Co-searching for Alpha

Boone Bowles Dora Horstman Xiaan Zhou *

June 2025

Abstract

We infer firm connections from hedge funds' co-search behavior in EDGAR and find that connected firms exhibit strong return predictability: annual alpha is 8.16%. Moreover, this effect is strongest for firms with complex disclosures and sparse analyst coverage, consistent with models of limited attention. We also show that hedge funds frequently churn their network of firm connections—responding to news and capturing fleeting ties among firms—and actively trade based on these connections. These findings highlight a distinct view of firm connectedness and uncover a novel channel of information flow that hedge funds translate into consistent alpha.

JEL classification: G12, G14, G23

Keywords: Co-search, Return predictability, Hedge fund, Information acquisition, EDGAR

*We thank Tyler Beason, Felipe Cabezon, Yong Chen, Lauren Cohen, Peter Cziraki, Roger Edelen, Jesse Ellis, William Grieser, Xing Huang, Adam Kolasinski, Philipp Illeditch, Greg Leonard, Jiangyuan Li, Kai Li, Markus Lithell, Bradley Paye, Alex Pecora, Ville Rantala, Vijay Singal, Ngoc-Khanh Tran, Jun Tu, Neng Wang, Pingle Wang, Mark Walker, Wei Wu, Yuqing Yang, Pengfei Ye and seminar and conference participants at NC State, Texas A&M University, Virginia Tech, 2024 Financial Management Association Meeting (Idea Session), 20th Annual Meeting of the Financial Research Association (Idea Panel), and Finance and Accounting 2025 Annual Research Symposium for their helpful comments and suggestions. All errors are our own. Contact information: Boone Bowles, Mays Business School, Texas A&M University, College Station, TX 77843; E-mail: boone.bowles@tamu.edu. Dora Horstman, Poole College of Management, North Carolina State University, Raleigh, NC 27695; E-mail: dlhorstm@ncsu.edu. Xiaan Zhou, Mays Business School, Texas A&M University, College Station, TX 77843; E-mail:xzhou520@tamu.edu.

1 Introduction

Firms are linked through various economic activities which have been shown to influence firm value.[1] Yet investors often overlook these connections, resulting in slow information diffusion and cross-asset return predictability. Motivated by this inefficiency, researchers have sought to understand how market participants gather and process inter-firm information.[2] But still, how investors learn from and trade on these connections is not well-understood.

Furthermore, while most investors may overlook these links, hedge funds are uniquely equipped to recognize and exploit them. Widely regarded as the market's most sophisticated traders (Brunnermeier and Nagel, 2004), hedge funds act as informed arbitrageurs, detecting and correcting mispricing at the individual stock level.[3] But inter-firm links—especially less-obvious links—are more complex than single-firm signals. If any investors are discovering them and exploiting them it should be hedge funds—yet, again, we know relatively little about which connections hedge funds study and whether those links feed into their trades and even market prices.[4]

This paper advances our understanding of both issues by directly observing hedge-fund research behavior on the SEC's EDGAR platform and identifying every instance of a hedge fund downloading filings from separate firms within days of each other (co-searching). We then construct a dynamic network of firms and show that it has strong predictive power for returns: a long-short strategy that buys firms whose network peers recently earned high returns and sells those whose peers lagged delivers risk-adjusted returns of 0.68% per month (8.16% per year). We also show that return predictability is three to five times stronger in settings magnifying limited attention—like when filings are complex or analyst coverage

[1] For example, supply-chain relationships and product-market competition have been well documented. See Cohen and Frazzini (2008), Menzly and Ozbas (2010), Ramírez (2024), Hoberg and Phillips (2016), and Dou and Wu (2023).

[2] E.g., Ali and Hirshleifer (2020) and Kaustia and Rantala (2021) examine the role of financial analysts in disseminating firm connection information.

[3] E.g., Akbas, Armstrong, Sorescu, and Subrahmanyam (2015), Chen, Da, and Huang (2019), Chen, Kelly, and Wu (2020).

[4] Two related papers are Gao and Huang (2016) and Crane, Crotty, and Umar (2023).

is thin—and the effect survives every common risk model. Moreover, linking our co-search data to 13F holdings reveals that hedge funds tilt their portfolios based on these networks, confirming that their research informs their trades.

Finally, we show that the inferred network evolves rapidly: over 30% of strong links disappear within a single quarter. This high churn rate coincides with news. In fact, only the links established during news days predict returns. Our findings reflect that it is the dynamic and real-time updating by hedge funds that predict returns. This distinction advances our view of hedge funds as fast information intermediaries and differentiates this new hedge fund learning network from other networks based on fundamentals or easily observable relationships.

A key feature of our paper is the novel dataset we construct capturing the research activity of hedge funds on the EDGAR platform. Specifically, using a hand-collected sample of unmasked hedge funds in the EDGAR Log Files, we focus on instances where the same hedge fund co-searches multiple public companies within days of each other. These co-searches, reflecting joint attention to potentially-related firms, are used to infer inter-firm links (or an investor-driven "learning network") from the investor's perspective. We then aggregate co-searches for every firm across all hedge funds, weighting a firm's link to its peers by the intensity of co-searches. The median firm in the sample has 276 peers across 98 different industries.

After identifying inter-firm links, our first set of analyses examines whether the network inferred from hedge funds' research contains information. We test this by sorting firms into portfolios based on the recent performance of their in-network peers. For example, we consider long positions in firms whose peers had high returns over the last quarter while taking short positions in firms whose peers performed the worst. The long-short return on this strategy over the next quarter is between 0.68% and 0.74% per month (between 8% and 9% per year), depending on the risk model.

We find consistent return results using Fama-MacBeth regressions and show return predictability persists over the next year—not just over one quarter—with no evidence of reversal. Taken together, our results suggests that there is a considerable amount of information contained in these networks and that hedge funds exploit informational inefficiencies rather than generate temporary price pressure.

Models of limited attention suggest that return predictability should be stronger in opaque information environments (Hirshleifer and Teoh, 2003). We use two prominent informational frictions from the literature—the complexity of information disclosures (You and Zhang, 2009; Loughran and McDonald, 2014) and the intensity of analyst coverage (Gibbons, Iliev, and Kalodimos, 2021; Chen et al., 2020)—to test whether the predictability in the network is related to limited attention. Our results confirm that it is: return predictability is most powerful—up to 5 times stronger than average—when information disclosures are more complex and among firms covered by fewer analysts. These findings not only support the limited attention hypothesis, but, more importantly, they highlight that hedge funds have a distinct advantage in both processing intricate disclosures and exploiting inefficiencies arising from investor inattention.

We also explore whether hedge funds' *trading* behavior mimics the firm connections uncovered by their own *research* behavior. After all, if their research predicts alpha, their trading should attempt to capture that alpha. Our tests reveal that hedge funds buy stocks whose co-searched peers have outperformed and sell stocks whose peers have underperformed. Further, these results are stronger when we incorporate short positions and consider hedge funds' net arbitrage activity.[5] Taken together, our findings show that hedge funds not only act as arbitrageurs at the individual-firm level, but also at the inter-firm level, exploiting cross-asset return predictability.

Next, we turn to examine the network itself, asking how dynamic this hedge fund learning network of firms is? The answer: remarkably dynamic. Tracking the retention

[5] Chen et al. (2019) constructs net arbitrage trading (NAT) to measure stock-level aggregate arbitrage trades. We consider aggregate short positions as a proxy for hedge funds short positions and use NAT to measure aggregate hedge fund trading including both long and short positions. See Appendix A.2.5 for regression results.

rate of co-searched firms across quarters reveals that 88% of a firm's closely-linked peers are new every quarter. In contrast, that same figure is 58% considering a network inferred from mutual fund co-searches, and is even lower for more traditional networks, such as supply-chain and product-market links. This finding highlights both the ephemeral nature of these connections and the unique adaptability of hedge funds to learn from fleeting firm connections.

To isolate the role of fleeting connections in hedge fund learning networks, we construct a set of pseudo-peer firms: firms that were co-searched by hedge funds in a previous quarter but are no longer classified as peers in the current quarter. Interestingly, portfolios formed using pseudo-peers exhibit no return predictability, suggesting that it is the fleeting ties—rather than historical co-search patterns—that capture economically meaningful information not yet fully incorporated into stock prices.

But what drives the fleeting connections? And how do hedge funds recognize it? To answer these questions we consider firm-news and construct two new sets of peer firms: one identified during *news* days and the other during *quiet* days (non-news days). Returning to our previous methods, we find that return predictability is driven only by peers linked during news days, not quiet days, suggesting that the information edge stems from transient links rather than persistent, structural relationships among firms. We further validate this by examining whether hedge fund co-searches are triggered by news events that mention related firms. Using RavenPack data to identify news stories that link two firms within the same event, we find a strong positive correlation between contemporaneous news coverage and hedge fund co-search activity. Importantly, we find no consistent association when news co-coverage occurs in the prior or subsequent quarters, reinforcing the notion that hedge funds quickly recognize emerging inter-firm connections.

In addition to our main results, we examine a series of alternative explanations that could account for the observed return predictability. One possibility is that the returns of peer firms simply proxy for a given firm's own returns or for broader industry return dynamics. If this were the case, the observed effects could be explained by standard stock-level momentum (Jegadeesh and Titman, 1993, 2002; Novy-Marx, 2012) or industry momentum (Moskowitz and Grinblatt, 1999; Hoberg and Phillips, 2018). To address this, we control for firms' own lagged returns as well as value-weighted returns of its three-digit SIC industry group. Furthermore, we exclude stocks from the same industry when constructing peer sets to further isolate the effects from industry-wide trends. Our results remain statistically significant and economically meaningful across all specifications, indicating that the return predictability captured by the hedge fund learning network is not explained by traditional momentum effects.

To better understand hedge funds' information advantage, we benchmark the hedge fund learning network against an analogous network constructed from mutual funds' co-search activity. While mutual funds face different institutional frictions and investment mandates, they nevertheless serve as a useful comparison group of sophisticated and informed investors. Peer firms based on mutual fund co-searches on the EDGAR platform exhibit no return predictability across all specifications and time horizons. This sharp contrast suggests that hedge funds have superior ability in discovering inter-firm links. Further, we follow Lee, Ma, and Wang (2015) and compare the explanatory power of conventional financial ratios of peer firms to those of the target firm. We find that mutual funds rely more on conventional characteristics to link firms. Together, our results underscore the special ability of hedge funds to detect unique and fleeting inter-firm connections.

Finally, we test the robustness of our network's return predictability. In our main analysis, we define peers as firms whose filings were co-searched within a ±7-day window. We adjust the search-window to ±15 days and ±3 days separately and confirm that the return predictability remains in both cases.[6] We also extend our data set to incorporate additional filing types.[7] Incorporating these additional filings into the hedge fund learning

[6]The learning network with ±3 days of time window shows the strongest return predictability. This is probably because a shorter time-window can capture stronger and more relevant information connections between firms.

[7]The filings we use include material event disclosures (8-K), initial public offering information (S-1), and

network construction, we find that the return predictability remains significant, confirming that the network captures predictive signals across a broad set of information sources.

Our paper makes several important contributions to the literature. First, it extends the research on firm networks by introducing a novel perspective to describe how firms are connected. Prior work has examined firm connections through one specific channel such as customer-supplier relationships (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Ramírez, 2024), firm employment ties (Giroud and Mueller, 2019), management connections (El-Khatib, Fogel, and Jandik, 2015), and product market competition (Hoberg and Phillips, 2016; Dou and Wu, 2023). Unlike documenting one specific way, our study introduces an investor learning network, which reflects overall how investors understand firm relationships. This approach aligns with the spirit of Ali and Hirshleifer (2020) and Kaustia and Rantala (2021), who also examine market participants' attention from an all-in-one perspective. By bridging the gap between economic firm linkages and investor behavior, our framework provides a view from market participants, shedding light on how firm connections are understood and acted upon by sophisticated investors.

Second, we show that inter-firm return predictability is driven by fleeting, dynamically updated connections. While prior research highlights that structural relationships, such as customer-supplier links, can predict stock returns (Cohen and Frazzini, 2008), we find that stale connections lack predictive power in our setting. This is because many inter-firm ties emerge from short-lived events—such as joint ventures, litigation, or market shocks—making them more transient than persistent firm-specific characteristics. Our findings suggest the ability to detect and respond to newly forming links matters. As emphasized in recent studies such as Bowles, Reed, Ringgenberg, and Thornock (2024), Ivkovich and Zekhnini (2024) and Bowles, Reed, Ringgenberg, and Thornock (2025), timely recognition and exploitation of informational signals are essential for generating alpha. By highlighting the rapid churn of peer firm networks, our study provides new insight into how sophisticated investors uncover

proxy statement (14A). Lee et al. (2015) use the same set of filings.

and trade on dynamic firm connections.

Finally, this paper expands the literature on the role of sophisticated investors in incorporating individual firm information into inter-firm information. The potential profits from trading provide incentives for arbitrageurs to gather and process information (Grossman and Stiglitz, 1980). Prior research has documented how sophisticated investors leverage information advantages to facilitate mispricing correction at the individual stock level. A vast body of literature explores the role of hedge funds as arbitrageurs in the stock market (Chen et al., 2020, 2019; Cao, Liang, Lo, and Petrasek, 2018b; Cao, Chen, Goetzmann, and Liang, 2018a; Calluzzo, Moneta, and Topaloglu, 2019; Akbas et al., 2015; Kokkonen and Suominen, 2015; Ben-David, Franzoni, and Moussawi, 2012; Agarwal, Daniel, and Naik, 2009; Stein, 2009). Building on the literature, our paper is the first to examine how arbitrageurs trade based on inter-firm information advantages rather than just firm-specific information. We find that hedge funds systematically arbitrage the well-documented lead-lag return relationship across firms, highlighting a distinct mechanism through which they facilitate price discovery and enhance market efficiency.

The remainder of the paper is organized as follows. Section 2 describes the data, the network construction method, and the summary statistics. Section 3 presents the main results on the network return predictability and hedge funds arbitrage trading on inter-firm links. Section 4 provides additional tests supporting our main hypotheses and results. Section 5 presents robustness tests. Finally, Section 6 concludes.

2 Data and Methodology

Our primary data source is the SEC's EDGAR Log Files, which we use to identify the information acquisition activity—and implied learning networks—of hedge funds and mutual funds. The EDGAR Log Files record every electronic "request" to view a public filing in the SEC's EDGAR database. Each observation lists the filing requested, the timestamp of the request, and the requesting IP address.[8] For each request, we link requesting IP addresses to hedge funds names obtained from hedge fund commercial databases and mutual funds in the CRSP Mutual Funds database. We also link filings to public firms in the Center for Research in Securities Prices database (CRSP).[9]

To construct the hedge fund learning network, we first link IP addresses in the EDGAR Log Files to hedge funds and match the filings to public firms. For each firm, we define its peer firms every quarter using a three-step process. Step 1, for every firm we identify the date on which a hedge fund requests the firm's 10-K or 10-Q filings, then—looking within the 7 days before and 7 days after this date—we identify the other public firms that are researched by the same hedge fund. Firms co-searched like this are designated as "peer firms."

Step 2, we measure co-search frequency for every firm with each of its peer firms by calculating the *minimum daily request count* between the two companies by a given hedge fund. This approach captures the frequency with which each pair of firms are researched together, while mitigating bias arising from differences in firm size or filing popularity.

Step 3, we aggregate co-search frequency by firm-quarter; summing the co-search frequencies across all hedge funds. This yields the total co-search frequency between each pair of firms for the quarter. This method ensures joint learning behavior by individual hedge funds rather than filing request activity driven by widespread investor attention or firm size.

Example (Tesla, 2015 Q2). On May 22, 2015, an IP address registered to hedge fund Point72 requests Tesla's 10-K/Q twelve times. Four days earlier the same IP requested MGM Resorts' filings twice, so MGM is a peer of Tesla that quarter with a co-search frequency of 2.

Figure 1 shows the learning network around Tesla in the second quarter of 2015. Hedge funds—including Citadel, Point72, and others—co-searched for Tesla alongside firms such

[8] E.g., The filing request is for the 2013 annual report (10-K) for IBM. The request was made at 10:14 am on March 1, 2014. The request originated from the IP address 123.123.123.*abc*.

[9] A full description of the EDGAR Log Files and the procedure to unmask IP addresses to identify mutual funds and hedge funds is provided in Appendix A.1.

as California Resources Corporation (energy), Delta Airlines (transportation), and MGM Resorts (hospitality). These co-searches form the edges of the network and capture moments when hedge funds evaluate Tesla in the context of economically or thematically related firms, even across different sectors. To capture the dynamic nature of investor attention, we update the network at a quarterly frequency. Figure 2 presents the learning network in the following quarter. While Tesla remains central, its co-searched peers have shifted. Delta and California Resources are no longer present, while new connections emerge with Amazon and Precision Castparts, suggesting an evolving set of perceived linkages. These changes suggest that hedge funds continuously adapt their information acquisition behavior in response to changing firm fundamentals or macroeconomic factors. Our network provides a real-time, behaviorally grounded lens on how sophisticated investors perceive economic relationships among firms.

2.1 Comparison With Other Networks

Our learning networks capture unique connections among firms. Consistent with Lee et al. (2015), linked firms seldom share an industry classification; the ten closest peers, on average, span eight distinct three-digit SIC codes. We also compare the hedge fund learning network to the TNIC3 network from Hoberg and Phillips (2016); a network based on firm similarity scores. The common linkages shared by the two networks only account for 1% of firm links in our learning network. This indicates that the hedge fund learning network is novel and does not simply reflect product market connections.

2.2 Other Data Sources

We rely on several data sources in addition to the EDGAR Log Files. Briefly, we use institutional holdings data from Thomson Reuters 13F, news data from RavenPack, monthly returns and prices from CRSP, accounting data from COMPUSTAT, measures of 10-K complexity obtained from The Notre Dame Software Repository for Accounting and Finance

(SRAF), and analyst coverage data from I/B/E/S.[10]

2.3 Summary Statistics

Table 1 reports firm-quarter-level summary statistics for the sample that covers 2009 Q1–2017 Q2. Panel A shows that the average firm has 374 peers across 98 three-digit SIC industries—only 4.02% of the peers belong to the same industry—while Panel B shows that the average firm has hedge fund ownership of 4.17%. Panel C summarizes 10-K filings complexity and analyst coverage.

3 Results

This section presents three main sets of empirical results that collectively demonstrate how hedge funds acquire and act on firm connections.

3.1 Return Predictability

In an efficient market, where information is rapidly incorporated into stock prices, return predictability based on firm connections would not be observed. However, in an imperfect market, return predictability arises if certain economic relationships between firms are slowly reflected in prices. If hedge funds uncover these neglected firm links through their research, peer-firm returns within learning networks should forecast future returns over and above what can be explained by temporary price pressures.

3.1.1 Portfolio Sorting Analysis

Each quarter we sort stocks into quintiles by their peers' prior-quarter returns and form a long-short portfolio that buys the top quintile and sells the bottom. The peer returns

[10]More details on these subsidiary data sources are found in Appendix A.3.

are aggregated by co-search frequency in the main results.[11] Panel A of Table 2 reports abnormal returns estimated across various risk models, including: excess market returns, Fama-French three factors (Fama and French, 1993), Carhart momentum factor (Carhart, 1997), Fama-French five factors (Fama and French, 2015), Pástor-Stambaugh liquidity factor (Pastor and Stambaugh, 2003), and mispricing factors (Stambaugh and Yuan, 2017).[12]

The results demonstrate significant return predictability within hedge fund learning networks. For a one-quarter holding period, value-weighted long-short strategy earns an average monthly abnormal return of 0.68% (approximately 8.16% annually).[13] The abnormal returns remain statistically significant across all risk models. Also, when tracking portfolio performance over the four subsequent quarters—as in Figure 3—we find that monthly returns concentrate in the first quarter following portfolio formation and have no significant effects in later quarters. This suggests that any information from peer firms is incorporated into prices within a quarter and, further, that the returns are not driven by price pressure.[14] Overall, our portfolio analysis highlights the superior ability of hedge funds to identify economically meaningful, yet often overlooked, firm links.

3.1.2 Fama-MacBeth Regressions

To further validate our results, we perform Fama-MacBeth regressions which test whether peer information carries incremental explanatory power after controlling for conventional risk betas. The dependent variable is excess firm returns. The main independent variable, $PeerRet_{i,t-1}$, is the lagged returns of peer firms and we control for various firm risk exposures.

[11] Appendix Table A1 shows the results of equally-weighting the co-searched firms. This weighting method allows us to differentiate the information contained in the selection of co-search firms from the collective wisdom of hedge funds. Our results show that the overall return predictability of peer firms is driven by both reasons.

[12] In Table 2, portfolio returns are shown for the quarter after the quarter of portfolio formation. Appendix Table A2 shows the results for longer holding periods.

[13] In addition, Appendix Table A3 reports results based on equal-weighted portfolios, which reveal a similar pattern of significant abnormal returns across all horizons, further validating the robustness of the findings.

[14] For specific statistics, see Appendix Table A4.

The regression model is specified as follows [15]:

$$Ret_{i,t} = \alpha_i + \alpha_t + \beta_1 PeerRet_{i,t-1} + \beta_2 Controls_{i,t-1} + \varepsilon_{i,t}. \tag{1}$$

Panel B of Table 2 presents the results of the Fama-MacBeth regressions. In Column (1) we report the result for the baseline regression without controlling for firm risk exposures while in Columns (2)–(7) we gradually add firm betas estimated from a 24-month rolling window.[16] The inclusion of risk exposure at the firm level ($\beta_{i,t-1}$) slightly reduces the magnitude of the coefficients, yet they remain statistically significant. In the full model—Column (7)—the coefficient remains at 0.068 (t-stat = 2.15), demonstrating robustness across different asset-pricing models.

3.1.3 Heterogeneity in Different Information Environments

Investor inattention delays the incorporation of information into stock prices. Information frictions, such as the complexity of firm disclosures or the opacity of information dissemination environment, pose challenges for investors and contribute to slower information diffusion. Building on this insight, we hypothesize that hedge funds gain a comparative advantage in environments characterized by greater information friction. These frictions create opportunities for sophisticated investors to extract value from firm disclosures that are overlooked or misinterpreted by the broader market. If this mechanism is at play, we should observe stronger return predictability within the hedge fund learning network for firms facing more information frictions.

Disclosure Complexity We begin by considering the complexity of the firm disclosures. Our hypothesis is that 10-K complexity creates hurdles and frictions for investors to un-

[15] Standard errors are corrected for autocorrelation using Newey-West adjustments with three lags.

[16] We use the following risk factors: excess market returns, Fama-French three factors (Fama and French, 1993), Carhart momentum factor (Carhart, 1997), Fama-French five factors (Fama and French, 2015), Pástor-Stambaugh liquidity factor (Pastor and Stambaugh, 2003), and mispricing factors (Stambaugh and Yuan, 2017).

derstand, thus the investors with better understanding and information processing ability should have superior information advantage. Using the complexity score from Loughran and McDonald (2024), we sort firms into quintiles based on 10-K complexity and define an indicator variable $HighComplexity$, which equals one if the 10-K complexity of a given firm is in the top quintile. Our model is specified as follows:

$$Ret_{i,t} = \alpha_i + \alpha_t + \beta_1 PeerRet_{i,t-1} + \beta_2 PeerRet_{i,t-1} \times HighComplexity_{i,t-1} + \beta_3 HighComplexity_{i,t-1} + \beta_4 Controls_{i,t-1} + \varepsilon_{i,t}. \quad (2)$$

We discover stronger return predictability of the hedge fund network for firms with more complex 10-K filings. The coefficient of interest is the interaction term, $PeerRet_{i,t-1} \times HighComplexity_{i,t-1}$, which captures the differential impact of high complexity filings on return predictability. In the baseline model—Column (1) of Table 3 Panel A—the coefficient on the interaction term is 0.154 and statistically significant. This interaction term remains positive and significant—between 0.175 and 0.204—even when gradually including risk factors. This suggests that for high-complexity firms, return predictability from peer firms is approximately 5 times stronger than for low-complexity firms. Hence, hedge funds have a stronger ability to predict returns for firms with high-complexity filings and our findings strongly support the hypothesis that hedge funds' ability to identify nuanced firm connections is amplified in environments where disclosure complexity deters other investors.

Analyst Coverage Next, we examine the network return predictability across firms with varying levels of analyst coverage. As shown in Chen et al. (2020), limited analyst coverage deters information dissemination and potentially results in more mispricing. We use analyst coverage as a proxy for the transparency of a firm's information environment, with lower coverage indicating a more opaque information environment. Similarly to the previous test, we sort firms into quintiles based on the number of analysts covering the firm. We define $LowCoverage_{i,t-1}$ equal to 1 if the given firm belongs to the bottom quintile with the least number of analysts. Our model is specified as follows:

$$Ret_{i,t} = \alpha_i + \alpha_t + \beta_1 PeerRet_{i,t-1} + \beta_2 PeerRet_{i,t-1} \times LowCoverage_{i,t-1} \\ + \beta_3 LowCoverage_{i,t-1} + \beta_4 Controls_{i,t-1} + \varepsilon_{i,t}. \quad (3)$$

Again, the coefficient of interest is the interaction term $PeerRet_{i,t-1} \times LowCoverage_{i,t-1}$, which is positive and significant with coefficients ranging from 0.238 and 0.373, as shown in Panel B of Table 3. This implies that for low-coverage firms, return predictability from peer firms is up to 3.5 times stronger than for high-coverage firms, and indicates that hedge funds capture stronger signals for firms with low analyst coverage. Again, this highlights the critical role of hedge funds as arbitrageurs in opaque markets.

3.2 Hedge Fund Trades

Having shown that hedge funds can identify meaningful inter-firm links, an important question remains: Do hedge funds act on this information when making investment decisions? This section directly addresses this by linking hedge fund trading behavior to their information acquisition. We infer trading activity from changes in hedge funds' quarterly holdings, as reported in 13F filings. Then, using these changes, we estimate the following linear probability model that relates the likelihood of a hedge fund buying (or selling) a firm to the past returns of its peer firms as identified in the network:

$$Buy(Sale)_{i,t} = \alpha_i + \alpha_t + \beta_1 PeerRet_{i,t-1} + \beta_2' Controls_{i,t-1} + \varepsilon_{i,t}, \quad (4)$$

where $Buy(Sale)_{i,t} = 1$ are at the firm-quarter level and indicate whether hedge funds increase (decrease) their holdings in aggregate.

Table 4 shows that changes in hedge fund holdings do indeed follow the returns of peer firms. Panel A shows that hedge funds are more likely to increase their holdings of a firm when its peers had recent positive returns while Panel B shows that hedge funds are likely to sell when peers had low recent returns. This trading pattern suggests that hedge funds systematically interpret peer-firm performance as a signal for future stock movements and adjust their portfolios accordingly.

Further, we also study whether the trading amount is related to peer-firm returns. Following Chen et al. (2019), we define abnormal holdings as the change in the current-quarter holding minus the previous-quarter holdings. Our model is specified as follows:

$$Ab\_HF_{i,t} = \alpha_i + \alpha_t + \beta_1 PeerRet_{i,t-1} + \beta_2 Controls_{i,t-1} + \varepsilon_{i,t}, \tag{5}$$

where $Ab\_HF_{i,t}$ stands for "abnormal aggregate hedge fund ownership" for firm i in quarter t, computed as hedge fund ownership at the current quarter t minus the hedge fund ownership of last quarter $t-1$.

The results in Panel C of Table 4 indicate that past peer performance translates into significant changes in aggregate hedge fund ownership. The coefficient on $PeerRet_{i,t1}$ is positive and significant, implying that hedge fund ownership of a given firm increases following strong past returns of peer firms. This further implies that a one-standard deviation increase in past peer-firm returns leads to a 6.6 basis points increase in abnormal holdings (44% of the mean abnormal holding of 15 basis points).[17]

Collectively, our evidence demonstrates that hedge funds actively exploit neglected firm-link information and adjust their positions based on insights extracted from co-searched firms. These results also reinforce the idea of hedge funds as sophisticated arbitrageurs.

[17]See Appendix Table A5 for the results incorporating short positions

3.3 Fleeting Ties

3.3.1 Network Update Frequency

The links between firms—links captured by hedge funds—are dynamic. To examine how dynamic, we measure the update frequencies (transition matrices) of learning networks inferred from hedge funds and compare them with networks inferred from mutual funds' research. Specifically, we sort peer firms into quintiles based on the co-search frequency at quarter $t-1$ and report the percentage of peer firms that stay in the same quintile at quarter t.

Table 5 shows that 11.2% of peer firms remain as the most searched (Quintile 5) peer firms next quarter. Also, the transition matrix from hedge funds shows relatively high dropout rates—*dropout* refers to a peer firm no longer being co-searched at quarter $t+1$—from 26.4% to 67.6% across all quintiles. Column (5) and (6) report the differences in the percentages in each quintile for hedge funds' and mutual funds' transition matrices. The results highlight that firms linked by mutual fund retain a significantly higher proportion of their peers compared to firms linked by hedge funds. The results show consistent positive differences in retention rates from 5.6% to 31.0% and consistent negative differences in dropout rates from −17.2% to −22.4%.

Figure 4 extends this analysis by presenting transition rates across the subsequent four quarters. The left panel presents the transition rates of top co-searched peer firms remaining in the highest co-searched quintile over next four quarters, while the right panel shows the percentage of these peers that drop out of the network entirely. The figure reflects the dynamic nature of hedge fund networks, which are updated more frequently to incorporate new information and evolving relationships, while mutual fund networks are more stable, potentially relying on static or slow-changing peer connections.

3.3.2 Network Update and Return Predictability

To examine whether return predictability is driven by timely information regarding inter-firm links rather than persistent static relationships, we construct a placebo test using a set of pseudo peer firms. Specifically, we define pseudo peers as firms that were co-searched by hedge funds in the previous quarter but are no longer considered peers in the current quarter. While these firms may remain economically related to the firm of interest, the absence of current co-search activity suggests a lack of new or relevant information flow between them. If return predictability were solely attributable to static economic ties, we would expect similar predictive power from pseudo peer returns. However, we find no evidence of return predictability using pseudo peer performance, supporting the conclusion that the hedge fund learning network captures timely signals about evolving inter-firm connections.

Table 6 presents the results of Fama-MacBeth regressions. The first set of tests examine the contemporaneous return co-movement between peer firms and a given firm. In Columns (1) and (2), the coefficients of $PeerRet_{i,t}$ and $PseudoPeerRet_{i,t}$ are positive and statistically significant, confirming the economic ties between peer firms and a given firm. However, as shown in Columns (3) and (4), for the future returns of a given firm, only the coefficient on $PeerRet_{i,t}$ remains statistically significant, while the coefficient on $PseudoPeerRet_{i,t}$ becomes insignificant. This indicates that the information is only contained in the peer firms identified by hedge funds in the current quarter.

3.3.3 News Days Versus Quiet Days

The dynamic nature of hedge fund co-searching network motivates us to study whether hedge funds recognize these firm links from news. We carefully separate peer firms of a given firm into the ones identified during news days or quiet days. If the return predictability is driven by peers identified during news days, it suggests that the information edge of hedge funds comes from fleeting firm connections. If predictability also appears using peer firms identified on quiet days, it suggests that hedge funds' learning networks reflect overlooked persistent and structural firm relationships.

Following Bowles et al. (2024), we define the three-day window around a news event in RavenPack as news days. All other periods are classified as quiet days. The portfolio sorting results are shown in Figure 5. We find significant positive alphas *exclusively* trading based on peers identified during news days. This observation suggests that the informational advantage stems from temporarily overlooked, event-driven firm connections. In contrast, the absence of significant alpha during target firms' quiet days indicates that hedge funds' co-search activity does not reflect persistent, structural firm relationships. Overall, our results highlight hedge funds' unique ability to rapidly identify and exploit transient inter-firm information, rather than merely capturing static or passive forms of firm linkage.

3.3.4 Co-coverage of News Events

Given the evidence that hedge fund information advantage stems from timely updates, we next compare the timing of hedge fund co-search activity with instances where two firms are jointly mentioned in news coverage. If hedge funds update their learning network contemporaneously with such news co-coverage, it would provide further support that they are actively and dynamically responding to public news evolving inter-firm information.

Specifically, we utilize news co-coverage of firms, capturing instances where firms are mentioned together in significant news events. We adopt a linear regression model and a Poisson regression model. The models are specified respectively as follows:

$$
\begin{aligned}
LnNumCoSearch_{i,j,t} = \alpha_{i,j} + \alpha_t + \beta_1 LnNewsCoCoverage_{i,j,t} + \beta_2 Controls_{i,t} \\
+ \beta_3' Controls_{j,t} + \varepsilon_{i,j,t}, \quad (6)
\end{aligned}
$$

$$
\begin{aligned}
log(NumCoSearch_{i,j,t}) = \alpha_{i,j} + \alpha_t + \beta_1 NewsCoCoverage_{i,j,t} + \beta_2 Controls_{i,t} \\
+ \beta_3' Controls_{j,t} + \varepsilon_{i,j,t}. \quad (7)
\end{aligned}
$$

Table 7 presents the relationship between hedge fund co-searches on EDGAR and news co-coverage of firm pairs. Columns (1) and (2) report results from linear regressions where the independent variable is the log-transformed co-search frequency. The coefficient of $LnNewsCoCoverage_{i,j,t}$ is positive and statistically significant, indicating that greater news co-coverage between two firms is associated with increased co-search behavior by hedge funds. Columns (3) and (4) estimate Poisson regressions with the raw co-search frequency as the independent variable, and the coefficient of $NewsCoCoverage_{i,j,t}$ remains positive and highly significant. Control variables are firm size and hedge fund ownership for both firms i and j in quarter t. The inclusion of control variables in Columns (2) and (4) does not materially change the results. These findings suggest that hedge funds are more likely to co-search EDGAR filings of firms that are jointly covered in the news.

To further investigate the timing of hedge fund responses to news-based firm linkages, Appendix Table A6 explores whether co-search behavior aligns with prior or subsequent news co-coverage. Panel A shows no significant relationship when news co-coverage occurs in the previous quarter, suggesting that hedge fund co-search is not driven by stale information. Panel B similarly shows no consistent relationship with news co-coverage in the next quarter across different models. These results reinforce the interpretation that hedge funds update their information networks contemporaneously with relevant news, rather than front running news or reacting to delayed news.

4 Additional Tests

4.1 Alternative Explanations

In addition to the primary analysis, we study several alternative factors that could explain the observed return predictability within the hedge fund learning network. One of these is stock momentum. If peer firm returns can serve as effective proxies for the contemporary performance of a given firm, the discovered return predictability could be driven by the firm's auto-correlation. To address this, we augment our Fama-MacBeth regressions to control for the given firm's own lagged returns.

As shown in Table 8 Column (2), the regression results demonstrate that the predictive power of peer firms' returns in hedge fund learning networks remains robust even after controlling for firm lagged return $Ret_{i,t-1}$. The coefficients of $PeerRet_{i,t-1}$ are consistently positive and statistically significant, while those of $Ret_{i,t-1}$ are insignificant, indicating that a firm's own past returns cannot predict the firm's return in the next quarter. These findings confirm that hedge fund learning networks identify meaningful inter-firm links beyond simple return auto-correlations and reveal hedge funds' ability to uncover firm link information.

We also consider whether industry momentum could drive the results. Prior research, such as Moskowitz and Grinblatt (1999) and Hoberg and Phillips (2018), highlights the role of industry-wide trends in stock returns, which could potentially explain the predictive power of peer firm returns. To test this, we include industry returns as controls in our models.

Table 8 Column (3) and (4) report the results of Fama-MacBeth regression controlling for the given firm's industry lagged return ($IndustryRet_{i,t-1}$). The industry is defined using 3 digits SIC code. Our results show that peer firms' returns significantly predict future returns. Industry-level returns also positively predict returns, but their impact does not undermine the effects of peers. The firm's own lagged returns are insignificant, confirming that return predictability is not driven by historical performance. These findings underscore the informational advantage of hedge fund learning networks in identifying economically relevant and nuanced inter-firm linkages, even across industry boundaries.

4.2 Comparison With Mutual Funds

To better understand the information content of firm links identified by hedge funds, we compare the differences in return predictability between the networks derived by hedge funds and by mutual funds. We construct mutual fund learning networks using the same methodology as for hedge funds. Specifically, we identify each firm's peers based on co-search patterns from mutual funds and test for return predictability within the network. Panel A of Table 9 presents summary statistics for co-search peers characteristics in the mutual fund learning network: Each firm has about 1128 peer firms, much greater than peer firms of each firm in the hedge fund learning network.

We conduct portfolio sorting analysis to test for return predictability within the mutual fund networks. Specifically, we sort stocks into quintiles each quarter based on the lagged portfolio returns of peer firms identified through mutual fund co-searches. The long-short portfolios reveal no statistically significant excess returns or alphas across any holding period or risk model. For one-quarter holding periods, the monthly excess return is 0.42% and the estimated alphas are between 0.40% and 0.48% across different risk factor models, but none are statistically significant. Results over two- to four-quarter horizons are shown in Appendix Table A8. The portfolio returns remain negligible or negative, with no meaningful improvement in predictive power. This contrasts sharply with hedge fund learning networks, which exhibit robust and significant return predictability.

To understand the nature of how firms are linked by institutional investors, we follow Lee et al. (2015) and document the explanatory power (R^2) of the financial characteristics of the peer firms to those of the target firm.[18] By comparing between hedge funds and mutual funds, we show that mutual funds consistently have higher R^2, suggesting that mutual funds rely more on conventional characteristics to define firm connections. The results are consistent

[18] See Appendix A9 for more details on each financial characteristic variable used.

with our finding that it is the fleeting information that drives the return predictability of hedge funds' learning network.

5 Robustness

We consider a few robustness tests for our main results on the return predictability within the hedge fund learning network.

5.1 Co-search Window

In our main analysis, we define a firm's co-search peers as those whose filings are searched within a ± 7-day window before or after the firm's own filing search by hedge funds. This window balances the trade-off between capturing accurate peer relationships and accommodating potential lags in information release between firms. A shorter window is likely to enhance the precision of peer identification by focusing on closely timed searches. However, it may fail to account for instances where significant information is released with a lag, such as when firms have different fiscal year-ends or face delays in filing disclosures. Thus, we use both ± 15-day window and ± 3-day window as robustness checks to ensure both accuracy and inclusiveness in the identification process.

We use the same method in our main result and conduct portfolio sorting analysis. The only difference is that we adjust the co-search time window from ± 7 days to a broader ± 15-day window and a narrower ± 3-day window around each firm's filing request. Table 10 Panel A and B report the abnormal returns of the long-short portfolios, adjusted by different risk factor models. For the one-quarter holding period, the long-short portfolio delivers statistically significant monthly excess returns of 0.33% for ± 15-day window and 0.76% for ± 3-day window, corresponding to approximately 3.96% and 9.12% annually. The higher abnormal return indicates a stronger predictive power of the hedge fund learning network over a shorter holding period. The results indicate that the hedge funds' informational advantage primarily exists within a quarter and diminishes as the market gradually incorporates the information over time.

5.2 Other Filings

Although 10-K and 10-Q filings provide critical information about public firms, other filings such as 8-K, S-1, and 14A also serve as valuable sources of information and may offer additional insights into firm connections. To account for these potential sources of learning, we extend our analysis to include these filings. The results show that the return predictability within the hedge fund learning network remains robust and significant based on portfolio sorting analysis.

We reconstruct hedge fund learning network by incorporating co-searches of additional filings, including 8-K, S-1, and 14A, along with 10-K and 10-Q filings. In other words, peer firms are defined as any firms whose filings (10-K, 10-Q, 8-K, S-1, or 14A) were searched within a ±7-day window of each firm's filings. Then, we redo the portfolio sorting analysis and report the long-short portfolio returns and alphas in Table 10 Panel C. For the one-quarter holding period, the portfolio delivers significant monthly excess returns of 0.70% (approximately 8.4% annually), The alphas range from 0.74% to 0.85% across risk models, indicating strong short-term predictability. We extend the holding period and find return predictability remains significant for two quarters, but the long-short portfolio returns drop to 0.29% per month. Consistent with our previous results, these findings highlight that hedge funds can utilize a broad source of information and discover the firm connections before the market adjusts.

6 Conclusion

In this study, we build firm connections using hedge fund co-search patterns of firm filings on the SEC's EDGAR system, to examine how sophisticated investors exploit information on economically linked firms. Our findings illustrate that hedge fund learning networks have substantial predictive power for stock returns. The long-short portfolio generates up to 8.16% annual alpha over the next quarter. This predictive capability is robust across various risk adjustments. Moreover, this predictability is especially pronounced when information frictions are high where corporate disclosures are complex or analyst coverage is sparse, consistent with limited investor attention delaying information diffusion.

We also provide direct evidence that hedge funds trade strategically on these inter-firm signals: they increase their positions in a stock following high returns from its connected peers and reduce positions after poor peer performance, thereby arbitraging cross-firm mispricing. Importantly, the hedge fund learning network is highly dynamic and time-sensitive, with linkages forming and dissolving rapidly around firm-specific news. In contrast, an analogous network constructed for mutual funds shows no such predictive power, underscoring the unique ability of hedge funds to exploit evolving inter-firm information.

Taken together, our findings indicate that hedge funds serve as agile information intermediaries by leveraging neglected public signals and accelerating their incorporation into stock prices, which in turn enhances price discovery and contributes to greater market efficiency. Overall, this study sheds new light on the process of investor learning in networks and highlights the need to account for inter-firm connections when evaluating asset prices.

References

Agarwal, V., Daniel, N. D., Naik, N. Y., 2009. Hedge funds and market efficiency. The Journal of Investment Management 7, 1–18.

Akbas, F., Armstrong, W. J., Sorescu, S., Subrahmanyam, A., 2015. Smart money, dumb money, and capital market anomalies. Journal of Financial Economics 118, 355–382.

Ali, U., Hirshleifer, D., 2020. Shared analyst coverage: Unifying momentum spillover effects. Journal of Financial Economics 136, 649–675.

Ben-David, I., Franzoni, F. A., Moussawi, R., 2012. Hedge fund stock trading in the financial crisis of 2007–2009. The Review of Financial Studies 25, 1–54.

Bowles, B., Reed, A. V., Ringgenberg, M. C., Thornock, J., 2025. Predicting anomalies. Available at SSRN 4939779 .

Bowles, B., Reed, A. V., Ringgenberg, M. C., Thornock, J. R., 2024. Anomaly time. The Journal of Finance 79, 3543–3579.

Brunnermeier, M. K., Nagel, S., 2004. Hedge funds and the technology bubble. The Journal of Finance 59, 2013–2040.

Calluzzo, P., Moneta, F., Topaloglu, S., 2019. When anomalies are publicized broadly, do institutions trade accordingly? Management Science 65, 4555–4574.

Cao, C., Chen, Y., Goetzmann, W. N., Liang, B., 2018a. Hedge funds and stock price formation. Financial Analysts Journal 74, 54–68.

Cao, C., Liang, B., Lo, A. W., Petrasek, L., 2018b. Hedge fund holdings and stock market efficiency. The Review of Asset Pricing Studies 8, 77–116.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Chen, Y., Da, Z., Huang, D., 2019. Arbitrage trading: The long and the short of it. The Review of Financial Studies 32, 1608–1646.

Chen, Y., Kelly, B., Wu, W., 2020. Sophisticated investors and market efficiency: Evidence from a natural experiment. Journal of Financial Economics 138, 316–341.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. The Journal of Finance 63, 1977–2011.

Crane, A., Crotty, K., Umar, T., 2023. Hedge funds and public information acquisition. Management Science 69, 3241–3262.

Dou, W. W., Wu, W., 2023. Competition network and predictable industry returns. Available at SSRN 4535226 .

El-Khatib, R., Fogel, K., Jandik, T., 2015. Ceo network centrality and merger performance. Journal of Financial Economics 116, 349–382.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Gao, M., Huang, J., 2016. Capitalizing on capitol hill: Informed trading by hedge fund managers. Journal of Financial Economics 121, 521–545.

Gibbons, B., Iliev, P., Kalodimos, J., 2021. Analyst information acquisition via edgar. Management Science 67, 769–793.

Giroud, X., Mueller, H. M., 2019. Firms' internal networks and local economic shocks. American Economic Review 109, 3617–3649.

Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. The American Economic Review 70, 393–408.

Hirshleifer, D., Teoh, S. H., 2003. Limited attention, information disclosure, and financial reporting. Journal of accounting and economics 36, 337–386.

Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. Journal of political economy 124, 1423–1465.

Hoberg, G., Phillips, G. M., 2018. Text-based industry momentum. Journal of Financial and Quantitative Analysis 53, 2355–2388.

Ivkovich, Z., Zekhnini, M., 2024. Dissecting the timing of anomalies. Available at SSRN 4606746 .

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance 48, 65–91.

Jegadeesh, N., Titman, S., 2002. Cross-sectional and time-series determinants of momentum returns. The Review of Financial Studies 15, 143–157.

Kaustia, M., Rantala, V., 2021. Common analysts: method for defining peer firms. Journal of financial and quantitative analysis 56, 1505–1536.

Kokkonen, J., Suominen, M., 2015. Hedge funds and stock market efficiency. Management Science 61, 2890–2904.

Lee, C. M., Ma, P., Wang, C. C., 2015. Search-based peer firms: Aggregating investor perceptions through internet co-searches. Journal of Financial Economics 116, 410–431.

Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. the Journal of Finance 69, 1643–1671.

Loughran, T., McDonald, B., 2024. Measuring firm complexity. Journal of Financial and Quantitative Analysis 59, 2487–2514.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. The Journal of Finance 65, 1555–1580.

Moskowitz, T. J., Grinblatt, M., 1999. Do industries explain momentum? The Journal of finance 54, 1249–1290.

Novy-Marx, R., 2012. Is momentum really momentum? Journal of financial economics 103, 429–453.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. The Journal of Political Economy 111, 642–685.

Ramírez, C. A., 2024. Firm networks and asset returns. The Review of Financial Studies 37, 3050–3091.

Stambaugh, R. F., Yuan, Y., 2017. Mispricing factors. The Review of Financial Studies 30, 1270–1315.

Stein, J. C., 2009. Presidential address: Sophisticated investors and market efficiency. The Journal of Finance 64, 1517–1548.

You, H., Zhang, X.-j., 2009. Financial reporting complexity and investor underreaction to 10-k information. Review of Accounting studies 14, 559–586.

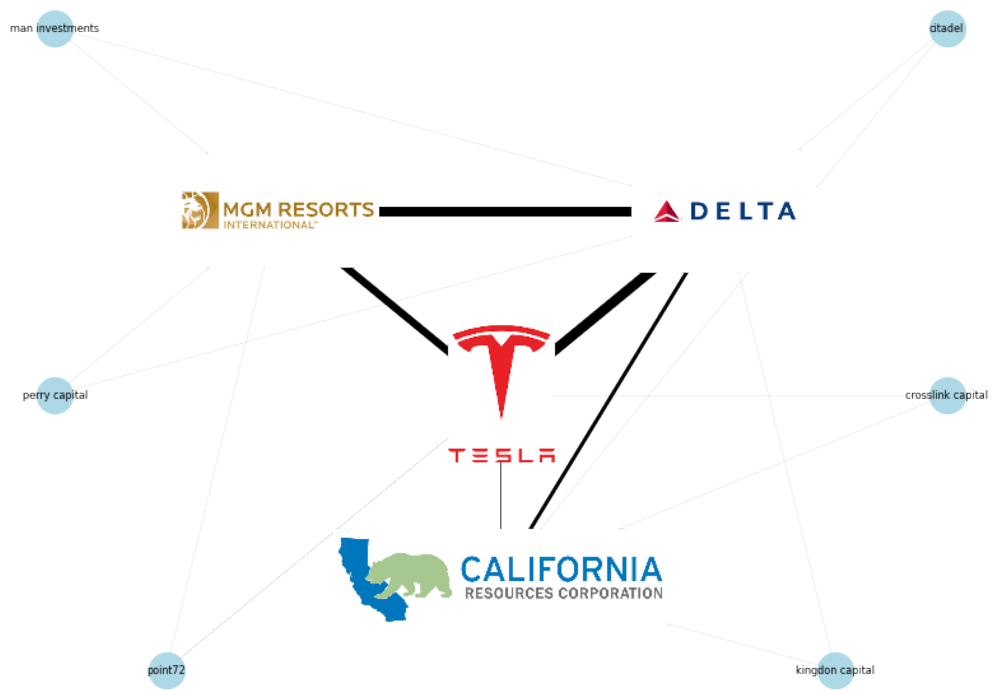

Figure 1: Example of Hedge Fund Co-search Learning Network in 2015Q2

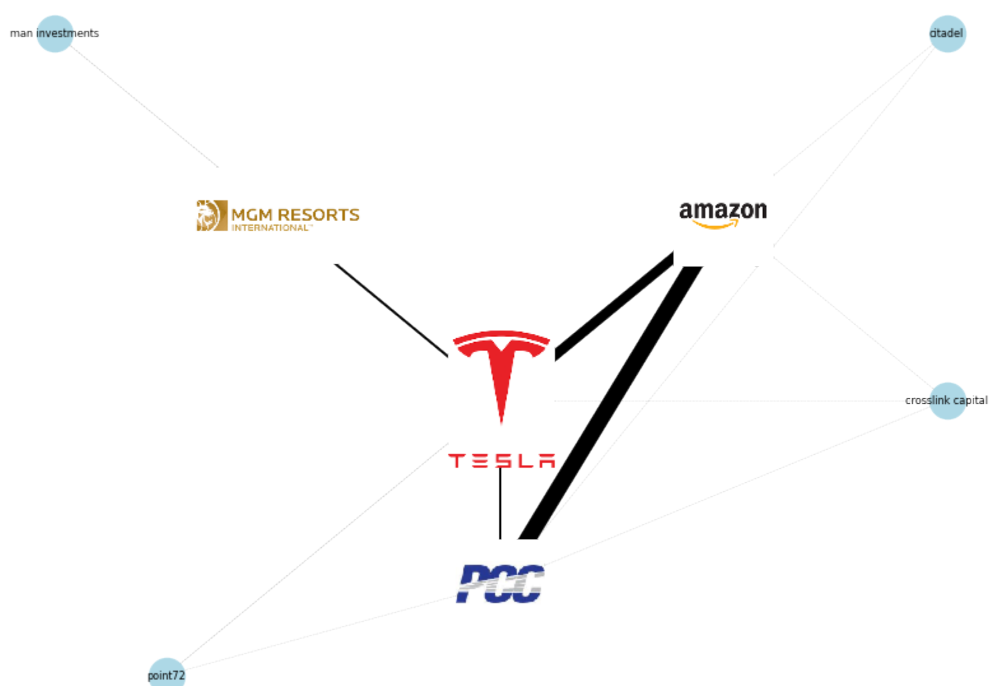

Figure 2: Example of Hedge Fund Co-search Learning Network in 2015Q3

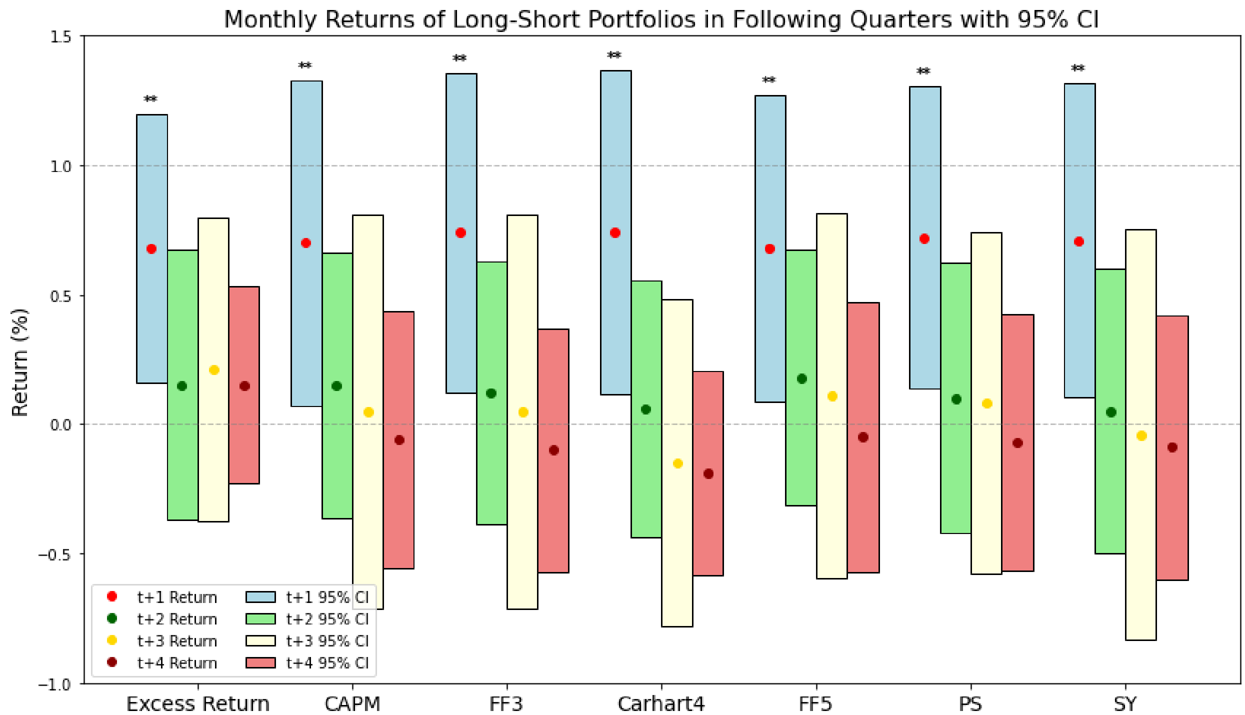

Figure 3: Long-Short Portfolio Returns Sorted by Past Peer Firm Returns over Subsequent Quarters

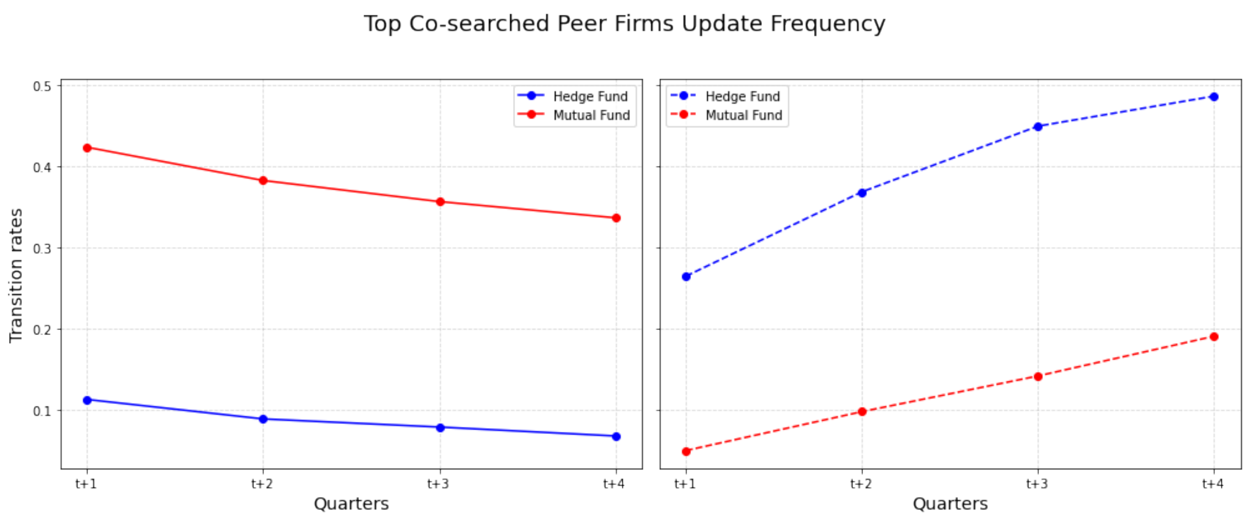

Figure 4: The left panel presents the transition rates of top co-searched peer firms remain in the same quintile over the next four quarters while the right panel presents transition rates of those that drop out of the network.

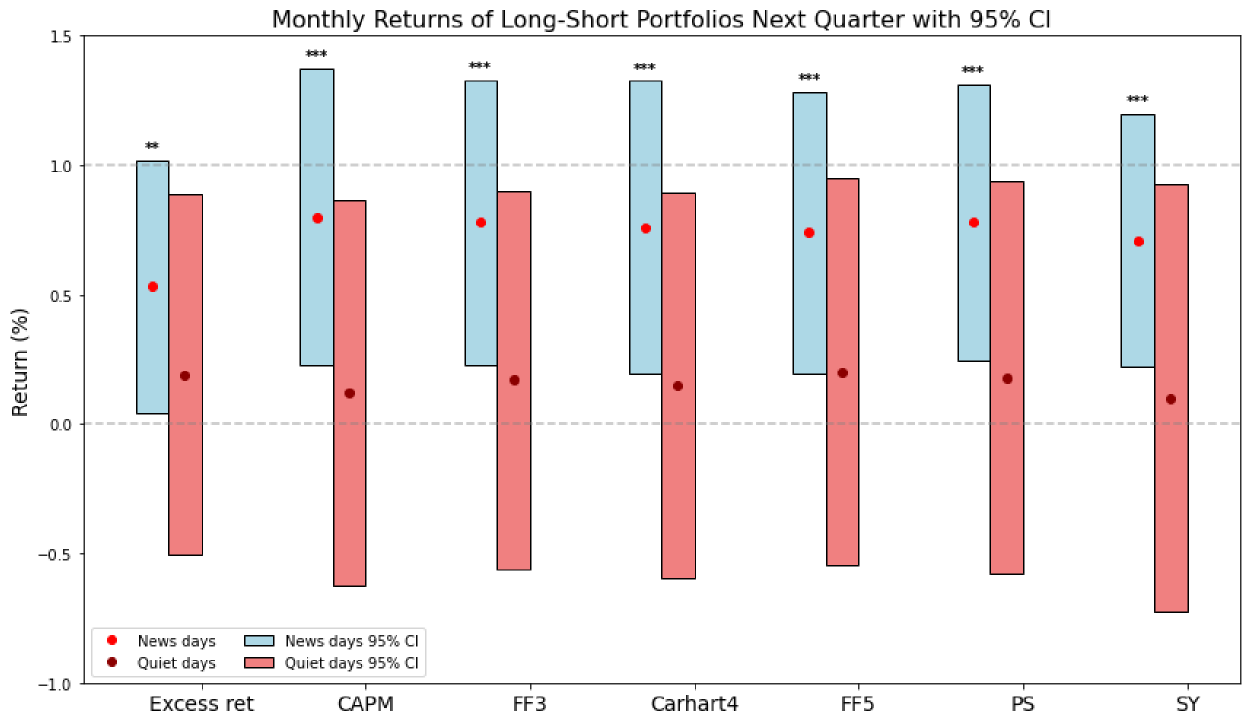

Figure 5: Long-Short Portfolio Returns During News Days and Quiet Days

Table 1: Summary Statistics

This table presents summary statistics calculated at the stock-quarter level. The sample period is from 2009 first quarter to 2017 second quarter. Panel A reports summary statistics related to the hedge fund learning network, including the number of co-searched peer firms, number and percentage of peer firms in the same 3 digits SIC industry group, number of industry groups peer firms belong to, raw return, co-search frequency-weighted peer firms raw return and value-weighted industry return. Panel B presents summary statistics on hedge fund ownership, buy (sale) indicator whether aggregate hedge fund holdings increase (decrease) or not, and abnormal hedge fund holding defined as the change in current-quarter ownership relative to the previous quarter. Panel C presents summary statistics on 10-K filings complexity scores and number of analyst coverage.

Panel A: Hedge Fund Learning Network (Firm i by Quarter t)

	Mean	Median	Std. Dev.	10th	90th	N
Peers (count)	373.63	276	338.92	24	896	61635
PeersSameIndustry (count)	12.34	4	21.88	0	38	61635
PctPeersSameIndustry (pct)	4.02	1.37	7.17	0	11.36	61635
PeersTotalIndustries (count)	97.77	98	58.47	16	178	61635
Ret (firm i return in pct)	6.12	4.07	27.05	-18.60	30.22	61635
PeerRet (peer firm return in pct)	5.84	5.28	11.04	-7.02	16.54	61635
IndustryRet (industry return in pct)	5.71	5.91	12.08	-8.49	18.55	59827

Panel B: Hedge Fund Holdings and Trades (Firm i by Quarter t)

	Mean	Median	Std. Dev.	10th	90th	N
HF_Ownership (pct)	4.17	2.32	9.18	0.41	9.03	56777
Buy (indicator)	0.52	1	0.50	0	1	58582
Sale (indicator)	0.47	0	0.50	0	1	58582
Ab_HF (HF abnormal holding in pct)	0.15	0.02	4.22	-1.21	1.51	58582

Panel C: Summary of Other Variables (Firm i by Quarter t)

	Mean	Median	Std. Dev.	10th	90th	N
Complexity	0.46	0.43	0.18	0.26	0.71	56743
Numest	6.77	5	6.69	0	16	55361

Table 2: Return predictability from Hedge Fund Learning Network

This table presents the return predictability in hedge fund learning network. Panel A shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio, formed based on lagged returns of peer firms in the hedge fund learning network. The returns of each quintile portfolio are the value-weighted returns across firms in this quintile portfolio. Panel B reports the coefficients and t-statistics from Fama-MacBeth regressions. The dependent variables are given firm returns in quarter t ($Ret_{i,t}$). The main independent variable is the co-search frequency-weighted peer firm returns in quarter $t-1$ ($PeerRet_{i,t-1}$). Newey-West standard errors are estimated with three lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Panel A: Long-Short Portfolios

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Monthly returns with one-quarter holding period (%)						
0.68**	0.70**	0.74**	0.74**	0.68**	0.72**	0.71**
(2.59)	(2.20)	(2.37)	(2.34)	(2.28)	(2.44)	(2.32)

Panel B: Fama-MacBeth Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					$\mathrm{Ret_{i,t}}$		
$\mathrm{PeerRet_{i,t\text{-}1}}$	0.090**	0.070**	0.061**	0.060*	0.064**	0.064**	0.068**
	(2.402)	(2.114)	(2.053)	(1.896)	(2.078)	(2.074)	(2.153)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.002	0.010	0.016	0.020	0.022	0.018	0.010
N	59839	44941	44941	44941	44941	44941	44941

Table 3: Information Friction and Return Predictability

This table examines how information friction, measured by disclosure complexity (Panel A) and analyst coverage (Panel B), affects return predictability within the hedge fund learning network using Fama-MacBeth regressions. We sort firms into quintiles based on 10-K complexity and analyst coverage, respectively. The main independent variables are interactions between information friction indicators and co-search frequency-weighted peer returns. Newey-West standard errors with three lags. t-statistics in parentheses. *, **, and *** denote significance at 10%, 5%, and 1% levels.

Panel A: Disclosure Complexity and Return Predictability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				$\text{Ret}_{i,t}$			
$\text{PeerRet}_{i,t-1}$	0.069**	0.050**	0.045*	0.042*	0.051**	0.045**	0.030
	(3.411)	(2.124)	(1.994)	(1.827)	(2.263)	(2.053)	(1.108)
$\text{PeerRet}_{i,t-1} * HighComplexity_{i,t-1}$	0.154**	0.188**	0.186***	0.188***	0.188***	0.175**	0.204***
	(2.143)	(2.342)	(2.451)	(2.511)	(2.471)	(2.302)	(2.730)
$\text{HighComplexity}_{i,t-1}$	-0.011**	-0.011**	-0.010*	-0.010*	-0.009*	-0.009*	-0.009*
	(-2.292)	(-2.114)	(-1.993)	(-1.807)	(-1.866)	(-1.807)	(-1.944)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.005	0.013	0.020	0.024	0.020	0.023	0.028
N	56743	51744	51744	51744	51744	51744	51744

Panel B: Analyst Coverage and Return Predictability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				$\text{Ret}_{i,t}$			
$\text{PeerRet}_{i,t-1}$	0.128**	0.120**	0.112**	0.112**	0.113**	0.112**	0.112**
	(2.045)	(2.173)	(2.144)	(2.124)	(2.153)	(2.114)	(2.064)
$\text{PeerRet}_{i,t-1} * LowCoverage_{i,t-1}$	0.373***	0.314***	0.299**	0.292**	0.294***	0.289**	0.238***
	(3.801)	(2.501)	(2.382)	(2.352)	(2.412)	(2.362)	(2.828)
$\text{LowCoverage}_{i,t-1}$	0.053***	0.050***	0.051***	0.050***	0.051***	0.050***	0.053***
	(8.561)	(8.611)	(8.311)	(8.572)	(8.556)	(8.637)	(8.417)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.034	0.036	0.046	0.050	0.052	0.051	0.058
N	54482	50565	50565	50565	50565	50565	50565

Table 4: Learning Network and Hedge Fund Trading

This table examines whether changes in each firm's aggregate hedge fund holdings follow lagged returns of peer firms return. In panel A, the dependent variable is a dummy variable for buy $Buy_{i,t}$ for firm i in quarter t in Panel A where $Buy_{i,t} = 1$ if hedge funds in aggregate increase holdings and $Buy_{i,t} = 0$ otherwise; the dependent variable is a dummy variable for sale $Sale_{i,t}$ for firm i in quarter t in Panel B where $Sale_{i,t} = 1$ if hedge funds in aggregate decrease holdings and $Sale_{i,t} = 0$ otherwise; the dependent variable in Panel C is the abnormal aggregate hedge fund ownership $Ab\_HF_{i,t}$ for firm i in quarter t, computed as hedge fund ownership at the current quarter t minus the hedge fund ownership of last quarter $t-1$. The main independent variable is peer firm returns in quarter $t-1$ ($PeerRet_{i,t-1}$). Newey-West standard errors are estimated with six lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Panel A: Aggregate hedge fund buy indicator							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\text{Buy}_{i,t}$	$\text{Buy}_{i,t}$	$\text{Buy}_{i,t}$	$\text{Buy}_{i,t}$	$\text{Buy}_{i,t}$	$\text{Buy}_{i,t}$	$\text{Buy}_{i,t}$
$\text{PeerRet}_{i,t-1}$	0.084**	0.109**	0.107***	0.102***	0.105***	0.105***	0.109**
	(2.162)	(2.739)	(2.835)	(2.853)	(2.796)	(2.766)	(2.698)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.001	0.002	0.003	0.003	0.004	0.003	0.003
N	57601	43269	43269	43269	43269	43269	43269
Panel B: Aggregate hedge fund sale indicator							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\text{Sale}_{i,t}$	$\text{Sale}_{i,t}$	$\text{Sale}_{i,t}$	$\text{Sale}_{i,t}$	$\text{Sale}_{i,t}$	$\text{Sale}_{i,t}$	$\text{Sale}_{i,t}$
$\text{PeerRet}_{i,t-1}$	-0.089**	-0.116**	-0.113**	-0.106**	-0.112**	-0.110**	-0.114**
	(-2.108)	(-2.557)	(-2.599)	(-2.607)	(-2.749)	(-2.519)	(-2.468)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.001	0.002	0.003	0.003	0.004	0.003	0.003
N	57601	43269	43269	43269	43269	43269	43269
Panel C: Aggregate hedge fund abnormal holding							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\text{Ab\_HF}_{i,t}$	$\text{Ab\_HF}_{i,t}$	$\text{Ab\_HF}_{i,t}$	$\text{Ab\_HF}_{i,t}$	$\text{Ab\_HF}_{i,t}$	$\text{Ab\_HF}_{i,t}$	$\text{Ab\_HF}_{i,t}$
$\text{PeerRet}_{i,t-1}$	0.006**	0.006*	0.006**	0.006**	0.000	0.003	0.006**
	(2.084)	(2.003)	(2.083)	(2.074)	(0.233)	(1.416)	(2.092)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.001	0.002	0.008	0.013	0.030	0.017	0.006
N	57352	42789	42789	42789	42789	42789	42789

Table 5: Comparison of Peer Firm Update Frequency

This table presents two measures of peer firm update frequency from last quarter $t-1$ to this quarter t in hedge fund and mutual fund learning network. For each firm, we show the percentages of peer firms that remain in the same quintile or drop out (i.e., are no longer searched) in the following quarter. Columns (5) and (6) present the differences between hedge funds and mutual funds. T-statistics are provided in parentheses, and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	Hedge Fund Network		Mutual Fund Network		Col (3) - Col (1)	Col (4) - Col (2)
	(1)	(2)	(3)	(4)	(5)	(6)
From/To	Remain in same quintile	Dropout	Remain in same quintile	Dropout	Remain in same quintile	Dropout
Q1(low)	0.189	0.676	0.270	0.504	0.082***	-0.172***
					(16.815)	(-16.437)
Q2	0.082	0.536	0.137	0.316	0.056***	-0.220***
					(20.256)	(-16.299)
Q3	0.119	0.443	0.185	0.219	0.067***	-0.224***
					(19.773)	(-15.779)
Q4	0.190	0.324	0.284	0.124	0.094***	-0.200***
					(13.749)	(-13.470)
Q5(high)	0.112	0.264	0.423	0.049	0.310***	-0.215***
					(21.429)	(-10.406)

Table 6: Return Predictability using Pseudo Peer Firms

This table reports the coefficients and t-statistics from Fama-MacBeth regressions. In Columns (1) and (2), we study the contemporaneous return co-movement between firm i and its peer firms in quarter t, between firm i and pseudo peer firms which were among peer firms in quarter $t-1$ but are not co-searched with firm i by hedge funds in quarter t, and the dependent variable is firm i return in quarter t ($Ret_{i,t}$). In Columns (3) and (4), we study the return predictability of peer and pseudo peer firm returns in quarter t on firm i return in quarter $t+1$. The dependent variable is firm i firm return in quarter $t+1$. The independent variables are peer firm returns ($PeerRet_{i,t}$) and pseudo peer firm returns ($PseudoPeerRet_{i,t}$) in quarter t. For pseudo peer firms, we assume that the co-search frequencies with firm i in quarter t remain unchanged from quarter $t-1$. Newey-West standard errors are estimated with three lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	(1)	(2)	(3)	(4)
	$\text{Ret}_{i,t}$		$\text{Ret}_{i,t+1}$	
$\text{PeerRet}_{i,t}$	0.433***		0.090**	
	(6.857)		(2.402)	
$\text{PseudoPeerRet}_{i,t}$		0.183***		-0.033
		(6.436)		(-1.620)
Avg R-squared	0.008	0.002	0.002	0.001
N	61635	53386	59839	52413

Table 7: News Co-coverage and Hedge Fund Co-search

This table examines the relationship between hedge fund co-searches and concurrent news co-coverage of both firms in a co-search pair. Columns (1) and (2) present coefficients and t-statistics from linear regressions, where the dependent variable is the log-transformed co-search frequency between firms i and j in quarter t, defined as $LnNumCoSearch_{i,j,t} = \ln(NumCoSearch_{i,j,t} + 1)$. The key independent variable is the log-transformed total number of news articles significantly related to both firms in quarter t, defined as $LnNewsCoCoverage_{i,j,t} = \ln(NewsCoCoverage_{i,j,t} + 1)$. Columns (3) and (4) report coefficients and z-statistics from Poisson regressions, where the dependent variable is the raw co-search frequency, $NumCoSearch_{i,j,t}$, and the main independent variable is $NewsCoCoverage_{i,j,t}$. Adjusted R-squared values are reported for Columns (1) and (2), while Pseudo R-squared values are provided for Columns (3) and (4). We include t-statistics and z-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	(1)	(2)	(3)	(4)
	LnNumCoSearch$_{i,j,t}$		NumCoSearch$_{i,j,t}$	
LnNewsCoCoverage$_{i,j,t}$	0.042***	0.041***		
	(4.224)	(4.200)		
NewsCoCoverage$_{i,j,t}$			0.005***	0.005***
			(4.636)	(4.839)
Control	No	Yes	No	Yes
Firm-pair fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Model	Linear	Linear	Poisson	Poisson
R-squared	0.205	0.206	0.268	0.268
N	10401809	10401809	10401809	10401809

Table 8: Return Predictability Controlling for Momentum

This table reports the coefficients and t-statistics from Fama-MacBeth regressions controlling for both firm-level and industry-level momentum. The dependent variable is firm i return in quarter t ($Ret_{i,t}$). The main independent variable is the co-search frequency-weighted peer firm returns in quarter t ($PeerRet_{i,t-1}$). The main control variables are firm i return in quarter $t-1$ ($Ret_{i,t-1}$) and firm i's 3 digits SIC industry group value-weighted return excluding firm i return itself in quarter $t-1$ ($IndustryRet_{i,t-1}$). Newey-West standard errors are estimated with six lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	(1)	(2)	(3)	(4)
		$\text{Ret}_{i,t}$		
$\text{PeerRet}_{i,t-1}$	0.090**	0.079**	0.067**	0.072**
	(2.321)	(2.478)	(2.226)	(2.386)
$\text{Ret}_{i,t-1}$		-0.009		-0.011
		(-1.091)		(-1.550)
$\text{IndustryRet}_{i,t-1}$			0.036**	0.040**
			(2.224)	(2.658)
Avg R-squared	0.002	0.008	0.007	0.013
N	59839	59839	58069	58069

Table 9: Mutual Fund Learning Network

This table presents summary statistics and return predictability in the mutual fund learning network. The sample period is from Jan 2009 to Jun 2017. Panel A reports summary statistics for the number of co-search peer firms, number of peer firms in the same 3 digits SIC industry group, percentage of peer firms in the same industry group, number of industry groups peer firms belong to, firm i raw return , co-search frequency-weighted peer firms raw return and value-weighted industry return. Panel B shows monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio. The holding periods are three months. Newey-West standard errors are estimated with three lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Panel A: Mutual Fund Learning Network (Firm i by Quarter t)

	Mean	Median	Std. Dev.	10th	90th	N
Peers (count)	1127.73	1219	548.34	291	1794	76226
PeersSameIndustry (count)	26.68	10	37.79	0	80	76226
PctPeersSameIndustry (pct)	2.63	0.99	3.61	0	7.65	76226
PeersTotalIndustries (count)	189.40	208	53.83	106	240	76226
Ret (firm i return in pct)	6.54	4.22	26.86	-18.35	31.32	76226
PeerRet (peer firm return in pct)	5.83	5.30	9.77	-5.70	16.30	76226
IndustryRet (industry return in pct)	6.14	6.42	12.14	-8.77	19.56	73861

Panel B: Long-Short Portfolios

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Monthly returns with one-quarter holding period (%)						
0.42	0.46	0.48	0.48	0.45	0.47	0.40
(1.34)	(1.11)	(1.16)	(1.14)	(1.11)	(1.18)	(1.05)

Table 10: Robustness Tests

This table shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio, formed based on lagged returns of peer firms in the hedge fund learning network with different length of co-search window and more Edgar filings included. Panel A presents long-short portfolio returns where the time window for co-search is relaxed to before and after 15 days of each firm's 10-K/Q filings request on Edgar. Panel B presents long-short portfolio returns where the time window for co-search is restricted to before and after 3 days of each firm's 10-K/Q filings request on Edgar. And Panel C shows long-short portfolio returns where search traffic 8-K, S-1, and 14A filings are also included. The returns of each quintile portfolio are the value-weighted returns across firms in this quintile portfolio. Newey-West standard errors are estimated with three lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Panel A: Long-Short Portfolios (Co-search using 15-Day Window)

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Monthly returns with one-quarter holding period (%)						
0.33***	0.34**	0.37**	0.37**	0.36**	0.44***	0.36**
(2.67)	(2.33)	(2.40)	(2.35)	(2.24)	(2.77)	(2.23)

Panel B: Long-Short Portfolios (Co-search using 3-Day Window)

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Monthly returns with one-quarter holding period (%)						
0.76***	0.85***	0.87***	0.87***	0.85***	0.80***	0.84***
(3.24)	(2.91)	(3.08)	(2.98)	(3.16)	(2.93)	(3.10)

Panel C: Long-Short Portfolios (Include Co-search of Other Edgar Filings)

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Monthly returns with one-quarter holding period (%)						
0.70**	0.81**	0.85***	0.84**	0.85***	0.74**	0.84***
(2.61)	(2.40)	(2.62)	(2.46)	(2.74)	(2.47)	(2.67)

Appendix

This appendix provides additional details and empirical evidence to supplement the main text. There are three primary sections of this appendix. Appendix A.1 contains details about the EDGAR Log Files and the process of unmasking mutual fund and hedge funds IP addresses. Appendix A.2 contains supplementary test results. Appendix A.3 provides details on data sources used in this study.

A.1 Unmasking IP addresses in EDGAR Log Files

The EDGAR Log Files contain billions of observations of "requests" or "requests to view a filing." Each observation details the filing requested (accession number), the date and time of the request, and the requester (the IP address making the electronic request). A snapshot of the raw EDGAR Log Files is shown below:

IP Address	Date	Time	CIK	Accession Number
38.97.91.ecg	20170531	09:47:33	051143	000104746917001061
38.65.241.fhf	20170531	11:07:28	274191	000002741917000008
67.199.249.igg	20170924	12:27:02	320193	000032019317000009
216.223.41.aah	20170924	16:12:55	831259	000083125917000016

Given our focus in this paper on stock (CIK) and the requester, linking the masked IP addresses to identifiable investors (e.g., mutual fund families) is pivotal to our study. To unmask the IP addresses, we first notice the fourth octet in the examples above.[1] In place of the actual digits of the requesting IP address, the fourth octet is reported as a set of three letters. However, organizations typically register blocks of IP addresses, with the most common block fixing the first three octets and containing all 256 versions of the fourth octet.[2] In other words, only the first three octets are necessary to identify the organization that has registered that block of IP addresses.

Using this insight, we downloaded historical IP address registration records from 2010 through 2017 to identify the blocks of IP addresses registered to investment firms.[3] Then, using this hand-collected mapping between investment firms and IP addresses, we unmask the requesters in the EDGAR Log Files. As a result, the snapshot of raw data from above has been transformed into the following.

[1]For IP addresses, an *octet* is a group of eight bits, or the one to three digit numbers (from 0 to 255) separated by periods in the examples above.

[2]For example, all 256 IP addresses beginning with 38.97.91 will be registered to the same organization.

[3]IP registration records were acquired from MaxMind, https://www.maxmind.com/en/home.

Investment Firm	Date	Time	Ticker	Filing
Abrams Capital	20170531	09:47:33	IBM	10-K for 2016
Harbor Capital	20170531	11:07:28	TGT	10-K for 2016
Crabel Capital	20170924	12:27:02	AAPL	10-Q for Q2 2017
Ronin Capital	20170924	16:12:55	FCX	Earnings for Q2 2017

Furthermore, the three letters used to mask the fourth octet is static, not dynamic. This means, for example, that def replaces the digits 146 for every instance of 146. This allows us to identify unique IP addresses. In other words, though an unmasked mutual fund may make 50 requests one day, we can observe how many different IP addresses made those requests. This insight is particularly important as it allows us to calculate the variable IPs that we use throughout the paper.

We have also adjusted the data to remove likely bots. As mentioned, the raw EDGAR Log Files contain billions of requests with many thousands of requests per day coming from single IP addresses. It is unlikely that these thousands of requests per day represent a human actually clicking on documents in EDGAR. It is much more likely that they represent computer programs (bots) downloading large quantities of data at a time. Given these IP addresses do not fit with the spirit of our research, we remove them from the data. The removal process is as follows: we remove IP addresses that either (i) make over $1,000$ requests in a day or (ii) make requests for over 100 different CIKs (i.e., firms).

A.2 Supplementary tests for main results

A.2.1 Equal-weighted peer firm return

Table A1: Long-Short Portfolio Returns Sorted by Equal-weighted Peer Firm Past Returns (over Subsequent Quarters)

This table shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio over the first to fourth subsequent quarters after portfolio formation. Portfolios are constructed based on the lagged equal-weighted returns of peer firms instead of co-search frequency within the hedge fund learning network. The returns of each quintile portfolio are the value-weighted returns across firms in this quintile portfolio. Newey-West standard errors are estimated with three lags, six lags, nine lags and twelve lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Panel A: Monthly returns in the quarter immediately after portfolio formation (%)						
0.58***	0.57**	0.62***	0.62***	0.60***	0.61***	0.64***
(3.06)	(2.94)	(3.15)	(3.10)	(3.07)	(3.22)	(3.26)
Panel B: Monthly returns in the second quarter after portfolio formation (%)						
0.16	0.10	0.10	0.07	0.17	0.08	0.23
(0.65)	(0.44)	(0.38)	(0.69)	(0.32)	(0.25)	(0.93)
Panel C: Monthly returns in the third quarter after portfolio formation (%)						
0.20	-0.06	-0.01	-0.18	0.07	0.01	-0.02
(0.91)	(-0.21)	(-0.18)	(-0.84)	(0.25)	(0.03)	(-0.10)
Panel D: Monthly returns in the fourth quarter after portfolio formation (%)						
0.22	0.00	-0.02	-0.09	0.06	0.01	-0.19
(1.17)	(0.02)	(-0.10)	(-0.49)	(0.21)	(0.07)	(-1.65)

A.2.2 Value weighted long-short portfolio

Table A2: Long-Short Portfolio Returns Sorted by Peer Firm Past Returns

This table shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio, formed based on lagged returns of peer firms in the hedge fund learning network. The returns of each quintile portfolio are the value-weighted returns across firms in this quintile portfolio. Newey-West standard errors are estimated with three lags, six lags, nine lags and twelve lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Panel A: Monthly returns with one-quarter holding period (%)						
0.68**	0.70**	0.74**	0.74**	0.68**	0.72**	0.71**
(2.59)	(2.20)	(2.37)	(2.34)	(2.28)	(2.44)	(2.32)
Panel B: Monthly returns with two-quarter holding period (%)						
0.40**	0.39**	0.41***	0.41**	0.39**	0.39**	0.38*
(2.12)	(2.53)	(2.63)	(2.51)	(2.28)	(2.39)	(1.95)
Panel C: Monthly returns with three-quarter holding period (%)						
0.40**	0.32**	0.32**	0.31**	0.33**	0.32**	0.28
(2.14)	(2.40)	(2.36)	(2.04)	(2.19)	(2.34)	(1.40)
Panel D: Monthly returns with four-quarter holding period (%)						
0.34**	0.20*	0.19	0.18	0.21*	0.19*	0.20
(1.99)	(1.72)	(1.55)	(1.29)	(1.75)	(1.67)	(1.20)

A.2.3 Equal weighted long-short portfolio

Table A3: Portfolio Returns Sorted by Past Peer Firm Returns (Equal Weighted)

This table shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio, formed based on lagged returns of peer firms in the hedge fund learning network. The returns of each quintile portfolio are the equal-weighted returns across firms in this quintile portfolio. Newey-West standard errors are estimated with three lags, six lags, nine lags and twelve lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Panel A: Monthly returns with one-quarter holding period (%)						
0.45**	0.48**	0.53**	0.52**	0.54**	0.51**	0.54**
(2.28)	(2.03)	(2.27)	(2.24)	(2.13)	(2.12)	(2.41)
Panel B: Monthly returns with two-quarter holding period (%)						
0.29**	0.29**	0.32***	0.31***	0.37***	0.31**	0.32**
(2.43)	(2.33)	(2.78)	(2.73)	(2.96)	(2.48)	(2.42)
Panel C: Monthly returns with three-quarter holding period (%)						
0.36***	0.31**	0.32**	0.30**	0.40***	0.32**	0.27*
(2.83)	(2.29)	(2.46)	(2.48)	(3.11)	(2.61)	(1.89)
Panel D: Monthly returns with four-quarter holding period (%)						
0.30***	0.22*	0.21*	0.19*	0.27**	0.21*	0.24**
(2.78)	(1.80)	(1.69)	(1.69)	(2.23)	(1.77)	(2.01)

A.2.4 Price pressure versus informative trades

Table A4: Portfolio Returns Sorted by Past Peer Firm Returns (over Subsequent Quarters)

This table shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio over the first to fourth subsequent quarters after portfolio formation. Portfolios are constructed based on the lagged returns of peer firms within the hedge fund learning network. The returns of each quintile portfolio are the value-weighted returns across firms in this quintile portfolio. Newey-West standard errors are estimated with three lags, six lags, nine lags and twelve lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Panel A: Monthly returns in the quarter immediately after portfolio formation (%)						
0.68**	0.70**	0.74**	0.74**	0.68**	0.72**	0.71**
(2.59)	(2.20)	(2.37)	(2.34)	(2.28)	(2.44)	(2.32)
Panel B: Monthly returns in the second quarter after portfolio formation (%)						
0.15	0.15	0.12	0.06	0.18	0.10	0.05
(0.57)	(0.58)	(0.47)	(0.24)	(0.72)	(0.38)	(0.18)
Panel C: Monthly returns in the third quarter after portfolio formation (%)						
0.21	0.05	0.05	-0.15	0.11	0.08	-0.04
(0.71)	(0.13)	(0.13)	(-0.47)	(0.31)	(0.24)	(-0.10)
Panel D: Monthly returns in the fourth quarter after portfolio formation (%)						
0.15	-0.06	-0.10	-0.19	-0.05	-0.07	-0.09
(0.78)	(-0.24)	(-0.42)	(-0.95)	(-0.19)	(-0.28)	(-0.35)

A.2.5 Net arbitrage trading and peer firm past returns

Table A5: Net arbitrage trading and Peer Firm Past Returns

This table examines whether changes in each firm's net arbitrage trading follow lagged returns of peer firms. The dependent variable is the net aggregate trading $NAT_{i,t}$. The main independent variable is peer firm returns in quarter $t-1$ ($PeerRet_{i,t-1}$). The sample period is from 2009 first quarter to 2015 third quarter. Newey-West standard errors are estimated with six lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	(1) $\text{NAT}_{i,t}$	(2) $\text{NAT}_{i,t}$	(3) $\text{NAT}_{i,t}$	(4) $\text{NAT}_{i,t}$	(5) $\text{NAT}_{i,t}$	(6) $\text{NAT}_{i,t}$	(7) $\text{NAT}_{i,t}$
$\text{PeerRet}_{i,t-1}$	0.009*	0.009*	0.011**	0.010**	0.012**	0.010**	0.011*
	(1.989)	(2.011)	(2.323)	(2.151)	(2.252)	(2.240)	(1.911)
Control	No	CAPM	FF3	Carhart4	FF5	PS Liquidity	SY Mispricing
Avg R-squared	0.001	0.003	0.011	0.012	0.015	0.013	0.013
N	27136	27114	27114	27114	27114	27114	27114

A.2.6 Co-coverage of news events

Table A6: Prior or Subsequent News Co-coverage and Hedge Fund Co-search

This table examines the relationship between hedge fund co-searches and prior or subsequent news co-coverage of both firms in a co-search pair. We include t-statistics for linear regressions and z-statistics for Poisson regressions in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Panel A: Prior Co-coverage News				
	(1)	(2)	(3)	(4)
	$\text{LnNumCoSearch}_{i,j,t}$		$\text{NumCoSearch}_{i,j,t}$	
$\text{LnNewsCoCoverage}_{i,j,t-1}$	0.007	0.007		
	(0.862)	(0.883)		
$\text{NewsCoCoverage}_{i,j,t-1}$			0.001	0.001
			(0.599)	(0.734)
Control	No	Yes	No	Yes
Firm-pair fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Model	Linear	Linear	Poisson	Poisson
R-squared	0.229	0.229	0.300	0.300
N	4428406	4428406	4428406	4428406
Panel B: Subsequent Co-coverage News				
	(1)	(2)	(3)	(4)
	$\text{LnNumCoSearch}_{i,j,t}$		$\text{NumCoSearch}_{i,j,t}$	
$\text{LnNewsCoCoverage}_{i,j,t+1}$	0.015*		-0.000	
	(1.842)		(-0.059)	
$\text{NewsCoCoverage}_{i,j,t+1}$		0.015*		0.000
		(1.846)		(0.052)
Control	No	Yes	No	Yes
Firm-pair fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Model	Linear	Linear	Poisson	Poisson
R-squared	0.219	0.220	0.289	0.289
N	4428406	4428406	4428406	4428406

A.2.7 Decomposing return predictability

Beyond examining discrete news events and fundamental firm characteristics, return predictability within hedge fund learning networks may also vary according to different dimensions of peer firm returns. To specifically highlight hedge funds' ability in identifying hidden inter-firm links, we further decompose peer firm returns into same industry versus different industry groups.

Table A7 decomposes peer firm returns into same 3 digits SIC industry versus different industry groups. The analysis reveals that return predictability predominantly arises from peer firms belonging to different industry groups rather than from within-industry peers. Specifically, peer firms from different industries significantly predict returns, highlighting hedge funds' ability to identify economically meaningful yet unconventional inter-industry relationships. This finding underscores the informational advantage that hedge funds hold by leveraging nontraditional and overlooked firm connections beyond standard industry classifications.

Table A7: Return Predictability (Same versus Different 3 Digits SIC Industry Groups)

This table reports the coefficients and t-statistics from Fama-MacBeth regressions. The dependent variable are firm i return in quarter t ($Ret_{i,t}$), and firm i return from quarter t to quarter t+1 ($Ret_{i,t\rightarrow t+1}$). The main independent variables are the peer firm returns in quarter $t-1$ ($PeerRet\_DiffSIC3_{i,t-1}$) where only peer firms from different 3 digits SIC industry groups are included, and peer firm returns in quarter $t-1$ ($PeerRet\_SameSIC3_{i,t-1}$) where only peer firms from the same 3 digits SIC industry group are included. The main control variable is firm i return in quarter $t-1$ ($Ret_{i,t-1}$). Newey-West standard errors are estimated with three lags and six lags, respectively. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	(1) $\text{Ret}_{i,t}$	(2) $\text{Ret}_{i,t\rightarrow t+1}$
$\text{PeerRet\_DiffSIC3}_{i,t-1}$	0.082**	0.146**
	(2.075)	(2.284)
$\text{PeerRet\_SameSIC3}_{i,t-1}$	0.031***	0.017
	(3.341)	(1.535)
$\text{Ret}_{i,t-1}$	-0.011*	-0.032**
	(-1.741)	(-2.530)
Avg R-squared	0.012	0.010
N	46482	46560

A.2.8 Compare with mutual funds: return predictability

Table A8: Long-Short Portfolio Returns Sorted by Peer Firm (in Mutual Fund Network) Past Returns

This table shows the monthly excess returns and alphas (generated from various risk factor models listed in the table) of the long-short quintile portfolio, formed based on lagged returns of peer firms in the mutual fund learning network. The returns of each quintile portfolio are the value-weighted returns across firms in this quintile portfolio. Newey-West standard errors are estimated with three lags, six lags, nine lags and twelve lags. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Excess return	CAPM	FF3	Carhart4	FF5	PS liquidity	SY mispricing
Panel A: Monthly returns with one-quarter holding period (%)						
0.42	0.46	0.48	0.48	0.45	0.47	0.40
(1.34)	(1.11)	(1.16)	(1.14)	(1.11)	(1.18)	(1.05)
Panel B: Monthly returns with two-quarter holding period (%)						
-0.16	-0.23	-0.20	-0.22	-0.22	-0.23	-0.27
(-0.73)	(-1.41)	(-1.19)	(-1.23)	(-1.15)	(-1.23)	(-1.36)
Panel C: Monthly returns with three-quarter holding period (%)						
0.02	-0.13	-0.10	-0.12	-0.08	-0.10	-0.18
(0.10)	(-0.77)	(-0.70)	(-0.86)	(-0.52)	(-0.70)	(-1.06)
Panel D: Monthly returns with four-quarter holding period (%)						
0.11	-0.08	-0.09	-0.11	-0.04	-0.08	-0.13
(0.53)	(-0.40)	(-0.56)	(-0.69)	(-0.24)	(-0.53)	(-0.76)

A.2.9 Compare with mutual funds: R^2 comparison

To better understand how hedge funds identify peer firms within their learning networks, we examine the extent to which peer firm characteristics explain the cross-sectional variation in a firm's key financial characteristics. In other words, we perform the following regressions cross-sectionally at each quarter and document the average R^2 for hedge funds and mutual funds, respectively. For a given firm characteristics, we consider the following variables including market capitalization (market cap), book-to-market ratio (book/market), valuation multiples including price-to-book ratio (pb) and price-to-earnings ratio (pe), two financial statement ratios including return on assets (ROA) and return on equity (ROE). Finally, we compare the differences in the average R^2 to gain insight into the distinctiveness of hedge funds' learning strategies.

$$Var_{i,t} = \alpha_t + \beta_t VarPeer_{i,t} + \epsilon_{i,t} \tag{8}$$

Table A9 compares the explanatory power (R^2) of firms' financial characteristic variables using the characteristics of the peers of the firm identified by hedge funds (HF) and mutual funds (MF). For example, if we look at the market capitalization, the peer firms' market capitalization identified by mutual funds exhibit a significantly higher R^2 (0.147) compared to hedge funds (0.108). The difference (0.039) is significant at the 1% level, suggesting mutual funds tend to pair a given firm with peer firms at similar size. Overall, by looking at different firm characteristics, we find that mutual funds consistently outperform hedge funds in explaining firms' valuation metrics using peer firm characteristics. This suggests that mutual funds rely on conventional and well-established financial and market-based metrics to define firm connections. The significant differences between Column (2) and Column (1) show that hedge fund focus on uncovering more unconventional, nuanced connections that are not directly captured by traditional valuation measures.

Table A9: Comparison of Average R^2 Obtained from Cross-sectional Regressions

This table compares the average R^2 from several quarterly cross-sectional regressions in the form of

$$Var_{i,t} = \alpha_t + \beta_t VarPeer_{i,t} + \epsilon_{i,t}, \tag{9}$$

using quarterly financial statement data from Compustat and market capitalization data from CRSP on March, June, September, and December of each year. Columns 1 and 2 report average R^2 from quarterly cross-sectional regressions, regressing firm i's characteristics $Var_{i,t}$ in a given quarter t on the concurrent co-search frequency-weighted peer firms characteristics $VarPeer_{i,t}$ in the hedge fund and mutual fund learning network, respectively. Each row considers a different Var: first is market capitalization (market cap), second is book-to-market ratio (book/market), third and fourth are two valuation multiples including price-to-book ratio (pb) and price-to-earnings ratio (pe), fifth and sixth are two financial statement ratios including return on assets (ROA) and return on equity (ROE). Column 3 test for the significance of the differences in average R^2 between hedge fund and mutual fund learning network. We include t-statistics in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	(1) R-squared for HF	(2) R-squared for MF	(3) (2)-(1)
Market cap	0.108	0.147	0.039***
			(3.934)
Book/Market	0.014	0.021	0.007**
			(2.432)
P/B	0.014	0.013	-0.001
			(-0.804)
P/E	0.002	0.012	0.011***
			(4.312)
ROA	0.044	0.073	0.029***
			(2.926)
ROE	0.002	0.004	0.002*
			(1.683)

A.3 Data Appendix

In this appendix we provide details on the subsidary data sources used in our study.

A.3.1 Hedge Fund Holdings

We analyze hedge fund trading activities on the long side using institutional holdings data from Thomson Reuters 13F. First, we identify hedge funds through EDGAR downloading IP addresses and manually match them by name to institutional holdings in Thomson Reuters 13F. We then aggregate hedge fund holdings at the stock level. For each stock i in quarter t, hedge fund holdings are calculated as the total number of shares held by all hedge funds divided by the total shares outstanding. Abnormal hedge fund holdings ($Ab\_HF_{i,t}$) are defined as the change in hedge fund holdings from the previous quarter. Further, we construct two binary variables to capture hedge fund trading behavior. We define $Buy_{i,t} = 1$ if the aggregate hedge fund holdings in stock i increase from the previous quarter and $Buy_{i,t} = 0$ otherwise. Similarly, we define $Sale_{i,t} = 1$ if the aggregate hedge fund holdings decrease and $Sale_{i,t} = 0$ otherwise. If a hedge fund does not hold a particular stock, its holdings for that stock are set as zero.

A.3.2 RavenPack News

We use data from RavenPack, which provides a history of news releases linked to specific firms starting from Jan 2009 to Jun 2017. We impose three primary filters on the RavenPack news release observations: (1) We restrict the sample to news sources from Barrons, Marketwatch, Wall Street Journal and Dow Jones Newswires; (2) We filter based on the RELEVANCE variable, a relevance score assigned by RavenPack that quantifies the connection between a news item and a specific firm. Following the RavenPack User Guide on WRDS, we include only news releases with a RELEVANCE score above 75, which are considered significantly relevant; (3) We retain only news releases that are significantly related to at least two firms. Finally, we construct the total number of news release observations that are significantly related to both firm i and firm j in quarter t, denoted as $NewsCoCoverage_{i,j,t}$.

A.3.3 Detail on CRSP

Monthly returns, prices, and shares outstanding are obtained from CRSP database. We start with the universe of U.S. common stocks with CRSP share codes of 10 or 11 and are traded on major exchanges NYSE, AMEX, and NASDAQ (exchange code 1, 2 or 3) over the period from Jan 2009 to June 2017. We exclude penny stocks (i.e., those with closing prices less than $5), financial firms (SIC codes between 6000 and 6999), and utilities (SIC codes between 4900 and 4999). Accounting data are obtained from COMPUSTAT. Data on 10-K complexity to measure firm-level information complexity by Loughran and McDonald (2024) are obtained from The Notre Dame Software Repository for Accounting and Finance (SRAF). Analyst coverage is obtained from I/B/E/S.