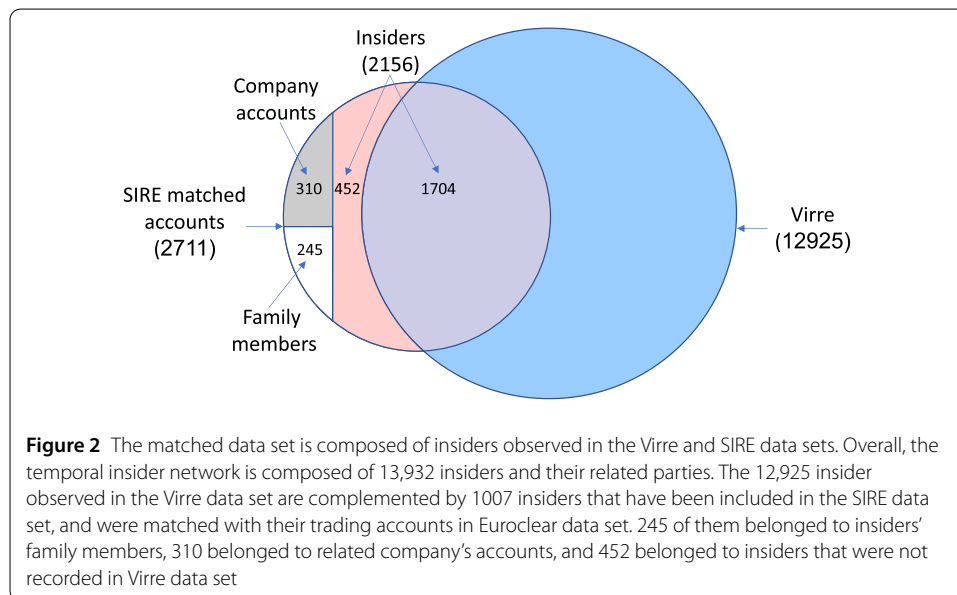
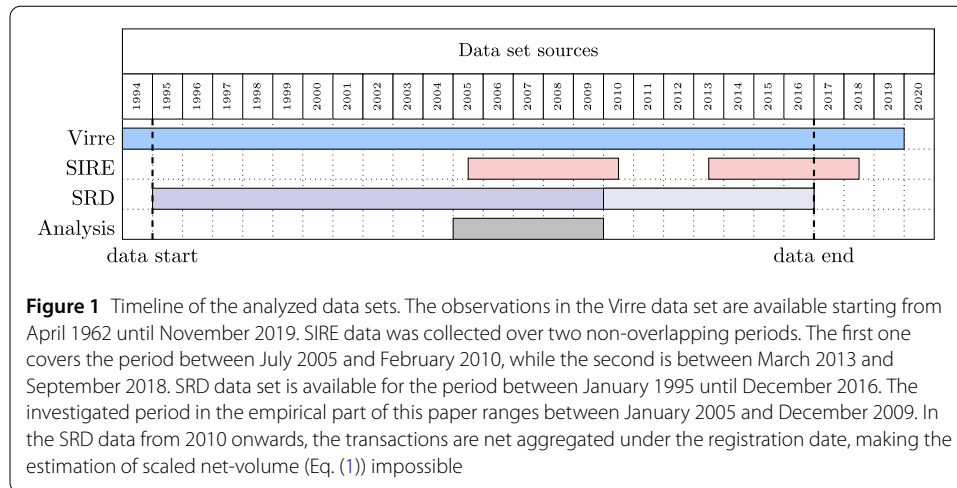
<sup>4</sup>Our complete data set includes double the information about insider trades used in Berkman et al. [39]. Because of this we were able to match a significantly larger number of insider accounts.

<sup>5</sup>Defined in Sect. 3 of Chap. 12 in Securities Markets Act.

<sup>6</sup>Inside information is defined in Article 7(1)(a) of the Market Abuse Regulation, <https://www.finanssivalvonta.fi/en/regulation/regulatory-framework/market-abuse-regulation/inside-information/>

<sup>7</sup>In our analysis we use a subset of this data set. In particular, we investigate the window between 2005 and 2009, see Sect. 3.3.

<sup>8</sup>See <https://www.prh.fi/en/kaupparekisteri/tietopalvelut/virre.html>



**Table 1** Summary statistics for insider data over all the years from April 1962 and November 2019 and the analyzed period between January 2005 and December 2009

	Total period (1962–2019)	Analyzed period (2005–2009)
# companies	153	150
# investors	13,932	6318
# investors who reported transactions	3514	1892
# matched accounts	2711	1756
# matched insiders	2156	1487
# matched family members	245	138
# matched related parties	310	131

our analysis is specifically focused on determining which attributes of the co-employment relationships are more likely to be associated with trade synchronization.

Second, to unveil the *entire trading history in the insiders' accounts over all the securities*, we match reported insiders' trades in SIRE data with the anonymized trading data in the SRD data set. The matching is done by looking up the reported insider's own company

trades in the SRD data set based on the trading date, trade direction, traded volume, traded security ID, the initial and closing balance, and the birth year of the account holder. In some cases, there were no trades executed and reported during the five year period, but we could match the accounts by the balance of the shareholdings reported in the SIRE data set. Once an insider is matched with his/her account in the SRD data set, we obtain access to information about all his/her transactions, i.e., the transactions in own company and all other transactions executed in HSE between 1995 and 2016. Hereafter, we will refer to the insiders with matched trading accounts as *matched insiders*.

The ideal match is when all mandatory trade notifications lead to one anonymized owner ID in the SRD data set. In some cases, even though some transactions in the SRD or the mandatory notifications were missing, we could uniquely match accounts based on account balances. Moreover, in some cases, to confirm an account match when not all trades or account balances have been matched, we used information about insiders' year of birth, gender, geographical location, nationality, and language to decide the account match.

We are not able to match insiders with their trading accounts either because the investors (1) do not trade at all and therefore do not have trading accounts; (2) they trade but have not traded companies where they need to make mandatory trade notifications, which makes their trading account identification impossible; or (3) we are not able to uniquely match their accounts. The latter can be related to identical trades in terms of volume, price, and direction executed by multiple investors on the same day. If neither the account balances nor the socioeconomic attributes help to identify the account holder, we cannot match the account without observing additional transactions.

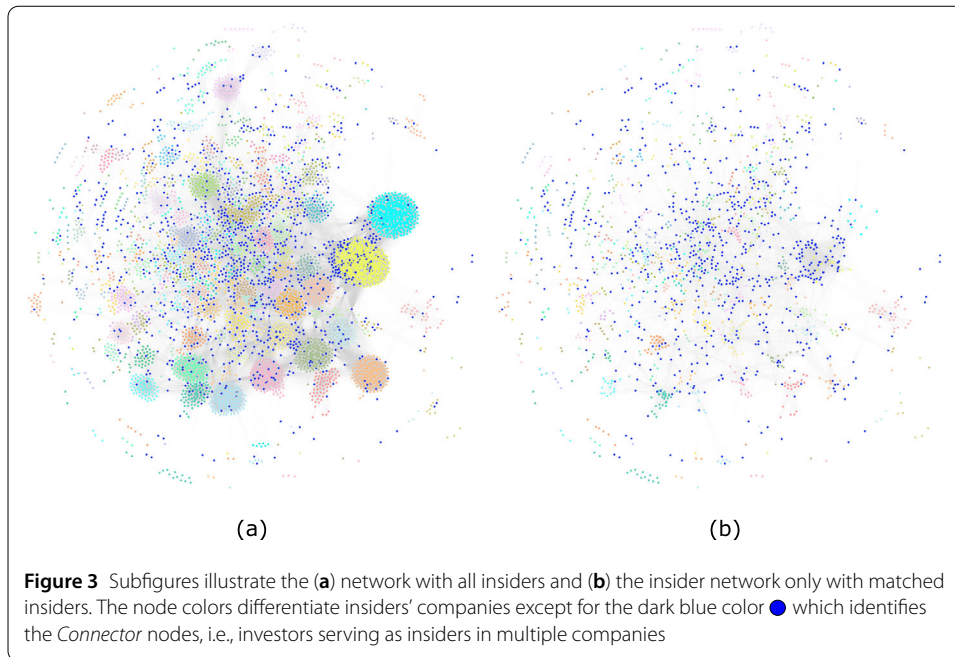
Out of 3514 insiders who have reported their transactions, we were able to match 2711 with their trading accounts in SRD data (a success ratio of 77%). The matched accounts include 2156 insiders, 245 family member accounts, and 310 accounts of other related third parties. The complementary cumulative distribution function of the ratio between the matched insiders and all insiders in a company is shown in Fig. 8 for all 153 companies. Appendix B provides more detailed descriptive statistics about the data set.

### 3.3 Analyzed data

In the empirical part, we limit our analysis between 2005 to 2009 for two reasons. First, SIRE data are not available prior to 2005. Second, from 2010 onwards the transactions in SRD data are net aggregated under the registration date. This makes the estimation of scaled net-volume (Eq. (1)) impossible after 2010. During this period, 6402 individuals were part of the insider network, composed of eight disconnected components. The largest connected component contains 6318 investors, while the size of the second-largest component is only 20 investors. The size of the remaining six components varies from 18 to two investors.

We focus our analysis on the largest connected component of the network, which is composed of insiders from 150 companies. The social investor network is defined as  $\mathcal{G} = (\mathcal{I}, \text{CON})$ , where  $\mathcal{I}$  is the set of investors, and CON is the set of co-employment connections. We denote a link between investors  $i \in \mathcal{I}$  and  $j \in \mathcal{I}$  as  $(i, j)$ , and say that they are linked in the social network, i.e.,  $(i, j) \in \text{CON}$  if both of them have been insiders in the same company at the same time. In total, there are  $|\mathcal{I}| = 6318$  insiders and  $|\text{CON}| = 500,766$  connections in the analyzed social network. The average degree  $\langle k \rangle = 158.52$ , and the net-





work's diameter  $d = 8$ .<sup>9</sup> The distribution of the number of insiders per company is shown in Fig. 14. We report the summary statistics for the analyzed network in the Appendix C.

Out of  $|\mathcal{I}| = 6318$  investors (nodes),  $|\mathcal{I}^m| = 1756$  are matched with their transactions in the SRD data over all securities. Even if the majority of the investors are not matched with the shareholder registration data, they are important for our analysis. Note that unmatched investors can trade both their own and other company shares. Insider network with all insiders and the network with only matched insiders are shown in Fig. 3 (a) and Fig. 3 (b).

## 4 Results

### 4.1 Trade synchronization and social distance

In this section, we investigate the association between trade synchronization and social distance for matched insiders. We start by retrieving investor trade synchronization network applying the SVN method [33] for a set of 30 most traded securities among the investors in the analyzed network (see Appendix Table 7). Here, we do not validate the links. Instead, we define the score that measures the level of synchronization for investors  $i$  and  $j$ ,

$$\text{SCORE}_{ij} = 1 - p_{ij}, \quad (6)$$

where  $p_{ij}$  is the minimum  $p$ -value obtained with Eq. (5). For investor pairs that have no overlapping trades we set  $\text{SCORE}_{ij} = 0$ .

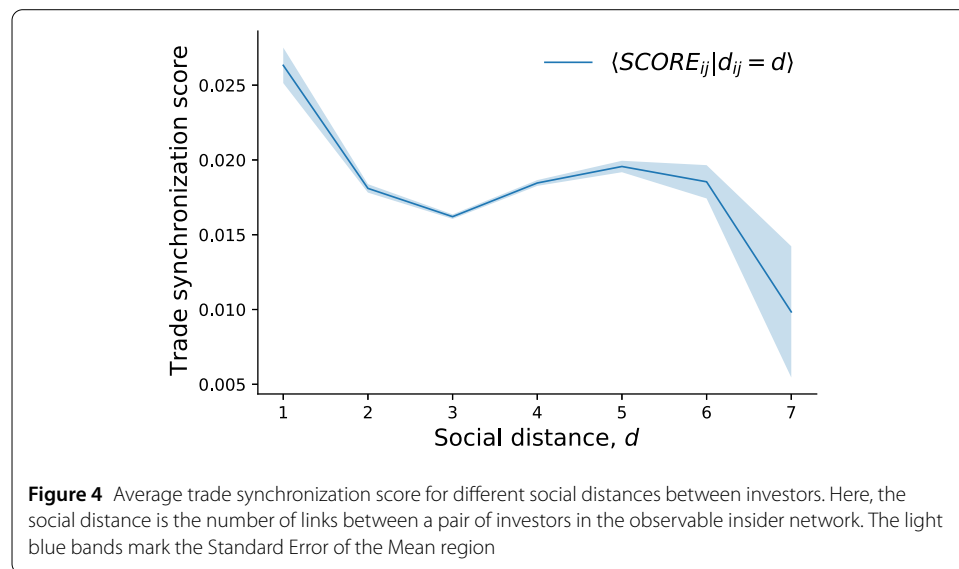
Generally, a  $p$ -value is a probability of having empirical observations under randomness –  $P(\text{data}|\text{null model})$ . We can say that the lower (higher) probability, the less (better) the random model explains an observed overlap between the transactions, and the more (less) abnormal the observed overlap is. From this point of view, the  $p$ -value, can be used

<sup>9</sup>The diameter is defined as the smallest number of hops needed to move between two nodes that are the farthest apart in the network.



to measure the strength of synchronization. Particularly,  $SCORE_{ij}$  is defined by the probability that the observed synchronization is not generated by the random null model, which assumes independence between traders.

Next, we calculate the social distances  $d_{ij}$  using our proxy for the investor social network. The social distance between two investors is defined as the length of the shortest path between them. The two most distant matched insiders are seven links apart.<sup>10</sup> Finally, we calculate  $\langle SCORE_{ij} | d_{ij} \rangle$  which defines the average trade synchronization score for investor pairs separated by  $d_{ij}$  social connections. Figure 4 shows that the highest scores of trade synchronization are between the directly connected nodes. There is a clear decay of the scores as the distance increases, suggesting, that synchronized trading may actually be associated with social proximity. We evaluate the significance in the decrease of the score with one-side independent two-sample test with unequal variance and sample sizes (Welch's  $t$ -test). The decreasing average score is statistically significant comparing scores at social distances 1 and 2, 2 and 3, and 6 and 7 (see Table 2). However, we observe a local peak at the social distance five. It can imply that information travels via roughly five links before both nodes at the ends of the chain act on it. Intuitively, in case of illegal information exchange, investors close to the source of information may be wary of acting



**Figure 4** Average trade synchronization score for different social distances between investors. Here, the social distance is the number of links between a pair of investors in the observable insider network. The light blue bands mark the Standard Error of the Mean region

**Table 2** One-sided independent two-sample Welch's  $t$ -test on the difference in trade synchronization score for pairwise consecutive social distance. Here,  $D_k$  is the set of the synchronization scores between investors at the distance  $k$

$l$ and $m$	avg. $D_l$	avg. $D_m$	$p$		$ D_l $	$ D_m $	$H_A$
1 and 2	0.02633	0.01810	7.59E-12	***	18,027	219,572	$D_1 > D_2$
2 and 3	0.01810	0.01621	1.67E-09	***	219,572	686,930	$D_1 > D_2$
3 and 4	0.01621	0.01846	1		686,930	473,756	$D_1 > D_2$
4 and 5	0.01846	0.01956	9.95E-01		473,756	127,752	$D_1 > D_2$
5 and 6	0.01956	0.01854	1.92E-01		127,752	14,355	$D_1 > D_2$
6 and 7	0.01854	0.00984	2.75E-02	*	14,355	498	$D_1 > D_2$

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$

<sup>10</sup>The shortest path between matched insiders can go via unmatched insiders.

on it due to the higher probability of being monitored for the abuse of insider information. For example, Ahern [6] finds that buy-side managers and analysts act as the tippers as they receive information in the fourth and later links. The results are robust with the Jaccard coefficient as the trade synchronization score, see Table 8 and Fig. 17 in Appendix F.

#### 4.2 Are investors more similar to their neighbors with whom they synchronize their trades?

In this section, we start to analyze the effects of investor socioeconomic attributes on the trade synchronization between socially connected investors. Here, we aim to understand why investors trade more similarly with some of their social connections. The underlying hypothesis is that investors are more similar to their social contacts with whom they synchronize their transactions. We investigate the strength of a tie between investors through three attributes: age, postal code, and investor's insidership periods. In particular, we hypothesize that a stronger similarity of investor attributes such as (i) smaller age difference, (ii) closer geographical proximity, and (iii) longer time of joint co-employment in the same company (co-insidership) facilitates trade synchronization.

To test Hypotheses (i)–(iii), we conduct one-side Welch's  $t$ -tests. We create an experiment set and a reference set of investor pairwise relationships. We remind that pairs of investors linked in the social network  $\mathcal{G}$  belong to the set CON. Social connections between investors with matched accounts are defined as  $\text{CON}^m = \{(i, j) : (\exists i, j \in \mathcal{I}^m)[(i, j) \in \text{CON}]\}$ . Next, we denote pairs of investors who synchronize their trades as SYNC and define it as

$$\text{SYNC} = \{(i, j) \mid p_{ij} < \alpha \cap i, j \in \mathcal{I}^m\}, \quad (7)$$

where  $\alpha$  is a chosen statistical significance threshold and  $p_{ij}$  is the  $p$ -value obtained with Eq. (5). Note that investors who synchronize their trades do not necessarily have a direct link between them. However, as we are analyzing a connected network, there is a path between each pair of investors.

The experiment set is the set of investor pairs who are both socially connected and synchronize their trades, defined as

$$\text{CON-SYNC} = \text{CON} \cap \text{SYNC} = \{(i, j) \mid (i, j) \in \text{CON} \cap (i, j) \in \text{SYNC}\}. \quad (8)$$

We find a relatively small number of social links that exhibit trade synchronization. For example, there are  $|\text{CON-SYNC}| = 336$  social links out of  $|\text{CON}^m| = 18,027$  social links between investors with matched accounts. This observation is, in fact, inline with the social networks literature. Specifically, it has been observed that usually the interaction with neighbors in a social network is low [42] and only a few social links are actively exploited for communication [43].

To understand whether there is a qualitative difference between the social connections over which investors synchronize and do not synchronize their trades, we define the reference set of links as

$$\text{CON-NSYNC} = \{(i, j) \mid (\exists k \in \mathcal{I}^m)[(i, k) \in \text{CON-SYNC} \cap (i, j) \in \text{CON} \setminus \text{SYNC}]\}. \quad (9)$$

The reference set CON-NSYNC includes pairs of investors where one of the investors synchronizes trades with at least one social connection and the second investor is socially

**Table 3** One-sided Welch's  $t$ -test on the difference in age, length of co-insidership and geographical distance between CON-NSYNC (Eq. (9)) and CON-SYNC (Eq. (8)) sets (Panel A), and CON-NSYNC\* (Eq. (10)) and CON-SYNC sets (Panel B). Trade synchronization between investors is validated at the significance level of  $\alpha = 0.01$ . Here,  $H_A$  is the direction of the alternative hypothesis. The units for the co-insidership are trading days, age difference in years, and geographical distance in kilometers

Panel A						
	avg. CON-NSYNC	avg. CON-SYNC	$p$	CON-NSYNC	CON-SYNC	$H_A$
Age difference	9.40	9.42	5.19E-01	35,271	312	CON-NSYNC $>$ CON-SYNC
Geographical distance	90.50	63.49	2.01E-05 ***	28,309	323	CON-NSYNC $>$ CON-SYNC
Co-insidership	513.90	642.52	8.95E-08 ***	35,546	336	CON-NSYNC $<$ CON-SYNC
Panel B						
	avg. CON-NSYNC*	avg. CON-SYNC	$p$	CON-NSYNC*	CON-SYNC	$H_A$
Age difference	9.50	9.42	4.36E-01	7422	312	CON-NSYNC* $>$ CON-SYNC
Geographical distance	67.52	63.49	2.70E-01	6610	323	CON-NSYNC* $>$ CON-SYNC
Co-insidership	536.14	642.52	9.00E-06 ***	7694	336	CON-NSYNC* $<$ CON-SYNC

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$

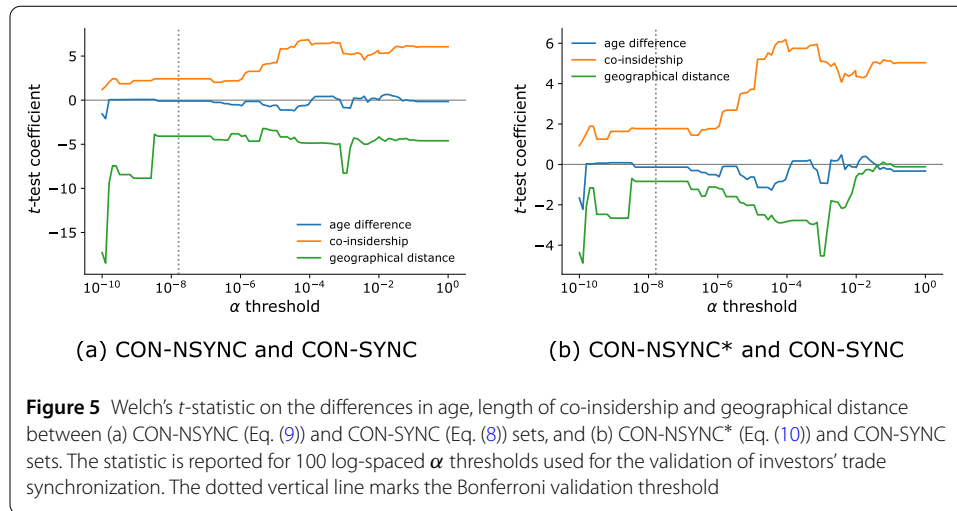
connected to the said investor but does not synchronize trades with him/her. For the illustration of the test and reference sets, see Fig. 16 (b) in Appendix E.

Table 3 (Panel A) summarizes the results of one-sided Welch's  $t$ -tests when the statistical link validation threshold  $\alpha = 0.01$ . We do not find evidence to support the Hypothesis (i), as the age difference is statistically insignificant comparing CON-SYNC and CON-NSYNC, but we provide evidence for Hypotheses (ii) and (iii). Particularly, the geographical distance is significantly shorter in CON-SYNC than in CON-NSYNC. Therefore, we reject the null hypothesis in favor of the alternative (Hypothesis (ii)). Thus socially connected insiders are more likely to synchronize trading if their geographical distance is shorter. This observation is inline with the findings in Baltakys et al. [11], that investors located closer to each other tend to synchronize their trades more than those who are farther away. The novelty of our results is that we establish this association between individuals with observable social connections. Moreover, in favor of the alternative Hypothesis (iii) we find that the length of the co-insidership is significantly longer for relationships in the set CON-SYNC than CON-NSYNC. In other words, the longer the work relationship is, the more likely it is for the colleagues to synchronize their trading.

For the robustness check we restrict the set of links by considering only those investors who have their trading accounts matched, i.e.,

$$\text{CON-NSYNC}^* = \{(i, j) \mid i \in \mathcal{I}^{\text{CON-SYNC}} \cap j \in \mathcal{I}^m \cap (i, j) \in \text{CON} \setminus \text{SYNC}\}. \quad (10)$$

We repeat the Welch's  $t$ -tests using CON-NSYNC\* as a reference set, see Table 3 (Panel B). The geographical distances are no longer statistically significantly shorter than in the reference set CON-NSYNC\*.



To see if the findings are generally true with various link validation threshold levels  $\alpha$ , we run another robustness check. We take 100 threshold levels  $\alpha$  equally spaced on a log scale between  $10^{-10}$  and 1. At each threshold level  $\alpha$ , a pair of investors  $i, j$  is said to be synchronizing their trades  $((i, j) \in \text{SYNC})$ , if the  $p$ -value obtained with the hypergeometric test (Eq. (4)) is less than  $p_{ij} < \alpha$ . The results are reported in Fig. 5 (a) and (b) for the reference sets CON-NSYNC and CON-NSYNC\*, respectively. We observe that across different validation thresholds the socially connected and synchronized investor pairs (CON-SYNC) exhibit stronger similarity of geographical distance (Hypothesis (ii)) and the length of co-insidership (Hypothesis (iii)). Figure 5 (b) shows that, in fact, the results for the geographical distance in general support Hypothesis (ii), even if results with higher threshold levels, such as  $\alpha = 0.01$ , are insignificant. At the same time, this analysis confirms that there is no statistically significant difference in terms of the age difference whether the socially connected investors synchronize their transactions or not. Overall, we provide strong evidence on Hypothesis (iii), partial evidence on Hypothesis (ii), and no evidence on Hypothesis (i).

### 4.3 Similarity on the shortest paths between synchronized investors

We continue our analysis of the effects of investor attributes' on trade synchronization. Here, our focus is on the shortest paths between investors who synchronize their trades but are not directly connected in the social network. This synchronization may partially result from information exchange and be driven by intermediate investors, i.e. *Connectors*. However, we do not expect investors to act on information each time they receive it. In fact, some of the private information may be classified, therefore acting on it would be illegal. Ahern [6] finds that on average, inside tips travel along three links in the network before they are acted upon. In this case, instead of acting on an inside tip, an investor may choose to pass it further along in the network, possibly to win favor with other investors. Therefore, it is plausible that *Connectors* can enable trade synchronization even if they do not trade themselves. On average, there are 3.3 hops in the shortest paths between the insiders with trading similarities who do not have a direct social link in our network.<sup>11</sup>

<sup>11</sup>Here we used  $\alpha = 0.01$  to validate the synchronization between investors.

Interestingly, Ahern [6] finds that on average, inside tips travel along three links in the network before they are acted upon.

In this section, our hypothesis is that trade synchronization of two indirectly linked investors is more likely to occur with a stronger similarity of attributes between *Connectors*. In particular, we expect that the social connections on the paths between investors who synchronize their trades are associated with (i\*) smaller age differences, (ii\*) closer geographical proximity, and (iii\*) longer time of joint co-insidership.

To test Hypotheses (i\*)–(iii\*), we conduct one-side Welch's  $t$ -tests. First, we denote the set of investor pairs who synchronize their trades but are not directly connected in the social network as  $\text{NCON-SYNC} = \text{SYNC} \setminus \text{CON}$ . The length of the shortest path between investors  $i, j \in \text{NCON-SYNC}$  is denoted as  $d_{ij}$ . The set of all the shortest paths between investors  $i$  and  $j$  is defined as

$$\text{PATH}_{ij} = \left\{ (k_0, k_1), \dots, (k_{d_{ij}-1}, k_{d_{ij}}) \mid k_0 = i \cap k_{d_{ij}} = j \cap (k_l, k_{l+1}) \in \text{CON} \right\}, \quad (11)$$

where  $0 \leq l \leq d_{ij} - 1$ . Then the set of all links on the shortest paths between investors  $i$  and  $j$  is defined as

$$\mathbf{P}_{ij} = \left\{ (k, l) \in \text{CON} \mid (\exists \text{path} \in \text{PATH}_{ij}) [(k, l) \in \text{path}] \right\}, \quad (12)$$

Our experiment set includes all the links on shortest paths between investors who synchronize their trades but are not directly connected, defined as

$$\text{PATH-SYNC} = \left\{ (k, l) \mid (\exists (i, j) \in \text{NCON-SYNC}) [(k, l) \in \mathbf{P}_{ij}] \right\}. \quad (13)$$

The reference set includes all remaining social connections except links between investors who are directly socially connected and synchronize trades,

$$\text{NPATH-SYNC} = \text{CON} \setminus (\text{PATH-SYNC} \cup \text{CON-SYNC}). \quad (14)$$

For the illustration of the test and reference sets, see Fig. 16 (c) in the Appendix E.

Table 4 (Panel A) summarizes the results of one-sided Welch's  $t$ -tests when the statistical link validation threshold  $\alpha = 0.01$ . Note that differently from the experiment described in the previous section, here we investigate social connections on the shortest paths between investors who are not directly socially connected. We provide evidence for Hypotheses (i\*) and (ii\*) that investor pairs with smaller age differences and shorter geographical distances are more likely to be located between investors who synchronize their trades. This result may signal a favorable setting for information transfer. At the same time, we find no evidence in support of Hypothesis (iii\*). In fact, if the one-sided Welch's test on the length of co-insidership were formulated in the other way around, the results would be highly significant. Given that trade synchronization results from information flows, this could be explained as follows. Even if *Connector* nodes participate in information transfer, they do not necessarily trade themselves. Their motivation for information transfer can be quite different. As Ahern [6] points out, one of the motives for information transfer is to gain favor. Therefore, information may travel through acquaintances in order to establish and strengthen relationships. The joint working history is important in considering whether

**Table 4** One-sided Welch's  $t$ -test on the differences in age, length of co-insidership and geographical distance between NPATH-SYNC (Eq. (14)) and PATH-SYNC (Eq. (13)) sets (Panel A), and NPATH-SYNC\* (Eq. (16)) and PATH-SYNC\* (Eq. (15)) sets (Panel B). Trade synchronization between investors is validated at the significance level of  $\alpha = 0.01$ . Here,  $H_A$  is the direction of the alternative hypothesis. The units for the co-insidership are trading days, age difference in years, and geographical distance in kilometers

Panel A							
	avg. NPATH-SYNC	avg. PATH-SYNC	$p$		NPATH-SYNC	PATH-SYNC	$H_A$
age difference	9.03	8.92	2.48E-03 **		465,655	33,960	NPATH-SYNC > PATH-SYNC
geographical distance	138.92	86.03	0 ***		299,877	26,905	NPATH-SYNC > PATH-SYNC
co-insidership	483.47	406.26	1		466,402	34,227	NPATH-SYNC < PATH-SYNC
Panel B							
	avg. NPATH-SYNC*	avg. PATH-SYNC*	$p$		NPATH-SYNC*	PATH-SYNC*	$H_A$
age difference	9.37	9.35	4.20E-01		7558	9338	NPATH-SYNC* > PATH-SYNC*
geographical distance	78.81	67.24	6.11E-10 ***		6128	8108	NPATH-SYNC* > PATH-SYNC*
co-insidership	606.78	441.41	1		8287	9603	NPATH-SYNC* < PATH-SYNC*

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$

you act on trading tips yourself or just pass them on to others. When you receive trading tips from a colleague you have known for a short time only, you are more likely to distribute information to others rather than act on it yourself.

For the robustness check we restrict the set of links further, by considering only those investors who have their trading accounts matched. The experiment set is redefined as

$$\text{PATH-SYNC}^* = \{(k, l) \mid (k, l) \in \text{PATH-SYNC} \cap k, l \in \mathcal{I}^m\}, \quad (15)$$

and the reference set as

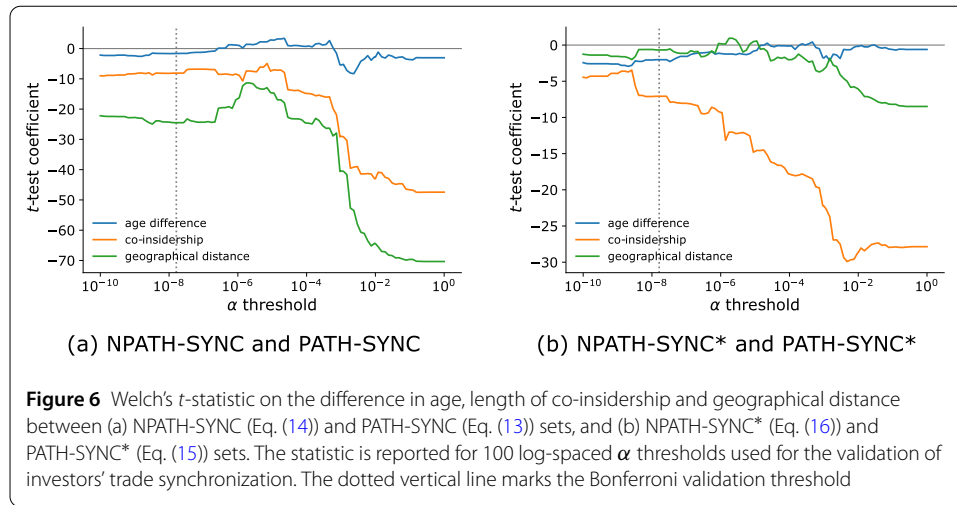
$$\text{NPATH-SYNC}^* = \{(i, j) \mid (i, j) \in \text{NPATH-SYNC} \cap i, j \in \mathcal{I}^m\}. \quad (16)$$

These results are reported in Table 4 (Panel B). The results for age differences become insignificant.

In addition, we apply 100 threshold levels for the validation of investor trade similarities equally spaced between  $10^{-10}$  and 1, see Fig. 6. The Welch's  $t$ -statistic for the age difference fluctuates around zero, yet is significant with  $\alpha = 0.01$ . At the same time, the difference in the length of the co-insidership is consistently negative and significant. The difference in the geographical distance is negative and significant with different thresholds  $\alpha$  in the baseline analysis, but fluctuates closer to zero with low  $\alpha$  values for robustness checks that include only matched investors.

#### 4.4 Are synchronizing investors more similar if they are socially connected?

In this section, we investigate whether investors who synchronize their trades are more similar in age and geographical location if they are socially connected in the co-



employment network. We select two sets: the test set CON-SYNC (defined by Eq. (8)) constituting of investor pairs who are both connected in the social network and synchronize their transactions, and the reference set NCON-SYNC comprised of the insider pairs that are not directly connected but synchronize their trades. Our hypothesis is that investors who synchronize their trades are (i\*\*) more similar in age, and (ii\*\*) geographically closer if they are also socially connected. Since by construction, investor pairs in NCON-SYNC are not socially connected, they also do not have a co-insidership attribute.<sup>12</sup> Therefore, we only look into the Welch's  $t$ -test for age difference and geographical distance. For the illustration of the test and reference sets, see Appendix Fig. 16 (d).

Assuming that information travels between investors who synchronize their trades, those investor pairs that are not socially connected (NCON-SYNC) can either have a social connection that we do not observe or exchange information indirectly through other investors. Suppose the main test set links (CON-SYNC) do not exhibit stronger ties quantified through attributes than the links in the reference set. In that case, we have a reason to suspect that we have missing observations for the social links between investors who synchronize their trades but are not socially connected. Otherwise, we anticipate that disconnected but synchronizing investors use the observable social network to acquire valuable information indirectly through other investors.<sup>13</sup>

Table 5 summarizes the results of one-sided Welch's  $t$ -tests when the statistical link validation threshold  $\alpha = 0.01$ . We provide evidence for Hypotheses (i\*\*) and (ii\*\*) that the age differences are smaller and geographical distances are shorter between investors who synchronize their trades and are socially connected. In addition, as a robustness check, we apply 100 threshold levels for the validation of investor trade similarities equally spaced between  $10^{-10}$  and 1, see Fig. 7. The results are rather consistent with different levels of  $\alpha$ , especially regarding geographical distance.

<sup>12</sup>If a pair of investors were insiders in the same company simultaneously, they would be connected in the social network and have an observation for the length of their joint co-insidership.

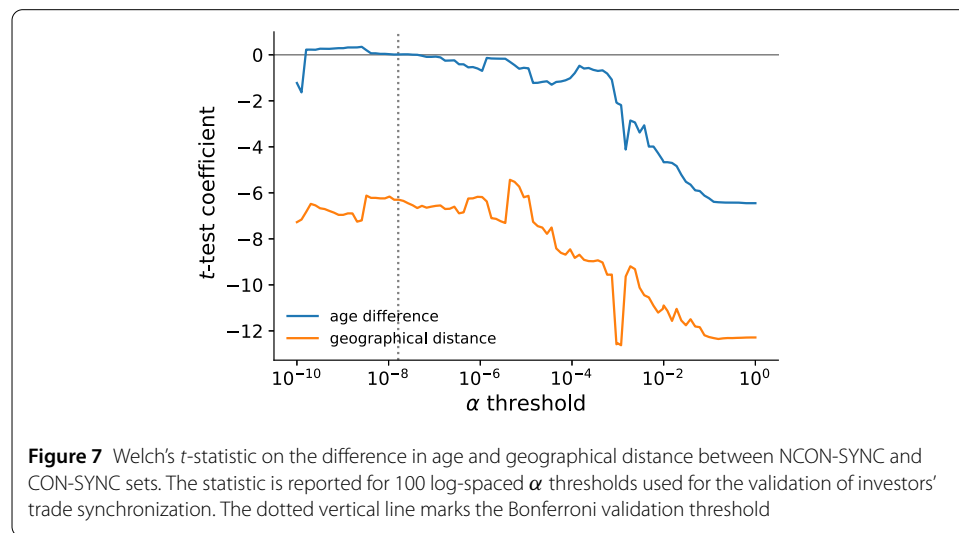
<sup>13</sup>Alternatively, if investor ties are stronger in terms of more similar age and closer geographical distance between pairs in the main set CON-SYNC, which contain pairs of colleagues in the same company, it can simply mean that the employees of the same company are relatively uniform in terms of measured attributes.



**Table 5** One-sided Welch's  $t$ -test on the differences in age and geographical distance between NCON-SYNC and CON-SYNC sets. Trade synchronization between investors is validated at the significance level of  $\alpha = 0.01$ . Here,  $H_A$  is the direction of the alternative hypothesis. The units for the co-insidership are trading days, age difference in years, and geographical distance in kilometers

	avg. NCON-SYNC	avg. CON-SYNC	$p$		NCON-SYNC	CON-SYNC	$H_A$
age difference	11.66	9.42	2.18E-06	***	11,574	312	NCON-SYNC > CON-SYNC
geographical distance	135.19	63.49	2.65E-24	***	13,287	323	NCON-SYNC > CON-SYNC

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$



## 5 Discussion and conclusions

In this paper, we provide evidence of statistical relation between the characteristics of social ties and investor trading behavior using three attributes: differences in age, geographic proximity, and length of joint co-employment. We show that investors that are closer in the social network also have a stronger trading similarity. We indicate that trade synchronization between an investor pair is more likely to happen when investors have worked together (serving as insiders) in the same company longer. This is an intuitive finding, because the time spent together could give more ground for more frequent interactions. This could be related to the development of trust over time. Moreover, partial evidence is provided that geographical proximity explains trading synchronization. In addition, partial evidence is provided that investors of the similar age are more likely to be found on the paths between investors who synchronize their trades.

One of the limitations of our study is that the observable social network is only a proxy of the complete social network between the investors in our analysis. While we observe the relationships between insiders that arise from co-employment, we do miss other types of social relationships. In this regard, alternative proxies for the social network and different investor socioeconomic attributes should be investigated. The other limitation is that we are not able to uniquely identify trading accounts for all investors in our social network proxy. To remedy this we have performed robustness checks to consider only investors with matched accounts.

A trade synchronization network could be used as a proxy for the information network [8]. From this point of view, one could conclude that our results reveal patterns in mutual information transfer between investors, but, possible implications to information trading should be considered with caution. In particular, trading decisions can be independent even if the information is shared between two investors. Moreover, if two investors synchronize their transactions, it does not necessarily mean that they have been exchanging information. For example, both investors can react to public news similarly [9], they may have developed similar trading strategies or consult the same financial advisors.

In our future research, we will analyze investors' positions in the insider network and how they relate to their stock market performance. Moreover, predictive analysis about investors' trading behavior using information about their social connections is in our interest.

## **Appendix A: Data sources**

### **A.1 Euroclear shareholder registration data set (SRD)**

This section describes the shareholder registration data set in more detail. This data contains complete trading records from all Finnish investors on publicly traded stocks in HSE along with the background information on traders' transactions and their attributes from 1995 to 2016. Each transaction in the data set is characterized by the investors' anonymized ID, trade and registration dates, security identifier (ISIN), traded volume, investor's sector code, postal code, and investor's birth year and gender for household investors. We use this data set not only to match insiders' IDs, but we also use it to analyze insider trade synchronization.<sup>14</sup> More information about the Euroclear data set can be found in other research publications (see, e.g., [9, 11, 16, 44–49]).

### **A.2 Insider register data and mandatory notification of trades (SIRE)**

The Euroclear Finland's insider register service (SIRE) data set contains information about mandatory notifications of trade by 3514 company insiders. It covers information from 153 companies listed in the HSE. The data set was manually collected from SIRE insider registry for two disjoint periods of roughly five years, from July 2005 to February 2010 and March 2013 to September 2018. The SIRE register reports historical data for the previous five years of insider holdings and transactions for the publicly listed companies. The data contains a list of insiders names and surnames and a list of third parties related to them, e.g., family members or related companies. Information on insiders includes their nationality, spoken language, the dates and the basis for considering them an insider in given companies. Each trade notification comes with an identifier, both for traded stocks and derivative instruments, trading date and volume, and information about whether the transaction was executed by himself/herself or by his/her family member or by a related third-party company account. If an insider did not trade their own company shares, s/he might not be included in this data set. In some cases, individuals who disclosed their trades in the SIRE data sets and, therefore, are considered insiders, were not included in the Virre data set. A part of those insiders and related parties, who's mandatory notifications of trade

---

<sup>14</sup>Note that for the selected course of analysis, specifically SVN method, we cannot use the trading data after the end of 2009 as it contains aggregated net daily trades, which makes impossible to estimate separate buy and sell volumes for an investor.

were successfully used to match them with their complete trading histories in Euroclear data set, is used to complement the insider network constructed from the Virre data (see Fig. 2). Out of 3514 insiders in SIRE data set 2711 were matched with their trading accounts in Euroclear data set (reaching 77% matching success ratio). 1007 of them were not found in the Virre data set, and were added to the set of nodes in the temporal insider network. The bigger part of them (555) belong to family members and related companies.

### A.3 Insiders' assignments and positions data (Virre)

The primary source of information about the insider network structure comes from the Virre data set acquired from the Virre Information Service. The data set contains information about historical insider positions in Finnish companies. Virre data set contains the information about most company insiders, irrespectively of whether they have traded or not own company shares. The data includes the name of the company, the name and the surname of the insider, his/her role in the company, e.g., board members, auditors, legal representatives, and other key employees, and all start and end dates of the insider-ships. Virre data alone best represents the complete insider network. The whole data set contains information about 12,925 insiders. When we construct the insider network, we connect by links individuals indicated as insiders in the same companies. The ties last as long as the connected nodes are both considered insiders in the same company. However, the data does not include information about family members and controlled companies of the insiders. Virre data set spans between 1962 and 2019 and contains information about 148 companies.

## Appendix B: Descriptive statistics

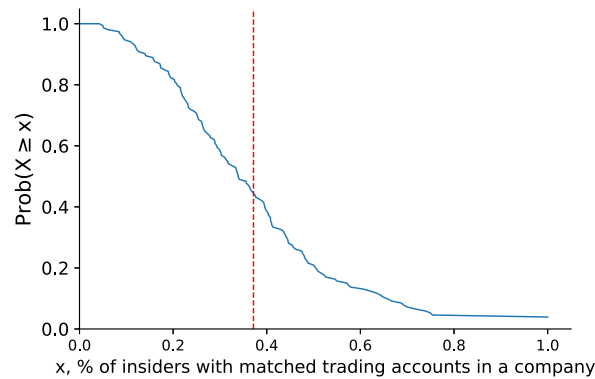
The list of company insiders is mainly comprised of employees in higher positions and board members. Typically key employees do not change too often, and board members can serve multiple terms. The stability of corporate structure allows us to have a sufficiently high coverage of insiders who trade in the stock market, even outside the two five-year periods for which SIRE data was collected.

In particular, in 2005, the start of the first SIRE data collection period, we were able to match 752 insiders, which corresponds to 22% of all insiders in the network at that moment (see Table 6). Looking at the network snapshots in 2000 and 2003 we observe only a marginal decrease in the fraction of matched insiders, correspondingly 14% and 19%. Figure 8 shows the complementary cumulative distribution function of the ratio between the matched insiders and all insiders in a company.

The insider network is dynamic, with daily snapshots between 1995 and 2016. Each snapshot on a day  $t$  contains a set of  $N_t$  nodes which represent insiders and related third parties as well as  $L_t$  links that represent relationships between them (see Table 6 for the global network statistics). The relationship between two nodes can indicate either both of the nodes being insiders in at least one same company, or one of them being an insider in some company and the other being a third party related to the insider (mostly leaf nodes), e.g., a family member, or a company, where s/he exerts control or has a significant influence on its investment decisions. Insiders with positions in multiple companies connect different insiders from different companies.

**Table 6** Global statistical properties of the insider network calculated on January 1st in a corresponding year  $t$ . The years were selected in order to illustrate network properties over the whole span of the SRD data as well as SIRE insider trade data set collection periods (see Fig. 1). Here,  $N_t$  denotes the number of nodes,  $N_{\text{MATCHED},t}$  – number of matched insiders,  $N_{\text{MATCHED},t}/N_t$  – ratio of matched insiders to all insiders in the network,  $L_t$  – number of links,  $\langle k_t \rangle$  – average degree,  $\langle k_t \rangle/N_t$  – normalized average degree  $N_t$ ,  $N_{c,t}$  – number of nodes in the maximal connected component,  $N_{c,t}/N_t$  – fraction of nodes belonging to the maximal connected component

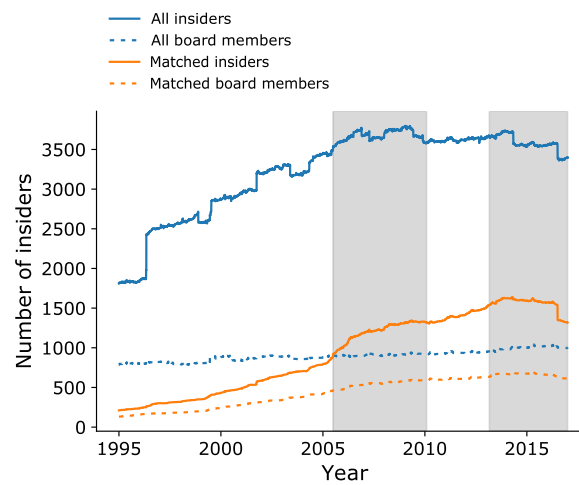
$t$	1995	2000	2003	2005	2008	2010	2013	2015	2016
$N_t$	1791	2838	3225	3410	3621	3515	3576	3458	3462
$N_{\text{MATCHED},t}$	189	407	616	752	1168	1257	1436	1516	1479
$N_{\text{MATCHED},t}/N_t$	0.11	0.14	0.19	0.22	0.32	0.36	0.40	0.44	0.43
$L_t$	54,079	160,500	200,447	224,317	173,952	138,267	114,347	86,332	86,261
$\langle k_t \rangle$	60.39	113.11	124.31	131.56	96.08	78.67	63.95	49.93	49.83
$\langle k_t \rangle/N_t$	0.03	0.04	0.04	0.04	0.03	0.02	0.02	0.01	0.01
$N_{c,t}$	1436	2725	3124	3137	3461	3415	3435	3256	3155
$N_{c,t}/N_t$	0.80	0.96	0.97	0.92	0.96	0.97	0.96	0.94	0.91



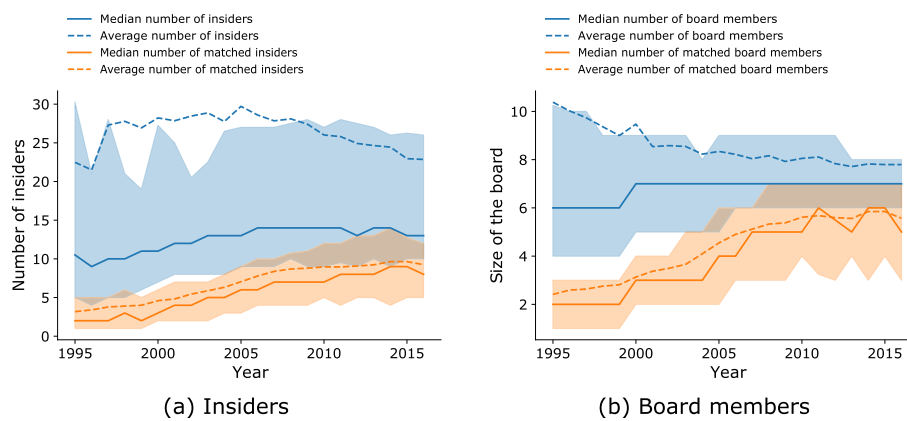
**Figure 8** The figure illustrates the complementary cumulative distribution function of the % of insiders with matched trading accounts over all 153 companies. Here we exclude family members and related company accounts. The percentage for each company is calculated by taking the ratio between the number of company insiders that have their trading accounts matched and the number of all insiders in the company. On average, 37% of a given company's insiders have their trading accounts identified (dashed red line). The average match ratio is higher than a global match ratio 19% (2711 matched accounts divided by 13,932 investors), because it is calculated as a mean of matching ratios for each company. Matched *Connector* insiders can be counted multiple times if they are insiders in more than one company

#### B.4 Number of insiders and board members

The daily number of insiders  $N_t$  and board members in our temporal network between 1995 and 2016 over all companies is shown in Fig. 9. The number of all insiders in the insider network (the blue solid curve) ranges from 1810 to 3798, while the number of matched insiders (the orange solid curve) ranges from 211 to 1636. For comparison, the dashed blue line shows the total number of board members in the insider network, ranging from 776 to 1044, and matched board member IDs ranging from 132 to 706 are shown by the orange dashed line. Note that we could only identify those insider accounts that traded own company stock within the two five-year periods (indicated by two grey colored bands in Fig. 9). This does not necessarily mean that other insiders were inactive investors, as they could have traded other stocks without being obliged to report them. While their trades are recorded in the SRD data set, we do not have sufficient information to identify those accounts.



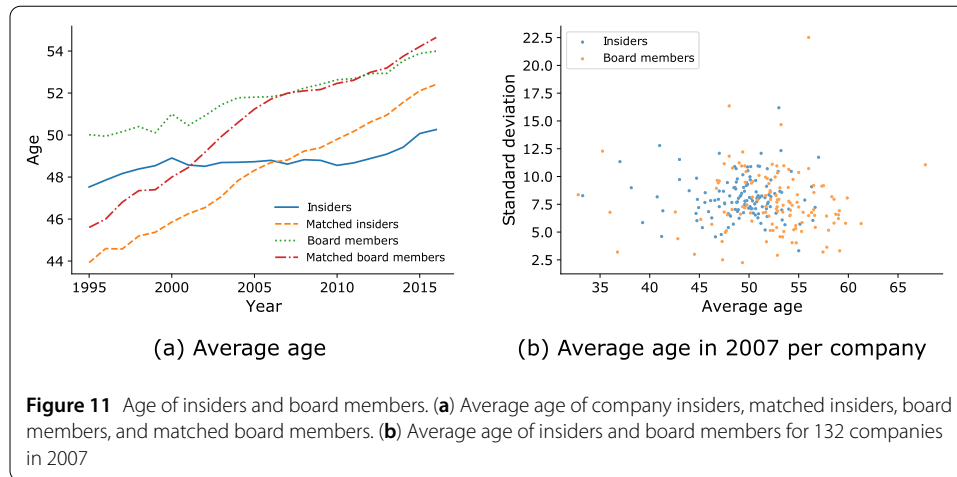
**Figure 9** Daily number of company insiders, matched insiders, board members, and matched board members. Solid lines represent insiders, while the dashed lines mark company board members. Grey bands mark two periods of SIRE data set. The difference between the blue and the orange solid (dashed) curves indicates the number of insiders (board members) that did not trade shares of their own securities during the observed period, i.e. the number of insiders and board members for which we had no data in order to match them with a trading account from the SRD data set. The difference is smaller when looking only at board members. This means that there are fewer unmatched board members than other kinds of unmatched insiders



**Figure 10** Number of insiders and board members. (a) Median and average number of insiders and matched insiders per company in a year. (b) Median and average number of board members and matched board members per company in a year. The bands around the medians are filled between 1st and 3rd quantiles

In our data set, there are approximately  $26.2 (\pm 2.4)$  insiders and  $6.9 (\pm 2.3)$  matched insiders<sup>15</sup> on average per company in a year (Fig. 10 (a)). The blue dashed line shows the average number of insiders per company in a year, ranging from 21.5 to 29.7 insiders. The orange dashed line in Fig. 10 (a) shows the average number of matched insiders and ranges from 3.2 to 9.6. Similarly, the average board size per company is composed of  $8.5 (\pm 0.8)$  board members in a company in a year, out of which on average we identified around  $4.4 (\pm 1.3)$  matched board members (Fig. 10 (b)).

<sup>15</sup>Here we excluded related family members and third-party company accounts.



### B.5 Age of insiders and board members

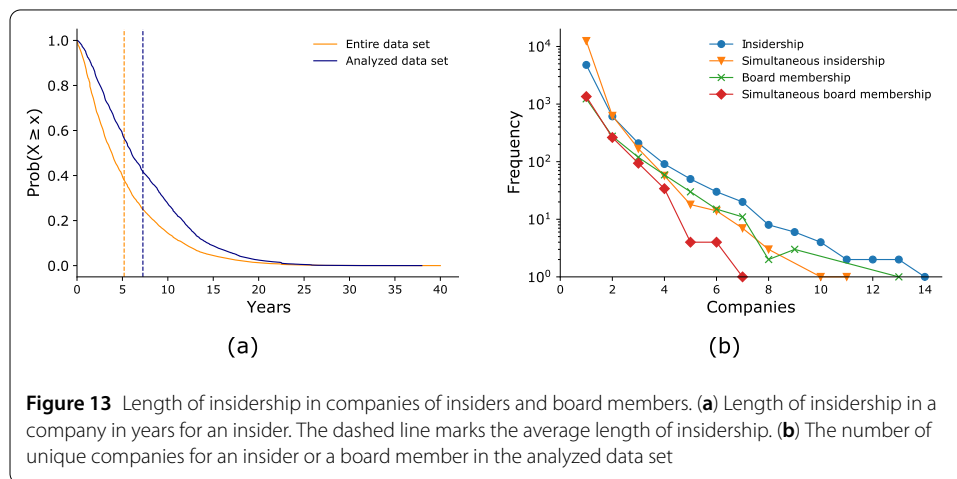
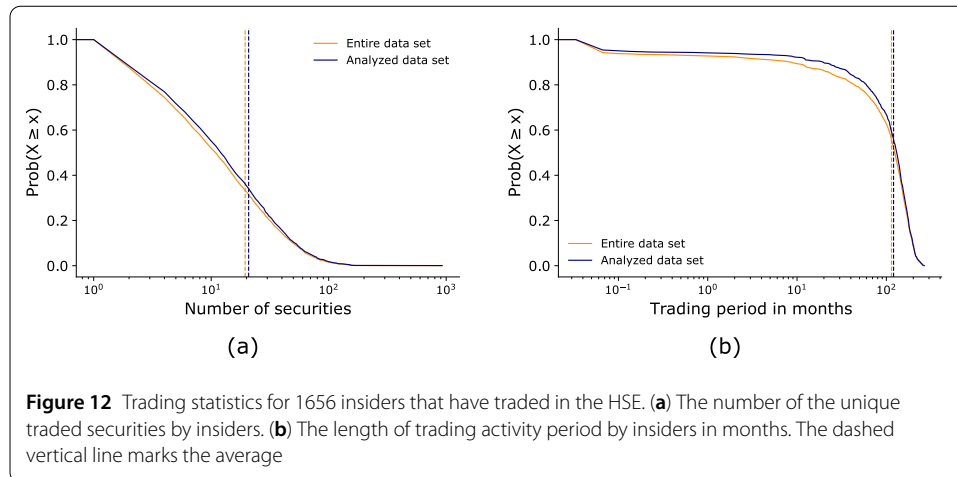
The average age of an insider is around 48.8 ( $\pm 0.6$ ) years, rising from 47.5 years in 1995 to 50.3 years by 2017 (see Fig. 11 (a)). The average age of a board member is around 51.8 ( $\pm 1.3$ ) years, rising from 49.9 years in 1995 to 54.0 years in the end of 2016 (the green line of the Fig. 11 (a)). At the end of the insider network data period, the average age of an insider (board member) was 50.5 (54.2) years, while the average age of male (female) investors was 55 (57) years and Finland's population age was correspondingly 41 (44) years [50]. This means that an average insider is younger than an average investor and older than an average Finn, while the average board member is very close in age to the average Finnish investor. Based on the average board member age in the data set, we can suspect that the older board-members from the 1995-2005 period retired and were not captured in the matched data sets during the collection period between 2005 and 2010, and 2013 and 2018. This can be seen by a rapidly increasing red line until 2005 (Fig. 11 (a)). Age preferences for different companies are shown on Fig. 11 (b). We chose the year 2007 as the mid point of our empirical analysis.

### Appendix C: Statistics for the analyzed network

In what follows, we provide the summary statistics on the insiders from the analyzed network. Note that while the subset of investor is restricted to the subset observed between 2005 and 2009, their activity statistics cover their entire data set. Regarding insiders' trading activity, 1656 out of 1756 matched insider accounts have traded stocks in HSE.<sup>16</sup> An insider traded 20.8 ( $\pm 32.9$ ) securities on average (Fig. 12 (a)). There were 192 insiders who traded only 1 security, and, at the same time, the investor with the most diversified portfolio traded 923 distinct securities. In our network, the average trading period for an insider network is 9.9 ( $\pm 5.2$ ) years (Fig. 12 (b)). This number is obtained by taking into account all the observations for the selected 1656 insiders in the entire SRD data set.

While most serve as insiders for up to two years, the longest observed length of insider-ship in the Nordic Aluminium company reaches almost 38 years (Fig. 13 (a)). An average

<sup>16</sup>It is possible that a matched investor never made a marketplace trade, because information about other type of transaction, such as receiving securities from the company, were used in the matching procedure.

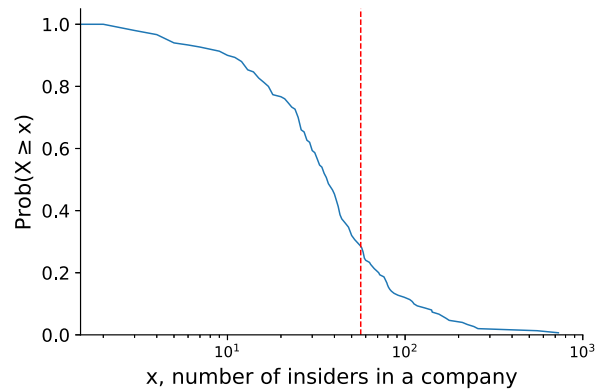


insider (board member) insidership length lasts  $7.25 \pm 5.26$  ( $6.59 \pm 5.3$ ) years in the analyzed data set and  $5.19 \pm 4.46$  ( $5.04 \pm 4.59$ ) years in the entire data set.

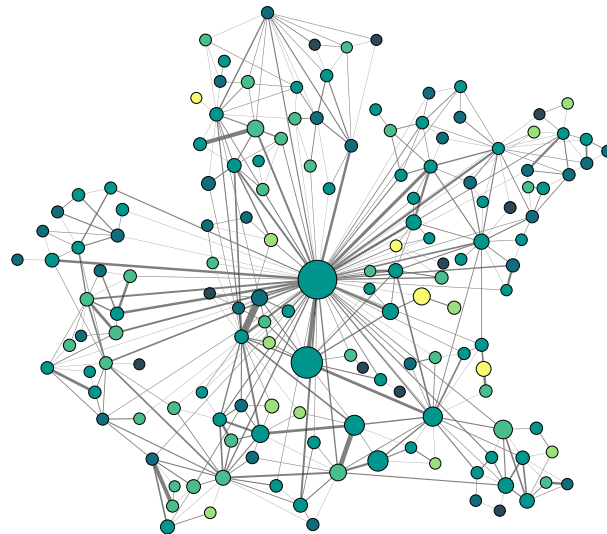
Figure 14 shows the distribution of the number of insiders per company. In the analyzed data set, an insider (board member) belongs to 1.3 (1.6) companies on average (Fig. 13 (b)). An insider with the maximum number of companies, worked in 14 companies in total and in seven simultaneously. A board member with the most companies has resided in 13 company boards in total and seven simultaneously. There are 1753 board members in total out of which there are 516 *Connector* board members, who belonged to more than one company board either changing companies sequentially or being a board member simultaneously in several companies. Out of 516 *Connector* board members, 400 were simultaneously on more than two companies' boards. These numbers are limited by the pool of insiders present between 2005 and 2009, but the observations for the selected insiders come from the total period (1962–2019).

In general, insider networks exhibit the small-world network properties [51, 52] and are comparable to the ones described in Battiston and Catanzaro [53]. More specifically, even if the networks are composed of thousands of nodes, it takes only a small number of leaps between nodes to traverse from any source to any target node. In the insider network,





**Figure 14** Complementary cumulative distribution function for the numbers of insiders in a company. Number of insiders ranges from 2 to 729 per company. The number of companies is 150. Red dashed line marks the average number of insiders per company = 56.25 ( $\pm 82.89$ ), the mode = 25, and the median = 36



**Figure 15** Company network. Each node represents a company. Links connect companies that share an insider. The thickness of a link is stronger when the companies share more insiders. For illustration purposes, the full network is reduced to Planar Maximally Filtered Graph. Number of nodes: 150, number of links: 386. The diameter of this network is 6 and the average degree is 5.15. The nodes are scaled per number of insiders in the company. The nodes are colored according to the proportion of male insiders. Percentage of male insiders in the company: ● - 100, ● - 90, ● - 80, ● - 70, ● - 60, ● - 50

all shortest paths travel through the *Connector* nodes, i.e., the nodes that have an insider position in two or more companies.

We illustrate the connections between company insiders through the *Connector* nodes in Fig. 15. Insiders, excluding the family members and third-party accounts, are aggregated into a single node representing the affiliated company. A link between two companies indicates that they have shared at least one insider. We can observe a high connectivity between the company nodes, with the biggest companies in the centre being Nokia and UPM-Kymmene. In the analyzed network, there are 1033 *Connector* nodes (out of which 575 are matched insiders). There are 200 (19%) women (out of which 104 are matched

insiders), and 833 (81%) are men (out of which 471 are matched insiders). Even though gender attribute is available in our data, it is omitted in this study due to high gender class imbalance demonstrated in Fig. 15. Overall, companies are dominated by male insiders, with only four gender-balanced companies. In fact, the male-dominant environment is a favorable setting for the information transfer channels, as both men and women prefer transferring to or receiving information from men [54–56].

#### Appendix D: Analyzed securities

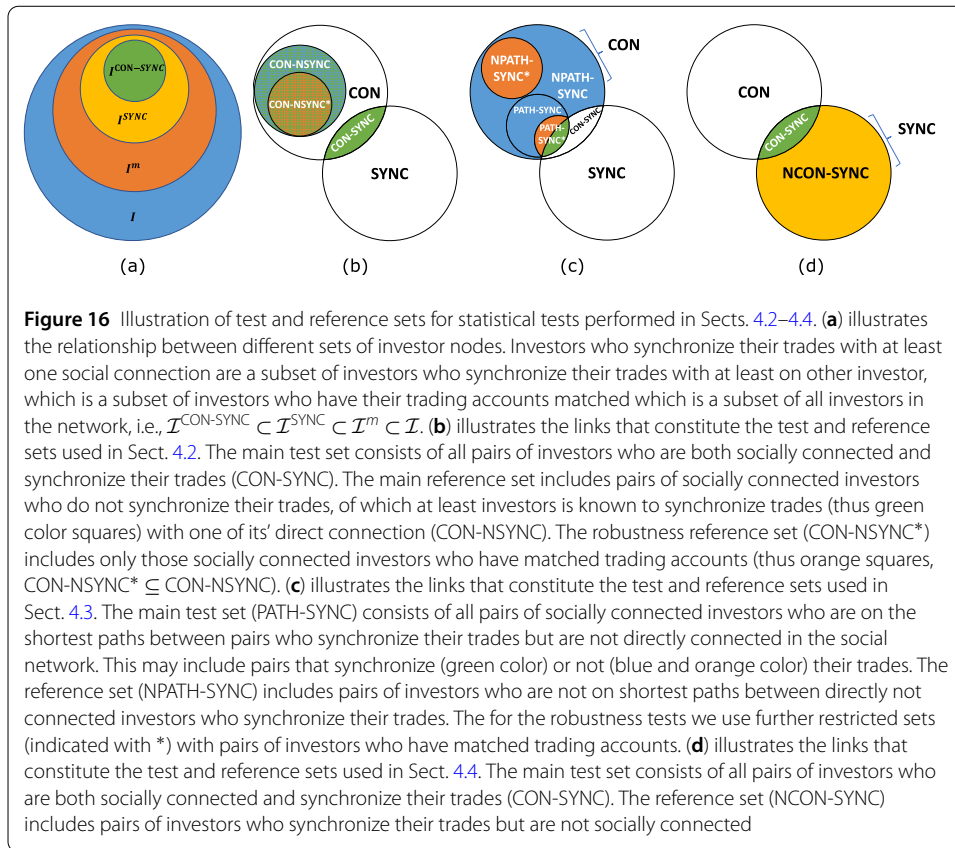
Table 7 shows 30 most traded securities by the investors analyzed in this paper.

**Table 7** Company name and security ID (ISIN)

Company	Security ID
Amer Sports	FI0009000285
Cargotec	FI0009013429
Cramo	FI0009900476
Elisa	FI0009007884
Fortum	FI0009007132
Huhtamäki	FI0009000459
Kemira	FI0009004824
Kesko	FI0009000202
KONE	FI0009013403
Metsä Board	FI0009000665
Metso	FI0009007835
Neste	FI0009013296
Nokia	FI0009000681
Nokian Renkaat	FI0009005318
Nordea Bank	FI0009902530
Outokumpu	FI0009002422
Perlos	FI0009007819
Pohjola Bank	FI0009003222
Ramirent	FI0009007066
Rautaruukki	FI0009003552
Sampo	FI0009003305
Sanoma	FI0009007694
Stockmann	FI0009000251
Stora Enso	FI0009005961
Telia	SE0000667925
TietoEVR	FI0009000277
UPM-Kymmene	FI0009005987
Uponor	FI0009002158
Wärtsilä	FI0009003727
YIT	FI0009800643

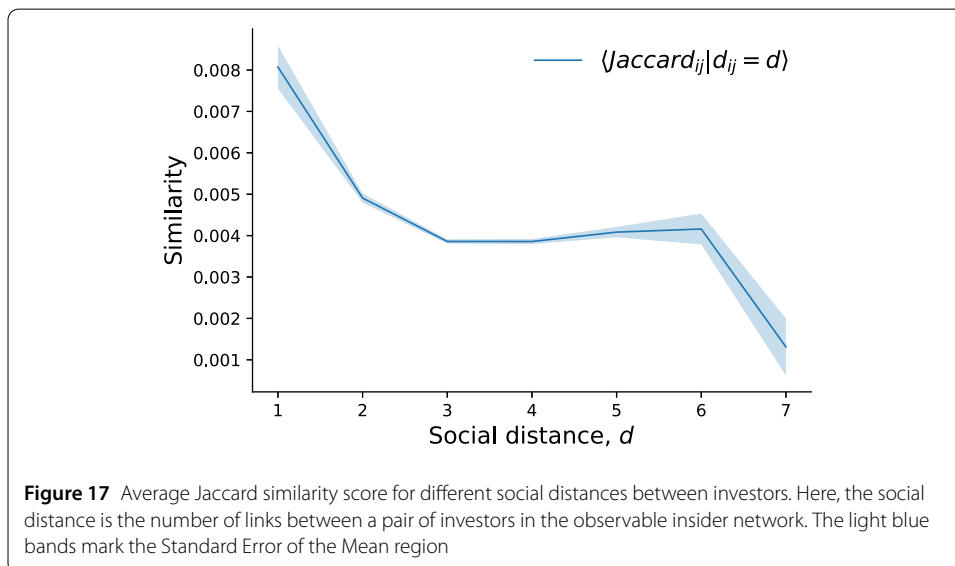
#### Appendix E: Investor nodes, investor links, test and reference sets

Here, we elaborate on the types of investor nodes and their notation. We also present different types of links and link sets used in our Welch's  $t$ -tests. Figure 16 illustrates the relationship between different sets of nodes and links used throughout the paper.



## Appendix F: Robustness check for the synchronization score and social distance

Here, we perform a robustness check for the average synchronization score and the social distance estimated with the Jaccard coefficient (Fig. 17 and Table 8).



**Table 8** One-sided independent two-sample Welch's *t*-test on the difference in trade synchronization score for pairwise consecutive social distance. Here,  $D_k$  is the set of the synchronization scores between investors at the distance  $k$

$l$ and $m$	avg. $D_l$	avg. $D_m$	$p$		$ D_l $	$ D_m $	$H_A$
1 and 2	0.00807	0.00491	1.39E-09	***	18,027	219,572	$D_1 > D_2$
2 and 3	0.00491	0.00386	9.64E-18	***	219,572	686,930	$D_1 > D_2$
3 and 4	0.00386	0.00386	4.92E-01		686,930	473,756	$D_1 > D_2$
4 and 5	0.00386	0.00408	9.49E-01		473,756	127,752	$D_1 > D_2$
5 and 6	0.00408	0.00416	5.76E-01		127,752	14,355	$D_1 > D_2$
6 and 7	0.00416	0.00131	1.44E-04	***	14,355	498	$D_1 > D_2$

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$

### Acknowledgements

Not applicable.

### Funding

M.B. received funding from the doctoral school of Tampere University. K.B. received OP Group Research Foundation Grant 20210143. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

### Abbreviations

HSE, Helsinki Stock Exchange; MTC, Multi-test correction; SRD, Shareholder registration data; SVN, Statistically Validated Network.

### Availability of data and materials

The data sets analyzed in this study are under the non-disclosure agreement and are not allowed to be distributed nor made open by the data provider.

### Declarations

#### Competing interests

The authors declare no competing interests.

#### Author contribution

M.B. collected and prepared the data set and conducted the empirical analysis. All authors analyzed the results and wrote and approved the manuscript.

#### Authors' information

Margarita Baltakienė, Kęstutis Baltakys, and Juho Kanninen are affiliated to Statistical Data Analytics, Unit of Computational Sciences, Tampere University, Finland.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 May 2022 Accepted: 29 October 2022 Published online: 17 November 2022

### References

- Campbell JY, Shiller RJ (1988) The dividend-price ratio and expectations of future dividends and discount factors. *Rev Financ Stud* 1(3):195–228
- Cochrane JH (2011) Presidential address: discount rates. *J Finance* 66(4):1047–1108
- Garrett I, Priestley R (2012) Dividend growth, cash flow, and discount rate news. *J Financ Quant Anal* 47(5):1003–1028
- Campbell JY (2003) Consumption-based asset pricing. *Handb Econ Finance* 1:803–887
- Fama EF, French KR (2015) A five-factor asset pricing model. *J Financ Econ* 116(1):1–22
- Ahern KR (2017) Information networks: evidence from illegal insider trading tips. *J Financ Econ* 125(1):26–47
- Saavedra S, Hagerty K, Uzzi B (2011) Synchronicity, instant messaging, and performance among financial traders. *Proc Natl Acad Sci USA* 108(13):5296–5301
- Ozsoylev HN, Walden J, Yavuz MD, Bildik R (2013) Investor networks in the stock market. *Rev Financ Stud* 27(5):1323–1366
- Baltakienė M, Kanninen J, Baltakys K (2021) Identification of information networks in stock markets. *J Econ Dyn Control* 131:104217
- Colla P, Mele A (2010) Information linkages and correlated trading. *Rev Financ Stud* 23(1):203–246
- Baltakys K, Baltakienė M, Kärkkäinen H, Kanninen J (2018) Neighbors matter: geographical distance and trade timing in the stock market. *Finance Res Lett*
- Shive S (2010) An epidemic model of investor behavior. *J Financ Quant Anal* 45(1):169–198

13. Brown JR, Ivković Z, Smith PA, Weisbenner S (2008) Neighbors matter: causal community effects and stock market participation. *J Finance* 63(3):1509–1531
14. Tumminello M, Lillo F, Piilo J, Mantegna RN (2012) Identification of clusters of investors from their real trading activity in a financial market. *New J Phys* 14(1):013041
15. Gutiérrez-Roig M, Borge-Holthoefer J, Arenas A, Perelló J (2019) Mapping individual behavior in financial markets: synchronization and anticipation. *EPJ Data Sci* 8(1):10
16. Baltakys K, Kannianen J, Emmert-Streib F (2018) Multilayer aggregation with statistical validation: application to investor networks. *Sci Rep* 8(1):8198
17. Musciotto F, Marotta L, Piilo J, Mantegna RN (2018) Long-term ecology of investors in a financial market. *Palgrave Commun* 4(1):92
18. Challet D, Chicheportiche R, Lallouache M, Kassibrakis S (2018) Statistically validated lead-lag networks and inventory prediction in the foreign exchange market. *Adv Complex Syst* 21(8):1850019
19. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27(1):415–444
20. Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network. *Am J Sociol* 115(2):405–450
21. Fischer CS (1982) To dwell among friends: personal networks in town and city. University of Chicago Press, Chicago
22. Marsden PV (1988) Homogeneity in confiding relations. *Soc Netw* 10(1):57–76
23. Verbrugge LM (1983) A research note on adult friendship contact: a dyadic perspective. *Soc Forces* 62:78
24. Festinger L, Schachter S, Back K (1950) Social pressures in informal groups; a study of human factors in housing
25. Aiello LM, Barrat A, Cattuto C, Schifanella R, Ruffo G (2012) Link creation and information spreading over social and communication ties in an interest-based online social network. *EPJ Data Sci* 1(1):12, 1–31
26. Lee S (2019) Learning-by-moving: can reconfiguring spatial proximity between organizational members promote individual-level exploration? *Organ Sci* 30(3):467–488
27. Burt RS (2000) Decay functions. *Soc Netw* 22(1):1–28
28. Kleinbaum AM, Stuart TE, Tushman ML (2013) Discretion within constraint: homophily and structure in a formal organization. *Organ Sci* 24(5):1316–1336
29. Kleinbaum AM (2018) Reorganization and tie decay choices. *Manag Sci* 64(5):2219–2237
30. Barone M, Coscia M (2018) Birds of a feather scam together: trustworthiness homophily in a business network. *Soc Netw* 54:228–237
31. Wong LHH, Gygas AF, Wang P (2015) Board interlocking network and the design of executive compensation packages. *Soc Netw* 41:85–100
32. McEvily B, Jaffee J, Tortoriello M (2012) Not all bridging ties are equal: network imprinting and firm growth in the Nashville legal industry, 1933–1978. *Organ Sci* 23(2):547–563
33. Tumminello M, Micciche S, Lillo F, Piilo J, Mantegna RN (2011) Statistically validated networks in bipartite complex systems. *PLoS ONE* 6(3):17994
34. Baker WE (1984) The social structure of a national securities market. *Am J Sociol* 89(4):775–811
35. Battiston S, Glattfelder JB, Garlaschelli D, Lillo F, Caldarelli G (2010) The structure of financial networks. In: *Network science*. Springer, Berlin, pp 131–163
36. Finger K, Lux T (2017) Network formation in the interbank money market: an application of the actor-oriented model. *Soc Netw* 48:237–249
37. Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D (2016) Statistically validated network of portfolio overlaps and systemic risk. *Sci Rep* 6(1):1–14
38. Miller RG (1981) Normal univariate techniques. In: *Simultaneous statistical inference*. Springer, Berlin, pp 37–108
39. Berkman H, Koch P, Westerholm PJ (2020) Inside the director network: when directors trade or hold inside, interlock, and unconnected stocks. *J Bank Finance* 118:105892
40. Grinblatt M, Keloharju M (2000) The investment behavior and performance of various investor types: a study of Finland's unique data set. *J Financ Econ* 55(1):43–67
41. Grinblatt M, Keloharju M (2001) What makes investors trade? *J Finance* 56(2):589–616
42. Benevenuto F, Rodrigues T, Cha M, Almeida V (2009) Characterizing user behavior in online social networks. In: *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, pp 49–62
43. Wilson C, Boe B, Sala A, Puttaswamy KP, Zhao BY (2009) User interactions in social networks and their implications. In: *Proceedings of the 4th ACM European conference on computer systems*, pp 205–218
44. Ilmanen M, Keloharju M (1999) Shareownership in Finland. *Fin J Bus Econ* 48(1):257–285
45. Ranganathan S, Kivelä M, Kannianen J (2018) Dynamics of investor spanning trees around dot-com bubble. *PLoS ONE* 13(6):0198807
46. Siikanen M, Baltakys K, Kannianen J, Vatrappu R, Mukkamala R, Hussain A (2018) Facebook drives behavior of passive households in stock markets. *Finance Res Lett* 27:208–213
47. Baltakienė M, Baltakys K, Kannianen J, Pedreschi D, Lillo F (2019) Clusters of investors around initial public offering. *Palgrave Commun* 5(1):1–14
48. Baltakys K (2019) Investor networks and information transfer in stock markets
49. Baltakys K, Kannianen J, Saramäki J, Kivela M (2020) Trading signatures: investor attention allocation in stock markets. Available at SSRN
50. Keloharju M, Lehtinen A (2017) A quarter century of shareholdings and trades of Finnish stocks. *Nord J Bus* 66(1):5
51. Milgram S (1967) The small world problem. *Psychol Today* 2(1):60–67
52. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442
53. Battiston S, Catanzaro M (2004) Statistical properties of corporate board and director networks. *Eur Phys J B* 38(2):345–352
54. Rantala V (2019) How do investment ideas spread through social interaction? Evidence from a ponzi scheme. *J Finance* 74(5):2349–2389
55. Aldrich H, Reese PR, Dubini P (1989) Women on the verge of a breakthrough: networking among entrepreneurs in the United States and Italy. *Entrep Reg Dev* 1(4):339–356
56. Bernard HR, Killworth PD, Evans MJ, McCarty C, Shelley GA (1988) Studying social relations cross-culturally. *Ethnology* 27(2):155–179