Information Channels and Portfolio Choices *

Adriana Cobas[†], Grace Weishi Gu[‡], and Zachary Stangebye[§]

February 28, 2025

Abstract

This paper studies how information flows from different sources (i.e., direct acquisition, bilateral connections, system-wide network diffusion) affect fund portfolio allocation. Using EDGAR log files and a novel network measurement, we identify funds' firsthand direct information acquisition and their information networks, respectively. We show that among all the channels, system-wide network flows have the largest impact and tend to make funds' portfolios more correlated, regardless of economic uncertainty. However, isolated bilateral information flows have the opposite effect, reducing portfolio correlations. Moreover, the impact of network on direct information acquisition is two-fold: Although network itself does crowd out direct information acquisition, it increases one's own information acquisition in reaction to the direct information acquisition by others in the network.

Keywords: network, information acquisition, bilateral, system-wide, portfolio allocation
JEL code: D85, D83, G11

[*]The opinions expressed in this article are the exclusive responsibility of the authors and do not necessarily represent those of the Central Bank of Chile or its Board.

[†]Central Bank of Chile
[‡]University of California Santa Cruz
[§]University of Notre Dame

1 Introduction

The speed at which information spreads is one of the most significant contributions of technological progress in recent decades. Especially in financial markets, information can disseminate quickly through multiple channels and reach many investors. Yet, it is unclear how different information channels contribute to investment decisions and how they can affect each other. In this paper, we categorize information channels into direct ones, through which information is collected firsthand, and indirect channels through network information diffusion. We investigate the effects of both channels on investors' portfolio decisions. Moreover, we attempt to understand the role of bilateral linkages versus the role of system-wide networks on portfolio correlation, as well as the impact of direct and indirect information acquisition on each other.

With these empirical questions in mind, we identify mutual funds specializing in US stocks from S&P Intelligence Unit and measure their firsthand direct information acquisition and secondhand indirect information acquisition through network channels. As a proxy for the first type of information acquisition, we use SEC Electronic Data Gathering Analysis and Retrieval (EDGAR) log files to capture funds' online access to listed companies' regulatory reports. This approach has been used in the finance literature extensively (e.g., Loughran and McDonald (2017); Gibbons et al. (2021)).

We measure funds' secondhand indirect information acquisition through network information diffusion following Diebold and Yilmaz's (2015) methodology to identify linkages between the funds. More specifically, using a database of daily fund share prices from Bloomberg, we assess the proportion of the variation in the forecast error of one fund's return that is explained by an independent shock affecting the return of another fund. The proportion indicates the magnitude of the network linkage.

We do not assume any particular channel by which information spreads through the network. The linkages we estimate could encapsulate multiple potential network formats. For instance, it could be that fund managers acquire information through conversations with each other (i.e., word-of-mouth format), by processing past investment decisions made by other funds (i.e., copycat format), by receiving the same market perspectives from private consultants (i.e., client network format), or through trading with each other (i.e., trade format). We conduct robustness checks to ensure the network linkages are not dominated by simultaneous direct information acquisition or price effect.

We construct bilateral linkages as the sum of two directional linkages between any two funds and system-wide total connectivity as the sum of all bilateral linkages in our sample. This allows us to identify our sample funds' bilateral information flows, as well as a well-specified notion of system-wide connectivity of an entire network. An important innovation in our approach is that our network connectivity measures are continuous, while measurements used in much of the existing work are binary. This allows us to obtain an intensity measure of the bilateral and the system-wide network.

We calculate funds' pairwise quarterly portfolio (absolute) correlations in 2004-2022 to test their correlation responses to our measures of funds' direct information acquisition and network connectivity. We find that bilateral information, whether acquired through the firsthand acquisition via EDGAR or through bilateral network, makes any two funds' portfolios less correlated on average. In contrast, system-wide information, especially through network diffusion, makes funds' portfolios more correlated. In particular, the system-wide network connectivity has the largest positive impact on portfolio correlations among all the channels.

Our interpretation of these results is that, relatively to system-wide network information, the bulk of bilateral information flows among funds are more likely to be idiosyncratic and out of liquidity needs, which reduces funds' portfolio correlations. The fact that system-wide network connectivity engenders significant and stronger correlations between fund portfolios suggests that after the idiosyncratic liquidity needs get washed out, there remains some information circulating in the system-wide network regarding asset payoffs that induces higher correlations across funds.

We conduct various robustness checks to verify that our benchmark regressions capture what is intended. In particular, given that our network measure using fund returns, we test to ensure the network effect in our benchmark results is not simply a price effect where funds sharing a common reaction to the market price, nor from funds' simultaneous direct information acquisition or asset holding overlaps. In addition, we also include additional controls to take into account the fact that a few of the sample funds belong to the same company group.

Once we establish the robustness of our benchmark results, we extend our analysis to various dimensions. First, we study asymmetric responses of negatively and positively correlated portfolios. We find that the benchmark results using the full sample are mainly driven by the fund pairs with non-negative correlations, even though they are a smaller share of the sample. In fact, total network connectivity becomes even more important for those non-negatively correlated funds, making them more positively correlated. This result is mainly driven by the larger magnitude on average for positive portfolio correlations (22%) than negative correlations (-0.65%). Second, we explore the interactive effects of information channels on portfolio correlations. No interactive effect is found between direct information acquisition and indirect network channels. However, we show that the system-wide network weakens the impact of bilateral network on portfolio correlations. Third, we examine the impact of uncertainty and show that the reduction effect of bilateral connectivity on portfolio correlation is stronger during high uncertainty periods.

Finally, when we examine how the information channels affect each other, we find evidence supporting network's crowding-out effect as well as its accelerator effect on direct information acquisition. That is, although network itself does crowd out direct information acquisition, it strengthens the amplification effect among funds' direct information acquisition (i.e., others' information acquisition increases one's own acquisition). This result reconciles two pieces of literature. One, using laboratory data, finds the two types of information channels are substitutes due to free riding (e.g., Halim et al. (2019); Ruiz-Buforn et al. (2021)). The other piece of literature, using theoretical models, shows that network information can be a strategic complement to direct information acquisition through various scenarios (e.g., Goldstein and Yang (2015), Ganguli and Yang (2009), Veldkamp (2006)). For instance, Cobas et al. (2023) shows that more direct information acquisition can be built on existing network information.

Contributions and Literature Our paper contributes to the understanding of how information flows from different sources (directly acquired and bilateral versus system-wide network diffusion) affect portfolio allocation, as well as how they may or may not interact with each other. Specifically, to our knowledge this is the first paper to distinguish firsthand direct acquisition from bilateral and system-wide network information flows and to examine their separate as well as interactive effects on portfolio allocation. In particular, other papers have focused either on the impact of private information acquisition (Cohen et al. (2008), Cohen et al. (2010)), *or* that of network (Hong et al. (2005), Ivković and Weisbenner (2007), Shive (2010), Pool et al. (2015), Ning Ding and Shen (2017), Rossi et al. (2018)) on portfolio allocation. But none has studied these information channels simultaneously nor distinguished bilateral versus system-wide networks. Our empirical results provide evidence for the importance of *system-wide* network information diffusion on investment behavior relative to other information channels.

Our work is closely related to the literature on the impact of networks on portfolio allocation. For instance, Shiller and Pound (1989) survey institutional investors and find that their portfolio choices are partly influenced by word-by-mouth communications. Feng and Seasholes (2004) exploit Chinese brokerage rules that investors must trade at their office and show that trades are highly correlated for those who are geographically close. Hong et al. (2005) find that a mutual fund manager is more likely to trade a particular stock in any quarter if other managers in the same city are trading that same stock. Ivković and Weisbenner (2007), using individual investor data from a single brokerage, find that a ten percentage point increase in neighbors' purchases of stocks from an industry is associated with a two percentage point increase in a household's own purchases of stocks from that industry. Shive (2010) using Finish data shows that individuals share useful information which induces financial trading.

In more recent works, Pool et al. (2015) show that the overlap of funds whose managers reside in the same neighborhood is considerably higher than that of funds whose managers live in the same city but in different neighborhoods. Heimer (2016) identifies traders' network using a dataset from an investment-specific social network flatform and finds that traders connected in the network develop correlated levels of the disposition effect. Ning Ding and Shen (2017) and Rossi et al. (2018) both utilize investment management mandates to identify fund manager networks and find that funds share more similar and higher returns, holdings, and trading with funds in their mandated networks than with funds outside the networks. In Di Maggio et al. (2019), investors who are connected through the same broker execute similar traders. Our paper also highlights the role of networks on portfolio allocation, like the above papers, but differs in that we bring in the roles of firsthand direct information acquisition and bilateral versus system-wide networks, as well as their interactions.

Another key difference is that, while all the papers above identify networks by a single channel through which investors connect, such as office locations, neighborhoods, brokers, social network platforms, investment mandates, and education institutions, this paper uses a different identification approach (Diebold and Yilmaz, 2015) to measure empirical investor network assuming that network information diffusion occurs simultaneously through multiple venues.[1] This measure values the aggregate strength of *all* network linkages between two funds over time using a continuous metric, while the papers in the last paragraph measure *one* particular linkage with a 0/1 dummy. It also allows us to compute not only bilateral but also *directional* and *system-wide* connectivity measures that assess different kinds of information diffusion in networks, which the previous papers cannot do with their network measures.

This paper is also related to a large number of papers on direct information acquisition and/or portfolio allocation. Many in this strand of literature examine the impact of direct information acquisition on return forecast accuracy and investment performance, to name a few, e.g., Gibbons et al. (2021), Chen et al. (2022a) and Chen et al. (2022b), or the impact of information acquisition cost on portfolio allocation (e.g., Van Nieuwerburgh and Veldkamp (2010)). Our paper focuses on the impact of direct information acquisition from public sources (with little to no cost) on portfolio allocation, in addition to the abovementioned network effect.

Regarding the impact of firsthand direct information acquisition and information diffusion through networks on each other, there is a separate relevant literature to which our work relates. For instance, in their theoretical paper Han and Yang (2013) show that social communication

[1] Ozsoylev et al. (2014) also uses an ex-post identification strategy for investor networks, but instead of using investors' realized returns they use investors' realized trades within a short period of time. Also unlike us, they do not include a direct information acquisition channel.

crowds out information production. Goyal et al. (2017) find that a decline in relative costs of connecting with others increases social networks and makes private investments more dispersed, which reduces aggregate investment. Halim et al. (2019) use a laboratory experiment to demonstrate that networks can crowd out information acquisition due to free-riding on the network information. However, there also exists a whole set of theoretical papers that consider direct information acquisition and network or public information complements (e.g., Goldstein and Yang (2015), Ganguli and Yang (2009), Veldkamp (2006), Cobas et al. (2023)). Our paper adds to this literature by examining field data and, to our knowledge, is the first to document the two-fold (both crowding-out and amplification) effect of network on direct information acquisition.

The rest of this paper is structured as follows. Section 2 describes the databases and measurements. Section 3 lays out our empirical strategy. Section 4 presents and discusses the benchmark results, as well as robustness checks and extensions. Section 5 concludes.

2 Data

Our empirical analysis combines three different data sets. First, we use quarterly portfolio data of US equity funds from the S&P Intelligence Unit to assess investment decisions over time. Second, we obtain daily share prices of the selected funds from Bloomberg to estimate funds' connectivity with each other. Finally, we use funds' internet access to company filings from the SEC's EDGAR log file database to indicate each fund's direct information acquisition activities about the assets traded in the market.

To maintain a tractable scale, we select from the S&P Intelligence Unit 20 mutual funds that invest in only US stocks with the biggest average portfolio sizes and fullest price history (by Bloomberg) in 2004-2022.[2] The number of funds we can include in our study is limited by our computational capacity to infer the funds' network. As detailed in Section 2.2, the methodology we use to infer funds' network requires, as a first step, to estimate a vector autoregression model (VAR) using the series of fund prices, that is, a twenty-variable VAR model for twenty funds. Moreover, as we intend to consider the evolution of funds' connectivity over time, we estimate one VAR per quarter using 1-year rolling windows of daily data including the quarter of interest. Hence, we have about 250 daily prices per fund in each VAR estimation. These make the VAR regressions computationally intensive, which limits our number of funds to 20. As a result, our portfolio sample consists of a set of 20 matrices, each with rows and columns reporting a fund's assets and quarterly positions, respectively. We observe a total of 2,805 different assets throughout the entire sample over 73 quarters in 2004q1-2022q1.

[2] However, we remove funds with more than 1000 asset holdings, as they could be index funds.

Table 1: Fund Portfolio Descriptive Statistics

Fund	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Size (USDmn)	167.762	94.721	92.911	85.341	72.295	56.344	55.312	52.131	50.197	45.792
In a group	1	No	No	2	No	No	No	2	No	No
Number of assets	99	85	222	168	462	117	92	223	269	74
Fund	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
Size (USDmn)	43.646	41.657	33.569	29.376	24.897	21.513	19.494	16.394	25.248	20.022
In a group	No	No	1	1	No	No	No	No	No	No
Number of assets	155	384	31	201	616	262	145	248	109	701

Notes: "In a group" indexes ownership of the funds to the same financial institution inside the sample.

Table 1 presents descriptive statistics for the funds in the sample. The portfolios present an important heterogeneity in both size and assets, which potentially gives them different levels of market power and visibility in the market. We see that the average size of the funds in our analysis ranges from 16 to 168 million dollars. The average number of different assets in the portfolio ranges from 31 to 701 and the average non-zero asset position is approximately 15 million dollars. Most of the funds are independent institutions, except for funds F1, F13, and F14, and F4 and F8 which belong to two different companies, respectively. No fund shares the same manager.

2.1 Portfolio Pairwise Correlation and Overlap

To assess the potential convergence in investor decisions, we use two alternative measures for portfolio similarity: correlation and overlap, and the former measure is our benchmark. To calculate quarterly correlations, we start by mapping each portfolio's relative positions, i.e., the percentage that each asset position represents in the total position of the fund that quarter, into a new matrix of dimensions $2,805 \times 73$, where rows represent the full set of assets for all funds in the dataset and the columns are quarters in the sample. For instance, if a fund does not have an asset that other funds have, then the relative position is zero. Having all the funds' portfolio relative positions represented in the same assets-quarters space, we then obtain the pairwise portfolio correlation using the pair of position vectors in quarter t of funds i and j, denoted as $\rho_{i,j,t} \in [-1, 1]$.

Table 2 panel 1) presents period-average correlations of a fund with others. There we see positive averages and negative medians, which suggests that the distribution of the correlations is right-skewed, where the majority of funds are negatively correlated and some funds have very strong positive correlations.

We measure portfolio overlap as in Equation 1, following Pool et al. (2015). We calculate funds i and j's portfolio weights w of an asset a, take their minimum, and repeat for all assets held by the two funds in the set $\mathcal{A}_t$.

Table 2: Fund Similarity Descriptive Statistics

Fund	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
1) Correlation										
Mean	0.016	0.021	0.032	0.154	0.005	-0.007	0.138	0.091	0.001	0.002
Median	-0.004	-0.006	-0.008	-0.005	-0.009	-0.009	-0.005	-0.006	-0.007	-0.004
2) Overlap										
Mean	0.009	0.014	0.025	0.124	0.005	0.001	0.112	0.084	0.004	0.002
Median	0	0	0	0	0	0	0	0	0	0
Fund	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
1) Correlation										
Mean	0.142	0.009	0.003	0.014	0.016	0.12	0.143	0.012	0.005	-0.007
Median	-0.005	-0.004	-0.003	-0.005	-0.006	-0.005	-0.005	-0.007	-0.002	-0.009
2) Overlap										
Mean	0.11	-0.015	0.003	0.01	0.016	0.098	0.114	0.013	0.007	0.002
Median	0	0	0	0	0.005	0	0	0	0	0

Notes: Mean and median portfolio correlations and overlap correspond to quarterly data in 2004-2022.

$$Portfolio\,Overlap_{i,j,t} = \sum_{a \in \mathcal{A}_t} min\{w_{a,i,t}, w_{a,j,t}\} \tag{1}$$

As shown in Table 2 panel 2), portfolio overlap is on average below portfolio correlations. The median is mostly at zero. Also, note that funds 3 and 12 have negative (short) positions on assets at times in our sample, but overall short positions are a small share of our sample.[3] This causes the mean of overlap for fund 12 to be negative. Hence, it is important to note that overlap and correlation are not the same, although the two portfolio-similarity measures are highly correlated with a correlation of 0.89 (see Table A.3).

2.2 Network Extraction

Our network extraction methodology and connectivity measures follow those proposed in Diebold and Yilmaz (2015). That is, the connectivity between two funds is the fraction of the forecast error variance in one of the fund's daily returns explained by an independent shock hitting the other fund's returns. The advantage of this approach is that pairwise connectivity is directional (non-symmetric), continuous, and varies for each pair of funds each period.

To estimate the network structure, we first calculate fund returns. For each fund in the sample

[3]In the case of fund 3, it has only 1 asset on one date that had a short position. In the case of fund 12, it has negative positions on 18 dates in 140 out of 384 assets. To deal with short positions, we employ the sum of the absolute values to calculate the total value of the portfolio used in the portfolio weight. This reduces the weight of assets with short positions.

we collect daily fund share prices from Bloomberg for the period of 2003-2022 and compute daily returns in period t for fund i as $r_{i,t} = 100\, ln(P_{i,t}/P_{i,t-1})$. The mean and median daily returns are reported in Table 3 Panel 1). The returns are slightly left skewed.

Table 3: Funds' Average Centrality and Connectivity Measures

Fund	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
1) Daily Return										
Mean (%)	0.000	0.000	0.008	0.003	-0.002	0.004	0.001	0.003	0.006	-0.002
Median(%)	0.000	0.021	0.018	0.016	0.000	0.000	0.000	0.013	0.022	0.000
2) Connectivity										
$C_{i\leftarrow\bullet}$	2.410	5.272	0.778	0.385	0.954	1.399	0.588	0.225	0.477	0.130
$C_{\bullet\leftarrow j}$	0.185	0.184	0.857	0.893	0.503	0.445	0.710	0.986	0.799	0.614
3) Centrality										
Eigen Vector	0.061	0.172	0.077	0.064	0.037	0.057	0.047	0.061	0.046	0.023
Fund	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
1) Daily Return										
Mean (%)	0.006	0.01	0.01	0	0.004	0.003	0.006	-0.002	0.004	0.002
Median (%)	0.020	0.016	0.024	0.000	0.000	0.000	0.026	0.000	0.000	0.031
2) Connectivity										
$C_{i\leftarrow\bullet}$	0.189	0.274	0.214	0.132	0.125	0.178	0.09	0.1	0.119	0.104
$C_{\bullet\leftarrow j}$	0.976	0.78	0.635	0.83	0.806	0.821	0.901	0.761	0.597	0.86
3) Centrality										
Eigen Vector	0.064	0.038	0.032	0.027	0.036	0.035	0.038	0.026	0.025	0.035

Notes: Average fund daily returns in 2003-2022. Average fund network features correspond to quarterly data in 2004-2022.

Then, we estimate a VAR(1) for each quarter using daily returns computed from the share prices of the 20 funds for the four quarters preceding the quarter under analysis.[4] We assume that the estimated linkages from the prior four quarters affect investors' decisions in the subsequent periods. We then use a 10-day projection forecast to obtain the corresponding quarterly orthogonalized forecast error variance decomposition matrix, from which we infer funds' connectivity measures.[5] In the matrix, each element off the diagonal $d_{i,j}$ with $i \neq j$ informs the percentage of the forecast error variance of returns in fund i that is produced by an independent shock from returns in fund j. Subsequently, we use five measures of connectivity that we can extract from the matrix, as in Diebold and Yilmaz (2015): pairwise directional connectivity ($C_{i\leftarrow j} = d_{i,j}$ and $C_{j\leftarrow i} = d_{j,i}$), pairwise bidirectional connectivity ($C_{i,j}$), total directional connectivity to others ($C_{\bullet\leftarrow j}$), total

[4]As we have a 20-variable VAR to be estimated with a series of about 250 daily prices (of 1-year rolling window) for each quarter, we simplify as much as possible by choosing a Var(1) model.

[5]For the selection of the projection horizon, we choose a value from which the connectivity measure starts to converge. In Figure A.1 in appendix A we show how connectivity varies with the days-in-projection parameter.

directional connectivity from others ($C_{i\to\bullet}$), and system-wide connectivity (C).

It is important to note that we are interested in measuring the connectivity over time. As a result, we will estimate the VAR model for each quarter t and then obtain the corresponding variance decomposition matrix from which we compute the full set of connectivity measures for the same quarter. In the following text, we omit the subscript t below to keep equations simple.

Directional and bidirectional pairwise connectivity isolate the mutual relationships between a pair of funds. The directional pairwise connectivities are denoted $C_{i\leftarrow j} = d_{i,j}$ and $C_{j\leftarrow i} = d_{j,i}$ for $i \neq j$ and measure the response of i to a shock in j and vice versa, respectively. Meanwhile, the bidirectional pairwise connectivity in Equation (2) aggregates the directed effects from both i to j and j to i (with $i \neq j$). It gives us the bilateral connectivity between two funds in each period.

$$C_{i,j} = d_{i,j} + d_{j,i} \text{ } (\textit{bidirectional connectivity}) \tag{2}$$

We then aggregate directional measures of connectivity for each fund. For instance, the total directed connectivity from fund j to others is computed using Equation (3). It represents the aggregated connectivity effect on all other funds from an independent shock in j.

$$C_{\bullet\leftarrow j} = \sum_{i=1, i\neq j}^{N} d_{i,j} \tag{3}$$

where N is the number of funds. The fourth measure is total directed connectivity to fund i from the rest of the network, i.e., Equation (4). It captures how responsive each fund is to the shocks in the rest of the network.

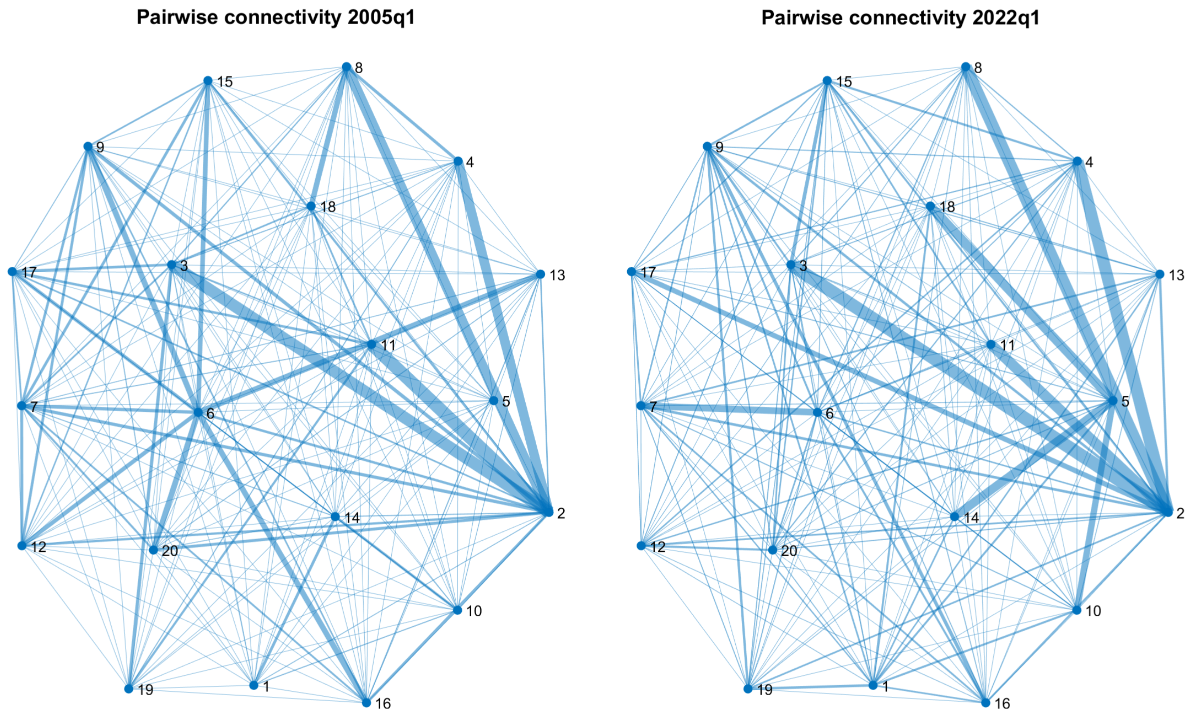
$$C_{i\leftarrow\bullet} = \sum_{j=1, j\neq i}^{N} d_{i,j} \tag{4}$$

Finally, system-wide connectivity in Equation (5) aggregates all directional bilateral effects. Although it does not provide information about the direction of information flows, it measures the degree of information sharing across the system in each period.

$$C = \frac{1}{N}\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} d_{i,j} \text{ } (\textit{total network}) \tag{5}$$

The pairwise bidirectional connectivity is plotted in Figure 1 for 2005q1 and 2022q1, respectively. Each node represents a mutual fund, and the width of the links between each pair of nodes represents the intensity of the connectivity between that pair of funds. The comparison between the two networks at the beginning and the end of the data sample shows some variations in the system

Figure 1: Pairwise Bidirectional Connectivity



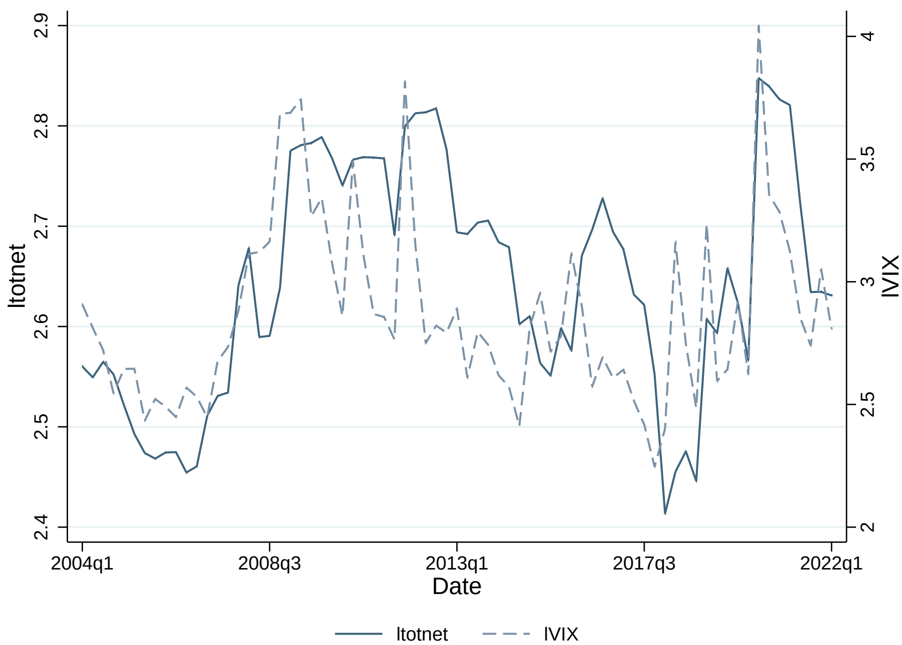
Notes: Each node represents one fund. The width of the links represents the intensity of the pairwise connection each period relative to the maximum link intensity in the same period.

connectivity over time, albeit with persistence.

Table 3 panel 2) shows the sample averages of directional connectivity measures. To the extent of receiving shocks from the others in the network, using the measure $C_{i\leftarrow\bullet}$, we can see that funds F1, F2, and F6 are the most connected. In other words, those funds have returns that are most affected by shocks from the other funds in the network. Meanwhile, according to the measure $C_{\bullet\leftarrow j}$, the funds that are most connected to transmit shocks to others are F8, F11, and F17. Comparing the two measures, we see that the directional effect *from* others has more high-value outliers than the directional effect *to* others, suggesting that there are a few funds much influenced by others but no exceptionally influential funds to others in our sample.

We also report the eigenvector centrality in Table 3 panel 3). It captures how connected a fund is to highly connected funds. According to this measure, the most central node is F2, which also happens to be the fund with the highest $C_{i\leftarrow\bullet}$ measure. In figure 1 we can see that F2 is strongly connected to funds F3, F4, F8, and F11, which, as expected, are also the most central among the funds.

Figure 2: System-wide Connectivity



Notes: System-wide connectivity is computed according to Equation (5). *ltotnet* is the logarithm of total network C_t, and $lVIX$ is log(VIX) plotted on the secondary axis.

Figure 2 shows the system-wide measure of total network C of the sample funds during our sample period. It has some variation over time and also appears to have a positive correlation with the VIX, which is 0.59. This suggests that connectivity increases during times of high uncertainty. One reason could be that investors connect more to gain more information during such periods. It is also possible that given the way we measure networks, the measurement itself may capture some common economic factors that are more prevalent during times of high uncertainty and affect many assets, rather than the actual network structure. In section 4.3.3, we address this issue.

2.3 Direct Information Acquisition

Finally, we use the EDGAR access log file database. It contains a daily log of the IP addresses that connect to the SEC site, reporting the date and time of users' connections, and the listed companies whose information the users have collected while there. We believe EDGAR log files provide a good measure for firsthand direct information acquisition by investors. Although there are other sources for investors to access firms' financial disclosures, such as Bloomberg terminals, the EDGAR filings are a first-source repository with the most detailed information. As noted in Loughran and McDonald (2017), non-EDGAR sources often process income statement and balance sheet information from the EDGAR filings and simplify it into a preset framework, and thus omit valuable information, such as off-balance sheet operations, and are also prone to errors. However, annual reports on EDGAR have all information, conveyed through the main text, plots, notes, and footnotes, which cannot be obtained from a Bloomberg terminal or a secondary website.[6]

We have created a database of all IP addresses used by the funds in our sample to connect to the Internet from the American Registry for Internet Numbers (ARIN). An IP address is a unique string of characters organized in four octets of numerical digits. In the database, each fund has one or more IP addresses, most of them (but not always) varying only in the last octet.

We then match our sample funds' IP database with the IP addresses in the EDGAR log files. To keep the privacy of information, EDGAR log files report only the first three octets of the IP addresses masking the last one with letters. This makes the identification of the fund's connection not perfect but noisy. Another company may share the first three octets of its IP address with some of our funds, which results in more counts to our funds' connections to the SEC filings than the actual number of counts. However, for such an overcount to happen, the company sharing the same first three octets with our sample funds also has to connect to the SEC filings for information in the same quarter, which may be uncommon. On an empirical assessment of the use of EDGAR data to measure information acquisition, Gibbons et al. (2021) exploit the fact that brokerage houses tend to register big blocks of IP addresses, which enables the authors to map the partially masked IP address to their sample. Similarly, here, we also find that mutual fund companies register big blocks of IP addresses. For instance, Vanguard registers to all IP addresses that begin with 192.175.128 (e.g., 192.175.128.0, 192.175.128.1, ..., 192.175.128.254, 192.175.128.255). This increases the probability of success in our noisy mapping of fund IP addresses to EDGAR log file's masked IP addresses.[7]

It is also important to note that many firms use automated connections to access the SEC filings. As a consequence, the same IP appears several times in a short period of time (even within a second), adding volume to the funds' EDGAR connection counts without reflecting actual additional information processing by the investors. Hence, we restrict our EDGAR access count to one per IP address a day. In the case that a fund manages more than one IP address (more than

[6] According to Loughran and McDonald (2017), the investor relations website for many firms link investors directly to the EDGAR website for any SEC filings, and these views and downloads are also counted in the EDGAR log.

[7] It is possible that funds that belong to the same group share IP addresses but it is not always the case. For instance, in our sample, Funds 4 and 8 in a group share the same IP addresses, while F1, F13, and F14, in a different group, do not.

one set of three octets), we consider more than one EDGAR connection by a fund a day, which would reflect more than one processing unit within the fund.

Then, we use the total number of accesses to EDGAR that each fund made each quarter during our sample period to proxy for their direct information acquisition activity to different assets.[8] In our sample, Only 9 out of the 20 funds accessed SEC filings through EDGAR. This may be due to the fact that there are many other direct sources for mutual funds to acquire information about listed companies, such as their website, shareholders roadshows, or direct visits to the company. As a result, our EDGAR access proxy may underestimate the direct information acquisition channel. However, for those 9 funds that accessed EDGAR, there are plenty of cross-sectional and time variations in the access counts for us to explore empirically.

2.4 Summary Statistics

We report in Table 4 the descriptive statistics of our outcome variables, direct and network information acquisition variables that we use in our benchmark analysis. We take the natural logarithm of the independent variables to keep their magnitudes more comparable with each other. And since direct information variables' data is sparser than the other variables, the number of observations in a regression will be constrained by those direct information variables.

Table 4: Summary Statistics

Variable	Definition	Obs	Mean	Std. dev.	Min	Max
Outcome:						
aCorrelation	Absolute value of fund-pair correlation	13,107	0.056	0.149	0	0.999
Direct information:						
lf1edgar	log(all EDGAR connections in a quarter by fund1 in the fund-pair +1)	4,563	0.854	1.424	0	6.979
lf2edgar	log(all EDGAR connections in a quarter by fund2 in the fund-pair +1)	4,880	1.551	1.964	0	6.979
lf1f2edgar	log(all EDGAR connections in a quarter by fund-pair +1)	2,226	2.058	2.026	0	7.076
ltotinfo	log(all EDGAR connections in a quarter by all sampled funds +1)	9,500	4.741	1.684	0	7.130
Networks:						
lbidirectionalconn	log(bilateral fund-pair network measure)	13,870	-3.662	1.469	-9.118	-0.090
ltotnet	log(system-wide network measure)	13,870	2.643	0.116	2.413	2.848

Notes: Sample period is 2004Q1-2022Q1, except that we only have EDGAR log file data up to 2016Q4.

3 Empirical Strategy

We now examine the effect of direct information acquisition and network connectivity on investor portfolio decisions using quarterly fund-pair panel data. In benchmark specifications, we aim to test how direct information acquisition (by the fund pair and other funds) and indirect information

[8]Our EDGAR access counts include a fund's access to any assets not just the assets they hold.

acquisition through (bilateral and system-wide) networks affect the strength of the correlation between investor portfolio decisions. We consider each possible combination (i, j) of the 20 mutual funds (with $i \neq j$) in our sample. Thus, we get a panel with 190 mutual fund pairs across 73 quarters in 2004Q1-2022Q1.

Our benchmark specification is:

$$\begin{aligned} aCorrelation_{ij,t} = \beta_0 + \beta_1 log(f1f2edgar_{ij,t}) + \beta_2 log(total\ information_t) \\ + \beta_3 log(bilateral\ connectivity_{ij,t}) + \beta_4 log(total\ network_t) + \lambda_{ij} + \mu_{ij,t} \end{aligned} \tag{6}$$

The dependent variable $aCorrelation_{ij,t}$ is the absolute value of pair-wise correlation between the two funds' (i, j) portfolios in quarter t. Control variable $f1f2edgar$ is fund pair-wise direct information acquisition through EDGAR and is measured as the sum of both funds' number of accesses to EDGAR. Variable $total\,information$ $(totinfo)$ is the direct information acquisition through EDGAR by all funds in our sample. Variable *bidirectional connectivity* is the pair-wise bilateral connectivity measure between the two funds, i.e., $C_{i,j}$ from Equation (2). Variable *total network* (*totnet*) is the system-wide connectivity measure among all funds in our sample, i.e., C over time from Equation (5). We also include pair-wise fixed effects, λ_{ij}. Again, notice that all of the raw values for the independent variables are logged.

In the robustness tests, we also consider *overlap* as the dependent variable and other specifications, which we will detail in Section 4.2. Finally, in all specifications using fund-pair data, standard errors are two-way clustered at the fund level for each fund in the pair, to control for possible correlations from repeated observations of a fund across pairs and quarters. Without the two-way clustering, the standard errors would be underestimated, although the coefficient estimates are not biased.

4 Results

This section presents the effects of direct and network information acquisition on funds' portfolio choice similarities. To highlight the different information channels, we use four metrics: pair-wise direct information acquisition, system-wide direct information acquisition, bilateral/bidirectional network, and system-wide network. Our results show that the system-wide network is essential for increasing fund portfolio similarities.

Next, We conduct various robustness tests to validate our benchmark results. Then, we extend our analysis to explore asymmetric responses of positive versus negative correlations, respectively, the interactive effects of the information channels on the correlations, and how the effects of information acquisition change during high uncertainty periods. Finally, we examine whether there exists any amplification or crowding-out effect between the four information channels and directional information flows.

4.1 Benchmark

We first present the effects of direct and network information acquisition on funds' portfolio choice correlation (Table 5). In column (1), we only consider the effect of direct information acquisition. According to the coefficient of $lf1f2edgar$, when a fund pair acquires more information, the absolute correlation between the funds' portfolios decreases, indicating that the funds may have different private information sets or different interpretations of their private information. However, when the total information directly acquired by the whole set of funds in the sample increases (see the coefficient of $ltotinfo$), the pair-wise absolute correlation increases. This implies an increase in the total information in the network leading fund pairs to make more correlated portfolio decisions.

Table 5: Direct and Network Information Acquisition

Absolute Correlation	(1) Direct	(2) Network	(3) Bilateral	(4) System-wide	(5) All
lf1f2edgar	-0.0052^{*}		-0.0015		-0.0051^{*}
	(0.0023)		(0.0011)		(0.0027)
ltotinfo	0.0059^{**}			0.0004	0.0028
	(0.0025)			(0.0005)	(0.0017)
lbidirectionalconn		-0.0019	-0.0062		-0.0052
		(0.0017)	(0.0043)		(0.0044)
ltotnet		0.0414^{**}		0.0510^{*}	0.1565^{*}
		(0.0196)		(0.0282)	(0.0735)
_cons	0.1123^{***}	-0.0602	0.1117^{***}	-0.0759	-0.3042
	(0.0078)	(0.0556)	(0.0123)	(0.0735)	(0.2009)
N	2128	13107	2128	8927	2128
Adj. R^2	0.875	0.848	0.875	0.864	0.881

Notes: Dependent variable is the absolute value of the portfolio correlation between a pair of two funds. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *ltotnet* is the log value of the network connectivity among all funds in the sample. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column (2) in Table 5 examines the network effect and shows a positive and significant coefficient on variable *ltotnet*. That is, when the network connectivity is stronger in the sample, the two funds in a pair make more correlated portfolio decisions. The bilateral/bidirectional connectivity has an insignificant negative effect. This suggests that on average funds rely less on bilateral networks and copy from *one* other fund's information or its investment choices, they learn more from the *system-wide* network.

More specifically, one possible interpretation of the above bilateral and system-wide network effect results is that the bulk of bilateral information flows among funds are likely to be idiosyncratic in nature and out of liquidity needs. As a consequence, such information flows typically result in trade either with each other or through a third party. For instance, when Funds A and B are strongly connected, Fund A learns that Fund B has excess liquidity needs to meet. They will then be able to trade in the market in opposite directions, pushing their portfolio correlations in a negative direction.

These liquidity needs, though, are idiosyncratic and washed out in the system-wide connectivity. The fact that system-wide network connectivity engenders a strong, significant, and positive correlation between fund portfolios suggests that after averaging out information regarding idiosyncratic liquidity needs, some information circulating in the network regarding asset payoffs remains. For instance, Fund A learns positive news about Apple stock from an EDGAR pull. This news leaks through network connections and induces positive correlation across funds as they all seek to purchase more Apple stock.

Column (3) in Table 5 considers only the effects of bilateral factors: pair-wise direct information acquisition and pair-wise connectivity. Both weaken the absolute correlation between two funds on average, confirming the results in columns (1) and (2). However, neither factor is statistically significant.

The effect of system-wide factors is highlighted in column (4): direct information acquisition by all funds and the entire network's connectivity. Their positive coefficients show that system-wide information acquisition and network connectivity strengthen the absolute correlation between two funds, confirming the prior results. Again, only the network variable (logged *totnet*) is statistically significant.

In column (5), we combine all four information variables into one specification. The coefficients are consistent with those in the previous columns. The coefficient of *ltotnet* is still the largest, highlighting the importance of the system-wide network effect in investment decisions.

4.2 Robustness

We test the robustness of the benchmark findings in this subsection. From our previous results, the system-wide network information diffusion effect is shown to be important for how correlated fund portfolios are. However, given the way we measure networks using fund returns, there are several concerns. First, it is possible that this network effect we measure is in fact a price effect. That is, instead of information diffusion through the network, what the *totnet*'s coefficient shows is funds reacting to observing the common market price. Second, if funds directly acquire information simultaneously and their information all conveys a similar story, which triggers similar reactions and fund returns, then this direct information acquisition could be mistakenly counted as bilateral connectivity by our network measurement. Third, when any two funds have overlapping portfolios, the overlap can mechanically raise both the portfolio asset position correlation and our bilateral/bidirectional connectivity measure, as the latter is derived from portfolio returns that depend on portfolio positions.

To address the above concerns, we conduct robustness checks in the next three subsections. In addition, in the last subsection, we conduct robustness checks by including a dummy variable for fund pairs in the same group and its interactions. All the robustness regression results are reported in the Appendix A.

4.2.1 Price effect

To distinguish price effect from network effect, we lag our independent variables by 1 quarter and 2 quarters, respectively, and run the same set of benchmark regressions as before. The intuition is that if it were price effect, the coefficient of lagged *totnet* would be different from that of not-lagged *totnet*, as the price effect should be simultaneous and not have any impact 1 or 2 quarters later. The results are reported in Table A.1. We can see that the coefficients of lagged *totnet* remain statistically significant and similar in magnitude. This result indicates that the network effect we have found previously is not a price effect but an effect with persistent impact like a network effect would be.

4.2.2 Simultaneous information acquisition

To estimate the possibility that our network measure captures the cases when two or more funds directly acquire similar information simultaneously, first, we test whether any two funds acquire information in the same quarter. Using the panel data of fund pairs over the sample period, we run a regression of a fund's EDGAR access ($lf1edgar$) on another fund's EDGAR access ($lf2edgar$) in the same period and vice versa. The results are reported in Table A.2 columns (1) and (2). We find significantly positive coefficients for the impact of a fund's direct information acquisition on another fund's direct information acquisition in the same quarter. Hence, we find evidence of simultaneous direct information acquisition.

However, will such incidences of simultaneous direct information acquisition dominate our measure of bilateral/bidirectional connectivity such that the measure does not capture connectivity but the simultaneous direct information acquisition? In a second test, we regress our bilateral/bidirectional connectivity measure on pair-wise direct information acquisition and vice versa (Table A.2 columns (3) and (4)). We do not find any evidence for pair-wise direct information acquisition ($lf1f2edgar$) correlating with our measure of bilateral/bidirectional connectivity ($lbidirectionalconn$). Hence, we conclude that our network measure is not dominated by cases where funds simultaneously and directly acquire similar information.

4.2.3 Fund portfolio overlap

This subsection examines whether overlapping portfolio positions artificially raise both the portfolio correlation and the bilateral/bidirectional connectivity measure in our sample. As in Pool et al. (2015), we compute the overlap between portfolios as in Equation (1).

Table A.3 shows the correlations among three variables: the absolute correlation used in the benchmark, the degree of portfolio overlap, and bilateral/bidirectional connectivity of a fund pair. It is not surprising to see a high correlation between the absolute correlation and portfolio overlap. And it is reassuring to see low correlations between bilateral/bidirectional connectivity and the other two variables. This implies that our absolute correlation variable captures what we need, i.e., the degree of portfolio overlap, while portfolio overlap does not influence our bilateral/bidirectional connectivity measure.

In addition, as an alternative measure for portfolio allocation decision similarity, we use portfolio overlap as the dependent variable, as in Pool et al. (2015). The results are shown in Table A.4 and similar to the benchmark results in Table 5. Bilateral variables ($lf1f2edge$ and $lbidirectionalconn$) negatively affect portfolio overlap, although insignificant. Among all of the variables, the system-wide network effect $ltotnet$ is still the strongest and most positive.

4.2.4 Funds in the same group

The last robustness issue is that some of the funds in our sample belong to the same company group, as shown in Table 1. Understandably, these funds in the same group on average show a higher bilateral/bidirectional connectivity than other unrelated funds in our sample, according to our t-test.[9] To eliminate the complications that can arise from funds in the same company group, we add to the benchmark regressions a dummy variable for fund pairs that belong to the same group and its interaction with key explanatory variables. Since fund managers working for

[9] The t-test result is not shown here but is available upon request.

the same group are very likely to know each other, this specification can also tell us how much our benchmark specification underestimates the bilateral connectivity effect. If *bidirectionalconn* perfectly captured a bilateral connection, the coefficient for $group \times lbidirectionalconn$ would be zero.

The results are reported in Table A.5. The estimated coefficient for $group \times lbidirectionalconn$ is statistically insignificant, which implies that *lbidirectionalconn* captures well the bilateral connection. The group dummy's interaction terms with other information channel variables are also mostly insignificant. The only exception is for $1.group \times lf1f2edgar$ which has a significantly positive coefficient across specifications. This means that for funds in the same groups, direct information acquisition leads to a stronger portfolio correlation, and the opposite case is true for funds not in the same groups. It is possible that communication to reach a concensus after processing directly acquired information is more common between fund managers within the same group than those that are not, which leads to our finding.

4.3 Extensions

Having established the robustness of our empirical measures and results, in the next several subsections we extend the analysis to examine asymmetric responses of positive and negative portfolio correlations, the interactive impact of information channels on portfolio correlations, and high uncertainty periods, as well as how information channels can affect each other.

4.3.1 Asymmetric Responses

In this subsection, we divide our sample by the sign of the correlation between any two funds' portfolios. Notice that the dependent variables are still the absolute value of the correlations regardless of the signs of the correlations. In Table 6, we report results for funds with non-negative correlation in the top panel and those for funds with negative correlation in the bottom panel.

The non-negative correlation results are similar to those in the benchmark in Table 5 but with larger coefficient magnitudes, highlighting the importance of network and information diffusion. It is worth emphasizing that the number of observations for the non-negative correlation (3,089) is about one third of that for the negative correlation (10,001). The median of the non-negative correlations is 0.095, while that of the negative correlations is -0.006. Although many funds are negatively correlated, the correlations are rather weak. For funds that are positively correlated, the correlations tend to be much stronger.

Despite that majority of funds are negatively correlated, the negative correlation results are nuanced because of the weak correlation. Bidirectional network in Table 6 column (6) makes fund portfolios less negatively correlated, while system-wide network is no longer significant in affecting funds whose portfolios are negatively correlated. Overall, the magnitude of the information effect for the negatively correlated funds is much smaller than that for the positively correlated funds.

4.3.2 Interactive Effects

To understand how the different information channels interact with each other in affecting fund portfolio decisions, we add their interaction terms to our benchmark specification and report the results in Table 7. Only column (2) shows a significant interaction effect. Although the bilateral network by itself strengthens portfolio correlation, its effect is weakened by the system-wide network, which is not surprising to learn. Moreover, at first it may appear odd that now the coefficient of *ltotnet* is insignificant. However, once we take into account the significant negative coefficient for the interaction term *lbidirectionalconn* × *ltotnet* and *lbidirectionalconn* being negative (see Table 4), the overall effect of *ltotnet* is still positive as in the benchmark results.

These results highlight the importance of network effects, whether being bilateral or system-wide. It is also reassuring to know that network variables do not have any interactive effect with direct information acquisition on portfolio correlations, and vice versa.

4.3.3 High Uncertainty Periods

Next, we consider possible state dependency by interacting our independent variables of interest with the quarterly volatility index (VIX). But before we run the regression, one may think that direct and network information acquisition can fluctuate with economic uncertainty. In our dataset, the only variable that has a relatively high correlation (50%) with VIX is *ltotnet*. This makes sense in that during such high uncertainty periods, prices are often driven by common macroeconomic and uncertainty factors, and given the way we have constructed the network measurement, during volatile periods *ltotnet* may capture more of those common factors rather than the actual network structure.

To tease out the component of macroeconomic and uncertainty effect on the network, we regress *ltotnet* on VIX with the following specification and then use the residual and constant $(\alpha_0 + e_{ij,t})$, which we call *Residual totnet*, to measure the system-wide network without the uncertainty factor influence.

$$ltotnet_{ij,t} = \alpha_0 + \alpha_1 VIX\ index_{ij,t} + e_{ij,t}$$

Table 6: Correlations: Non-negative vs Negative

Non-negative: Absolute Correlation	(1) Direct	(2) Network	(3) Bilateral	(4) System-wide	(5) All
lf1f2edgar	-0.0160^{***}		-0.0078		-0.0174^{*}
	(0.0034)		(0.0057)		(0.0075)
ltotinfo	0.0141^{***}			-0.0004	0.0058
	(0.0021)			(0.0027)	(0.0040)
lbidirectionalconn		-0.0047	-0.0121^{**}		0.0039
		(0.0066)	(0.0045)		(0.0082)
ltotnet		0.1814^{**}		0.2271^{*}	0.5962^{**}
		(0.0819)		(0.1294)	(0.1886)
Constant	0.3444^{***}	-0.2802	0.3518^{***}	-0.3780	-1.1755^{*}
	(0.0130)	(0.2281)	(0.0071)	(0.3356)	(0.4796)
N	691	3089	691	2241	691
Adj. R^2	0.746	0.799	0.743	0.837	0.799
Negative: Absolute Correlation	**(5) Direct**	**(6) Network**	**(7) Bilateral**	**(8) System-wide**	**(5) All**
lf1f2edgar	-0.0000		-0.0000		-0.0000
	(0.0001)		(0.0001)		(0.0001)
ltotinfo	0.0000			0.0001	0.0000
	(0.0001)			(0.0001)	(0.0001)
lbidirectionalconn		-0.0002^{**}	-0.0000		-0.0000
		(0.0001)	(0.0001)		(0.0001)
ltotnet		-0.0020		-0.0023	-0.0009
		(0.0012)		(0.0016)	(0.0011)
Constant	0.0075^{***}	0.0111^{***}	0.0074^{***}	0.0126^{***}	0.0096^{***}
	(0.0004)	(0.0031)	(0.0003)	(0.0042)	(0.0027)
N	1428	10001	1428	6661	1428
Adj. R^2	0.804	0.339	0.804	0.359	0.804

Notes: Dependent variable is the absolute value of the portfolio correlation between a pair of two funds only if the correlation is non-negative (top panel) or negative (bottom panel). $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *ltotnet* is the log value of the network connectivity among all funds in the sample. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Interactions Between Direct and Network Information Acquisition

Absolute Correlation	(1) Direct	(2) Network	(3) Bilteral	(4) System-wide
lf1f2edgar	−0.0071		0.0036	
	(0.0057)		(0.0026)	
ltotinfo	0.0056*			−0.0023
	(0.0028)			(0.0100)
lf1f2edgar × ltotinfo	0.0003			
	(0.0010)			
lbidirectionalconn		0.0467*	−0.0099	
		(0.0223)	(0.0059)	
ltotnet		−0.0258		0.0460
		(0.0221)		(0.0433)
lbidirectionalconn × ltotnet		−0.0184**		
		(0.0087)		
lf1f2edgar × lbidirectionalconn			0.0019	
			(0.0012)	
ltotnet × ltotinfo				0.0010
				(0.0040)
Constant	0.1140***	0.1175*	0.1004***	−0.0630
	(0.0099)	(0.0582)	(0.0177)	(0.1121)
N	2128	13107	2128	8927
Adj. R^2	0.875	0.848	0.875	0.864

Notes: Dependent variable is the absolute value of the portfolio correlation between a pair of two funds only if the correlation is non-negative. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *ltotnet* is the log value of the network connectivity among all funds in the sample. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Uncertainty and Information Acquisition

Absolute Correlation	(1) All	(2) Non-negative	(3) Negative
lf1f2edgar	0.0042	0.0093	−0.0000
	(0.0054)	(0.0269)	(0.0001)
ltotinfo	0.0017	0.0173^{**}	0.0001
	(0.0034)	(0.0058)	(0.0001)
lbidirectionalconn	0.0043	0.0371	-0.0003^{**}
	(0.0068)	(0.0231)	(0.0001)
Residual totnet	0.2101	0.6683^{*}	-0.0078^{**}
	(0.1131)	(0.3285)	(0.0033)
VIX index	0.0112	0.0275	-0.0009^{**}
	(0.0092)	(0.0373)	(0.0004)
lf1f2edgar × VIX index	−0.0006	−0.0016	0.0000
	(0.0004)	(0.0015)	(0.0000)
ltotinfo × VIX index	0.0001	−0.0005	−0.0000
	(0.0002)	(0.0003)	(0.0000)
lbidirectionalconn × VIX index	-0.0006^{*}	−0.0020	0.0000^{**}
	(0.0003)	(0.0012)	(0.0000)
Residual totnet × VIX index	−0.0045	−0.0103	0.0004^{**}
	(0.0035)	(0.0150)	(0.0002)
Constant	−0.4218	−1.3369	0.0255^{**}
	(0.2887)	(0.8281)	(0.0084)
N	2128	691	1428
Adj. R^2	0.883	0.810	0.807

Notes: Dependent variable is the absolute value of the portfolio correlation between a pair of two funds. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *Residual totnet* is the residual of log value of the network connectivity among all funds in the sample teasing out the VIX index effect. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8 shows how direct information acquisition and bilateral network affect fund portfolio correlations depending on the economic volatility. During more volatile periods, bilateral networks reduce slightly the portfolio correlation. The impact of network *Residual network* remains positive and significant for non-negatively correlated funds, and its effect does not vary with VIX.

For those negatively correlated funds, information diffusion through the network during volatile periods also matters. In particular, during normal times the system-wide network will make the portfolio correlation less negative. This effect is dampened during economic volatile times.

4.3.4 Amplification or Crowding Out

So far we have examined the information channels' impact on fund portfolio correlations and overlaps. In this section, we turn to examine how the information channels affect each other. We first focus on individual funds' direct information acquisition behavior and then shift our attention to fund pairs and their bilateral and system-wide networks.

Individual fund direct information acquisition

First, we examine how an individual fund's direct information acquisition can be affected by the direct acquisition of the other fund in a pair and by their bilateral connection. Table 9 columns (1) and (2) show the results. Each fund directly acquires more information from EDGAR when the other fund in the pair does the same, and this effect is strengthened when the two funds have a stronger bilateral network tie. However, the bilateral tie itself reduces the direct information acquisition. These results are interesting in light of the free-riding issue that is widely discussed in the literature. We show that what discourages a fund's information acquisition is not the information acquisition of the other fund in the same network, but the fact that the fund is in a network itself. That is, fund managers may choose direct information acquisition less often in general because they rely on a network (of indirect information); but once other managers in their network have acquired information directly, they will do so as well.

In Table 9 columns (3) and (4), we replace the bilateral network variable with VIX index to study the effect of uncertainty. We confirm our prior result that funds positively affect each other's direct information acquisition from EDGAR. We also find that uncertainty reduces slightly but significantly the direct information acquisition and weakens the positive impact of the other fund's direct information acquisition on one's own acquisition. Since in our data uncertainty is associated with a stronger system-wide network, it is possible that during such volatile times, funds resort more to network information instead. However, a key difference between uncertainty and network is that uncertainty makes the positive effect of funds' direct information acquisition on each other weaker, while network makes it stronger. It may be due to the fact that uncertainty is also associated with low information quality regardless of its sources as found by previous research (e.g., Chahine et al. (2021)).

Table 9: Individual Fund Direct Information Acquisition: By Pair

	(1) lf1edgar	(2) lf2edgar	(3) lf1edgar	(4) lf2edgar
lf2edgar	0.2600***		0.1946**	
	(0.0573)		(0.0713)	
lbidirectionalconn	−0.0678**	−0.1365**		
	(0.0275)	(0.0414)		
lf2edgar × lbidirectionalconn	0.0318*			
	(0.0143)			
lf1edgar		0.9205***		0.5194**
		(0.1769)		(0.1837)
lf1edgar × lbidirectionalconn		0.1475***		
		(0.0392)		
VIX index			−0.0142**	−0.0311***
			(0.0054)	(0.0092)
lf2edgar × VIX index			−0.0016**	
			(0.0005)	
lf1edgar × VIX index				−0.0067
				(0.0039)
Constant	0.3082	0.7795***	0.8057***	1.8579***
	(0.1690)	(0.0816)	(0.1350)	(0.1074)
N	2226	2226	2226	2226
Adj. R^2	0.506	0.440	0.509	0.437

Notes: Dependent variable is the direct informational acquisition of one fund in a fund pair. $lf1(f2)edgar$ is the log value of the direct information acquisition from accessing EDGAR by a fund in a pair plus one. $lbidirectionalconn$ is the log value of the network connectivity between the two funds in a pair. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In addition, we construct a new dataset of individual funds instead of fund pairs used in the previous analysis to examine how the network affects a fund's direct information acquisition. As shown in Table 10 columns (1)-(3), we find no evidence for the impact of centrality or directional information flows from or to a fund. However, when a fund is more connected to other funds overall (column 4) and when there is more direct information acquisition by other funds in the network (column 5), a fund's direct information acquisition increases. Column (5) result is consistent with the prior result in Table 9 in that others' direct information acquisition increases one's own acquisition, but column (4) result appears contradictory to the prior result that network itself reduces one's own direct information acquisition while increasing it when others also directly acquire information. Hence, we add the interaction term between *Othersnet* and *lOthersinfo* in column (6), and the result becomes consistent with that in Table 9.

Table 10: Individual Fund Direct Information Acquisition: By Individual

	(1) lfedgar	(2) lfedgar	(3) lfedgar	(4) lfedgar	(5) lfedgar	(6) lfedgar
EVCentrality	-16.7814					
	(12.9724)					
FromOthers ($C_{i\leftarrow\bullet}$)		-0.4046				
		(1.5372)				
ToOthers ($C_{\bullet\leftarrow j}$)			0.1171			
			(0.0715)			
Othersnet ($C_{i\leftarrow\bullet} + C_{\bullet\leftarrow j}$)				0.0972^{**}		-0.7588^{***}
				(0.0321)		(0.2073)
lOthersinfo (log(totinfo-fedgar))					0.2764^{**}	0.0318
					(0.0879)	(0.0899)
lOthersinfo × Othersnet						0.1532^{***}
						(0.0244)
Constant	2.2162^{**}	1.5060	1.1022^{***}	1.0511^{***}	-0.0488	1.1707^{*}
	(0.7745)	(1.1087)	(0.0684)	(0.0538)	(0.4065)	(0.5309)
Observations	497	497	497	497	485	485
R^2	0.015	0.001	0.002	0.001	0.121	0.193
Adjusted R^2	0.013	-0.001	0.000	-0.001	0.119	0.188
F	1.6735	0.0693	2.6840	9.1887	9.8947	752.1984

Notes: Dependent variable is the log value of the direct information acquisition from accessing EDGAR by a fund. $EVCentrality$ is the eigen vector measure of the fund's centrality in the network. $FromOthers$ is the information flow from other funds in the network. $ToOthers$ is the information flow to other funds in the network. *Othersnet* is the sum of the above two variables. $lOthersInfo$ is the logged value of total direct information acquisition by all other funds in the sample except the fund in question. We also include fund fixed effects. Robust standard errors are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fund-pair and system-wide direct information acquisition

Now we return to the fund-pair dataset as used in the benchmark regressions to examine the information acquisition of fund pairs, as well as that of the network. First, we construct a new variable $lothertotinfo$. It is the logged value of total direct information from EDGAR excluding the direct information of the fund-pair of interest. It measures the other funds' (less the fund-pair in question) direct information acquisition from EDGAR in our system. Then we construct another new variable *lothernet*. It is the logged value of all the information flows from and to the fund pair less the bilateral connectivity between the pair, i.e., $C_{1\leftarrow\bullet} + C_{\bullet\leftarrow 1} + C_{2\leftarrow\bullet} + C_{\bullet\leftarrow 2} - C_{1,2}$.

Table 11: Fund-pair and System-wide Direct Information Acquisition

	(1) lf1f2edgar	(2) lf1f2edgar	(3) lf1f2edgar	(4) ltotinfo	(5) ltotinfo
lothertotinfo	0.4070***		0.0315		
	(0.0736)		(0.0774)		
lothernet		0.7145	−1.5229**		
		(0.3995)	(0.5566)		
c.lothertotinfo × c.lothernet			0.3757***		
			(0.0515)		
lbidirectionalconn				−0.0096	
				(0.0778)	
ltotnet					4.0104*
					(2.0963)
Constant	0.2926	1.3238**	1.8200**	4.7049***	−5.8576
	(0.3194)	(0.4106)	(0.6371)	(0.2905)	(5.5451)
Observations	2226	2226	2226	9500	50
R^2	0.485	0.363	0.508	0.000	0.071
Adjusted R^2	0.474	0.350	0.497	−0.020	0.051
F	30.5624	3.1987	.	0.0154	3.6597

Notes: Dependent variable is information acquisition variable of interest. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. $lothertotinfo$ is the log value of $totinfo - f1f2edgar + 1$. *lothernet* is the log value of all the information flows from and to the fund pair less the bilateral connectivity between the funds in the pair. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *ltotnet* is the log value of the network connectivity among all funds in the sample. We also include fund-pair fixed effects, and standard errors, two-way clustered by each fund in the pair, are in parentheses, except for the regression of $ltotinfo$ in column (4) with only system-wide variables and many fewer observations. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11 columns (1)-(3) show that others' direct information acquisition increases a fund-pair's direct information acquisition but the effect is mostly through network, and the network itself reduces direction acquisition. These results are consistent with our previous results on individual funds. When we turn to the total information acquisition by all funds, we find that the system-wide direct information is not affected by information flow through the bilateral network (column 3), but is positively affected by the system-wide network (column 4).

With all of the above results, we conclude that the effect of networks on information acquisition is not a simple crowding out or amplification but both: Although network itself does crowd out direct information acquisition, it strengthens the amplification effect among funds' direct information acquisition (i.e., others' information acquisition increases one's own acquisition). This further highlights the importance of network in information acquisition: Networks not only disseminate information but also accelerate direct information acquisition across funds. Hence, networks play an essential role in the estimation of the total amount of information in the system.

Fund-pair directional information flows

At last, we explore what factors affect directional information flows between two funds. We find that a fund's centrality, which captures how connected a fund is to highly connected funds, positively affects the fund's impact on another fund (Table 12 columns (1) and (3)). This effect is stronger if the two funds have a stronger bilateral network. More interestingly, column (4) shows that a fund's centrality has a negative impact on its reception of influence from another fund (although not always significant, e.g., in column 3), and this effect is also further strengthened by bilateral network.

Table 12: Directional Information Flows

	(1) $Eff1f2$	(2) $Eff2f1$	(3) $Eff2f1$	(4) $Eff1f2$
EVCenF1	0.7100***	1.1798		
	(0.2065)	(0.9013)		
lbidirectionalconn	−0.0015	0.0297***	−0.0058	0.0052***
	(0.0016)	(0.0099)	(0.0044)	(0.0005)
EVCenF1 × lbidirectionalconn	0.1283***	0.3287		
	(0.0379)	(0.2097)		
EVCenF2			2.8318***	−0.1384***
			(0.2047)	(0.0211)
EVCenF2 × lbidirectionalconn			0.6852***	−0.0167**
			(0.0547)	(0.0066)
Constant	−0.0088	0.1766***	0.0039	0.0308***
	(0.0082)	(0.0392)	(0.0226)	(0.0025)
N	13870	13870	13870	13870
Adj. R^2	0.198	0.805	0.891	0.155

Notes: $Eff1f2$ is the bilateral directional influence from fund 1 to fund2, vice versa for $Eff2f1$. $EVCenF1(2)$ is the centrality measure of the fund 1(2). *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Conclusion

Information technology has increased both direct information acquisition and network connectivity in financial markets. In this paper, we propose an empirical assessment of the effects of direct information acquisition and network connectivity on the allocation decisions of a group of institutional investors.

We implement a novel application for the network extraction methodology from Diebold and Yilmaz (2015). Their framework allows us to compute bilateral and system-wide network connections. We also use SEC EDGAR log files to measure funds' firsthand direct information acquisition. Using panel fixed effect regressions, we find evidence that system-wide network connectivity strongly increases institutional investors' portfolio correlation and is consistently important.

Moreover, our data allows us to examine the impact of information sources on each other in a network. Others' direct information acquisition amplifies one's own acquisition. Network accelerates such acquisition amplification between funds, although by itself network also reduces direct information acquisition in general.

Our results call for new theoretical models that can combine both direct information acquisition and network and examine their joint impact on portfolio allocation and returns. Han and Yang (2013) is one of the first theoretical papers to combine the two information channels, although they focus more on market efficiency and welfare, rather than asset allocation. More recently, Cobas et al. (2023) also explore a competitive financial market with networks and costly information acquisition, where the direct information acquisition by others in the network can assist one's own acquisition, and show that portfolio correlation increases with the total amount of information in a system that includes both direct and network information acquisition, consistent with our empirical findings. Another implication of our result is to market volatility, which is well known to be affected by asset allocation.

This paper's empirical results also highlight the importance of network and the role it plays in affecting market information. The literature has been dominated by empirical evidence on strong free-riding incentives and the crowding-out effect of network or public information on private information. This paper peels back the well-studied crowding-out layer and reveals the positive impact of network on information acquisition. Policymakers should take into account such network effects when considering financial market information flows and the subsequent influence on market volatility and liquidity. Future studies can also dig deeper into the impact of information channels on fund performances and the characteristics of their asset holdings using our measurements of information channels.

References

Chahine, Salim, Mai Daher, and Samer Saade, "Doing good in periods of high uncertainty: Economic policy uncertainty, corporate social responsibility, and analyst forecast error," *Journal of Financial Stability*, 2021, *56*, 100919.

Chen, Deqiu, Yujing Ma, Xiumin Martin, and Roni Michaely, "On the fast track: Information acquisition costs and information production," *Journal of Financial Economics*, 2022, *143* (2), 794–823.

Chen, Honghui, Yuanyu Qu, Tao Shen, Qinghai Wang, and David X. Xu, "The Geography of Information Acquisition," *Journal of Financial and Quantitative Analysis*, 2022, *57* (6), 2251–2285.

Cobas, Adriana, Grace Weishi Gu, and Zachary Stangebye, "Social Networks and Financial Markets," *mimeo*, 2023.

Cohen, Lauren, Andrea Franzzini, and Christopher Malloy, "Sell-Side School Ties," *The Journal of Finance*, 2010, *65* (4), 1409–1437.

_ , Andrea Frazzini, and Christopher Malloy, "The Small World of Investing: Board Connections and Mutual Fund Returns," *Journal of Political Economy*, 2008, *116* (5), 951–979.

Di Maggio, Marco, Francesco Franzoni, Amir Kermani, and Carlo Sommavilla, "The relevance of broker networks for information diffusion in the stock market," *Journal of Financial Economics*, 2019, *134* (2), 419–446.

Diebold, Francis X. and Kamil Yilmaz, *Financial and Macroeconomic Connectedness. A network approach to measurement and monitoring*, Oxford University Press, 2015.

Ding, Jerry Parwada Ning and Jianfeng Shen, "Information Sharing within the Networks of Delegated Portfolio Managers: Evidence from Plan Sponsors and Their Subadvisers," *Journal of Behavioral Finance*, 2017, *18* (1), 99–113.

Feng, Lei and Mark S. Seasholes, "Correlated Trading and Location," *The Journal of Finance*, 2004, *59* (5), 2117–2144.

Ganguli, Jayant Vivek and Liyan Yang, "Complementarities, Multiplicity, and Supply Information," *Journal of the European Economic Association*, 2009, *7* (1), 90–115.

Gibbons, Brian, Peter Iliev, and Jonathan Kalodimos, "Analyst Information Acquisition via EDGAR," *Management Science*, 2021, *67* (2), 661–1328.

Goldstein, Itay and Liyan Yang, "Information Diversity and Complementarities in Trading and Information Acquisition," *The Journal of Finance*, 2015, *70* (4), 1723–1765.

Goyal, Sanjeev, Stephanie Rosenkranz, Utz Weitzel, and Vincent Buskens, "Information Acquisition and Exchange in Social Networks," *The Economic Journal*, 2017, *127* (606), 2302–2331.

Halim, Edward, Yohanes E. Riyanto, and Nilanjan Roy, "Costly Information Acquisition, Social Networks, and Asset Prices: Experimental Evidence," *The Journal of Finance*, 2019, *74* (4), 1975–2010.

Han, Bing and Liyan Yang, "Social networks, Information Acquisition and asset prices," *Management Science*, 2013, *59* (6), 1444–1457.

Heimer, Rawley Z., "Peer Pressure: Social Interaction and the Disposition Effect," *The Review of Financial Studies*, 07 2016, *29* (11), 3177–3209.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, "Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers," *The Journal of Finance*, 2005, *60* (6), 2801–2824.

Ivković, Zoran and Scott Weisbenner, "Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices," *The Review of Financial Studies*, 2007, *20* (4), 1327–1357.

Loughran, Tim and Bill McDonald, "The Use of EDGAR Filings by Investors," *Journal of Behavioral Finance*, 2017, *18* (2), 231–248.

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Information Acquisition and Under-Diversification," *The Review of Economic Studies*, 04 2010, *77* (2), 779–805.

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

Pool, Veronika K., Noah Stoffman, and Scott E. Yonker, "The people in your neighborhood: social interactions and mutual fund portfolios," *The Journal of Finance*, 2015, *70* (6), 2679–2732.

Rossi, Alberto G., David Blake, Allan Timmermann, Ian Tonks, and Russ Wermers, "Network centrality and delegated investment performance," *Journal of Financial Economics*, 2018, *128* (1), 183–206.

Ruiz-Buforn, Alba, Eva Camacho-Cuena, Andrea Morone, and Simone Alfarano, "Overweighting of public information in financial markets: A lesson from the lab," *Journal of Banking & Finance*, 2021, *133*, 106298.

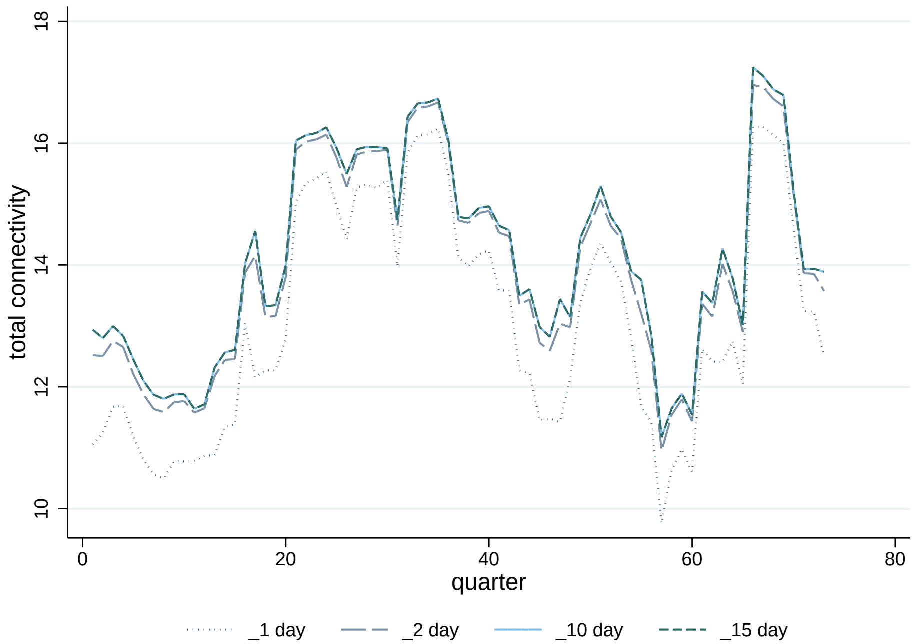
Shiller, Robert J. and John Pound, "Survey evidence on diffusion of interest and information among investors," *Journal of Economic Behavior & Organization*, 1989, *12* (1), 47–66.

Shive, Sophie, "An Epidemic Model of Investor Behavior," *The Journal of Financial and Quantitative Analysis*, 2010, *45* (1), 169–198.

Veldkamp, Laura L., "Information Markets and the Comovement of Asset Prices," *The Review of Economic Studies*, 2006, *73* (3), 823–845.

Appendix A Robustness Checks

Figure A.1: Connectivity measure across alternative projection horizon levels.



Notes: The quarterly connectivity measure corresponds to the sum of all the estimated linkages among the nodes each period.

Table A.1: Direct and Network Information Acquisition: Lag(s)

One Lag: Absolute Correlation	(1) Direct	(2) Network	(3) Bilateral	(4) System-wide	(5) All
L.lf1f2edgar	−0.00434*		−0.00215		−0.00415
	(0.00225)		(0.00206)		(0.00244)
L.ltotinfo	0.00352***			−0.000355	0.000380
	(0.000987)			(0.000749)	(0.00120)
L.lbidirectionalconn		−0.00145	−0.00330		−0.00254
		(0.00148)	(0.00227)		(0.00323)
L.ltotnet		0.0417**		0.0548*	0.160*
		(0.0196)		(0.0283)	(0.0770)
Constant	0.121***	−0.0593	0.122***	−0.0827	−0.296
	(0.0000269)	(0.0555)	(0.00347)	(0.0722)	(0.203)
Observations	2145	13041	2145	9051	2145
Adjusted R^2	0.877	0.848	0.877	0.865	0.883
r2	0.879	0.850	0.879	0.868	0.885

Two Lags: Absolute Correlation	(1) Direct	(2) Network	(3) Bilateral	(4) System-wide	(5) All
L2.lf1f2edgar	−0.00324		−0.00315		−0.00321
	(0.00219)		(0.00239)		(0.00201)
L2.ltotinfo	0.000219			−0.00148	−0.00298
	(0.000306)			(0.00117)	(0.00193)
L2.lbidirectionalconn		−0.00144	−0.00371		−0.00326
		(0.00149)	(0.00207)		(0.00251)
L2.ltotnet		0.0405*		0.0590*	0.170*
		(0.0195)		(0.0296)	(0.0833)
Constant	0.135***	−0.0561	0.123***	−0.0884	−0.311
	(0.00314)	(0.0561)	(0.00208)	(0.0734)	(0.213)
Observations	2145	12870	2145	9070	2145
Adjusted R^2	0.876	0.848	0.876	0.864	0.883
r2	0.879	0.850	0.879	0.867	0.885

Notes: Dependent variable is the absolute value of the portfolio correlation between a pair of two funds. All main controls are lagged by one quarter in the top panel and by two quarters in the bottom panel. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *ltotnet* is the log value of the network connectivity among all funds in the sample. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Funds' Simultaneous Direct Information Acquisition

	(1) lf1edgar	(2) lf2edgar	(3) lbidirectionalconn	(4) lf1f2edgar
lf2edgar	0.1825**			
	(0.0759)			
lf1edgar		0.4520**		
		(0.1469)		
lf1f2edgar			−0.0173	
			(0.0175)	
lbidirectionalconn				−0.0624
				(0.0397)
Constant	0.5191***	1.2449***	−3.3921***	1.8441***
	(0.1223)	(0.1195)	(0.0360)	(0.1361)
N	2226	2226	2226	2226
R^2	0.512	0.434	0.747	0.361

Notes: Dependent variable is the information variable of interest. $lf1(f2)edgar$ is the log value of the direct information acquisition from accessing EDGAR by a fund in a pair plus one. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair plus one. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Overlapping Portfolios

	aCorrelation	overlapping	bidirectionalconn
aCorrelation	1.0000		
overlapping	0.8944	1.0000	
bidirectionalconn	-0.0647	-0.0504	1.0000

Table A.4: Direct and Network Information Acquisition: Portfolio Overlap

Portfolio Overlap	(1) Direct	(2) Network	(3) Bilateral	(4) System-wide	(5) All
lf1f2edgar	−0.0031		−0.0010		−0.0030
	(0.0017)		(0.0010)		(0.0018)
ltotinfo	0.0033**			−0.0012	0.0014*
	(0.0010)			(0.0008)	(0.0007)
lbidirectionalconn		−0.0030	−0.0031		−0.0025
		(0.0018)	(0.0028)		(0.0029)
ltotnet		0.0781**		0.0487**	0.1015*
		(0.0319)		(0.0186)	(0.0498)
_cons	0.0895***	−0.1843*	0.0901***	−0.0822*	−0.1784
	(0.0014)	(0.0893)	(0.0075)	(0.0472)	(0.1336)
N	2226	13870	2226	9500	2226
Adj. R^2	0.906	0.746	0.905	0.868	0.909

Notes: Dependent variable is the degree of portfolio overlapping between a pair of two funds. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. *lbidirectionalconn* is the log value of the network connectivity between the two funds in a pair. *ltotnet* is the log value of the network connectivity among all funds in the sample. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Direct and Network Information Acquisition: Dummy for Funds in the Same Group

Absolute Correlation	(1) Direct	(2) Network	(3) Bilateral	(4) System-wide	(5) All
lf1f2edgar	-0.0050^{**}		-0.0024		-0.0050^{*}
	(0.0021)		(0.0019)		(0.0024)
ltotinfo	0.0043^{*}			0.0002	0.0016
	(0.0019)			(0.0007)	(0.0014)
1.group $\times lf1f2edgar$	0.0509^{***}		0.0682^{***}		0.0639^{***}
	(0.0013)		(0.0191)		(0.0110)
1.group $\times ltotinfo$	0.0302			0.0120	0.0163
	(0.0207)			(0.0160)	(0.0093)
lbidirectionalconn		-0.0016	-0.0055		-0.0050
		(0.0016)	(0.0045)		(0.0047)
ltotnet		0.0378^{*}		0.0458^{*}	0.1353^{*}
		(0.0182)		(0.0258)	(0.0626)
1.group $\times lbidirectionalconn$		-0.0130	-0.0156		0.0061
		(0.0076)	(0.0110)		(0.0296)
1.group $\times ltotnet$		0.1747		0.2306	0.5773
		(0.2009)		(0.3424)	(0.4933)
_cons	0.1120^{***}	-0.0598	0.1117^{***}	-0.0748	-0.3117
	(0.0086)	(0.0551)	(0.0135)	(0.0748)	(0.2008)
N	2128	13107	2128	8927	2128
Adj. R^2	0.880	0.848	0.878	0.865	0.888

Notes: Dependent variable is the absolute value of the portfolio correlation between a pair of two funds. $lf1f2edgar$ is the log value of the direct information acquisition from accessing EDGAR by the fund pair. $ltotinfo$ is the log value of the direct information acquisition from accessing EDGAR by all funds in the sample. $lbidirectionalconn$ is the log value of the network connectivity between the two funds in a pair. $ltotnet$ is the log value of the network connectivity among all funds in the sample. $group = 1$ when two funds are in the same group, otherwise zero. We also include fund-pair fixed effects. Standard errors, two-way clustered by each fund in the pair, are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.