Internet Appendix to
Investor Networks in the Stock Market

Han Ozsoylev Johan Walden M. Deniz Yavuz Recep Bildik

A Proof of Proposition 1

Consider a realization of trades, $\tau = (\tau_1, \tau_2, \ldots, \tau_K)$, where $\tau_k = (t_k, p_k, a_k)$ is represented by a time of the trade, a portfolio and an agent, and w.l.o.g. we assume that the sequence t_k is ordered, i.e., that it is weakly increasing in k. Given the assumptions of our stylized framework, these trades can be partitioned into Q similar trade types, representing information diffusion for different information events, i.e., a partition of the set $\{1, \ldots, K\}$ into $\Gamma_1, \ldots, \Gamma_Q$, with $\Gamma_i = \{v_1^i, \ldots, v_{r_i}^i\}$, $r_i \geq 1$, $\sum_i r_i = K$, and we assume (again, w.l.o.g.) that the v_j^i's are weakly increasing in j for all i. Now, v_1^i represents the agent who received the initial signal in each event, who was then followed by trades by agents $v_2^i, \ldots v_{r_i}^i$. For each trade, v_j^i, define the set of trades that occur within the next period $q\Delta t$ as Z_j^i, and the corresponding EIN, $\mathcal{E}^{\Delta t,1}$ which has a link between agents m and n if and only if $a_{v_j^i} = m$ and $a_{v_{j'}^i} = n$ for some $v_{j'}^i \in Z_j^i$, or $a_{v_j^i} = n$ and $a_{v_{j'}^i} = m$ for some $v_{j'}^i \in Z_j^i$. Define $N_j^i = |Z_j^i|$ to be the number of trades in the same information event within the time period Δt after trade v_j^i. Finally, define D_j^i to be the number of agents who are connected to agent $a_{v_j^i}$ in the estimated network, $\hat{\mathcal{E}}$, but did not trade either before that agent in the information event Γ_i, or within time period Δt after that agent, i.e., agents, a, such that $\hat{\mathcal{E}}_{a_{v_j^i},a} = 0$, $a_{v_{j'}^i} \neq a$ for any $j' < j$, and $a_{v_{j'}^i} \neq a$ for any $v_{j'}^i \in Z_j^i$.

Using standard combinatorial analysis, given the estimated network $\mathcal{E}^{\Delta t,1}$, it then follows that the likelihood for the realization of trades τ is

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{Q} \left[\prod_{k=1}^{r_i} \left(1 - (1 - q\Delta t)^{N_k^i}\right) \times (1 - q\Delta t)^{D_k^i} \right] (1 + O(q\Delta t)) \\
&= \prod_{i=1}^{Q} \left[\prod_{k=1}^{r_i} N_k^i q\Delta t \times (1 - D_k^i q\Delta t) \right] (1 + O(q\Delta t)).
\end{aligned}
$$

This is simply the product of the likelihood of all the observed trades (the first, N_k^i, part) times the likelihood of not observing trades although agents are linked (the second, D_k^i, part), across all (Q) information events.

Now, consider the effect on the likelihood function of deleting links in the EIN. The deleted links decrease N_k^i which has a decreasing effect on the likelihood function because there were fewer agents who could have stimulated any agent's trade. This affects the first part of the product. They also increase the D_k^i's which has a positive effect on the likelihood function, because there were fewer agents who could have influenced agents who did not trade. This affects the second part of the product. However, for small $q\Delta t$, the first effect dominates the second, given the form of the likelihood function, so by deleting links the likelihood function decreases.

The EIN chooses the maximum number of links between agents who trade within the time window of information diffusion, so there are no such links that can be added to further increase the likelihood function. However, links can be added between agents who never trade within the same window. In the case when they do trade in the same event, but outside of the window, this strictly decreases the likelihood function, since it has no effect on the first part, but strictly decreases the second part of the product. Similarly, adding a link between an agent who trades and one who does not trade in an event will strictly decrease the likelihood function. Finally, adding a link between two agents who never traded will not change the likelihood function (since the realized trades provide no information about potential links between agents who never traded), but it does increase the number of links in the network.

Thus, $\mathcal{E}^{\Delta t,1}$ is a maximum likelihood estimator for the agents in the information network, $\mathcal{E}_I$, and it is the unique such estimator with a minimum number of links.

B Degree adjusted measure of the expected number of overlaps

We describe how to adjust the test of overlapping links in Section 4.1 to account for the degree distribution.

Obviously, our computed EINs are not completely random; if they were, the degree distributions would be Poisson distributed. However, a detailed analysis of the degree distribution (available upon request) also shows that the true distribution has heavier tails. A better specified test for stability is therefore to study the number of overlaps, given the (heavy-tailed) degree distributions observed in practice.

We proceed as follows: We take the degree distribution from the EIN generated under the first six months as given, representing the true data generation process, and then compare the actual and predicted number of overlaps, given this data generating process. Specifically, given that the degree distribution of the EIN is D, and that links in the second EIN are formed such that the probability that a link added to the second network involves investor i is proportional to investor i's degree in the first EIN (i.e., the probability is $\frac{D_i-1}{k_1-N}$), the probability of an overlap, if the second EIN contains only one link, is

$$\begin{aligned}\mathbb{P}(\text{Overlap}) &= \frac{1}{2}\sum_{i=1}^{N}\mathbb{P}(\text{Investor } i \text{ is linked}) \times \mathbb{P}(\text{Investor } j \text{ is linked with } i|\text{Investor } i \text{ is linked}) \\ &= \frac{1}{2}\sum_{i=1}^{N}\frac{D_i-1}{k_1-N}\times\frac{D_i-1}{N}.\end{aligned}$$

Here, the factor $\frac{1}{2}$ arises to avoid double counting links (which are bidirectional). Finally, given that $k_2 << K$, the expected number of overlaps, when k_2 links are chosen in the second EIN, is approximately $k_2 \times \mathbb{P}(\text{Overlap})$, leading to the following approximate formula for the expected numbers of overlaps

$$E_{\text{Degree adjusted}}[y] \approx \frac{k_2}{2k_1N}\sum_{i=1}^{N}(D_i-1)^2. \tag{1}$$

We compare the expected and true number of overlaps for the 1-minute and 5-minute windows, under this assumption of a *degree adjusted* network generating process. The over-representation of overlaps, although not as high as under the assumption of a completely random data generation process, is substantial: For the 1-minute window, overlaps are 7.55 times as frequent as what would be expected under null hypothesis of a random degree adjusted network generating process. For the 5-minute window, the factor is 6.09. This is for the standard definition of degrees, with the connection threshold (M) set to one, in which case the comparable numbers for the completely random data generating process were 113.2 and 72.2, respectively.

We note in passing that the statistical significance of this over-representation is huge, which is why we do not provide significance levels. For example, under the first hypothesis of a completely random data generation process with a 1-minute window, the expected number of overlaps is 104,750. The standard deviation is 325, so rejection at the 5% significant level would be reached if there were 105,385 overlaps, whereas the true number of overlaps is about 11.9 million. The EINs are thus relatively stable over time, which is consistent with the intuition that they are proxies of information networks.

C Description of Information Events

Table 1 in this Internet Appendix describes the eleven information events used in Section 4.3 of the main paper.

In the extended set with 27 events, the additional events used are: ASELS 3/9/2005, DOHOL 1/14/2005, NTHOL 4/7/2005, GIMA 4/22/2005, KARTN 1/25/2005, TEBNK 1/12/2005, KRDMA 5/17/2005, TSKB 1/14/2005, TUPRS 3/18/2005, ULKER 2/14/2005, AKGRT 5/2/2005, VESTL 2/14/2005, BFREN 2/5/2005, BFREN 1/25/2005, BOYNR 1/18/2005, and ECILC 3/14/2005.

D Alternative profit measures

Our return measures are value weighted in that large trades get more weight than small trades when calculating investor returns. Average return measures, i.e., measures that give all trades equal weight, have also been used in the literature. We repeat our main tests using average returns measures where each trade is equally weighted. The results are very similar to what we obtained with value weighted returns, as seen in Table 2 in this Internet Appendix. Specifically, the table shows similar coefficients, t-statistics and economic significance in all tests, as what we obtained in Table 4 of the main paper.

E Conditioning on degree

To further verify that centrality above and beyond degree matters, we fix D within a limited range and study how returns are related to centrality within this range. The results are shown in Table 3 in this Internet Appendix, panel A. We see that centrality is still an economically and statistically significant factor in all tests. The economic significance of a one standard deviation increase in centrality is somewhat lower than in the full test, being about 0.1%-0.3%. This is mainly because the variation—and thereby the standard deviation—of c is much lower for any given value of d, than the unconditional standard deviation. Also, the statistical significance is of course lower, given the much smaller sample size for the conditional tests. We also do the same tests out-of-sample in panel B of the table, basing the centrality measure on the first 6 months of trades, and the profit measure on the remaining six months, with similar results.

Ticker	*Industry*	*Magnitude*	*Movement date*	*Information event(s)*	*News source(s) and date(s)*
BJKAS	Soccer	+19.4%	7/26/2005	The company is close to a deal regarding the transfer of Brazilian soccer player Ailton Goncalves da Silva playing in Schalke 04 in the German league.	Reported in the Istanbul Stock Exchange Bulletin on 7/26/2005.
BJKAS	Soccer	-33.5%	12/26/2005	There were negative news and comments regarding the company's debts in some of the daily newspapers on 12/26/2005. The company issued a statement denying these news on that day. Also, on the same day, the company announced that no profit had been generated in the previous fiscal year and therefore no dividends would be paid.	Reported in the Istanbul Stock Exchange Bulletin on 12/26/2005.
DEVA	Pharmaceuticals	+15.8%	7/12/2005	The number of Deva Holding shares in the portfolios of the emerging market funds managed by GEM Global Equities Management S.A. reached to an amount corresponding to 5.27% of the company capital. The company announced that it plans to construct new facilities in the Gebze Organized Industrial Zone to satisfy the increasing demand in the domestic market.	Reported in the Istanbul Stock Exchange Bulletin on 7/13/2005. Reported in the magazine Ekonomist on 7/14/2005.
DYOBY	Paint and chemicals	+12.25%	7/14/2005	Ahmet Yigitbasi, a senior executive of the company, told that they aim to achieve 20% revenue growth in 2005.	Reported in the magazine Tek Borsa on 7/11/2005.
EREGL	Steel manufactturing	+10.56%	9/12/2005	In a press meeting, the management announced that the company's profit increased by 34.4% in the first half of the year compared to the same period in the previous year. The general manager of the company, Fadil Demirel, told that the valuations mentioned in the media in relation to the forthcoming privatization of the company were too low given the company's financial performance and production capacity.	Reported in the daily newspapers Aksam, Milliyet, Radikal, Sabah on 9/14/2005 and 9/15/2005.
SAHOL	Conglomerate	+6.99%	10/4/2005	The company executives met Toyota officials in a four-day-long visit to Japan. Both companies declared their intention to strengthen their business partnership.	Reported in the daily newspaper Aksam on 10/7/2005.
SEKFK	Leasing	-21.27%	7/8/2005	Rabobank, one of the leading Dutch retail banks, signed a share purchasing agreement on July 7th with Sekerbank Personeli Munzam Sosyal Guvenlik and Yardimlasma Vakfi, a major shareholder of the company. The transaction price was 43% lower than the market price.	Reported in the magazine Ekonomist on 7/11/2005.
SISE	Glass manufacturing	+7.08%	7/29/2005	The company put the world's biggest glassware oven into service in Bulgaria. The company is accelerating its investments in order to realize its vision to become the world leader in glass production.	Reported in the daily newspapers Hurriyet and Vatan on 8/1/2005.
SKBNK	Banking	+14.55%	7/12/2005	The negotiations regarding the sale of Sekerbank shares from Sekerbank Personeli Munzam Sosyal Guvenlik and Yardimlasma Vakfi to Rabobank were completed and a share purchasing agreement was signed on 7/7/2005. Rabobank will make calls to other shareholders of Sekerbank following the realization of the share transfer. The parties agreed that, at the end of the call period, Rabobank will be the majority shareholder of Sekerbank.	Reported in the Istanbul Stock Exchange Bulletin on 7/13/2005.
TEKST	Banking	+12.27%	8/16/2005	In an announcement made by Fitch Ratings, Tekstilbank's good credit quality and strong risk management system were emphasized. It is rumored that the Hungarian OTP Bank is interested in acquiring Tekstilbank.	Reported in the Istanbul Stock Exchange Bulletin on 8/15/2005. Reported in the magazine Ekonomist on 8/15/2005.
TNSAS	Retail	+12.12%	8/19/2005	Koc Holding is interested in acquiring Tansas and signed an exclusivity agreement with Dogus Holding, the current majority shareholder of the company. Koc aims to be the largest retailer in the country with this acquisition. The company's first half-year profit is 1.9 million YTL. The company made a loss of 30.2 million YTL in the first half of the previous year. Net sales of the company increased by 19%.	Reported in the Istanbul Stock Exchange Bulletin and the daily newspapers Radikal, Vatan between 8/17/2005 and 8/20/2005. Reported in the magazine Para on 8/19/2005.

Table 1: Description of events. The table describes the 11 information events that were reported in public news in association with large stock movement. Each event is considered to be idiosyncratic to the stock and significant enough to cause a large stock price movement.

A. Returns

	1	2	3	4	5	6	7	8	9	10	11
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	$t-error$	$t-error$	Ramsey	Ramsey
Centrality (c)	0.0030					0.0062		0.0033		0.0061	
	> 20					15.4		4.0		15.0	
Degree (d)		0.0030				-0.0090		-0.0062		-0.0090	
		> 20				-19.6		-6.7		-19.4	
Rescaled Centrality ($c-d$)			-0.00035				0.0038		0.0008		0.0038
			-9.3				10.6		1.30		10.2
# of trades (n)				0.0041		0.0093	0.0063	0.0072	0.0041	0.0093	0.0062
				> 20		> 20	> 20	19.9	> 20	> 20	> 20
Quantity (q)					0.0016	-0.0017	-0.0019	-0.0012	-0.0015	-0.0017	-0.0015
					> 20	< -20	< -20	-9.0	-10.5	< -20	< -20
$\bar{R}^2$	0.0058	0.0059	0.00002	0.0090	0.0033	0.011	0.011				
$\Delta\mu$	0.6%	0.6%	0.01%	0.4%	0.5%	1.3%	0.1%	0.7%	0.003%	1.3%	0.1%

B. Excess returns

	1	2	3	4	5	6	7	8	9	10	11
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	$t-error$	$t-error$	Ramsey	Ramsey
Centrality (c)	0.0001					0.0089		0.0065		0.0088	
	3.3					> 20		9.4		> 20	
Degree (d)		0.0001				-0.0135		-0.0114		-0.0135	
		1.7				< -20		-14.5		< -20	
Rescaled Centrality ($c-d$)			0.0037				0.0055		0.0030		0.0055
			11.5				16.6		4.5		16.2
# of trades (n)				0.0008		0.0063	0.0019	0.0055	0.0007	0.0063	0.0018
				16.1		> 20	> 20	18.9	4.0	> 20	> 20
Quantity (q)					0.0002	-0.0004	-0.0007	-0.0002	-0.0005	-0.0003	-0.0007
					6.8	-5.5	-11.9	-3.4	-4.7	-5.6	-12.0
$\bar{R}^2$	0.000020	3.4E-6	0.000024	0.000046	8.2E-5	0.0039	0.0012				
$\Delta\mu$	0.03%	0.02%	0.1%	0.1%	0.06%	1.8%	0.2%	1.3%	0.1%	1.8%	0.2%

Table 2: Centrality and equally weighted returns. The table repeats the tests in Table 4 of the main paper, using equally weighted returns in which each trade is equally weighted when calculating investor returns. The dependent variable is equally weighted returns in Panel A and equally weighted excess returns in Panel B. Each column represents a regression. The first row displays coefficients while the second row displays t-statistics. The first 7 columns display results from OLS regressions, columns 8-9 display results from a regression that is robust to heavy-tailed error terms, and columns 10-11 display results from iteratively re-weighted least square regression (using Ramsey's E-function). Returns and excess returns are calculated as each investor's average return across trades, where the return is corrected for market returns in the case of excess returns. The variable $\Delta\mu$ highlights the economic significance of the results by showing the change in returns, given a one standard deviation increase of the variable in univariate regressions and centrality or rescaled centrality in multivariate regressions, all else equal. Detailed definitions of the other variables are provided in Table 4 of the main paper.

A. In sample

Range of degree, D		20-200	200-2,000	2,000-20,000
Centrality, c	β_{OLS}	0.0018	0.0039	0.029
	t_{OLS}	5.72	> 20	> 20
	$\Delta\mu_{OLS}$	0.13%	0.3%	0.2%
	$\beta_{t-error}$	0.0001	0.0019	0.0009
	$t_{t-error}$	0.12	5.1	4.3
	$\Delta\mu_{t-error}$	0.01%	0.1%	0.1%
	β_{Ramsey}	0.0016	0.0038	0.0029
	t_{Ramsey}	5.1	19.7	> 20
	$\Delta\mu_{Ramsey}$	0.11%	0.2%	0.2%

B. Out-of-sample

Range of degree, D		20-200	200-2,000	2,000-20,000
Centrality, c	β_{OLS}	0.0010	0.0020	0.035
	t_{OLS}	1.9	5.4	12.1
	$\Delta\mu_{OLS}$	0.08%	0.1%	0.2%
	$\beta_{t-error}$	0.0004	0.0011	0.0019
	$t_{t-error}$	0.34	1.5	3.2
	$\Delta\mu_{t-error}$	0.03%	0.07%	0.1%
	β_{Ramsey}	0.0011	0.0020	0.0035
	t_{Ramsey}	2.0	5.3	11.9
	$\Delta\mu_{Ramsey}$	0.08%	0.1%	0.2%

Table 3: Regressions conditioned on degree. The table shows regressions results of returns on centrality for subsamples of the investor network consisting of all investors with degrees between 20-200, 200-2,000, and 2,000-20,000. Panel A show the results using the whole year of trades to define centrality and profits. The first 3 rows display results from OLS regressions, the next three rows display results from a regression that is robust to heavy-tailed error terms, and the final three rows display results from iteratively re-weighted least square regressions. The variable $\Delta\mu$ highlights the economic significance of the results by showing the change in returns, given a one standard deviation increase of centrality, all else equal. Detailed definitions of the variables are provided in Table 4. Panel B shows the same regressions when centrality is defined over the first six months of trades, and returns over the latter six months.