ARTICLE

Clusters of investors around initial public offering

Margarita Baltakienė [1]*, Kęstutis Baltakys[1], Juho Kanniainen[1], Dino Pedreschi[2] & Fabrizio Lillo[3]

ABSTRACT   The complex networks approach has been gaining popularity in analysing investor behaviour and stock markets, but within this approach, initial public offerings (IPOs) have barely been explored. We fill this gap in the literature by analysing investor clusters in the first two years after the IPO filing in the Helsinki Stock Exchange by using a statistically validated network method to infer investor links based on the co-occurrences of investors' trade timing for 69 IPO stocks. Our findings show that a rather large part of statistically similar network structures form in different securities and persist in time for mature and IPO companies. We also find evidence of institutional herding.

[1] Unit of Computational Sciences, Tampere University, Korkeakoulunkatu 1, 33720 Tampere, Finland. [2] Department of Computer Science, University of Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy. [3] Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy. *email: margarita.baltakiene@tuni.fi

Introduction

Initial public offerings (IPOs) play an important role in financial markets because they open new investment opportunities, redistribute funds' allocations and attract new investors to the market. An IPO is usually a long-awaited event in the life of a privately held company, both for the current stockholders and the public exchange investors, giving the owners the opportunity to cash in and giving the investors a chance to gain from potential underpricing and future returns. Here, numerous financial studies have addressed various behavioural biases in relation to IPOs: Ljungqvist and Wilhelm Jr (2005) analysed the satisfaction with an IPO underwriter's performance, Ljungqvist and Wilhelm Jr (2003) indicated a unique pricing behaviour around the dot-com bubble, while Kaustia and Knüpfer (2008) found that investors' personal experiences and previous IPO returns have a significant impact on future IPO subscriptions. Other studies have analysed IPO investments (Karhunen and Keloharju, 2001), IPO earnings (Spohr, 2004) and IPO underpricing (Keloharju, 1993) in financial markets on an aggregated level.

Financial markets, in turn, are complex systems comprised of financial decisions, information flows and direct and indirect investor interactions. A typical aspect of a financial market is multidimensionality and agent heterogeneity (Lakonishok and Maberly, 1990; Musciotto et al., 2016). Making an investment decision is a complex procedure because it is layered with different choices that are influenced by various market factors, investors' experiences, wealth and investors' stage of life. It is crucial to understand the characteristics of the underlying investor behaviour patterns because these, when combined with their behaviours, shape the dynamics of the whole market and thus are important factors in explaining the booms and bubbles in the financial markets (Ranganathan et al., 2018). Because investors seek higher returns, one possibility is to use social networks and other private information channels to follow other investors' strategies and to exploit privately channelled information in stock markets. Recently, Baltakys et al. (2018a) provided evidence of the negative relationship between distance and trade timing similarity for household investors, indicating that face-to-face communication is still important in financial decision making. According to Ozsoylev et al. (2013), information links can be identified from realised trades because investors who are directly linked in the information network tend to time their transactions similarly. We follow this idea and use observations on investor-level transactions from shareholder registration data to identify the links between investors, here with a special focus on identifying investor clusters. Prior studies have investigated the structures of investor networks in different contexts (Ozsoylev et al., 2013; Tumminello et al., 2012; Gualdi et al., 2016; Musciotto et al., 2018; Ranganathan et al., 2018; Baltakys et al., 2018b), but investor clusters around IPOs have barely been explored.

We address this research gap by performing a broad multistock exploratory analysis of investor clusters over 69 stocks in the first two years of their IPO. In particular, we seek to establish whether the identified investor clusters are persistent over the first two years of the IPOs and appear across multiple IPO securities, as well as with existing, mature stocks in the market. Our analysis unveils statistically robust investor clusters that form simultaneously in various securities, and that persist over time.

Most of the earlier papers perform analyses on an aggregated category level (Karhunen and Keloharju, 2001; Grinblatt and Keloharju, 2001; Lillo et al., 2015; Siikanen et al., 2018) or concentrate on a single highly liquid stock (Tumminello et al., 2012; Musciotto et al., 2018). Even though earlier studies might have included nearly all market participants (Tumminello et al., 2011a; Musciotto et al., 2018), due to the focus on a single most liquid security, the results were limited and insufficient to conclude what strategies investors employ when trading over multiple securities. In contrast to previous research in the IPO literature, the current study is the first one on early-stage trading behaviour patterns on an individual investor account level. On the other hand, in opposition to the existing research on investor networks, in the current paper, instead of focusing of heavily capitalised stocks we analyse collective investor trading strategies that emerge after IPOs in the Helsinki Stock Exchange (HSE).

With the growing amounts of data and the availability of new datasets, the network theory has become a popular approach in analysing financial complex systems (e.g., Emmert-Streib et al., 2018). Notwithstanding the high interest in the market structure, investor networks and the complexity of investor behavioural interrelationships remain weakly explored. Indeed, high precision financial investor-level datasets covering years of historical data and containing information about the social links are very rare and expensive because of their sensitive nature. Moreover, transactional data often have no explicit or implicit links between investors. As a consequence, the network inference methodologies have gained much interest in recent research (Ozsoylev et al., 2013; Gualdi et al., 2016). Similar to Musciotto et al. (2018), we use the statistical validation method proposed by Tumminello et al. (2011a), which best suits our objectives and the available dataset.

In the current paper, we infer investor networks based on the investors' trading co-occurrences for 69 securities that had their IPOs between the years 1995 and 2007, and we obtain multilink networks covering two years after their IPOs. Further, by applying the Infomap algorithm (Rosvall and Bergstrom, 2008) on the investor networks, we obtain clusters of investors that share high trade-timing synchronisation. With the obtained network partitioned into clusters, we detect statistically robust clusters that persist in the networks between the first and the second years after the IPO. We also find clusters that form and re-occur over multiple securities. Finally, by cross-validating investor clusters on IPO securities with the investor clusters of more mature stocks, we conclude that the phenomenon of persistent clusters observed in earlier studies (see e.g. Musciotto et al., 2018) is not limited to mature companies but is also observable in young securities during the first years after their IPO.

Dataset and methodology

Dataset. In this paper, we use a unique database provided by Euroclear Finland. The dataset contains all transactions executed in the HSE by Finnish stocks shareholders between 1995 and 2009 on a daily basis. The data records represent the official certificates of ownership and include all the transactions executed in the HSE that change an ownership of assets. Each transaction in the dataset has a rich set of attributes—such as investor sector code, investor birth year, gender and postal code—that we make use of in our analysis to identify and characterise the investor groups. The dataset classifies investors into six main categories: households; nonfinancial corporations; financial and insurance corporations; government; nonprofit institutions; and the rest of the world. Finnish domestic investors correspond to a separate account ID, while foreign investors can choose the nominee registration for the trades. However, the analysis cannot be conducted for nominee-registered transactions because individual nominee investors cannot be uniquely identified. Rather, the nominee investors are pooled together under the custodian's nominee trading account. Therefore, a single nominee-registered investor's account holdings may correspond to a large aggregated ownership of several foreign investors. So to avoid inconsistencies in the results, we eliminated nominee transactions from our analysis. This dataset has been also analysed and described in

previous research (e.g., Ilmanen and Keloharju, 1999; Baltakys et al., 2018a, 2018b; Ranganathan et al., 2018; Siikanen et al., 2018).

The analysed data are restricted to marketplace transactions for securities that had their IPO listing in the HSE between 1995 and 2009. The official listing dates were provided by NASDAQ OMX Nordic explicitly for the current research. We analyse 69[1,2] stocks in total that were listed in Finland on the Main Exchange or First North in the given time period (Table 1). Some companies (e.g. Oriola) have two share classes with different voting rights. Class

Table 1 Summary of IPO stocks

ISIN	Company name	Industry	Total # of transactions	# of unique investors	IPO date
FI0009004881	Aspoyhtymä	Industrials	13,157	2070	1995-01-12
FI0009800346	Orion B	Basic materials	399,268	45,588	1995-05-11
FI0009800320	Orion A	Basic materials	116,334	18,132	1995-05-11
FI0009900336	Lemminkäinen	Industrials	94,849	13,269	1995-06-01
FI0009005318	Nokian Renkaat	Consumer goods	1,152,852	60,476	1995-06-01
FI0009800643	YIT	Industrials	896,718	54,808	1995-09-04
FI0009005870	Konecranes	Industrials	715,306	26,940	1996-03-27
FI0009005953	Stora Enso A	Basic materials	73,993	14,816	1996-05-02
FI0009005961	Stora Enso R	Basic materials	1,514,604	52,567	1996-05-02
FI0009005987	UPM-Kymmene	Basic materials	2,323,897	118,769	1996-05-02
FI0009006381	PKC Group	Industrials	194,480	24,624	1997-04-03
FI0009006415	Nordic Aluminium	Basic materials	19,012	4291	1997-04-24
FI0009005805	Kyro	Consumer services	44,418	9100	1997-06-09
FI0009006589	Rocla	Basic materials	15,415	3918	1997-06-17
FI0009006621	Helsingin Puhelin	Telecommunications	116,532	32,367	1997-11-25
FI0009006738	Elcoteq	Technology	503,265	43,323	1997-11-26
FI0009006696	Pöyry	Industrials	125,202	14,135	1997-12-02
FI0009006761	Metsä Tissue	Basic materials	11,286	3725	1997-12-09
FI0009007017	Alma Media I	Consumer services	10,673	2472	1998-04-01
FI0009007025	Alma Media II	Consumer services	30,500	5383	1998-04-01
FI0009007066	Ramirent	Industrials	295,726	21,662	1998-04-30
FI0009006829	Sponda	Financials	213977	19,500	1998-06-01
FI0009007215	Mandatum Pankki	Financials	25,732	6430	1998-08-03
FI0009007264	Elektrobit	Technology	681,542	74,839	1998-09-15
FI0009007371	Sonera	Telecommunications	1,504,103	140,253	1998-11-17
FI0009007355	Rapala VMC	Consumer goods	30,739	5202	1998-12-04
FI0009007132	Fortum	Utilities	2,068,556	120,902	1998-12-18
FI0009007629	Conventum	Financials	13,395	2736	1999-03-01
FI0009801286	Janton	Consumer services	22,946	5418	1999-03-15
FI0009007553	Eimo	Telecommunications	187,912	24,664	1999-03-23
FI0009007728	Teleste	Technology	209,132	22,537	1999-04-06
FI0009007546	Keskisuomalainen	Consumer services	11,019	2046	1999-04-19
FI0009007686	SanomaWSOY A	Consumer services	10,784	2438	1999-05-03
FI0009007694	Sanoma	Consumer services	458,541	33,242	1999-05-03
FI0009006886	Technopolis	Financials	85,510	8892	1999-06-08
FI0009007819	Perlos	Telecommunications	520,835	44,281	1999-06-28
FI0009007835	Metso	Industrials	1,528,914	69,361	1999-07-01
FI0009007884	Elisa	Telecommunications	1,209,330	199,530	1999-07-01
FI0009008080	Aspocomp Group	Industrials	99,023	10,948	1999-10-01
FI0009007918	Aldata Solution	Technology	253,021	22,840	1999-10-27
FI0009801310	F-Secure	Technology	578,978	70,994	1999-11-09
FI0009008221	Comptel	Telecommunications	529,255	65,050	1999-12-13
FI0009902530	Nordea Bank	Financials	1,081,900	149,790	2000-01-31
FI0009008924	Sievi Capital	Financials	91,541	12,109	2000-05-24
FI0009008833	Tekla	Telecommunications	73,328	8581	2000-05-24
FI0009009146	Tecnomen	Telecommunications	19,745	4532	2000-07-04
FI0009009054	Okmetic	Telecommunications	75,944	10,430	2000-07-05
FI0009009633	Evox Rifa Group	Telecommunications	51,493	10,203	2000-11-01
FI0009009567	Vacon	Telecommunications	80,081	10,770	2000-12-19
FI0009008270	SSH Comm. Security	Technology	112,633	16,433	2000-12-22
FI0009009674	AvestaPolarit	Basic materials	24,752	4299	2001-01-30
FI0009009377	CapMan	Financials	74,153	11,279	2001-04-02
FI0009010219	Glaston	Industrials	47,748	8174	2001-04-02
FI0009010854	Lassila & Tikanoja	Industrials	120,822	13,385	2001-10-01
FI0009010862	Suominen	Consumer goods	51,734	7052	2001-10-01
SE0000667925	Telia	Telecommunications	870,709	107,088	2002-12-09
SE0000110165	OMX	Financials	8721	1851	2003-09-04
FI0009012843	Kemira GrowHow	Basic materials	142,417	25,253	2004-10-18
FI0009013296	Neste Oil	Oil & gas	1,387,293	81,750	2005-04-21
FI0009013429	Cargotec	Industrials	474,949	29,210	2005-06-01
FI0009013312	Affecto	Technology	40,635	5726	2005-06-01
FI0009013403	Kone	Industrials	618,717	30,192	2005-06-01
FI0009013924	Salcomp	Industrials	28,721	3688	2006-03-17
FI0009010391	Ahlstrom	Basic materials	87,853	16,594	2006-03-17
FI0009013593	FIM Group	Financials	11,379	3084	2006-04-21
FI0009014344	Oriola A	Health care	25,922	5595	2006-07-03
FI0009014351	Oriola B	Health care	116,890	19,279	2006-07-03
FI0009012413	Terveystalo Health	Health care	35,203	8946	2007-04-10
FI0009015309	SRV Yhtiöt	Industrials	56,384	9579	2007-06-15

International Securities Identification Number (ISIN), company, industry, total number of transactions, total number of unique investors and the IPO day of the security. ISINs from the error-free set are marked in bold

Table 2 Summary of the number of investors, absolute exchanged shares volume and the number of transactions

Investor category	# ids	Volume	# transactions
Non-financial corporations	29,008	10,492,715,279	3,678,419
Financial and insurance corporations	827	350,594,504,886	55,735,780
Government	277	7,279,324,503	298,434
Households	532,387	8,984,345,323	12,965,717
Non-profit institutions	3407	937,609,174	291,922
Rest of the world	4133	12,505,262,104	3,534,817
Total	**570,039**	**390,793,761,269**	**76,505,089**
Nominee registered	89	331,154,383,799	51,782,691
Non-nominee registered	569,993	59,639,377,470	24,722,398

Note that the total volume in the table is counted twice, both for the selling and buying transactions. Here, 43 out of 89 investors with a nominee-registered holding type also made transactions with a non-nominee-registered holding type

A shares give the owner more voting rights than Class B and hence potentially falls under a separate group of investors. Therefore, the comparison or a direct substitution of shares with one another seems improper, and we consider the securities with different voting classes as separate stocks.

Table 2 gives the number of investors, the number of transactions and the traded volume for the entire set of 69 IPO stocks. The total number of investors who traded an IPO security is 570,039, and the total number of transactions is 76,505,089. The table also shows the number of nominee and non-nominee-registered investors. As shown, a few nominee accounts perform roughly twice as many trades as the non-nominee accounts.

Methodology. The given dataset is composed of transaction data where investors' social links are not explicitly given, nor can they be directly obtained from other sources because of data anonymisation. However, given that investors must individually react and adapt to a quickly changing environment, they should identify and follow the best trading strategies. To detect investors with *similar* trading strategies or, more precisely, trade timing similarity, we take a look at the pairwise investors' trading co-occurrences. In the current paper, we use a statistically validated network (SVN) method first introduced by Tumminello et al. (2011a). This method, briefly presented below, has been demonstrated to be effective in investigating financial, biological and social systems (Tumminello et al., 2011a, 2012).

To compare the trading position taken by an investor on a given day, irrespective of the absolute volume traded, a categorical variable is introduced that describes the investor's trading activity. For each investor i and each trading day t having the volume sold of a security $V_s(i, t)$ and the volume bought of a security $V_b(i, t)$, we calculate the scaled net volume ratio as follows:

$$r(i,t) = \frac{V_b(i,t) - V_s(i,t)}{V_b(i,t) + V_s(i,t)} \qquad (1)$$

Then, a daily trading state can be assigned for an investor after having selected a threshold θ, as follows:

$$\begin{cases} b - \text{primarily buying state, when } r(i,t) > \theta \\ s - \text{primarily selling state, when } r(i,t) < -\theta \\ bs - \text{buying and selling state, when } -\theta \leq r(i,t) \leq \theta \end{cases}$$

Note that $r(i, t)$ is not defined for day t that had no trading activity, and therefore, no trading state is assigned. In our analysis, much like in Musciotto et al. (2016), we set $\theta = 0.25$. We have verified that the calculations are not sensitive to θ selection:

the results do not vary significantly for the θ threshold ranging from 0.01 to 0.25. With this categorisation, the system can be mapped into a bipartite network. We will take one set of nodes composed of investors and the other set composed of the trading days.

The states b, s and bs of investor i are indicated as i_b, i_s and i_{bs}, respectively. There are nine possible combinations of the three trading states between investors i and j: (i_b, j_b), (i_b, j_s), (i_b, j_{bs}), (i_s, j_b), (i_s, j_s), (i_s, j_{bs}), (i_{bs}, j_b), (i_{bs}, j_s) and (i_{bs}, j_{bs}). Because we are focusing on the positive relationship between investors' trading strategies, we further analyse only the situations where both investors have been in a buy state (i_b, j_b), both investors have been in the sell state (i_s, j_s), and both investors have been day traders (i_{bs}, j_{bs}), thus excluding the other six trading state co-occurrences.

Statistically validated networks. With the categorical variables on the trading states, the co-occurrence of the trading states of investors i and j can be identified and statistically validated. First, for each investor, her or his activity period is identified. Second, for an investor pair, the length of a joint trading period is determined, T, which is equal to the number of trading days in an annual data sample for a given security (≈ 250). Then, in the intersection periods of a trader's activity, N_i^P (N_j^P) denotes the number of days when investor i (j) is in a given state {b, s, bs}. Moreover, $N_{i,j}^P$ denotes the number of days when we observe the co-occurrence of the given states for investors i and j. Under the null hypothesis of the random co-occurrences of a state for investors i and j, the probability of observing X co-occurrences of the investigated states for two investors in T observations can be expressed by the hypergeometric distribution $H(X|T, N_i^P, N_j^P)$ (Tumminello et al., 2011a). For each trading state $P = \{b, s, bs\}$, a p-value can be associated as follows:

$$p\left(N_{i,j}^P\right) = 1 - \sum_{X=0}^{N_{i,j}^P - 1} H(X|T, N_i^P, N_j^P) \qquad (2)$$

Using the SVN method, for each security we construct two subsequent year networks. The analysis for each security spans from the initial listing day up to the second year after the IPO. We assign the categorical variables that define the investor's daily trading state, and we select only domestic Finnish investors who have traded an IPO stock at least five days during the first or second year. For each analysed security, we take two consecutive one-year periods of categorised trading states for investors. Taking the projection of the investor set in a year, we obtain an annual monopartite investor network, and two investor networks for consecutive years are obtained for each security.

We adjust the p-thresholds using a false discovery rate (FDR) correction (Benjamini and Hochberg, 1995) by taking the sorted p-values $p_1 < p_2 < \ldots < p_{n_{\text{tests}}}$ in an increasing order and retain those that satisfy $p_i < \alpha \cdot i / n_{\text{tests}}$, $i = 1, \ldots, n_{\text{tests}}$. Here, we apply $\alpha = 0.05$, and n_{tests} equals the total number of observed relationships in a year. All networks are essentially multilink networks, where each link describes the type of trading co-occurrence between an investor pair. This adjustment is needed because there are multiple links and thus multiple tests with a given network. The link between investors i and j is considered to be statistically significant and thus existing if the corresponding p-value, $(N_{i,j}^P)$, is below the FDR-adjusted p-threshold. In this way, we obtain validated networks for the first and second years. As an example, Fig. C.1 in Appendix C shows the first year sorted p-values and the FDR thresholds for Kemira GrowHow links.

Statistically validated clusters: persistence in time. We are interested in the investors' cluster evolution over time. In other words, we want to verify whether investors systematically synchronise their trading strategies with other investors and if such behaviour can be detected in the subsequent year networks. With the community partition for each network, we identify persistent clusters (i.e., clusters that share the same statistically significant component of investors in both the first and the second years after the IPO). Further, we briefly present the method from Marotta et al. (2015).

We are interested in identifying statistically similar clusters that emerged in both years (i.e., clusters with the overexpression of the same investor composition in both clusters, which share nonrandom elements). The probability that X elements in the cluster C_1 of the first year network composed of N_{C_1} elements also appear in the cluster C_2 of the second year composed of N_{C_2} elements under the null hypothesis that the elements in each cluster are randomly selected is given by the hypergeomteric distribution $H(X|N, N_{C_1}, N_{C_2})$, where N is the total number of unique elements over 2 years. By using this distribution, a p-value can be associated with the observed number $N_{C_1 C_2}$ of elements of the cluster C_1 reoccurring in C_2 according to the following equation:

$$p(N_{C_1 C_2}) = 1 - \sum_{X=0}^{N_{C_1 C_2} - 1} H(X|N, N_{C_1}, N_{C_2}) \qquad (3)$$

We reject the null hypothesis if $p(N_{C_1 C_2})$ is smaller than a given adjusted threshold, in which case we say that the cluster C_1 is statistically similar with the cluster C_2. We adjust the statistical threshold using the FDR correction with $\alpha = 0.05$ and the number of tests being equal to the total number of cluster pairs over 2 years that shared at least one common element.

Statistically validated clusters: similarity across securities. Additionally, to check if the same cluster exists over multiple securities, we expand the analysis and further look for statistically significant overlapping clusters across all investigated securities. Because the IPO event is the alignment point in our analysis, we look for the overlapping clusters in the set of first-year networks and the set of second-year networks separately. We again use the method (Eq. (3)) for the cluster overlaps to detect clusters with nonrandomly overlapping elements (investors). To calculate the p-values, we take N equal to the total number of unique investors across all investigated securities in the same year, where N_{C_1} is the number of investors in the cluster C_1, N_{C_2} is the number of investors in the cluster C_2, and $N_{C_1 C_2}$ is the number of common investors in both C_1 and C_2. Again, we adjust the statistical threshold using the FDR correction, where $\alpha = 0.05$ and the number of tests is equal to the total number of cluster pairs within the same year that shared at least one common element.

Overexpression and underexpression of the characterising investor attributes. To describe the investor clusters from the perspective of the attributes, such as postal code, age, gender or the type of organisation, we again use the hypergeometric test for identifying nonrandom overlap (Tumminello et al., 2011b). Once we obtain a system of N elements partitioned into clusters (communities), we want to characterise each cluster C of N_C elements. Each element of the system has a certain number of attributes from a specific class. Here, we want to see if the number of elements in the cluster with a specific attribute value is significantly larger than randomly selecting the elements from the total system elements. For each attribute Q of the system, we test if Q is over-expressed in the cluster C. The probability that X elements in cluster C have the attribute Q under the null hypothesis that the elements in the cluster are randomly selected is given by the hypergeomteric distribution $H(X|N, N_C, N_Q)$, where N_Q is the total number of elements in the system with attribute Q. By using this distribution, a p-value can be associated with the observed number $N_{C,Q}$ of elements in cluster C that have the attribute Q analogously with Eq. (3). We reject the null hypothesis if the p-value is smaller than a given FRD-adjusted p-threshold, and we then say that the attribute Q is overexpressed in cluster C. In the FDR-adjustment, the number of tests is equal to the total number of unique attribute values over all attribute classes and all clusters in a network.

Alternatively, the attribute's Q underexpression can also be tested. Here, we want to see if the number of elements in the cluster with a specific attribute value is significantly lower than randomly selecting the elements from the total system elements. The probability under the null hypothesis that the value of an attribute Q in a cluster C is smaller than the observed value in the system can be obtained from the left tail of the hypergeometric distribution, as follows:

$$p_u(N_{C,Q}) = \sum_{X=0}^{N_{C,Q}} H(X|N, N_C, N_Q) \qquad (4)$$

Again, if $p_u(N_{C,Q})$ is smaller than a given FDR-adjusted p-threshold, we say that the attribute Q is underexpressed in cluster C. We used the same setting for the FDR correction.

Results
Using the SVN methodology, for each of the 69 securities we infer b, s and bs trading state networks for the first and the second year after their IPO dates. In order to identify investor clusters we start by aggregating the networks for all three possible joint-trading states into one weighted network. Each link in the network is given the weight $w \in \{1, 2, 3\}$ depending on how many validated trading states have been observed for a given investor pair[3]. Finally, for each weighted network we identify clusters using Infomap community detection algorithm[4] (Rosvall and Bergstrom, 2008). Identified communities are locally dense connected subgraphs in a network that play an important role in understanding a system's topology. In the current paper, communities represent investor clusters that are timing their trades synchronously throughout the year. Table 3 summarises the number of observed clusters during the first and the second year. For example, during the first year, 54 investor clusters were identified in the security's Kemira GrowHow (FI0009012843) networks, while during the second year 64 clusters were formed. Figure 1a, b visualise the later Infomap clusters for the first-year and second-year networks.

5

Table 3 Investor network clusters' statistics

ISIN	Unique clusters Y1	Unique clusters Y2	Persisting clusters Y1 → Y2	Unique investors Y1	Active investors Y1	Unique investors Y2	Active investors Y2	Active investors Y1 ∩ Y2	Median cluster size
FI0009004881	1/14 (7%)	1/10 (10%)	10 → 10	714	107	875	111	38	3 (2)
FI0009800346	0/28 (0%)	4/40 (10%)	3 → 3	3747	251	4410	316	117	3 (4)
FI0009800320	1/8 (13%)	0/21 (0%)		1867	106	2365	146	48	4 (3)
FI0009900336	0/0 (0%)	0/11 (0%)		441	22	1465	90	10	0 (4)
FI0009005318	0/5 (0%)	0/13 (0%)	1 → 1	545	51	653	74	19	2 (2)
FI0009800643	0/5 (0%)	0/32 (0%)	1 → 1	536	50	2730	261	23	3 (3)
FI0009005870	0/6 (0%)	0/14 (0%)	1 → 1	947	68	734	67	22	2 (2)
FI0009005953	1/14 (7%)	0/10 (0%)	1 → 1	2108	131	2509	104	46	2 (2)
FI0009005961	0/30 (0%)	4/65 (6%)	11 → 13	3570	280	5555	501	159	3 (4)
FI0009005987	8/82 (10%)	11/110 (10%)	29 → 32	11,093	678	15,139	906	314	3 (4)
FI0009006381	0/32 (0%)	0/38 (0%)	2 → 2	5277	226	4085	316	74	3 (3)
FI0009006415	0/11 (0%)	0/5 (0%)		1258	85	601	35	9	3 (2)
FI0009005805	1/39 (3%)	1/12 (8%)		4835	294	1560	102	42	3 (3)
FI0009006589	0/11 (0%)	0/1 (0%)		1853	92	274	20	7	3 (2)
FI0009006621	2/66 (3%)	1/63 (2%)	6 → 7	14,372	469	11,033	565	155	4 (5)
FI0009006738	0/38 (0%)	3/71 (4%)	2 → 2	5789	305	7261	542	84	2 (3)
FI0009006696	0/5 (0%)	0/5 (0%)	1 → 1	1073	72	672	39	15	3 (3)
FI0009006761	0/8 (0%)	0/7 (0%)		1000	52	1252	55	17	4 (3)
FI0009007017	0/0 (0%)	1/4 (25%)		534	29	888	38	8	0 (2)
FI0009007025	0/9 (0%)	0/20 (0%)	1 → 1	1025	68	1951	125	29	2 (2)
FI0009007066	0/8 (0%)	0/3 (0%)		1984	60	341	28	14	3 (2)
FI0009006829	0/5 (0%)	1/15 (7%)	1 → 1	1902	56	3212	136	27	2 (3)
FI0009007215	3/10 (30%)	1/25 (4%)	1 → 1	1673	83	2952	156	21	5 (2)
FI0009007264	7/113 (6%)	52/475 (11%)	68 → 98	8067	854	43745	4288	482	3 (4)
FI0009007371	20/272 (7%)	136/818 (17%)	227 → 389	33,419	2633	82,702	10,050	1467	6 (7)
FI0009007355	0/1 (0%)	0/0 (0%)		747	13	774	32	6	2 (0)
FI0009007132	8/111 (7%)	0/54 (0%)	5 → 6	22,617	943	18,156	514	218	5 (6)
FI0009007629	0/2 (0%)	0/12 (0%)		596	53	1426	91	14	2 (3)
FI0009801286	0/15 (0%)	0/1 (0%)		3191	115	968	38	16	3 (2)
FI0009007553	6/81 (7%)	2/99 (2%)	17 → 16	9492	657	9449	997	182	4 (3)
FI0009007728	1/44 (2%)	1/31 (3%)	1 → 1	7219	303	3355	322	76	4 (2)
FI0009007546	0/12 (0%)	0/0 (0%)		232	26	97	7	5	2 (0)
FI0009007686	2/5 (40%)	0/1 (0%)		753	45	417	22	3	2 (2)
FI0009007694	4/27 (15%)	0/2 (0%)		2774	176	1909	88	28	3 (2)
FI0009006886	2/11 (18%)	0/2 (0%)		1849	103	819	39	11	2 (2)
FI0009007819	4/135 (3%)	2/123 (2%)	13 → 14	16608	1223	8287	1117	329	6 (5)
FI0009007835	2/41 (5%)	0/34 (0%)	4 → 4	6320	283	3910	235	85	3 (3)
FI0009007884	11/136 (8%)	2/100 (2%)	4 → 3	58,326	1049	20,940	934	277	3 (5)
FI0009008080	1/11 (9%)	0/9 (0%)	1 → 1	1296	91	1094	102	27	2 (3)
FI0009007918	1/87 (1%)	1/79 (1%)	9 → 10	7136	802	7199	1051	256	6 (6)
FI0009801310	15/169 (9%)	4/218 (2%)	79 → 84	30,706	2328	20,898	2497	672	2 (7)
FI0009008221	38/337 (11%)	8/252 (3%)	198 → 172	35,617	3454	17,235	2541	985	7 (6)
FI0009902530	6/62 (10%)	6/62 (10%)	8 → 8	25,808	572	12,223	614	200	6 (5)
FI0009008924	0/12 (0%)	1/2 (50%)		2644	151	1070	74	27	4 (2)
FI0009008833	0/2 (0%)	0/3 (0%)		674	35	615	45	14	2 (2)
FI0009009146	2/28 (7%)	0/1 (0%)		3444	285	1188	70	34	2 (2)
FI0009009054	1/8 (13%)	1/8 (13%)	1 → 1	1832	135	1120	112	34	2 (2)
FI0009009633	4/19 (21%)	1/14 (7%)		2847	224	1771	107	33	2 (3)
FI0009009567	4/12 (33%)	0/8 (0%)		1614	112	1380	96	28	2 (2)
FI0009008270	6/53 (11%)	0/12 (0%)		5743	442	2558	137	64	3 (2)
FI0009009674	1/24 (4%)	1/13 (8%)	4 → 4	2033	226	3746	161	88	4 (6)
FI0009009377	1/5 (20%)	1/6 (17%)		2779	151	2329	133	26	2 (2)
FI0009010219	1/8 (13%)	1/5 (20%)		1438	84	1078	66	26	4 (2)
FI0009010854	0/2 (0%)	0/14 (0%)		573	41	1164	114	19	4 (4)
FI0009010862	0/5 (0%)	1/10 (10%)		879	66	1604	99	16	2 (4)
SE0000667925	22/120 (18%)	7/129 (5%)	8 → 9	17,759	1186	21,725	1580	476	4 (7)
FI0009011005	1/4 (25%)	0/0 (0%)		576	43	176	9	3	2 (0)
FI0009012843	5/54 (9%)	2/64 (3%)	5 → 5	8047	464	9609	818	183	5 (6)
FI0009013296	33/262 (13%)	42/336 (13%)	180 → 221	24,350	3518	22,421	3603	1555	7 (7)
FI0009013429	12/133 (9%)	5/89 (6%)	26 → 24	9945	1016	6012	691	326	3 (3)
FI0009013312	3/26 (12%)	0/13 (0%)	2 → 1	2667	224	1204	125	52	4 (4)

Table 3 (continued)

ISIN	Unique clusters Y1	Unique clusters Y2	Persisting clusters Y1 → Y2	Unique investors Y1	Unique investors Y2	Active investors Y1	Active investors Y2	Unique investors Y1 ∩ Y2	Active investors Y1 ∩ Y2	Median cluster size
FI0009013403	11/112 (10%)	4/92 (4%)	45 → 44	10234	7952	1084	769	7952	409	2 (3)
FI0009013924	3/29 (10%)	1/11 (9%)		1804	2104	192	235	2104	43	2 (2)
FI0009010391	3/37 (8%)	6/50 (12%)	3 → 3	8822	5915	306	434	5915	114	3 (5)
FI0009013593	3/17 (18%)	0/0 (0%)		2345	870	162	9	870	4	3 (0)
FI0009014344	2/3 (67%)	0/0 (0%)		2815	1128	83	34	1128	15	2 (0)
FI0009014351	11/56 (20%)	0/6 (0%)		10,338	3267	399	135	3267	69	2 (2)
FI0009012413	11/22 (50%)	5/22 (23%)	1 → 1	4788	6627	243	237	6627	62	2 (5)
FI0009015309	9/64 (14%)	3/29 (10%)	1 → 1	6748	2208	521	187	2208	95	3 (4)

Columns 'Unique clusters Y1 (Y2)' show the number of asset-specific investor clusters over all clusters observed in the first (second) year networks. Here, asset-specific investor clusters are defined as those that were not observed in other IPO networks. The number in the brackets () shows the ratio in percentage. 'Persisting clusters Y1 → Y2' shows the number of clusters with statistically significant overlaps in the first and the second years. Note that clusters split and merge, and thus the number of persisted clusters is not necessarily the same for both years. Columns 'Unique investors Y1 (Y2)' show the total number of investors per ISIN in a year. Columns 'Active investors Y1 (Y2)' show the total number of investors who traded at least 5 days per ISIN in a year. The column 'Active investors Y1 ∩ Y2' shows the total number of investors who traded at least 5 days per ISIN in both first and second year after IPO. The column 'Median cluster size' shows the number of investors in a median-sized cluster in the first (second) year network. ISINs from the error-free set are marked in bold

Next, for each security, we detect clusters with a statistically significant investor overlap between the first and second year. The summary of statistically validated cluster time persistence for all 69 securities is presented in the fourth column of Table 3. For example, in the Kemira GrowHow networks, only 5 of the 54, i.e. 9% of clusters identified in the first year were observed in the second year. Figure 1c, d display those five clusters that persisted over the first two years after the IPO. The observation in the example that only a small number of clusters persist into the second year is consistent for the majority of the analysed IPO securities. However, there are several securities for which more than a half of the first year clusters persist into the following year. A sample of time persistent clusters and their composition in terms of investor attributes are visualised in the Appendix Figs. A.1 and A.2.

By calculating the fraction of clusters that do not persist into the second year, we observe that over all 69 securities on average 88% of the first-year clusters are not observed in the following year, while the same number falls to 78% for mature company networks inferred during the same periods (more details about the comparison to mature companies are provided in the following section). This observation can suggest the existence of IPO trading strategy-related clusters that form exclusively during the first year after the IPO date and break up in the following year.

Additionally, we analyse cluster overlap across multiple securities, separately for the first-year and second-year networks. The second and third columns in Table 3 show the number of asset-specific clusters over the total number of communities in the first and second year. Here, by asset-specific clusters, we refer to the clusters that are not observable within investor networks of the same year for other IPO securities in our investigated 69 security universe. The number of observed asset-specific clusters is rather small and is around 15% (9%) during the first (second) year averaged over all 69 securities. This means that the majority of investor clusters are found to be present in multiple securities, i.e. they execute synchronised trading strategies over multiple IPOs. Note that this cluster synchronisation is observed even though the network inference periods are not aligned in time. The observed decrease in the overall percentage of asset-specific clusters hints that during the second year after IPO more clusters use non-IPO related trading strategies. This is later supported by the mature security analysis (see the next section and Tables 4 and 5). Figure A.3 in Appendix A shows a sample of clusters with statistically significant investor overlap across multiple securities.

Combining the previous results together, we observe persistent clusters that emerge in investor networks over multiple securities. Figure 2 explains the visualisation of a cluster in this study and Fig. 3 shows a sample of clusters that both, overlap over time and over multiple securities. In the figure, the top (bottom) row of the group refers to the first- (second-) year clusters. Moreover, the downward arrows associate statistically similar clusters in the first-year and second-year networks. The arrows between the clusters in the same year after IPO are omitted for the simplification of the visualisation. Notably, even if some of the clusters are not persistent over time, quite often they appear over different securities.

Next, we analyse the overexpression and underexpression of the investor attributes in the identified investor clusters. We say that a cluster is overexpressing (underexpressing) an attribute if the number of investors in the cluster with that particular attribute is significantly higher (lower) than could be expected under the null model defined in the "Dataset and Methodology" section. We are primarily interested in the sector code attribute analysis, where investors can be assigned households, nonfinancial
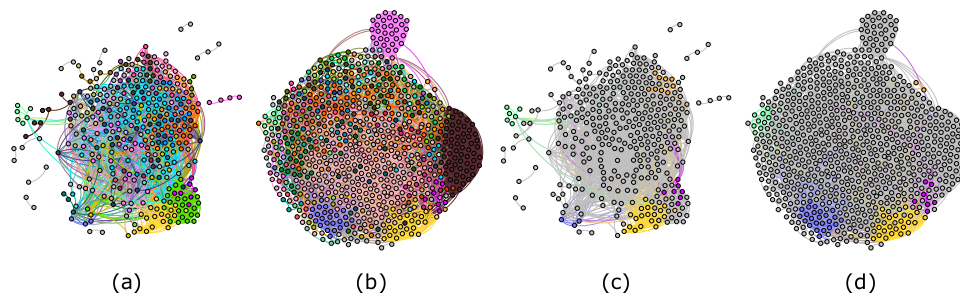
Fig. 1 Infomap clusters and their evolution for Kemira GrowHow (FI0009012843). Community detection is used with weighted links based on the total number of buy state, sell state, and day trade link types between two investors. **a** FDR: 54 clusters, first year after IPO, **b** FDR: 64 clusters, second year after IPO, **c**, **d** show five statistically significant overlapping clusters in both years. Node position is fixed. The colours of reoccurring clusters in all graphs coincide. In **a**, **b**, each cluster has a unique colour, with the exception of those with fewer than four elements, which are coloured in grey

Table 4 Five mature companies with the highest number of transactions in HSE		
ISIN	**Company name**	**IPO date**
FI0009000681	Nokia	1981-04-01
FI0009000277	Tieto	1984-06-01
FI0009000665	Metsä Board B	1987-01-02
FI0009002943	Raisio V	1989-04-25
FI0009003727	Wärtsilä	1991-01-17

corporations, financial and insurance corporations, government, nonprofit institutions, and the rest of the world attribute. Additionally, we test whether or not attributes related to gender, age or geographical location are over expressed or underexpressed[5].

Over all 69 securities, we identify 115 (28) investor clusters with 182 (40) overexpressed (underexpressed) attributes during the first year after the IPO, and 130 (44) investor clusters with 236 (70) overexpressed (underexpressed) attributes during the second year. The number of overexpressed (underexpressed) attributes is larger than the number of investor clusters, because each cluster can overexpress (underexpress) more than one attribute. The overexpressed clusters are observed over 28 different securities during the first year after IPO and for 27 different securities during the second year after IPO. As for the underexpressed clusters, they are observed over 16 securities during the first year and 20 securities during the second year after IPO.

In order to present the attribute analysis in a concise way, we use the fact that the same clusters appear over multiple securities and assign overexpressed (underexpressed) investor clusters into groups if they are statistically similar. Figure 4 presents the resulting sector code attribute overexpressing investor cluster networks for the first and second years after respective IPOs. In the figure, nodes on the left (right) hand side of the vertical dashed line represent investor clusters observed in the first (second) year after IPO. Statistically similar cluster nodes are connected with links and dotted lines circle network components. Each connected component in the network relates to a group of clusters with a statistically similar investor composition. The dashed lines crossing from the left to the right-hand-side indicate that there is a statistical similarity for some of the clusters in the components between the first and the second year.

Tables B.1 and B.2 in the Appendix summarise the overexpressed and underexpressed cluster attributes for each investor cluster component in Figs. 4 and 5. The largest first and second year components in Fig. 4 are over-represented by finance-insurance and general government institutions, as well as nonprofit organisations. Moreover, the same components

underexpress Household sector (see Fig. 5), further supporting their institutional profile. In addition, the same components overexpress location attributes, in particular Helsinki and South-West regions (see Fig. B.1 in the Appendix). Investor clusters with an overexpression of a geographical attribute could be observed because of some locally present investment strategy, for example an investor club, or some other means of local information transfer. Overall, the results show that the largest cluster components mainly contain institutions that are timing their trades similarly in a year. Compared with household investors, institutional traders form robust clusters, that execute similar trade-timing strategies over multiple IPOs, both during the first and the second year after the IPO date. Our findings thus support the studies that provide evidence of institutional herding (Nofsinger and Sias, 1999; Sias, 2004). Some of the financial institutions, such as pension insurance companies, are driven by the same legislation and portfolio restrictions, which can lead to the same trading strategies. Alternatively, traders working for financial institutions have mutual and/or joint private information channels, leading to similar trade timing. The third explanation is that they react to public news in similar ways.

Do clusters of IPO investors exist with mature companies?. To verify if our identified clusters are just IPO-related or if they exist with mature companies[6] as well, we compare the clusters of the new-to-the-market stocks with five mature companies (see Table 4). For each mature security, just like previously for IPOs, we construct SVNs and identify investor clusters with Infomap algorithm. When constructing the first-year and second-year networks, the periods are aligned with respective IPO dates. This way we construct 345 (69 × 5) networks for each year. Next, we analysed the overlaps between mature security investor clusters and the investor clusters inferred with the data from IPOs, to answer the question if the investor clusters identified with IPO securities exist with a mature company. When statistically validating overlaps between mature and IPO security investor network clusters, we use the total number of cluster pairs with at least one investor in common between an IPO and all five mature securities as the number of tests for the FDR correction. Table 5 shows the number of statistically similar clusters between the IPO and mature securities, as well as the total number of clusters observed in the IPO and the mature security during the exactly same period. Here we observe that on average over all investigated IPO securities only 16% of IPO clusters are not observed in one of the five investigated mature securities during the first year after IPO, and 13% during the second year. By looking at the same table, we can see that only a fraction of total clusters observed in mature securities are also observed in IPO security networks. It can be because not all investors who trade mature

Table 5 IPO and mature companies investor clusters overlap

ISIN	IPO date	Year	FI0009000681	FI0009003727	FI0009000665	FI0009000277	FI0009002943	# Unique cl.
FI0009004881	1995-01-12	Y1	2 (2) {118}	1 (1) {17}	2 (2) {11}	{8}	3 (2) {11}	9/14 (64%)
		Y2	9 (11) {142}	6 (8) {33}	7 (9) {34}	1 (1) {5}	6 (6) {57}	1/10 (10%)
FI0009800346	1995-05-11	Y1	21 (25) {146}	13 (12) {19}	12 (12) {20}	7 (7) {15}	16 (14) {29}	3/28 (11%)
		Y2	26 (32) {133}	18 (16) {30}	23 (23) {46}	1 (1) {2}	15 (19) {67}	6/40 (15%)
FI0009800320	1995-05-11	Y1	5 (7) {146}	3 (3) {19}	2 (3) {20}	1 (2) {15}	4 (5) {29}	3/8 (38%)
		Y2	10 (14) {133}	12 (14) {30}	7 (8) {46}	1 (1) {2}	11 (10) {67}	5/21 (24%)
FI0009905318	1995-06-01	Y1	3 (3) {152}	{21}	2 (3) {21}	2 (3) {15}	1 (1) {29}	2/5 (40%)
		Y2	9 (10) {123}	7 (8) {34}	5 (6) {38}	1 (1) {3}	8 (9) {62}	1/13 (8%)
FI0009900336	1995-06-01	Y1	{152}	{21}	{21}	{15}	{29}	0
		Y2	6 (8) {123}	6 (7) {34}	5 (8) {38}	{3}	6 (6) {62}	1/11 (9%)
FI0009800643	1995-09-04	Y1	5 (8) {176}	1 (2) {21}	2 (3) {27}	{10}	3 (3) {45}	0/5 (0%)
		Y2	19 (31) {133}	17 (19) {38}	21 (22) {38}	1 (1) {2}	14 (15) {67}	6/32 (19%)
FI0009005870	1996-03-27	Y1	4 (4) {136}	2 (3) {32}	2 (2) {41}	1 (1) {2}	2 (3) {65}	1/6 (17%)
		Y2	13 (15) {190}	6 (7) {26}	10 (9) {30}	1 (1) {4}	9 (10) {56}	0/14 (0%)
FI0009005987	1996-05-02	Y1	56 (54) {128}	23 (24) {30}	26 (24) {43}	{2}	33 (28) {72}	18/82 (22%)
		Y2	82 (91) {225}	27 (22) {34}	40 (30) {33}	6 (4) {9}	43 (37) {58}	20/110 (18%)
FI0009005961	1996-05-02	Y1	21 (29) {128}	13 (18) {30}	19 (19) {43}	1 (1) {2}	19 (18) {72}	2/30 (7%)
		Y2	51 (73) {225}	29 (26) {34}	36 (26) {33}	3 (3) {9}	31 (33) {58}	7/65 (11%)
FI0009005953	1996-05-02	Y1	11 (12) {128}	7 (8) {30}	9 (10) {43}	{2}	8 (9) {72}	2/14 (14%)
		Y2	9 (10) {225}	4 (6) {34}	8 (7) {33}	1 (1) {9}	6 (8) {58}	0/10 (0%)
FI0009006381	1997-04-03	Y1	26 (34) {179}	5 (5) {27}	14 (13) {33}	2 (2) {4}	19 (18) {54}	4/32 (12%)
		Y2	31 (40) {312}	13 (11) {25}	13 (15) {61}	19 (19) {80}	20 (28) {118}	2/38 (5%)
FI0009006415	1997-04-24	Y1	6 (8) {210}	3 (4) {32}	5 (6) {32}	{4}	8 (9) {56}	3/11 (27%)
		Y2	4 (6) {333}	3 (3) {24}	2 (2) {54}	4 (4) {30}	4 (4) {30}	0/5 (0%)
FI0009005805	1997-06-09	Y1	28 (44) {240}	16 (16) {35}	22 (19) {39}	8 (7) {14}	19 (22) {53}	4/39 (10%)
		Y2	9 (13) {356}	3 (3) {29}	4 (7) {62}	5 (7) {93}	5 (9) {160}	2/12 (17%)
FI0009006589	1997-06-17	Y1	9 (14) {245}	7 (6) {39}	6 (7) {37}	4 (4) {14}	6 (11) {51}	2/11 (18%)
		Y2	1 (1) {377}	{29}	{62}	{91}	{162}	0/1 (0%)
FI0009006621	1997-11-25	Y1	60 (81) {273}	23 (19) {27}	24 (20) {46}	31 (24) {43}	44 (39) {62}	3/66 (5%)
		Y2	59 (123) {547}	26 (20) {27}	24 (21) {46}	50 (55) {90}	53 (101) {236}	2/63 (3%)
FI0009006738	1997-11-26	Y1	30 (46) {275}	11 (10) {28}	18 (16) {48}	16 (15) {45}	20 (21) {62}	3/38 (8%)
		Y2	54 (107) {557}	12 (12) {27}	14 (14) {47}	43 (48) {97}	57 (86) {241}	5/71 (7%)
FI0009006696	1997-12-02	Y1	4 (5) {274}	1 (1) {28}	1 (1) {48}	2 (3) {46}	4 (5) {65}	0/5 (0%)
		Y2	5 (8) {602}	3 (3) {25}	4 (4) {50}	3 (3) {88}	5 (9) {210}	0/5 (0%)
FI0009006761	1997-12-09	Y1	8 (14) {291}	6 (8) {26}	8 (10) {49}	7 (9) {45}	6 (8) {68}	0/8 (0%)
		Y2	7 (15) {606}	6 (5) {25}	4 (4) {41}	6 (10) {90}	7 (18) {229}	0/7 (0%)
FI0009007017	1998-04-01	Y1	{316}	{28}	{62}	{81}	{117}	0
		Y2	3 (4) {765}	2 (2) {32}	{44}	1 (1) {87}	3 (4) {232}	1/4 (25%)
FI0009007025	1998-04-01	Y1	7 (9) {316}	6 (6) {28}	5 (6) {62}	6 (6) {81}	6 (7) {117}	2/9 (22%)
		Y2	15 (24) {765}	10 (11) {32}	10 (9) {44}	10 (16) {87}	14 (22) {232}	3/20 (15%)
FI0009007066	1998-04-30	Y1	6 (9) {326}	5 (7) {29}	4 (5) {56}	5 (8) {87}	6 (11) {142}	0/8 (0%)
		Y2	3 (2) {836}	2 (1) {35}	2 (3) {42}	2 (2) {102}	2 (2) {223}	0/3 (0%)
FI0009006829	1998-06-01	Y1	4 (5) {341}	3 (3) {30}	2 (2) {60}	2 (2) {89}	1 (1) {153}	0/5 (0%)
		Y2	9 (14) {906}	4 (4) {38}	7 (12) {46}	5 (6) {96}	7 (7) {210}	5/15 (33%)
FI0009007215	1998-08-03	Y1	6 (12) {437}	1 (2) {30}	5 (4) {57}	4 (6) {93}	6 (11) {179}	3/10 (30%)
		Y2	15 (24) {615}	6 (6) {39}	5 (6) {45}	13 (14) {103}	11 (12) {185}	6/25 (24%)
FI0009007264	1998-09-15	Y1	94 (134) {469}	16 (13) {25}	15 (15) {48}	53 (45) {94}	75 (81) {204}	9/113 (8%)
		Y2	376 (361) {587}	37 (31) {41}	32 (23) {41}	163 (105) {125}	237 (144) {200}	64/475 (13%)
FI0009007371	1998-11-17	Y1	248 (350) {515}	41 (24) {26}	49 (35) {45}	148 (83) {90}	206 (187) {233}	13/272 (5%)

Table 5 (continued)

ISIN	IPO date	Year	FI0009000681	FI0009003727	FI0009000665	FI0009000277	FI0009002943	# Unique cl.
FI0009007355	1998-12-04	Y2	683 (550) {645}	47 (31) {36}	49 (39) {48}	252 (108) {112}	300 (156) {183}	99/818 (12%)
		Y1	1 (1) {612}	{26}	{47}	{91}	1 (1) {226}	0/1 (0%)
FI0009007132	1998-12-18	Y2	{611}	{33}	{44}	{115}	{175}	0
		Y1	91 (153) {629}	26 (16) {24}	32 (23) {40}	67 (60) {87}	85 (125) {250}	9/111 (8%)
FI0009007629	1999-03-01	Y2	35 (45) {523}	18 (19) {33}	29 (25) {49}	31 (34) {112}	38 (58) {166}	4/54 (7%)
		Y1	2 (3) {689}	1 (1) {33}	{41}	{81}	1 (1) {26}	0/2 (0%)
FI0009801286	1999-03-15	Y2	6 (7) {553}	3 (3) {34}	4 (4) {56}	8 (9) {97}	4 (4) {130}	3/12 (25%)
		Y1	14 (23) {731}	6 (5) {33}	5 (5) {41}	10 (13) {81}	15 (28) {262}	0/15 (0%)
FI0009007553	1999-03-23	Y2	1 (1) {580}	1 (1) {32}	{54}	1 (1) {96}	1 (1) {124}	0/1 (0%)
		Y1	65 (108) {745}	28 (19) {35}	20 (17) {48}	44 (44) {83}	61 (77) {242}	8/81 (10%)
FI0009007728	1999-04-06	Y2	65 (70) {551}	12 (10) {36}	14 (13) {52}	56 (48) {100}	42 (26) {128}	11/99 (11%)
		Y1	36 (78) {811}	16 (20) {33}	8 (10) {46}	27 (32) {92}	36 (59) {230}	5/44 (11%)
FI0009007546	1999-04-19	Y2	21 (22) {530}	10 (11) {32}	10 (10) {53}	18 (22) {90}	11 (11) {125}	5/31 (16%)
		Y1	1 (1) {850}	1 (1) {35}	1 (1) {46}	{100}	1 (1) {225}	0/1 (0%)
FI0009007694	1999-05-03	Y2	{447}	{29}	{53}	{84}	{124}	0
		Y1	23 (43) {855}	11 (13) {37}	4 (6) {44}	18 (28) {99}	19 (29) {228}	4/27 (15%)
FI0009007686	1999-05-03	Y2	1 (2) {454}	1 (2) {34}	{51}	2 (3) {93}	{131}	0/2 (0%)
		Y1	3 (7) {855}	{37}	{44}	3 (3) {99}	4 (7) {228}	1/5 (20%)
FI0009006886	1999-06-08	Y2	{454}	{34}	{51}	{93}	{131}	1/1 (100%)
		Y1	6 (6) {898}	2 (2) {33}	3 (3) {47}	5 (6) {97}	8 (10) {210}	2/11 (18%)
FI0009007819	1999-06-28	Y2	2 (2) {433}	1 (1) {34}	1 (1) {53}	1 (1) {90}	1 (1) {130}	0/2 (0%)
		Y1	122 (277) {859}	38 (23) {36}	29 (28) {47}	86 (63) {92}	98 (115) {203}	5/135 (4%)
FI0009007835	1999-07-01	Y2	101 (124) {513}	31 (20) {34}	35 (33) {51}	81 (67) {98}	63 (58) {123}	6/123 (5%)
		Y1	35 (64) {791}	22 (22) {34}	16 (18) {48}	30 (38) {95}	30 (44) {186}	4/41 (10%)
FI0009007884	1999-07-01	Y2	30 (42) {596}	19 (17) {34}	16 (16) {51}	23 (31) {103}	18 (27) {124}	1/34 (3%)
		Y1	104 (171) {791}	31 (23) {34}	27 (26) {48}	74 (57) {95}	82 (75) {186}	13/136 (10%)
FI0009008080	1999-10-01	Y2	82 (145) {596}	32 (24) {34}	36 (29) {51}	73 (73) {103}	57 (61) {124}	6/100 (6%)
		Y1	9 (16) {597}	3 (4) {41}	2 (1) {42}	6 (7) {120}	7 (8) {197}	2/11 (18%)
FI0009007918	1999-10-27	Y2	9 (15) {775}	5 (7) {34}	5 (6) {53}	7 (10) {89}	7 (7) {111}	0/9 (0%)
		Y1	81 (167) {631}	20 (21) {41}	16 (16) {47}	58 (64) {120}	58 (64) {192}	2/87 (2%)
FI0009801310	1999-11-09	Y2	67 (150) {763}	21 (22) {37}	29 (21) {51}	51 (61) {91}	33 (44) {101}	6/79 (8%)
		Y1	136 (149) {630}	22 (18) {38}	20 (16) {50}	85 (58) {113}	69 (67) {195}	14/169 (8%)
FI0009008221	1999-12-13	Y2	183 (318) {765}	34 (21) {38}	35 (30) {45}	90 (67) {86}	93 (70) {108}	22/218 (10%)
		Y1	280 (292) {536}	37 (23) {34}	39 (29) {46}	151 (91) {112}	146 (108) {171}	32/337 (9%)
FI0009902530	2000-01-31	Y2	215 (356) {843}	50 (27) {39}	52 (30) {42}	115 (74) {89}	78 (60) {104}	22/252 (9%)
		Y1	46 (61) {489}	17 (22) {37}	21 (21) {55}	39 (49) {115}	29 (46) {158}	5/62 (8%)
FI0009008924	2000-05-24	Y2	55 (118) {915}	34 (29) {49}	37 (31) {46}	40 (47) {90}	20 (19) {83}	2/65 (3%)
		Y1	10 (15) {439}	4 (4) {33}	3 (3) {58}	6 (7) {86}	7 (8) {130}	0/12 (0%)
FI0009008833	2000-05-24	Y2	1 (1) {856}	{61}	{47}	{80}	{70}	1/2 (50%)
		Y1	{439}	{33}	{33}	{130}	{130}	1/2 (50%)
FI0009009146	2000-07-04	Y2	2 (2) {856}	2 (2) {61}	2 (3) {47}	1 (1) {86}	{70}	0/3 (0%)
		Y1	18 (28) {586}	4 (4) {35}	7 (8) {54}	15 (19) {98}	13 (15) {123}	5/28 (18%)
FI0009009054	2000-07-05	Y2	1 (1) {849}	1 (1) {61}	1 (1) {46}	1 (1) {78}	{66}	0/1 (0%)
		Y1	5 (8) {600}	2 (4) {35}	1 (1) {54}	4 (6) {102}	3 (3) {128}	1/8 (12%)
FI0009009633	2000-11-01	Y2	3 (4) {849}	3 (4) {60}	3 (4) {47}	2 (5) {78}	4 (4) {69}	0/8 (0%)
		Y1	12 (16) {762}	3 (4) {37}	3 (4) {47}	6 (6) {92}	7 (6) {107}	5/19 (26%)
FI0009009567	2000-12-19	Y2	7 (9) {548}	2 (2) {64}	1 (1) {50}	4 (4) {93}	2 (2) {53}	5/14 (36%)
		Y1	8 (9) {818}	5 (5) {40}	4 (2) {47}	8 (8) {93}	1 (1) {101}	3/12 (25%)
FI0009008270	2000-12-22	Y2	2 (4) {468}	4 (4) {58}	3 (3) {59}	5 (5) {101}	2 (2) {52}	0/8 (0%)
		Y1	45 (68) {858}	16 (12) {39}	17 (11) {48}	23 (26) {91}	17 (18) {91}	5/53 (9%)

Table 5 (continued)

ISIN	IPO date	Year	FI0009000681	FI0009003727	FI0009000665	FI0009000277	FI0009002943	# Unique cl.
FI0009009674	2001-01-30	Y2	4 (4) {477}	5 (5) {59}	2 (3) {58}	6 (5) {95}	3 (3) {52}	2/12 (17%)
		Y1	18 (36) {900}	19 (24) {47}	19 (25) {52}	19 (31) {91}	11 (11) {81}	1/24 (4%)
FI0009009377	2001-04-02	Y2	11 (24) {499}	11 (19) {58}	11 (22) {52}	12 (35) {105}	4 (6) {41}	1/13 (8%)
		Y1	2 (2) {795}	2 (2) {54}	1 (1) {49}	2 (2) {88}	1 (2) {102}	2/5 (40%)
FI0009010219	2001-04-02	Y2	1 (1) {526}	{42}	2 (2) {53}	1 (1) {118}	{14}	3/6 (50%)
		Y1	5 (11) {795}	4 (4) {54}	4 (5) {49}	5 (5) {88}	4 (6) {102}	1/8 (12%)
FI0009010862	2001-10-01	Y2	2 (2) {526}	1 (1) {42}	{53}	1 (1) {118}	{14}	2/5 (40%)
		Y1	1 (1) {639}	{164}	4 (5) {55}	3 (4) {101}	1 (1) {51}	0/5 (0%)
FI0009010854	2001-10-01	Y2	2 (2) {639}	3 (3) {39}	5 (8) {67}	7 (8) {94}	{14}	2/10 (20%)
		Y1	3 (3) {164}	2 (3) {62}	2 (2) {55}	2 (3) {101}	{51}	0/2 (0%)
SE0000667925	2002-12-09	Y1	42 (44) {248}	18 (18) {47}	29 (27) {57}	57 (51) {91}	20 (18) {35}	33/120 (28%)
		Y2	96 (125) {327}	52 (45) {58}	60 (49) {70}	66 (63) {86}	45 (37) {73}	11/129 (9%)
SE0000110165	2003-09-04	Y1	1 (1) {339}	{44}	2 (3) {66}	1 (1) {93}	1 (1) {77}	2/4 (50%)
		Y2	{173}	{56}	{105}	{100}	{73}	0
FI0009012843	2004-10-18	Y1	23 (41) {187}	37 (42) {71}	34 (44) {107}	36 (53) {99}	20 (27) {71}	7/54 (13%)
		Y2	41 (65) {438}	36 (42) {159}	42 (54) {133}	31 (33) {113}	6 (5) {25}	6/64 (9%)
FI0009013296	2005-04-21	Y1	153 (138) {211}	144 (87) {102}	76 (60) {95}	109 (65) {81}	31 (28) {63}	36/262 (14%)
		Y2	237 (207) {273}	165 (134) {185}	106 (76) {107}	148 (89) {110}	34 (30) {60}	41/336 (12%)
FI0009013312	2005-06-01	Y1	19 (28) {323}	13 (14) {129}	18 (17) {107}	15 (13) {82}	9 (12) {68}	1/26 (4%)
		Y2	6 (9) {266}	6 (7) {171}	7 (8) {95}	8 (10) {94}	5 (4) {50}	1/13 (8%)
FI0009013403	2005-06-01	Y1	75 (85) {323}	81 (75) {129}	39 (37) {107}	56 (45) {82}	22 (15) {68}	10/112 (9%)
		Y2	64 (64) {266}	66 (80) {171}	34 (31) {95}	55 (49) {94}	7 (8) {50}	10/92 (11%)
FI0009013429	2005-06-01	Y1	88 (105) {323}	85 (77) {129}	51 (44) {107}	71 (54) {82}	21 (16) {68}	18/133 (14%)
		Y2	61 (73) {266}	64 (69) {171}	25 (26) {95}	48 (39) {94}	9 (10) {50}	9/89 (10%)
FI0009010391	2006-03-17	Y1	21 (25) {243}	16 (17) {182}	22 (26) {108}	20 (21) {119}	12 (14) {68}	4/37 (11%)
		Y2	22 (26) {370}	17 (19) {121}	28 (37) {118}	26 (34) {101}	9 (8) {36}	6/50 (12%)
FI0009013924	2006-03-17	Y1	19 (22) {243}	13 (12) {182}	17 (22) {108}	16 (17) {119}	5 (5) {68}	3/29 (10%)
		Y2	4 (4) {370}	4 (4) {121}	2 (2) {118}	4 (3) {101}	1 (1) {36}	2/11 (18%)
FI0009013593	2006-04-21	Y1	10 (14) {268}	6 (7) {184}	5 (9) {102}	10 (14) {112}	2 (2) {60}	2/17 (12%)
		Y2	{337}	{127}	{124}	{90}	{32}	0
FI0009014351	2006-07-03	Y1	28 (32) {263}	21 (24) {163}	25 (27) {93}	26 (25) {90}	6 (6) {50}	16/56 (29%)
		Y2	3 (4) {336}	4 (4) {147}	5 (5) {142}	4 (4) {94}	{34}	0/6 (0%)
FI0009014344	2006-07-03	Y1	{263}	{163}	{93}	{90}	{50}	3/3 (100%)
		Y2	{336}	{147}	{142}	{94}	{34}	0
FI0009012413	2007-04-10	Y1	4 (4) {360}	2 (3) {122}	9 (9) {122}	6 (4) {95}	3 (3) {38}	8/22 (36%)
		Y2	6 (7) {315}	11 (13) {143}	9 (16) {161}	10 (13) {62}	{7}	6/22 (27%)
FI0009015309	2007-06-15	Y1	32 (37) {329}	26 (26) {139}	31 (35) {136}	34 (32) {92}	6 (5) {37}	12/64 (19%)
		Y2	19 (26) {380}	23 (30) {211}	16 (24) {198}	14 (17) {67}	2 (2) {16}	2/29 (7%)
Median		Y1						11%
		Y2						9%
Average		Y1						16%
		Y2						13%

The overlap is given as A (B) {C}, where A is the number of the overlapping clusters in IPO stock, (B) is the number of the overlapping clusters in the mature stock, and {C} is the total number of clusters of a mature stock. A does not necessarily equal B because any cluster can be statistically similar to more than one cluster in another security. '# unique cl.' is the total number of unique IPO investor clusters over the total number of clusters, where the percentage ratio is given in brackets (). Median and average percentage of the unique clusters over all IPO stocks are given in the bottom of the table

securities trade recently issued securities, and if they do, not all of them might apply the same trading strategies and, therefore, not form similar synchronised clusters as in mature securities.
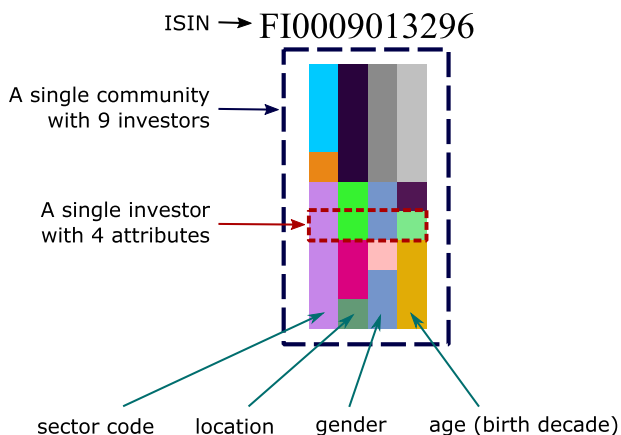

Fig. 2 Graphical representation of the clusters. A single cluster is visualised as a rectangle block, where a row represents one investor with four attributes: sector code, location, gender and birth year decade. *Sector code*: ▨—Households, ▨—Non-financial, ▨—Financial-insurance, ▨—General-government, ▨—Non-profit, ▨—Rest-world. *Geographic location*: ▨—Helsinki, ▨—South-West, ▨—Western-Tavastia, ▨—Central-Finland, ▨—Northern-Finland, ▨—Ostrobothnia, ▨—Rest-Uusimaa, ▨—Eastern-Tavastia, ▨—Eastern-Finland, ▨—South-East, ▨—Northern-Savonia. *Gender*: ▨—Male, ▨—Female, ▨—No-Gender. *Decade*: ▨—No-age, ▨—1910, ▨—1920, ▨—1930, ▨—1940, ▨—1950, ▨—1960, ▨—1970, ▨—1980, ▨—1990, ▨—2000

Conclusions

In the current paper, we analysed investor interactions and behaviours using a unique dataset of all Finnish investors' transactions in the HSE. Our selected set of 69 securities is aligned to an IPO event, which occurs when a company first starts publicly trading its securities. We performed an analysis for multiple securities on an individual investor account level by constructing the networks from the statistically validated trading co-occurrences. Our main focus was on the newly emerging market networks and their common and persistent market-driven structures with the other mature and new stocks.

Applying a community detection algorithm, we found statistically similar investor clusters with synchronised trading strategies that were forming repeatedly over several years and for multiple securities. We detected statistically robust clusters between the first and second year after an IPO. We also found clusters that could be found within other securities. By investigating cluster attribute overexpression and underexpression, we find a highly persistent institutional investor cluster. This finding provides further evidence about institutional herding. Comparing the findings with the clusters on mature securities, we observe that the majority of clusters can also be observed with a mature security.

Our results show that some synchronised trading strategies in financial markets span across multiple stocks, are persistent over time and occur with both newly issued and mature stocks. However, this analysis applies to the HSE only and does not generalise to all markets. Further research should check if this phenomenon also exists in other stock exchanges with a larger amount of IPOs; however, to the best of our knowledge, these investor-level data are not available, for example, from the U.S. markets.

Fig. 3 Statistically significant cluster overlaps across multiple securities and over time. The figure contains many subfigures separated by borders. Each subfigure presents a cluster of investors that spans over multiple securities and persists in time. The row alignment shows statistically similar clusters in the same year: the top row is the first after the IPO, and the bottom row is the second year after the IPO. The downward arrows show the cluster timewise evolution from the first to the second year for the same security. A cluster is represented by the rectangle. Each cluster is composed of investors with four attributes: sector code, geographic location, gender and decade. See the attribute colour mapping in Fig. 2
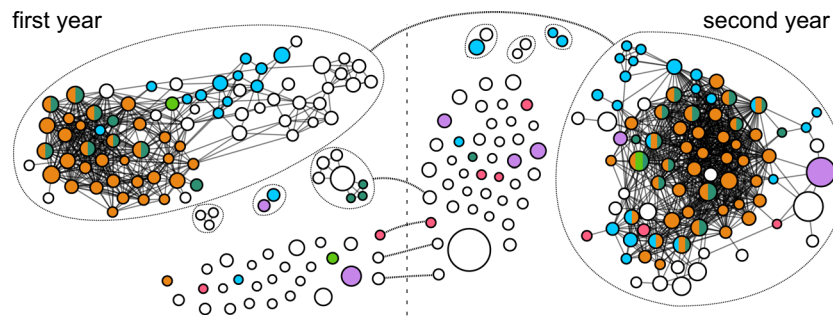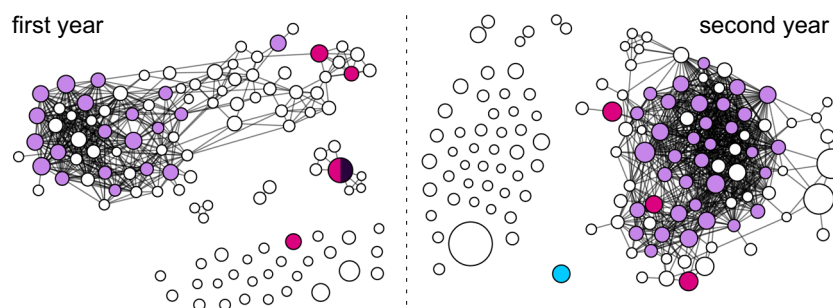
Fig. 4 Network of investor clusters with overexpressed attributes. On the left-hand-side are the clusters observed in the first year after respective IPOs and right-hand-side, in the second year. Investor cluster nodes are connected with continuous links if they share statistically significant number of individual investors. Dashed links represent statistical similarity between some of the connected cluster components in the first and the second year after the IPOs. Node colours identify overexpressed sector codes within clusters. For overexpressed geographical location see Appendix Fig. B.1, for underexpressed attributes see Fig. 5 and for all overexpressed and underexpressed attributes see Appendix Tables B.1 and B.2. *Sector code*: ● - Households, ●—Non-financial, ●—Financial-insurance, ●—General-government, ●—Non-profit, ●—Rest-world

Fig. 5 Network of investor clusters with underexpressed attributes. On the left-hand-side are the clusters observed in the first year after respective IPOs and right-hand-side, in the second year. Investor cluster nodes are connected with continuous links if they share statistically significant number of individual investors. Node colours identify underexpressed sector code and geographical location attributes within clusters. *Sector code*: ●—Households, —Financial-insurance. *Geographic location*: ●—Helsinki, ●—South-West

Traditional financial research assumes that investors are rational and hold optimal portfolios. However, actual investors have information, intellectual and computational limitations, and they satisfice[7] when making decisions. The systematic reoccurrence of the clusters gives a notion of possible stronger information connections that the investors share. For example, they may be consistently following the same public information sources or have mutual private information channels. However, with the current research, we do not try to explain the direction or the publicity of the information transfer. On the other hand, according to Ozsoylev et al. (2013), investor networks can be considered proxies of information networks if they are fairly stable over time. In light of this argument, the persistent and security-wide investor clusters can represent the mutual information channels that exist for both new IPO securities and mature stocks (e.g., Nokia).

Data availability

The dataset analysed in the current study is not publicly available and cannot be distributed by the authors because it is a proprietary database of Euroclear Finland. The database can be accessed for research purposes under the nondisclosure agreement by asking permission from Euroclear Finland.

Notes

1 In total, 75 securities had their IPOs during our analysis period. In this study we estimate investor networks during a 2-year period after their IPO date; therefore, we discarded ISIN FI0009014716 because its 2-year period falls out of our analysis period. Additionally, five ISINs (FI0009015580, FI0009015291, FI0009015713, FI0009005250 and FI0009902514) were discarded from the analysis, because no networks were estimated for them.

2 Unfortunately, the data appear to have issues with the trading date attribute for some securities, particularly for the transactions between 1998 and 2004. The net trading volumes on a daily resolution do not reconcile to 0 for all trading dates, while the volume sold should be equal to the volume bought per each stock during each day across all investors. This suggests that some transactions in the dataset were misplaced timewise because of incorrectly recorded trading dates. Only 14 of 69 securities fall into the completely error-free data period, and are marked in bold in Table 1.

3 For example, if the given investors were timing their buy transactions similarly so that they have a statistically validated link in the buying state but there were no statistical associations with the sell and buy–sell states, then the weight of the link between the investors would be 1.

4 We use *igraph* implementation of the Infomap algorithm with 100 as the parameter for the number of trials.

5 No-Gender and No-Age attributes refer to the institutional investors, but also to the individual investors who had no gender and/or birth year indicated in the data.

6 Recently, the long-term evolution of the clusters of the most capitalised stock in the HSE—Nokia—has been analysed in Musciotto et al. (2018).

7 The term *satisfice* refers to making optimal decisions under the limited resources. It was first defined in Simon and Barnard (1947).

References

Baltakys K, Baltakienė M, Kärkkäinen H, Kanniainen J (2018a) Neighbors matter: Geographical distance and trade timing in the stock market. Finance Res Lett https://doi.org/10.1016/j.frl.2018.11.013

Baltakys K, Kanniainen J, Emmert-Streib F (2018b) Multilayer aggregation with statistical validation: application to investor networks. Sci Rep 8(1):8198

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodol) 57 (1):289–300

Emmert-Streib F, Musa A, Baltakys K, Kanniainen J, Tripathi S, Yli-Harja O, Jodlbauer H, Dehmer M (2018) Computational analysis of the structural properties of economic and financial networks. J Netw Theory Financ 4 (3):1–32

Grinblatt M, Keloharju M (2001) How distance, language, and culture influence stockholdings and trades. J Financ 56(3):1053–1073

Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D (2016) Statistically validated network of portfolio overlaps and systemic risk. Sci Rep 6:39467

Ilmanen M, Keloharju M (1999) Shareownership in Finland. Finn J Bus Econ 48 (1):257–285

Karhunen J, Keloharju M (2001) Shareownership in Finland 2000. Finn J Bus Econ 50(2):188–226

Kaustia M, Knüpfer S (2008) Do investors overweight personal experience? evidence from ipo subscriptions. J Financ 63(6):2679–2702

Keloharju M (1993) The winner's curse, legal liability, and the long-run price performance of initial public offerings in finland. J Financ Econ 34 (2):251–277

Lakonishok J, Maberly E (1990) The weekend effect: trading patterns of individual and institutional investors. J Financ 45(1):231–243

Lillo F, Micciché S, Tumminello M, Piilo J, Mantegna RN (2015) How news affects the trading behaviour of different categories of investors in a financial market. Quant Financ 15(2):213–229

Ljungqvist A, Wilhelm Jr WJ (2003) Ipo pricing in the dot-com bubble. J Financ 58 (2):723–752

Ljungqvist A, Wilhelm Jr WJ (2005) Does prospect theory explain ipo market behavior? J Financ 60(4):1759–1790

Marotta L, Micciche S, Fujiwara Y, Iyetomi H, Aoyama H, Gallegati M, Mantegna RN (2015) Bank-firm credit network in japan: an analysis of a bipartite network. PloS one 10(5):e0123079

Musciotto F, Marotta L, Micciché S, Piilo J, Mantegna RN (2016) Patterns of trading profiles at the nordic stock exchange. a correlation-based approach. Chaos Solitons Fractals 88:267–278

Musciotto F, Marotta L, Piilo J, Mantegna RN (2018) Long-term ecology of investors in a financial market. Palgrave Commun 4(1):92

Nofsinger JR, Sias RW (1999) Herding and feedback trading by institutional and individual investors. J Financ 54(6):2263–2295

Ozsoylev HN, Walden J, Yavuz MD, Bildik R (2013) Investor networks in the stock market. Rev Financ Stud 27(5):1323–1366

Ranganathan S, Kivelä M, Kanniainen J (2018) Dynamics of investor spanning trees around dot-com bubble. PLoS One 13(6):e0198807

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA 105(4):1118–1123

Sias RW (2004) Institutional herding. Rev Financ Stud 17(1):165–206

Siikanen M, Baltakys K, Kanniainen J, Vatrapu R, Mukkamala R, Hussain A (2018) Facebook drives behavior of passive households in stock markets. Finance Res Lett 27:208–213

Simon HA, Barnard CI (1947) Administrative behavior: a study of decision-making processes in administrative organization. Macmillan

Spohr J (2004) Earnings management and ipos—evidence from Finland. Finnish J Bus Econ 53(2):157–172

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

Tumminello M, Micciche S, Lillo F, Piilo J, Mantegna RN (2011a) Statistically validated networks in bipartite complex systems. PLoS One 6(3):e17994

Tumminello M, Micciche S, Lillo F, Varho J, Piilo J, Mantegna RN (2011b) Community characterization of heterogeneous complex systems. J Stat Mech: Theory Exp 2011(01):P01019

Acknowledgements

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1057/s41599-019-0342-6.

Correspondence and requests for materials should be addressed to M.B.

Reprints and permission information is available at http://www.nature.com/reprints

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