It's a Small World: Social Ties, Comovements, and Predictable Returns*

Lin Peng Sheridan Titman Muhammed Yönaç Dexin Zhou

July 21, 2023

Abstract

We identify a new dimension of cross-firm linkages by exploring the social connectedness between firms' geographical locations. Industry peers located in regions with strong social ties tend to adopt similar strategies and exhibit strong co-movements in both fundamentals and returns. However, this information is not immediately reflected in stock prices and can be exploited using information contained in social peer returns (*SPFRET*). The predictability of *SPFRET* lasts for up to a year and forecasts future earnings surprises, analysts' forecast errors, and returns around earnings announcements. The effect is particularly strong for low-visibility firms and those located outside of industry clusters.

JEL Codes: G14, D85
Keywords: social networks, strategy, comovements, return predictability, machine learning

---

*The paper was previously circulated under the title "Social Ties and Predictable Returns." Lin Peng: Zicklin School of Business, Baruch College/CUNY, email: lin.peng@baruch.cuny.edu. Sheridan Titman: McCombs School of Business, The University of Texas at Austin, e-mail: sheridan.titman@mccombs.utexas.edu. Muhammed Yönaç: University of Bristol Business School, email: muhammed.yonac@bristol.ac.uk. Dexin Zhou: Zicklin School of Business, Baruch College, email: dexin.zhou@baruch.cuny.edu. We thank Jawad Addoum (discussant), Nicholas Barberis, Vasco Carvalho, Tiago Cavalcanti, Sudheer Chava, Meredith Crowley, Jared DeLisle (discussant), Xi Dong, Elisa Faraglia, Diego Garcia (discussant), Abhinav Goyal (discussant), Xing Huang (discussant), Ted Joyce, Xin Li (discussant), Dong Lou, Sebastiano Manzan, Pierre Magontier (discussant), Javier Mella (discussant), Ramana Nanda, Quoc Nguyen (discussant), Jun Wang, Junbo Wang (discussant), Wei Xiong, Jing Zhao (discussant), and conference and seminar participants at CICF 2021, UEA 2021, FMA 2021, MFA 2022, 13th FSU SunTrust Beach Conference, E(astern)FA 2022, NBER BF Working Group Spring 2022, AsianFA 2022 Meeting, CFEA 2022, SFW, SFS Cavalcade Asia-Pacific 2022, 2023 FMA Consortium on Asset Management, Baruch College, Capital University of Economics and Business, CEIBS, Durham University, Georgia Institute of Technology, the IAQF's Thalesians Seminar Series, Ozyegin University, Peking University, Shanghai Advanced Institute of Finance, Stockholm Business School, Tsinghua University, the University of Birmingham, the University of Bristol, and the University of Cambridge for their helpful comments and suggestions. Lin Peng gratefully acknowledges financial support from the Krell research grant and the Keynes Fund. All remaining errors are our own.

1 Introduction

A growing body of literature highlights the interconnected nature of firms in the economy. For example, Acemoglu et al. (2012) and McNerney et al. (2022) find that production networks and input-output linkages facilitate the propagation of firm-level shocks, amplify economic growth, and contribute to aggregate fluctuations. This paper studies a source of cross-firm linkages that arise because of social ties that exist between the managers of different firms. Our analysis, which focuses on the spatial nature of these connections, builds on recent research that shows that social connections between regions are associated with important economic exchanges, such as domestic and international trade flows, migration patterns, and knowledge spillovers (Breschi and Lenzi, 2016; Cohen et al., 2017; Bailey et al., 2018a, 2021).

This paper explores the link between social connections and various aspects of corporate strategy. The basic idea is that when managers interact and share ideas, their world views become more similar, and as a result, their strategies become more similar. We find that this is indeed the case, however, our evidence suggests that stock market participants do not pay sufficient attention to the fact that industry peers located in socially connected regions tend to be more similar. Specifically, we show that a lead-lag strategy that exploits the underreaction of stock returns to the returns of their socially connected industry peers generates excess returns.

We start by establishing that the strategies of industry peers are more similar if they are located in regions that are socially connected. To measure these social ties, we use Facebook's Social Connectedness Index (SCI) (see Bailey et al., 2018a), which measures the social connections of individuals both within and across U.S. counties.[1] We then analyze pairs of industry peers, and investigate whether pairs of firms headquartered in socially connected locations tend to have more similar strategies.

The analysis considers three measures of strategic similarity. The first, which is constructed by applying textual analysis to the firms' 10-K filings, measures the similarity of the paragraphs that correspond to business operations and strategies. The second and

[1] As the world's largest online social networking service, Facebook's enormous scale and coverage (over 258 million active users in the United States as of 2020) and the relative representatives of its user body makes SCI a unique measure of the real-world geographic structure of U.S. social networks at population scale. See, Bailey et al. (2018a) and Kuchler et al. (2020) for further discussion on these points.

third strategic similarity measures, which are taken from the existing literature, focus on dimensions of product market similarity (Hoberg and Phillips 2016) and technology similarity (Lee et al. 2019). We find that each of these pairwise similarity measures exhibits a significant positive correlation with the social connectedness of the firms' headquarters. For example, the 10-K-based strategic similarity is 3.5 percentage points higher for firm pairs with SCI in the top quartile than for those in the bottom quartile, which corresponds to 11.4% of the standard deviation of the strategy similarity measure.

Motivated by these pair-wise results, we construct SCI-weighted industry portfolios that are specially tailored for each firm in our sample. Relative to equally-weighted and value-weighted industry portfolios, these tailored portfolios place more weight on the focal firm's industry peers headquartered in counties with strong social ties to their home counties. The idea is that industry peers in socially connected counties are more similar, and as a result, changes in both a firm's fundamentals and its stock price are likely to be more closely related to the SCI-weighted averages than to the equal- or value-weighted averages of their industry peers. Our tests indicate that this is indeed the case. We find that while both the equal-weighted and SCI-weighted industry indexes explain movements in both firm fundamentals and stock returns, these movements tend to be better explained by the SCI-weighted indexes.

We then ask whether market prices fully reflect the importance of these social ties. We do this by calculating *SPFRET*, the SCI-weighted returns of each focal firm's industry peers, and examine the extent to which these returns predict the focal firm's future returns. If market prices fail to fully reflect the similarity of socially-connected industry peers, the SCI-weighted returns should lead the returns of the focal firms. Our evidence indicates that they do. Specifically, we show that a trading strategy that exploits these lead-lag relationships earn significant abnormal returns of 84 basis points per month.

The SCI-weighted industry returns continue to predict a firm's future returns for up to the next 12 months. The one-year cumulative long-short portfolio alpha is 4.54% and 5.17% for equal-weighted portfolios sorted by *SPFRET* and *SPFMOM* (i.e., the past one-year cumulative social peer firm returns), respectively. The corresponding value-weighted alphas are somewhat smaller at 2.12% and 3.60%, but still significantly different than zero. These longer-term results indicate that the observed predictability is driven by underreaction to information revealed by a firm's socially connected industry peers.

We conjecture that these excess returns are generated, at least in part, because of investor inattention to the implications of strategic similarity. To further explore this possibility, we stratify the sample by different measures of the firms' visibility. Consistent with the attention hypothesis, we find that return predictability is stronger for smaller firms, firms with low institutional ownership, and for firms with low analyst coverage, i.e., for firms that attract less attention. Similarly, our results reveal that information from social peer firms is more quickly reflected in the prices of firms located in industry clusters. This last observation is consistent with the findings of Engelberg et al. (2018) that analyst coverage is greater for stocks inside industry clusters.

We also find that social peer firm returns predict a firm's future earnings growth. Firms with the highest past cumulative social peer firm returns (*SPFMOM*) have, on average, standardized unexpected earnings (*SUE*) that are 34 percentage points higher in the next quarter than those with the lowest *SPFMOM* (or 24% of the standard deviation of *SUE*). Similarly, firms with the highest *SPFRET* exhibit 10.9 percentage points higher *SUE* in the following quarter. Consistent with the idea that social peer firm returns convey information that is underappreciated by market participants, we find that sell-side analysts and other market participants substantially underestimate the future earnings of firms with high social peer firm returns, i.e., analysts' forecast errors and returns around future earnings surprises can be predicted using these returns.

Our paper contributes to a number of distinct areas of research. First, our study adds to the growing literature on social economics and social finance. The existing body of research in economics studies how social ties influence economic exchanges across regions (e.g., Cohen et al., 2017; Bailey et al., 2018b, 2021). While a large part of this literature describes how social ties facilitate interactions, we believe we are the first to demonstrate that social ties between individuals in different locations are linked to the adoption of similar corporate strategies, and that this, in turn, leads to the co-movement of firms' fundamentals and stock returns.

Previous papers in social finance study the effect of peers on both retail investing and on the performances of professional investors.[2] Several more recent papers focus on

[2]Several papers find that social ties between institutional managers can generate valuable information and improve investment performances (e.g., Cohen et al., 2008; Pool et al., 2015; Hong and Xu, 2019). But other papers suggest that social effects are associated with the propagation of sentiment that influences both investment and housing decisions of individual investors (e.g., Shiller, 2010; Beshears et al., 2015; Hvide

firm-level outcomes. For example, Shue (2013) finds that executives with social ties (i.e., enrolled in the same sections of the MBA program) tend to adopt more similar corporate policies, and Kuchler et al. (2021) show that a firm's social proximity to institutional capital influences the valuation and liquidity of its stock. Our analysis of social peer effects is more closely tied to the corporate finance side of this literature, but our evidence of underreaction contributes to the investment side of the literature.

The second literature we contribute to examines how firm locations influence the co-movement of firm fundamentals and stock returns (e.g., Pirinsky and Wang, 2006; Parsons et al., 2020; Jin and Li, 2020). These papers suggest that in addition to common labor markets, the co-movements could be due to factors directly related to the locations, such as weather and political conditions. Our analysis provides further evidence on the importance of labor markets, and emphasizes the role of social ties that extend beyond a city's geographic boundaries.

Lastly, this paper adds to the extensive literature that studies return predictability in financial markets.[3] Our analysis closely follows the literature that examines economic linkages that generate lead-lag predictability, most notably, the industry lead-lag literature starting with Moskowitz and Grinblatt (1999).[4] We provide additional evidence that investors fail to fully account for the co-movement of similar stocks – we show that the excess returns of the lead-lag strategies that exploit this inattention can be improved when the lead portfolio is designed to be more similar to the stock whose return is predicted. Moreover, we provide a number of tests that shows that the information embedded in *SPFRET* is not subsumed by any of the sources of lead-lag explored in the previous literature.

The rest of the paper is organized as follows. In Section 2, we provide a detailed description of the data used in this study. In Section 3, we investigate how social connect-

and Östberg, 2015; Bailey et al., 2018c), exacerbate behavioral biases for retail traders in foreign exchange markets (e.g., Heimer, 2016), correspond to sub-optimal decisions of mutual fund managers (e.g., Pool et al., 2012; Au et al., 2023), and potentially facilitate illegal insider trading (Ahern, 2017).

[3] See McLean and Pontiff (2016) and Harvey (2017) for recent meta studies.

[4] In addition to industry momentum, we also include product market-based lead-lag effect (e.g., Hoberg and Phillips, 2018), geographic lead-lag effects (e.g., Parsons et al., 2020), customer-supplier lead-lag effects (e.g., Cohen and Frazzini, 2008), lead-lag effects based on technology similarity (Lee et al., 2019), the complicated firm effect (Cohen and Lou, 2012), and the lead-lag control based on shared analyst coverage (Ali and Hirshleifer, 2020).

edness between locations relates to firms' strategy similarity and comovement in fundamentals and returns. In Section 4, we discuss whether social peer firm returns can predict focal firms' returns. In Section 5, the returns of socially connected peer firms capture earnings-related information about the focal firm. In Section 6, we present robustness tests to the earlier empirical results. Section 7 concludes.

2 Data and Variable Definitions

Our sample includes all common stocks traded on the NYSE, AMEX, and NASDAQ exchanges, covering the period from July 1963 through December 2019. The Center for Research in Security Prices (CRSP) provides daily and monthly return and volume data. The accounting variables, including earnings, are obtained from the CRSP-Compustat merged database. Analyst earnings forecasts and institutional ownership data are from the Institutional Brokers' Estimate System (I/B/E/S) database and Thomson Reuters institutional (13F) holdings database, respectively. Data on the Fama-French five-factor and the Fama-French 48-industry classification are obtained from Kenneth French's data library. We eliminate stocks with a price per share less than \$5 or more than \$1,000. We require a minimum of 24 monthly observations for variables created using monthly data and 15 daily observations for those created using daily data. Unless otherwise stated, all variables are measured as of the end of the portfolio formation month. The variables and the corresponding definitions are summarized in Table A.1.

2.1 Key Variables

We follow Bailey et al. (2018a) and measure social connectedness between two U.S. counties using Facebook's *Social Connectedness Index* (SCI). The measure is the total number of Facebook friendship links between two U.S. counties (as of April 2016), divided by the product of the populations of the two counties. As the world's largest online social networking service, Facebook's scale and the relative representativeness of its user body make SCI a comprehensive measure of the geographic structure of the U.S. social networks.[5]

[5] Facebook had more than 2.1 billion monthly active users globally and 239 million active users in the United States and Canada as of 2017. This represents 68% of the adult population and 79% of online adults in the United States (Duggan et al., 2016). Facebook usage rates among U.S.-based online adults were relatively constant across various demographics and locations. Bailey et al. (2018a,b, 2019a,b,c, 2021);

Figure 1 plots heat maps of social connectedness, measured with SCI, for Cook County, IL, in Panel A, and Bartholomew, IN, in Panel B. The focal counties are colored in red and darker colors indicate higher social connectedness to the focal counties. Although the two focal counties are located in two adjacent states, the maps show differences in their relative connectedness to other regions. Cook County, which includes the city of Chicago, is strongly connected to nearby counties as well as to counties in the southern states along the Mississippi River. The pattern has been documented by Bailey et al. (2018a) and Kuchler et al. (2021) and is attributed to the large-scale migration of African Americans from southern states to northern industrial cities during the Great Migration of 1916–1970. Bartholomew County is strongly connected to nearby counties in the state of Indiana, counties in the neighboring state of Illinois, and counties in distant states such as Kentucky, Tennessee, North Dakota, Texas, and Florida. Hence, these plots show that there are substantial differences between geographic proximity and social connectedness across regions in the United States. These differences will help us identify the incremental information that social ties contain that is different from the effects of physical proximity.

[Insert Figure 1 here]

Our main variable of interest is the social peer firm return (*SPFRET*). For a given firm i and for a given month t, *SPFRET* is the average returns of i's same-industry peer firms that are located outside of the focal firm's headquarters states, weighted by the SCI between the headquarters locations of firm i and j. Formally, *SPFRET* is given by

$$SPFRET_{i,t} \equiv \frac{\sum_{j \in I_i} SCI_{i,j} RET_{j,t}}{\sum_{j \in I_i} SCI_{i,j}},$$

where I_i is the Fama-French 48-industry (FF48) to which firm i belongs, and S_i is the state where firm i's headquarters are located.[6] We exclude peer firms from the same state as

Kuchler et al. (2021, 2020), and Chetty et al. (2022) have all provided evidence that social connections observed on Facebook can serve as a reliable proxy for real-world connections, even in eras prior to the widespread use of computers and the internet.

[6]We obtain firm headquarters location data from the "Augmented 10-X Header Data" for the period between 1994 and 2018, supplemented by data parsed from 10-Q and 10-K filings on SEC's Edgar database for 2019. For observations prior to 1994, for which firms' electronic filings are unavailable, we follow Parsons et al. (2020) and use the first available headquarters location in a firm's post-1994 SEC filings. As argued by Parsons et al. (2020), such measurement error would only create noise and bias against find-

the focal firm, thus alleviating the concerns that our results are driven by firms in close geographic proximity (e.g., Pirinsky and Wang, 2006; Parsons et al., 2020). For our long-run prediction analysis, we also consider *SPFMOM*, the long-run version of *SPFRET*, by compounding the variable from month $t - 11$ to $t - 1$.

We illustrate the construction of *SPFRET* for Cummins Incorporated (ticker: CMI), a machinery industry (FF48 code Mach) company that produces engines, filtration, and power generation products and is headquartered in Bartholomew County, Indiana. Figure 1 Panel C displays a heat map of SCI with Bartholomew County as the focal county, highlighting only the counties with the presence of firms in the machinery industry, such as Caterpillar (ticker: CAT) and Clarcor (ticker: CLC) located in Peoria, IL and Williamson, TN, respectively, which share high social ties with Bartholomew. Stanley Black & Decker (ticker: SWK) is another industry peer located in Hartford, CT, a county with low SCI with Bartholomew.[7] Counties without the presence of such firms are presented in dark grey. When calculating *SPFRET* for CMI, peer firms headquartered in counties with higher SCI with Bartholomew are assigned higher weights. Consequently, CAT and CLC have corresponding weights of 3.44% and 4.32%, respectively, while the weight on SWK is only 0.57%. By comparison, the value-based weights for CAT, CLC, and SWK would have been 13.66%, 0.83%, and 4.32%, respectively. Therefore, the SCI-based weighting of industry peers is notably different from the equal- or value-weighted industry portfolio.

2.2 Firm Fundamentals

Following Parsons et al. (2020), we examine the comovement between focal firms and their peer firms' fundamental changes. We examine four firm fundamental variables, including ΔEPS, $\Delta Sales$, $\Delta Employees$, and $NewCapital$, all obtained using the annual fundamentals data set in Compustat. These variables are defined as follows: ΔEPS is the change in EPS scaled by lagged stock price (as in Kothari et al., 2006), $\Delta Sales$ is the

ing significant results. We thank Bill McDonald for providing the pre-1998 data through the Notre Dame Software Repository for Accounting and Finance (SRAF).

[7] Caterpillar designs, develops, engineers, manufactures, markets, and sells machinery, engines, financial products, and insurance to customers via a worldwide dealer network. It is the world's largest construction equipment manufacturer. Clarcor, Inc. manufactures engine filtration and industrial/environmental filtration systems. Stanley Black & Decker, Inc. manufactures industrial tools, household hardware, and security products. The weights are measured as of April 2016.

percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value [8].

2.3 Measuring Firm Strategy Similarities

As firm strategy is a complex concept that can influence many aspects of the firms' decisions, we use multiple measures to capture different facets of their strategy choices. First, we propose a novel and holistic measure of strategy similarity by applying textual analysis to the entire text of firms' 10-K filings. In addition to this new measure, we corroborate our results using two existing similarity measures that specifically focus on technology and product market strategies.

10-K Based Strategy Similarity We construct a 10-K-based strategy similarity measure using textual analysis of firms' 10-K filings. We do this in two steps. In the first step, we use the Latent Dirichlet Allocation (LDA) model estimated by Dyer et al. (2017) to identify paragraphs associated with corporate strategy. Specifically, for each word in a given paragraph, we assign the probability that the word belongs to a specific topic using the probability parameters provided by Dyer et al. (2017).[9] Next, we multiply the word's topic-specific probability with the frequency the word appears in the paragraph and sum up the product across all words in the paragraph to obtain a paragraph-level score for each of the topics. We then assign the topic with the highest score as the topic of that paragraph and extract paragraphs associated with topics under the "Business Operation & Strategy" category as defined in Dyer et al. (2017).[10] In the second step, we follow Hoberg and Phillips (2016) and classify a pair of firms as similar when these paragraphs include a common set of words. Specifically, 10-K strategy similarity (*STRATSIM*) is defined as the cosine of the word vectors of the strategy-related 10-K paragraphs of the two firms over the past five years.

[8]Following Loughran and Wellman (2011), enterprise value is defined as the market value of equity (Compustat data item MKVALT) plus short-term debt (DLC) plus long-term debt (DLTT) plus preferred stock value (PSTK) minus cash and short-term investments (CHE).

[9]We drop the commonly used words that appear in more than 25% of the 10-K documents.

[10]Business Operations & Strategy category includes topics such as advertising, environmental impact, software and system. The link table between topics and categories is available at: http://tinyurl.com/hrep57s.

We provide a number of examples in Appendix I to demonstrate how we construct this similarity measure. The first two examples are from pairs of firms with medium to high strategy similarity. The first pair features two companies in the construction industry: Granite Construction and Sterling Construction. Both companies discuss the importance of forming joint ventures to reduce risks in their project development process. Their strategy similarity score is above 0.4, which indicates that they share a high degree of similarity in their strategies. The second example comes from two firms that produce beverages (i.e., National Beverage Corporation and Coca-Cola). They have a moderate level of similarity (similarity score is 0.131). The excerpts from 10-K filings indicate that both companies emphasize developing and marketing products to consumers.

The next two examples come from the business services industry. ACI Worldwide has put much discussion into developing new technologies, while Healthcare Services Group focuses on office leasing and developing menus. As a result, there is little overlap in the words they used in their strategic discussions. Thus, these two companies exhibit a low level of strategy similarity with a numerical score of 0.029. The final pair of firms also comes from business service industries. One company (Landauer Inc) emphasizes its unique technology in measurement devices while the other company (ArcSight) seems to focus on solutions for enterprises. As a result, these firms also exhibit a low level of strategy similarity. Similar to the previous example, their similarity score is also 0.029.

Technology Strategy Similarity We expect that firm success is likely to be strongly influenced by the extent to which they adopt the right technologies. Thus, we expect firms to exhibit similar levels of success when they have similar technology strategies. Following Lee et al. (2019), we measure the technology similarity of firms, $STRATSIM^{TECH}$, by their patent overlap. Each patent falls into a unique USPTO class. Based on the patents granted in a rolling five-year window, one can form a vector where each dimension represents the number of patents filed in a specific USPTO class. The similarity in technology strategy is then computed based on the cosine similarity of these two vectors based on the past five years of patent data.

Product Market Strategy Similarity Product market similarity is measured using the approach proposed in Hoberg and Phillips (2016).[11] They parse the product description section of 10-K filings. Specifically, they form a list of words based on all 10-K product market descriptions in a given year. For each firm, they represent the product market description based on a vector of this word list. The similarity between firms' product strategies, $STRATSIM^{PROD}$, is then calculated using the cosine similarity approach based on the 10-K data released in the past year.

2.4 Controls

Lead-Lag Return Predictors We control for a list of variables that previous studies have shown to have cross-firm return predictability. The data are available for the entire sample period from July 1963 through December 2019 unless otherwise mentioned.

For a given stock and for a given month t, we define *INDRET* as the month-t equal-weighted average return of stocks from the same industry as the focal stock (see Moskowitz and Grinblatt, 1999).[12] *TNICRET*, available starting July 1989, is the equal-weighted stock return of peer firms identified through 10-K product text (Hoberg and Phillips, 2018). *GEORET* is the month t equal-weighted average return of all stocks from the same economic area (EA) as the focal stock, excluding same-industry stocks (Parsons et al., 2020). *CFRET*, available since January 1982, is the month t weighted average return of stocks that share at least one analyst with the focal stock over the previous 12 months, where weights are the number of shared analysts between stocks.

CRET, available starting December 1976, is the equal-weighted average stock return of the main customers of the focal firm, where a six-month gap is required between the fiscal year-end of the supplier and stock returns (Cohen and Frazzini, 2008). *TECHRET*, available between July 1963 and July 2012, is the weighted average stock return of technology-linked peer firms, where the weights are the technology closeness between the peer firm and the focal firm, determined by the similarities between patent distributions across different technology categories (Lee et al., 2019).[13] *CONGRET* is the pseudo-conglomerate

[11] We obtain this data from Hoberg-Phillips Data library: https://hobergphillips.tuck.dartmouth.edu/.

[12] We also consider a long-run version by compounding the variable from month $t-11$ to $t-1$.

[13] The economic area-ZIP Code link file is obtained from Riccardo Sabbatucci's website. The textual network industry relatedness classification (TNIC) for individual firms is obtained from the online data library of Gerard Hoberg and Gordon Phillips. Customer-supplier firm links are obtained through the Linking Suite

return, defined as the sales-weighted return of (value-weighted) single-segment firm portfolios, formed for each segment in which a conglomerate firm operates (Cohen and Lou, 2012); data are obtained from the Compustat segment files and starts from July 1977.

Other Controls We also control for a stock characteristics that have been show in the literature to predict returns. The variables are measured as of the end of month t unless otherwise stated and are described below.

RET is the monthly stock return, adjusted for delisting to avoid survivorship bias (Shumway, 1997). We estimate firm size (*SIZE*) and book-to-market ratio (*BMKT*) following Fama and French (1992). *MOM* is obtained by compounding *RET* from month $t-11$ to $t-1$ (Jegadeesh and Titman, 1993). *IVOL* is the monthly idiosyncratic volatility, computed as the standard deviation of the daily residuals obtained by regressing the stock return on the Fama and French (1992) three factors over the previous month (Ang et al., 2006). *ILLIQ* is Amihud's illiquidity (Amihud, 2002), defined as the average daily ratio of the absolute stock return to the dollar trading volume within the previous month. *MAX* is the maximum daily stock return realized over the previous month (Bali et al., 2011). *SKEW* is the sample skewness of the daily stock returns from the previous month (Bali and Hovakimian, 2009). *COSKEW* is the stock's monthly coskewness, following (Harvey and Siddique, 2000).

2.5 Summary Statistics

Table **1** provides summary statistics (Panel A) and correlations (Panel B) for the pairwise variables used in our analyses, including SCI, strategy similarity measures, shared analyst coverage, and customer-supplier linkages. First, we notice that our 10-K based strategy similarity measure is positively correlated with technology and product market similarity measures, confirming that the 10-K based measure captures information related to firms' strategy choices. Second, we find that SCI is indeed positively related to many measures of firm-pair similarity measures, such as strategy, product, and technology similarity.

Appendix Table A.2 presents the summary statistics of *SPFRET*, together with stock returns and return predictors established in the previous studies. The distributions of

by WRDS. We thank Usman Ali and Stephen Teng Sun for providing their data on technological overlap among firms.

these variables are consistent with earlier studies such as Ali and Hirshleifer (2020). In terms of correlation, *SPFRET* and *INDRET* are both positively correlated with contemporaneous returns (*RET*) with correlation coefficients of 0.166 and 0.204 respectively. *SPFRET* is highly correlated with *CFRET* of Ali and Hirshleifer (2020), with a coefficient of 0.494. Furthermore, the average cross-sectional correlation between *SPFRET* and *INDRET* is considerable, at 0.642. This is likely due to the construction of *SPFRET*, which relies on the SCI-weighted, same-industry firm returns.

[Insert Table **1** here]

3 Social Ties and Firm Comovements

A growing body of literature shows that social ties between regions foster economic interactions (e.g., Bailey et al., 2018b), suggesting that social ties can potentially foster the exchange of ideas between firms located in these regions. As a result, firms tend to adopt similar strategies, which in turn result in firms headquartered in socially connected areas exhibiting stronger comovements both in their fundamentals and returns. In this section, we empirically test these predictions.

3.1 Social Ties and Strategy Similarity

We first formally examine the extent to which social connectedness between firms' headquarters relates to firms' strategy similarity. We conduct pairwise regression analyses with the following specification:

$$Similarity_{i,j,t} = \beta SCI_{i,j} + \gamma Same\ County_{i,j} + \alpha_t + \delta_{Ind} + \zeta_{County,i} + \eta_{County,j} + \epsilon_{i,j,t}, \quad (1)$$

where *Similarity* is one of the three strategy similarity measures between the two firms, *STRATSIM*, $STRATSIM^{TECH}$, or $STRATSIM^{PROD}$. SCI measures the social connectedness between the headquarters locations of the two firms. We control for the industry affiliation of two companies, the headquarters counties, and time fixed effects. We limit our firm pairs so that they are in the same industry.[14]

[14]We follow Gu et al. (2020) and apply a rank-based standardization to the independent variables in all our regressions, unless otherwise stated. Specifically, in each period and for a given independent variable, we cross-sectionally rank the observed values from lowest to highest and map the ranks into the $[0, 1]$ interval

This set of results is reported in Table 2. In Panel A, we focus on the relationship between SCI and similarity measures. In column 1, we find a strong positive relationship between SCI and the 10-K-based strategy similarity measure. This result confirms that stronger connections between firms' headquarters locations are associated with a higher level of overall firm strategy. Specifically, comparing firm pairs with SCI in the top quartile with those in the bottom quartile, the strategy similarity is 3.5 percentage points higher, which corresponds to 11.4% of the standard deviation of the dependent variable.

Next, we focus on two specific dimensions of firm strategy similarity. In column 2, we report the results with technology similarity as the dependent variable. Again, SCI significantly relates to technology similarity, indicating that firms with high social connectedness tend to invest in and adopt similar technologies. Finally, we consider companies' product market similarities. As reported in column 3, we also find a positive and significant relationship between firms' social connectedness and product market similarity. Specifically, comparing firm pairs with SCI in the top quartile with those in the bottom quartile, the technology similarity and product market similarity are 3.5 and 3.7 percentage points higher, which correspond to 12.0% and 18.0% of the respective standard deviations.

In Panel B, we account for the effect of firms being colocated in the same county by including an indicator variable, *Same County*, that takes a value of one if both firms are located in the same county and zero otherwise. As Parsons et al. (2020) have shown, firms located in the same counties tend to share many similarities. Controlling for the same county effect ensures that the relationship between SCI and firm similarity is not solely driven by the firms' shared locations. The results in Panel B demonstrate that all the findings reported in Panel A remain robust. Collectively, while our results do not allow us to make any claims about causality, they do provide prima facie evidence that social connections foster strategy similarity.

[Insert Table 2 here]

(ranking is done within industry in Eq.1). This non-parametric transformation allows the analysis to focus on the ordering of the variables as opposed to the magnitudes, which makes the estimates less sensitive to outliers (e.g., Kelly et al., 2019; Freyberger et al., 2020).

3.2 The Comovements of Firm Fundamentals and Stock Return

Having shown that firms located in socially connected areas tend to adopt similar strategies, we next focus on understanding the implications for firms' fundamentals. Specifically, we examine whether firms' socially connected areas tend to comove more strongly in their fundamentals and returns. Following Parsons et al. (2020), we conduct the panel regressions:

$$Fundamental_{i,j,t} = \beta_1 Fundamental_{SPF,i,j,t} + \beta_2 Fundamental_{j,t} + \alpha_t + \epsilon_{i,j,t} \quad (2)$$

$$RET_{i,j,t} = \beta_1 RET_{SPF,i,j,t} + \beta_2 RET_{j,t} + \alpha_t + \epsilon_{i,j,t} \quad (3)$$

The dependent variable $Fundamental_{i,j,t}$ is a fundamental measure for a focal firm i, in the industry j, and at time t. We consider four fundamental measures, ΔEPS, $\Delta Sales$, $\Delta Employees$, and $NewCapital$, measured at the annual frequency. We also consider monthly firm returns as a key dependent variable.[15] The key independent variable $Fundamental_{SPF}$ (RET_{SPF}) is the change in the corresponding fundamental measure (returns) of industry peers, weighted by their SCI to the focal firm's headquarters. We exclude firms from the focal firm's headquarters states to reduce the influence of geographically co-located firms studied in Pirinsky and Wang (2006) and Parsons et al. (2020). To ensure that our results are not solely driven by fundamental comovement due to industry affiliation, we control for the average change in fundamentals and returns. We consider both equal and value-weighted industry portfolios.[16] We additionally control for time fixed effects.

We start our analyses with univariate analyses. In Table **3** Panel A, we focus on the relationship between the SCI-weighted portfolio and the focal firm. Column 1 presents the result based on ΔEPS. We find that focal firms' EPS growth significantly comove with that of social peer firms. Specifically, focal firms with the highest social peer firm EPS growth exhibit 2.44 percentage points higher ΔEPS than those firms with low social peer firm ΔEPS. Similarly, as reported in column 2, firms with the highest social peer $\Delta Sales$ are 30.85 percentage points higher than those firms with the lowest social peer firm sales

[15] In columns 1-4, for each fiscal year end-focal firm pair we look for peer firms that have the same fiscal year-end as the focal firm to ensure that the fundamentals are calculated over the same time period. In column 5, this constraint is not imposed.

[16] We exclude the focal firm when forming the equal-weighted industry peer.

growth. We also find significant comovement between focal firms and their social peers in both employee number and newly raised capital as reported in columns 3 and 4. Given strong comovements between focal firms' and social peer firms' fundamentals, we expect that focal firms' returns to comove strongly with their social peer firms. To investigate this hypothesis, we conduct the regression 2 with monthly returns as the variable of interest. We report this result in column 5. We find that firms with the highest social peer firm returns outperform those with the lowest social peer firm returns by 7.91%. This shows that firms headquartered in locations with close social ties to the focal firm county are more informative about the focal firm than average industry returns. As a comparison, in Appendix Table A.3, we report the univariate relationship between equal- and value-weighted industry portfolio and the focal firm. While we find a strong positive comovement between the equal-weighted industry peers and the focal firm, the economic magnitudes of these coefficients are smaller than those reported in Table **3** Panel A.

In Panel B, we further conduct multivariate regression analyses by including both SCI-weighted and equal-weighted portfolios. This analysis further highlights that the SCI-weighted portfolio tends to comove more with the focal firm compared with the equal-weighted portfolio. Out of the four fundamental performance variables, the SCI-weighted peer firm portfolio consistently exhibits strong and positive coefficients, while the equal-weighted portfolio only positively and significantly relates to focal firms' sales growths. Similarly, we also find that SCI-weighted returns exhibit strong return comovement with focal firms, but equal-weighted variables do not exhibit significant comovement with that of focal firms.

In Panel C, we conduct multivariate analyses with the value-weighted industry portfolio instead of the equal-weighted industry portfolio and show that our results are robust. These results suggest that the SCI-weighted peer portfolio is more tightly connected with focal firms in their fundamentals than equal- or value-weighted industry peers.

[Insert Table **3** here]

4 Social Ties and Return Predictions

As shown in Section 3, there is a significant correlation between a firm's fundamentals and stock returns and those of its social peers. However, it remains a question of whether the cross-firm linkages are promptly reflected in stock prices. If investors fail to fully incorporate information from social peer firms, it is possible that their returns could predict the future performances of focal firms. In this section, we conduct a formal analysis to explore whether the returns of social peer firms can serve as a predictor of focal firms' returns. We begin by presenting evidence that a long-short portfolio sorted by *SPFRET* yields significant abnormal returns. We then examine the extent to which any abnormal returns can be explained by the existing predictive relationships.

4.1 Portfolio Analysis

We first perform portfolio analysis and investigate the returns to a long-short portfolio sorted by *SPFRET*. Specifically, for each month, we sort our sample firms into deciles based on *SPFRET*. We then hold the portfolios for a month and calculate returns for each portfolio and the return differential between the portfolios with the highest and the lowest *SPFRET*.

Panel A of Table 4 reports the results of the univariate portfolio analysis. The first row shows the raw returns for the equal-weighted portfolios. To ensure that our results are not driven by risk factors, we also report the Fama-French five-factor alphas (FF5) for equal-weighted portfolios in the second row. The third and fourth row report the raw returns and FF5 alphas for value-weighted portfolios, respectively. We find that the SPFRET-sorted long-short portfolio produces economically large and highly significant excess returns. For example, the FF5 alpha is 157 basis points per month (t = 7.06) for the equal-weighted portfolio and 84 basis points (t = 5.22) for the value-weighted portfolio.

[Insert Table 4 here]

Figure 2 presents the average monthly long-short portfolio alpha over time. Panel A shows that a positive alpha exists for most of the years in our sample. Given that *SPFRET* is highly correlated with industry momentum, we examine the incremental return predictability of *SPFRET* by examining the abnormal returns of portfolios sorted by $SPFRET_{\perp}$ (which is obtained by orthogonalizing the *SPFRET* measure by industry momentum). Panel B shows that the $SPFRET_{\perp}$-based portfolios also consistently produce positive returns. This suggests that *SPFRET* captures additional information that is distinct from the average industry momentum effects.

[Insert Figure 2 here]

Given the finding that industry momentum strongly predicts a stock's future returns (e.g., Moskowitz and Grinblatt, 1999), we next conduct a two-way sorted portfolio analysis to further isolate the effect of *SPFRET* from the effects of industry momentum. At the end of each month, stocks are first sorted into quintiles based on *INDRET*. Then, within each *INDRET* quintile, stocks are further sorted into quintiles based on *SPFRET*. We calculate the returns for each of the 25 portfolios in the month that follows, along with the returns of five high-minus-low portfolios based on *SPFRET* for each *INDRET* quintile.

We report the results of a bivariate portfolio sort with equal-weighted and value-weighted Fama and French (2015) abnormal returns in Panel B of Table **4**. The final row reports the average alphas for the *SPFRET* quintiles. The high-minus-low average alpha of the equal-weighted portfolio is 56 basis points per month with a *t*-statistic of 7.38 whereas the economic magnitude under value-weight is 39 basis points with a *t*-statistic of 4.84. Hence, the bivariate sort results further confirm that *SPFRET* has economically large return predictability that is above and beyond the effect of the traditional industry momentum of Moskowitz and Grinblatt (1999).[17]

4.2 Social Peer Firm Returns and Other Lead-lag Predictors

Our analysis has thus far demonstrated the significant predictability of *SPFRET* in forecasting future returns. In this subsection, we conduct additional tests to enhance our understanding of the economic foundations of the lead-lag effect observed in the preceding section. Specifically, given that prior research has identified several other lead-lag variables capable of predicting firm returns, we examine the extent to which the predictability of *SPFRET* is attributable to these established mechanisms. In addition to the predictors already established in the literature, we introduce a new predictor that takes into account strategy similarity between firms and explore whether strategy similarity can help explain our findings. Furthermore, we investigate the robustness of our findings by controlling for these alternative cross-sectional predictors and assess the incremental contribution of *SPFRET* in a machine learning-based composite predictor that encompasses

[17] Another way to control for industry momentum is to sort *SPFRET* within each industry. We report these results in Table A.4. Consistent with the double-sort results presented in Table **4**, *SPFRET*-based strategy still generates significant abnormal returns.

all individual predictors.

4.2.1 Spanning Tests

We first conduct a portfolio spanning test to understand the extent to which the *SPFRET*-based return predictability is explained by other well-documented lead-lag relationships documented in the existing literature.

Specifically, we follow Ali and Hirshleifer (2020) and estimate the following time-series regression:

$$RET_{SPFRET,t} = \alpha + \beta F_t + \epsilon_t, \tag{4}$$

where RET_{SPFRET} is the value-weighted long-short portfolio returns constructed in Section 4.1. F represents a vector of long-short portfolios. We consider the (Fama and French, 2015) five-factor returns and the returns of portfolios based on firms' economic linkages as identified by previous studies. To ensure that the coefficients are comparable across different columns, we focus on a sample period of July 1989 through November 2019, so that return information exists for all the portfolios considered in our test. We report the portfolio alpha of the lead-lag portfolios considered in Figure 3. The figure shows that these lead-lag portfolios exhibit positive returns for our sample period and many of them exhibit considerable statistical significance.

[Insert Figure 3 here]

Table **5** reports the results for a sample period in which all portfolios under our consideration have valid return data to facilitate the comparison across different specifications. Column 1 is the baseline specification with a constant and the Fama-French five-factor as the explanatory variables. The column shows that the coefficient on the constant, which captures the Fama-French five-factor alpha of the *SPFRET*-based portfolio strategy, is 41.6 basis points per month and is significant at the 10% level.[18]

Next, we gradually include additional long-short portfolios one at a time to examine the extent that the alpha of a *SPFRET* portfolio can be explained by other lead-lag effects. We first consider the industry momentum strategy based on the FF 48 industry

[18] The magnitude is somewhat different from Table **4** because we only include the observations that can be used in subsequent spanning tests.

(i.e., *INDRET*) in column 2. Industry momentum is highly relevant as *SPFRET* is constructed using firms from the same industry. Indeed, we find that the *SPFRET* strategy loads highly positively on the industry momentum portfolio. However, after controlling for *INDRET*, the alpha of *SPFRET* is 44 basis points per month and remains significant at the 1% level. This result also confirms our earlier test that shows industry momentum cannot explain the *SPFRET* strategy.

We next include the geographic momentum documented in Parsons et al. (2020). Column 3 shows that the coefficient of *GEORET* is positive but insignificant, suggesting that geographic momentum is not a major contributor to the alpha of *SPFRET*. This may be because we exclude industry peers that are headquartered in the focal firm's headquarters county when constructing the *SPFRET* variable. After controlling for *GEORET*, we have an alpha of 41 basis points.

Cohen and Lou (2012) document that a conglomerate firm's returns can be predicted by the return of its industry segments. While it is unclear how this effect can directly explain our results, we nonetheless consider it as an additional control. Again, as reported in column 4, we find that our portfolio loads on *CONGRET* positively and significantly, but our alpha remains highly significant.

In column 5, we additionally include the lead-lag effect due to the customer-supplier relationship (Cohen and Frazzini, 2008). While it is possible that firms located in socially connected areas are more likely to form customer-supplier relationships, this idea is not supported in the correlation matrix reported in Panel B of Table **1**. The *SPFRET* alpha remains highly significant with an alpha of 36 basis points per month.

In column 6, we further include the lead-lag effect due to product market similarities (see Hoberg and Phillips, 2018, for more detail). As we show in Section 3, SCI is strongly positively related to firms' product market similarity. Thus, the lead-lag effect due to product market similarities may explain the returns of *SPFRET* strategy. Empirically, we find that the *SPFRET* portfolio loads significantly on the returns of the product market lead-lag effect. However, the alpha of the *SPFRET* portfolio remains highly significant, at 34 basis points.

We also add the lead-lag effect due to technology similarity between firms (Lee et al., 2019), as Bailey et al. (2018a) find evidence consistent with social connectedness facilitating technological spillover across regions. Moreover, as shown in Table **2** we find that firms headquartered in socially connected areas tend to exhibit high degrees of technology similarity. As reported in column 7, the variable does not fully subsume the alpha of the *SPFRET* portfolio.

Next, we control for portfolio returns based on *CFRET*, which captures returns of lagged returns of firms with shared analyst coverage (e.g., Ali and Hirshleifer, 2020). This strategy is not intended to capture a specific economic relationship, but it is shown that it accurately summarizes economic relevance between two firms and thus subsumes many existing lead-lag relationships. We include the return based on *CFRET* in column 8. While *SPFRET*-based portfolio significantly loads on the *CFRET*-based portfolio returns, we find that *SPFRET*-based portfolio still generates a 31 basis points alpha in the following month. The takeaway from our analysis, so far, shows that none of the existing lead-lag variables fully subsume the lead-lag relationship due to social ties.

To further understand what explains the remaining lead-lag effect, we revisit our analyses of firm strategy similarities. As we show in Table 2, social connectedness between firms' headquarters locations is strongly positively related to a new measure of strategy similarity estimated from the entire 10-K filing texts of a firm. Thus, it is possible that our results are driven by a previously undocumented firm strategy similarity-related lead-lag relationship.

To examine this conjecture, we form a firm strategy similarity-weighted portfolio using the focal firm's industry peers' returns (*STRATRET*). Specifically, for each focal firm, we construct the portfolio by including all firms in the focal firms' Fama-French 48 industry, excluding the firms that are headquartered in the same state as the focal firm. The peer firms' weight in the portfolio is the 10-K based strategy similarity used in our earlier analyses. As reported in column 9, we find that the strategy similarity return is strongly positively related to the return based on the *SPFRET* signal. Moreover, including the *STRATRET* portfolio reduces the economic and statistical significance of some of the existing lead-lag effects, including strategies based on product market similarity and overlapped analyst coverage. Finally, including the strategy similarity provides further explanatory power to the alpha associated with *SPFRET* and renders it insignificant. These observations are consistent with the idea that a substantial portion of *SPFRET*'s return predictability is, indeed, attributable to firms' strategy similarities.

[Insert Table 5 here]

Figure 4 further illustrates the relative contributions of the economic linkage-based variables, using coefficient estimates from the last column of Table **5** and the average returns of the corresponding long-short portfolios. The figure shows that the lead-lag effects documented in the existing literature explain 15 basis points of SPFRET-based returns and *STRATRET* is the most prominent variable in this comparison, which explains roughly 7.5 basis points. Thus, we confirm that *STRATRET* is at least as important as other well-documented lead-lag effects.

[Insert Figure 4 here]

4.2.2 Fama-MacBeth Regression Analysis

While spanning tests help us identify potential explanations for the strategy returns generated by *SPFRET*, it does not provide a horserace test of *SPFRET* and the competing predictors. Thus, we conduct further analysis using Fama-MacBeth regression, which allows us to examine whether social peer firm returns predict future stock returns at the stock level while accounting for a comprehensive set of controls.

We estimate the following regression:

$$RET_{i,t+1} = \alpha + \beta SPFRET_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1}, \tag{5}$$

where $RET_{i,t+1}$ is the one-month-ahead firm return and *SPFRET* is the social peer firm returns. X is a vector of control variables that includes short-term industry momentum (*INDRET*), the geographic lead-lag effect (*GEORET*), the shared analyst coverage lead-lag effect (*CFRET*), the customer-supplier lead-lag effect (*CRET*), the technological spillover effect (*TECHRET*), the complicated firm effect (*CONGRET*), and the text-based industry momentum (*TNICRET*), and the strategy similarity-based return (*STRATRET*). We also include a battery of well-known cross-sectional predictive variables, including *RET*, *SIZE*, *BMKT*, *BM*, *MOM*, *IVOL*, *ILLIQ*, *MAX*, *SKEW*, and *COSKEW*. We standardize the independent variables following Gu et al. (2020) and Kelly et al. (2019).

Table **6** reports the results of Fama-MacBeth regressions. In column 1, we only include *SPFRET* and a vector of firm characteristics unrelated to a between-firm lead-lag relationship. We find that *SPFRET* is positive and significant at the 1% level. Moving from the lowest to the highest *SPFRET* leads to a 1.28% higher average return in the following month. Next, we gradually include additional lead-lag controls in columns 2 through 9.

In column 2, we include the industry momentum effect documented in Moskowitz and Grinblatt (1999). As expected, the industry momentum variables are highly significant and their inclusion attenuates the effect of *SPFRET* to some extent, due to the strong positive correlation between these variables. Nevertheless, the coefficient of *SPFRET* remains large and highly significant, at 0.71, with a *t*-statistic of 3.8.

Column 3 additionally controls for the geographic lead-lag effects and shows that the coefficient of *GEORET* is positive and highly significant, consistent with Parsons et al. (2020). More importantly, the coefficient of *SPFRET* remains very similar to the corresponding value in column 2, suggesting that the geographic lead-lag effect does not contribute significantly to the effect of *SPFRET*.

Similarly, in columns 4 through 7, we further control for the "complicated firm" effect (Cohen and Lou, 2012) by including *CRET*, the customer-supplier effect (Cohen and Frazzini, 2008), the product market similarity effect (Hoberg and Phillips, 2018) (*TNICRET*), and the effect of technology linkage (Lee et al., 2019) (*TECHRET*), respectively. We show that these variables do not significantly affect the economic magnitude and the statistical significance of *SPFRET*.

Next, we control for *CFRET*, which is a proxy for the analyst lead-lag effect. This variable is known to explain many other lead-lag effects in the literature (Ali and Hirshleifer, 2020). Column 8 shows that the return predictability of *SPFRET* remains strong and economically meaningful even after including this control. To further ensure the validity of these key empirical analyses, we present a comprehensive set of robustness tests in subsection 6.2.

Finally, we include *STRATRET*, which is a lead-lag signal based on strategy similarity. As shown in the spanning test analysis, we find that this variable explains a substantial fraction of the returns related to *SPFRET* portfolio. However, it is unclear if *SPFRET* still exhibits the power to predict future returns after controlling for *STRATRET* in a cross-sectional regression. We find that *SPFRET* remains highly significant. In contrast, *STRATRET* does not exhibit independent power in predicting future returns. This result shows that, while *SPFRET* and *STRATRET* are highly related (with a correlation of 0.749), *SPFRET* better captures the multi-faceted similarity across pairs of firms. As a result, including *SPFRET* drives out the predictive power of *STRATRET*.

In sum, our results show that *SPFRET* continues to significantly predict next month's return despite including a comprehensive set of predictive variables associated with specific economic linkages across firms. The economic magnitude of *SPFRET* is substantial. Firms with the highest *SPFRET* outperform firms with the lowest *SPFRET* by 43 basis points in the following month. This set of results is consistent with the analyses presented in Table 5, supporting the notion that industry peer firms located in socially proximate areas contain important information about the focal firm that is not fully incorporated into the focal firm's stock prices.

The comprehensive set of economic linkages that we consider here partially explains the predictability of *SPFRET*, suggesting that social ties between firms' headquarters locations capture some of these known economic linkages such as shared analyst coverage and strategy similarity. More importantly, our results also show that these economic linkages do not fully subsume the predictability of *SPFRET*. Hence, our evidence suggests that the cross-firm linkages are multifaceted and that our social tie–based peer returns help provide incremental information regarding some of these nuanced linkages.

[Insert Table 6 here]

4.3 Relative Importance in the Composite Return Predictor

So far we have shown that our main variable, *SPFRET*, can significantly predict one-month-ahead focal stock returns in a way that is incremental to 17 other predictors used in the literature. Next, we assess the extent to which our variable can help enhance the joint return predictability of all the existing predictors using a machine-learning approach. Specifically, following Kelly et al. (2019) and Gu et al. (2020), we evaluate the incremental contribution of *SPFRET* relative to a composite predictor that aggregates all 18 individual predictors included in equation (5) through partial least squares (PLS).[19]

To obtain our results, starting with July 2000, we train a new partial least squares (PLS) model every twelve months, using all the data available up to that point. Then, for each training set, we obtain the in-sample R^2s from the full PLS model and from a restricted model that mutes one of the eight key cross-firm lead-lag return predictors (namely, *SPFRET*, *INDRET*, *GEORET*, *CONGRET*, *CRET*, *TNICRET*, *TECHRET*, and *CFRET*). For each muted variable, we calculate the R^2 difference between the full model

[19]We choose PLS because it is parsimonious in terms of the number of its hyperparameters, and it is also less prone to multicollinearity than OLS (Abdi, 2010).

and the restricted model.[20] We then define the relative importance of the variable as the corresponding R^2 difference divided by the sum of the R^2 differences associated with each of the eight predictors.

Figure 5 presents the monthly average relative importance of the nine cross-firm lead-lag return predictors for the 2000–2019 sample period. It shows that, among the eight key predictors, *SPFRET* contributes substantially to the predictive power of the composite predictor, with a relative importance of 26%. In particular, *SPFRET* is more important than *INDRET*, with the relative importance of *SPFRET* exceeding that of *INDRET* by 20 percentage points. In sum, our analysis indicates that *SPFRET* substantially improves the predictive power of the PLS-based composite return predictor, beyond the variables analyzed in the prior literature.

[Insert Figure 5 here]

To illustrate the economic magnitude of *SPFRET*'s contribution to the composite predictor, we compare the performance of portfolios based on the composite predictor that includes information from all 18 predictive variables (i.e., full model predictor) and an alternative composite predictor, "No SPFRET", that excludes *SPFRET*. Each month, we sort stocks into deciles based on the two predictors respectively and obtain the one-month-ahead value-weighted returns of the corresponding portfolios. We then compute the cumulative log returns over the period of August 2000 through December 2019.

Figure 6 presents the cumulative log returns of long-short portfolios based on the full model predictor (in orange) and the "No SPFRET" predictor (in blue). The cumulative return of the portfolio based on the composite predictor substantially outperforms that of the alternative "No SPFRET" portfolio by 72 percentage points by the end of 2019. Hence, the result suggests that *SPFRET* contains additional information that helps enhance the predictive power of the existing cross-sectional predictors in an economically meaningful way.

[Insert Figure 6 here]

[20] We follow Kelly et al. (2019) and Gu et al. (2020) and compute the restricted R^2 for a certain input variable by setting all values of that variable to zero in the trained complete model.

5 Drivers of Return Predictability: Underreaction v.s. Overreaction

So far, we have confirmed that *SPFRET* is a new and robust predictor of future returns. The analysis that follows explores potential behavioral biases that may be driving these results. We consider two potential explanations. First, our results may be driven by investors' underreaction to value-relevant information. A large literature shows that investors' underreaction can lead to prices not fully incorporating recent value-relevant information, leading to a lead-lag pattern in firms' returns. The second explanation is that investors may overreact, with a lag, to recent news from social peer firms.[21] Our next set of tests helps us distinguish between these two explanations.

5.1 Heterogeneity in Return Predictability

This subsection explores the relevance of a firm's information environment, which we measure with firm characteristics and whether its headquarter is located in an industry cluster.

5.1.1 Information Environment by Firms' Characteristics

One explanation for the return predictability that we document is slow information diffusion (e.g., Hong and Stein, 1999), due to investors' limited cognitive resources (e.g., Hirshleifer and Teoh, 2003; Peng and Xiong, 2006). The limited attention mechanism further predicts that the predictability should be stronger for firms with poor information environments. Thus, we examine how our results vary across firms with different information environments using three common proxies in the literature: firm size, analyst coverage, and institutional ownership.

We form portfolios based on double-sorts by *SPFRET* and the three information environment proxies. We report the value-weighted results for the short and the long legs of the strategy, as well as the long-short returns in Table 7 Panel A. We find that the return predictability is mainly driven by firms with low market capitalization. Long-short returns are 181 basis points per month for small firms, which is over three times as large

[21] For example, Lou and Polk (2022) find that comomentum, defined as the return correlation among stocks, signals overreaction and is associated with a greater reversal of momentum strategy returns.

as that for large firms. Similarly, we find that *SPFRET*-based strategy generates more positive returns for firms with low institutional ownership and analyst coverage.

Taken together, these results support the slow information diffusion hypothesis and suggest that a firm's visibility to its investors influences the speed at which information about the firm's peer returns is incorporated into the firm's stock prices.

[Insert Table 7 here]

5.1.2 Firms in Industry Clusters

The next hypothesis is motivated by Engelberg et al. (2018) that firms located inside industry clusters are observed more closely by market participants, implying that social peer firm information is more quickly incorporated into prices. As a result, return predictability of *SPFRET* should be weaker for firms located in industry clusters. To test this, we define a county as an industry cluster if the county is ranked among the top 20% or 10% in the market capitalization for a given Fama-French 48 industry. We then define an indicator variable, *OUTCLS*, as one if a firm's headquarters is located in one of the industry cluster counties and zero otherwise.

We double-sort the firms based on their *SPFRET* and *OUTCLS* and report the portfolio returns in 7 Panel B. The first three columns define industry cluster counties using the top 20% of counties, and the next two columns define the top 10% of counties as industry cluster counties. In both specifications, we find that *SPFRET* generates stronger long-short returns for firms that are located outside of industry clusters compared with firms headquartered inside industry clusters. The relationship suggests that stock prices of firms located outside of industry clusters are slow to incorporate important information in social peer firm returns.

5.2 Predicting Long-Run Returns and Fundamentals

This subsection investigates the extent to which social peer firm returns forecast firm stock returns and their fundamentals over longer horizons. We then examine the extent to which investors and analysts incorporate this information into their expectations. Our long-horizon analysis considers *SPFRET* as well as these returns cumulated over longer horizons, *SPFMOM*, following previous studies (e.g., Moskowitz and Grinblatt, 1999; Parsons et al., 2020; Ali and Hirshleifer, 2020).

5.2.1 Predicting Long-run Returns

Table 8 reports the long-horizon return predictability results using a calendar time portfolio approach following Jegadeesh and Titman (1993). Panel A presents the five-factor alpha of the equal-weighted calendar time portfolios. Both *SPFRET* and *SPFMOM* generate high return predictability in the month after portfolio formation (157 basis points based on *SPFRET* and 69 basis points based on *SPFMOM* per month). The long-short strategy continues to generate positive returns in months 2-3 after the portfolio formation month, generating an additional return differential of 24 and 61 basis points, respectively. The effect of *SPFRET* remains significant for months 4–6 and 7–12, and both variables lose their significance in generating excess returns after the first twelve months.[22]

The 12-month cumulative returns for the long-short portfolio sorted by *SPFRET* is 4.54%. As indicated in Table 3 Panel A, column (5), the long-short strategy based on *SPFRET* corresponds to a contemporaneous return of 7.52%. Compared to the contemporaneous effects, the cumulative return indicates that over 37% of the information in *SPFRET* is associated with a delayed reaction.[23] Panel B presents the Fama and French (2015) five-factor alpha for value-weighted portfolios. The patterns are similar to those presented in Panel A. In particular, we find that both *SPFRET* and *SPFMOM* strongly predict returns in the short run and there is no evidence for long-run reversals.

Overall, the findings show that the return predictability of social peer firm returns lasts for up to one year and the lack of reversal indicates that the excess returns reflect underreaction rather than a delayed overreaction.

[Insert Table 8 here]

5.2.2 Predicting Future Fundamental Performances and Market Participant Reactions

The premise of the underreaction hypothesis is that *SPFRET* captures information about a firms' future performances that is not captured in current prices. To provide further evidence, we next investigate the nature of the information that social peer firm returns capture by considering whether such information helps predict a firm's future fundamen-

[22] We conduct these analyses using the orthogonalized versions of *SPFRET* and *SPFMOM*, after controlling for all other variables in Panel A of Table A.5. The results are similar to those obtained by using raw *SPFRET* and *SPFMOM*.

[23] The delayed reaction is calculated as $4.54\%/(7.52\% + 4.54\%)$.

tal performances. Specifically, we focus on three earnings-related variables that include standardized unexpected earnings, forecast errors, and earnings announcement returns.

Empirically, we estimate the following panel regressions:

$$Fundamental_{i,t+s} = \alpha + \beta_1 SPFRET_{i,t} + \beta_2 SPFMOM_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1}, \tag{6}$$

where *Fundamental* represents a fundamental-related variable. *SPFRET* and *SPFMOM* are social peer firm returns and momentum, respectively.[24] X is a vector of control variables, including time and firm fixed effects. We report these results in Table **9**. In Panel A, we do not include additional controls. In Panel B, we additionally control for long-term and short-term industry momentum *INDMOM* and *INDRET*. Standard errors are clustered by month and firm.

Predicting Earnings Growth We first examine standardized unexpected earnings (SUE) in quarters 1–4 following the portfolio formation month.[25] Columns 1-4 of Table **9** Panel A reveal that *SPFMOM* and *SPFRET* strongly predicts future SUEs. Specifically, an increase from the lowest *SPFMOM* to the highest leads to a 34 percentage points increase in the following quarter's SUE, which corresponds to 24% of the average cross-sectional standard deviation of SUE. In the following three quarters, we continue to find significant, albeit smaller coefficients, indicating that both variables exhibit significant power in predicting earnings growth one-year in the future. Panel B, which includes additionally controls for industry returns and industry momentum, generates similar results indicating that social peer firm returns positively predict future earnings growth.

[Insert Table **9** here]

[24] We conduct robustness checks of results with $SPFRET_{\perp}$ and $SPFMOM_{\perp}$, the versions of *SPFRET* and *SPFMOM* that are orthogonalized to the other predictive variables. Table A.5 presents the results and shows that our findings are robust.

[25] The standardized unexpected earnings (constructed following Bernard and Thomas (1990) and Mendenhall (1991)). We calculate the unexpected earnings (UE) of stock i in calendar quarter q of an earnings announcement as $UE_{i,q} = EPS_{i,q} - EPS_{i,q-4}$, where $EPS_{i,q}$ and $EPS_{i,q-4}$ are the stock's basic earnings per share (EPS) excluding extraordinary items in quarters q and $q-4$ respectively. EPS is adjusted for stock splits and reverse splits through division by the Compustat item AJEXQ, the quarterly cumulative adjustment factor. Standardized unexpected earnings in quarter q ($SUE_{i,q}$) is defined as $UE_{i,q}$ scaled by its standard deviation over the past eight quarters, with a minimum of four UE observations available.

Do Investors Incorporate Information on Social Ties? Next, we investigate the extent to which market participants incorporate information regarding the performance of social peer firms. We first focus on sell-side analysts and ask whether their forecasts fully capture the information in social peer firm returns. Specifically, we ask whether social peer firm returns forecast the analysts' future forecast errors.

Following DellaVigna and Pollet (2009), we consider the analyst forecast error in quarters $t+1$ to $t+4$ as the key dependent variables.[26] As reported in columns 5-8 of Table 9 Panel A, both *SPFMOM* and *SPFRET* predict analyst forecast errors over the following year and this remains robust after including industry momentum controls (as reported in columns 5-8 of Panel B). These findings suggest that professional analysts do not fully incorporate information from the returns of social peer firms.

An alternative method to assess whether relevant fundamental information of social peer firms has been incorporated by investors is to examine returns around future earnings announcements. If investors fail to consider such information in a timely manner, the underreaction will be corrected around future earnings announcements. To examine this hypothesis, we estimate panel regressions with the dependent variable CAR, which represents the market-adjusted cumulative abnormal returns computed for three-day windows around earnings announcements in the next four quarters. As reported in Table 9 Panel A columns 9-12, we find that both *SPFRET* and *SPFMOM* are strongly related to earnings announcement returns in the following quarter. In fact, *SPFMOM* can predict earnings announcement CAR in two quarters. In terms of the economic magnitude, companies with the highest *SPFMOM* or *SPFRET* have 14 basis points higher returns in the following quarter's CAR compared with those with the lowest *SPFMOM* (*SPFRET*). As shown in columns 9-12 of Panel B, while controlling for industry momentum reduces the economic significance of the predictive coefficients in the first two quarters, it leads to more prolonged return predictability for *SPFMOM*.

[26]Forecast errors are defined as the difference between the announced earnings for that quarter and analyst consensus forecast (i.e., the median forecast), scaled by the stock price five trading days before the earnings announcement date. We adjust actuals, forecasts, and prices for stock splits and reverse splits by scaling them by the item CFACSHR—the cumulative adjustment factor from the daily CRSP file. We only use the forecasts for the next quarter and within 90 days of the earnings announcement date. If an analyst makes multiple forecasts within this time window, we use the latest forecast.

6 Further Analyses

We have shown that *SPFRET* strongly predicts future stock returns. This section examines these results in a more recent time period; other lead-lag strategies have weakened considerably in the post-2000 time period. We then conduct an extensive set of robustness checks of the main results presented in Table **6**. We end the section with further analysis on alternative specifications.

6.1 Predictability Post 2000

Consistent with a weakening of many stock return predictors in recent years, (e.g., McLean and Pontiff, 2016), the *SPFRET* strategy generates somewhat weaker returns in the post-2000 period. Indeed, as shown in Figure A.1, this is the case for all of the lead-lag predictors we studied.[27] One reason for this is that industry momentum (*INDRET*) is substantially weaker and becomes insignificant post-2000 (Ali and Hirshleifer, 2020). However, in contrast to the value-weighted industry momentum portfolio, the *SPFRET* portfolio still exhibits significant abnormal returns of 75 basis points, which is only slightly smaller than the full sample alpha of 84 basis points as reported in Table **4**.

6.2 Robustness Tests for Fama-MacBeth Analyses

Next, we analyze whether our regression analyses are sensitive to changes in the empirical model or variable specifications. We start by replicating the full Fama-MacBeth model. Appendix Table A.6 presents the results, with column 1 the same as column 9 of Table **6** to facilitate the comparison.

Industry Returns and Alternative Industry Classifications In our main analysis, we control for the equal-weighted industry returns (*INDRET*), which is highly correlated with *SPFRET*, with a coefficient of 64.2%. The correlation between *SPFRET* and the value-weighted *INDRET*, on the other hand, is 52.8%. Hence the equal-weighted *INDRET* provides a more stringent control and as such, presumably provides more conservative re-

[27] We exclude the firms with the largest market capitalization (i.e., the top 100 firms in market capitalization). These large firms are usually efficiently priced, and because they carry disproportionately large weights in the value-weighted portfolios post-2000, they make many strategies appear to be unprofitable in the post-2000 period. Our approach is similar to Jensen et al. (2021), who winsorize the market value of firms at 80% of the NYSE market capitalization when they construct value-weighted portfolios.

sults. Column 2 presents the results when value-weighted *INDRET* are used instead. In Column 3, we use an alternative industry classification that is based on TNIC (Hoberg and Phillips, 2018) and apply the corresponding *SPFRET* and *INDRET* in the regression. In both columns, the coefficients of *SPFRET* are substantially larger, both in magnitude and in statistical significance, confirming that our main findings are robust to the industry effects analyzed in the previous studies.

SCI Measured as Geographic Proximity One concern regarding the use of *SPFRET* is that the measure is based on Facebook's SCI as of 2016, and thus creates a potential look ahead bias in our earlier subperiod. We believe that SCI serves as a useful proxy that reflects stable historical social ties between regions. As shown in Bailey et al. (2018a), the Facebook SCI measure can be mapped to labor migration patterns dating back to the 1930s, suggesting that the SCI measure closely corresponds to historical social connections between regions. Nevertheless, there is a legitimate concern that traders may not have been able to use this measure in real time prior to 2016. There is also a concern that SCI between counties is determined endogenously. For example, the presence of economically linked firms might generate higher social connectedness between locations.

To address these concerns, we consider an alternative measure of social ties between two locations. Given that the Facebook SCI is highly correlated with geographic proximity (Bailey et al., 2018a), we use geographic proximity as an alternative SCI measure and define $SPFRET_{DIST}$ as the inverse distance–weighted industry peer firm returns. Thus, the alternative *SPFRET* measure relies only on the historically available information and is free of reverse causality concerns. Column 4 shows that $SPFRET_{DIST}$ significantly predicts future returns, with a comparable coefficient of 0.431. This suggests that one can use geographic proximity to proxy for social ties and generate significant abnormal profits in real-time portfolios.[28]

[28]The average cross-sectional correlation between *SPFRET* and $SPFRET_{DIST}$ is 45%. Despite this relatively high correlation, when we include both *SPFRET* and $SPFRET_{DIST}$ in the same regression, *SPFRET* remains highly significant at 0.450. In comparison, the coefficient of $SPFRET_{DIST}$ is smaller, at 0.343, and is also significant. Thus, consistent with Kuchler et al. (2021), our result shows that SCI captures the component of social connectedness between two locations that are distinctly different from geographic proximity.

Firms' Economic Presence Beyond Headquarters In column 5, we investigate whether the predictability of *SPFRET* can be explained by firms' economic presences in socially connected locations. We follow Garcia and Norli (2012) and Bernile et al. (2015) to extract the frequency of states in firms' 10-K filings. We exclude peer firms that are headquartered in a state in which the focal firm has an economic presence. This exercise is highly conservative, as it eliminates 28% of peer firms. We find that even under this very restrictive specification, *SPFRET* still delivers considerable predictive power.

Accounting for Homophily Another potential explanation for our result is that socially connected counties tend to have similar socioeconomic characteristics. Thus, firms located in connected counties are more likely to face correlated economic shocks. Thus, we control for the lagged portfolio returns that are weighted based on the similarity between the focal county and peer firms' counties with respect to four county characteristics: population density, education, political inclination, and income.[29] In column 6, we report the Fama-MacBeth regression that includes this control and finds that *SPFRET* still exhibits significant return predictability. In contrast, *SIMRET* does not exhibit a significant predictive coefficient, suggesting that our results are not driven by county homophily.

Additional Results We also explore a battery of alternative specifications and show that our main results remain robust. First, to ensure that our results are not driven by geographic proximity, we have excluded industry peer firms that are located in the focal firm's headquarter state in constructing *SPFRET*. In column 7, we further exclude peer firms that are located within a 100-mile radius of the focal firm's headquarters to construct *SPFRET*. Column 8 then examines a a ZIP code–based SCI measure, which defines geographic regions more granularly. To address the concern that the coefficient of *SPFRET* may be influenced by its correlations with these other control variables, column 9 the orthogonalized *SPFRET* as the main independent variable, where we define $SPFRET_{\perp}$ as the regression residual of $SPFRET$ on all the control variables. In column 10, we use an alternative standardization procedure in which independent variables are demeaned and

[29] For each county, we form a four-dimensional vector based on these four dimensions. Each dimension is normalized to a value between 0 and 1 following Gu et al. (2020). For a pair of counties, we calculate the Euclidean distance based on their county characteristic vectors. When calculating *SIMRET*, we weigh out-of-state industry peer firm returns based on the inverse Euclidean distance.

then divided by the variable's standard deviation and find the results to be robust.[30] In column 11, we use a panel regression instead of the Fama-MacBeth regression specification, where we include time fixed effects. The coefficient of *SPFRET* is highly significant, and the magnitude is also large.

We also delve deeper into the construction of *SPFRET* and evaluate the predictive power of two alternative variables. Our main measure *SPFRET* excludes returns of same-state peers, alleviating the concern that our results are driven by geographic proximity (e.g., Parsons et al., 2020). To directly account for the returns of same-state industry peers, we define $SPFRET_{STATE}$ as the SCI-weighted returns of same-state industry peers as the focal firm. Appendix Table A.7, column 1, presents the results. We find that the coefficient of *SPFRET* remains highly significant and comparable to the corresponding coefficients in Table **6**. The coefficient on $SPFRET_{STATE}$ is also positive and significant, although somewhat smaller than *SPFRET*. This result further confirms that our results are not driven solely by the return predictability of geographically proximate industry firms.

Similarly, to examine whether returns from socially connected firms from other industries exhibit predictive power, we define *NPFRET* as the SCI-weighted returns of firms from other industries. Appendix Table A.7, column 2 presents the results and shows that *SPFRET* remains robust, whereas *NPFRET* is insignificant. This suggests that the power of *SPFRET* mostly comes from industry-based fundamental linkages between firms.

7 Conclusion

The premise of this study is that when individuals interact, their views at least partially converge. As a result, if the managers of two firms interact with an overlapping circle of acquaintances, the strategies of the two firms are likely to be more similar. We present evidence that this is indeed the case. Specifically, within industry groups, firm strategies, across three different dimensions, are more similar if the firms are located in counties where Facebook's Social Connectedness index is high. Consistent with these firms having similar strategies, we show that the fundamentals of these connected firms tend to move together, and the monthly returns of socially connected firms tend to be more highly correlated than the returns of a typical pair of firms within an industry.

[30] Note that the coefficient in this column is not directly comparable to those in the other columns due to the scale differences.

The second part of this study asks whether the U.S. stock market efficiently accounts for the co-movement of socially connected firms. Given the novelty of our information about social connectiveness, we suspect that investors may not appropriately account for this channel of fundamental correlation. In particular, we might expect industry peers in socially connected locations to under react to the information conveyed by each other's stock returns.

Our evidence supports this limited attention hypothesis. The historical excess returns generated by buying (selling) stocks in the month following high (low) returns of social connected industry peers is quite significant. Consistent with investor inattention generating sluggish price adjustments, the results are stronger for firms with low visibility (measured by market capitalization, institutional ownership, or low analyst coverage), as well as for firms located outside of industry clusters. The predictability generated by socially connected industry peer returns lasts for up to one year and does not reverse in the long run. In addition, social peer firm returns help predict focal firms' future earnings, analyst forecast errors, and future earnings announcement returns.

Our findings raise a number of issues that suggest future avenues of continuing research. Most importantly, it illustrates that the nature of the individuals that make up a firm's organization has an important influence on its stock return patterns. Given this evidence, we cannot simply view firms as bundles of assets and technologies - the people matter as well. Indeed, one interpretation of the strong evidence of predictability after 2000, when the profitability of most lead-lag strategies waned, is that the quality of the people employed has, over time, become a more important determinant of the long-term success of firms. When the nature of the people in an organization becomes more important, the social ties between the people become more relevant.

The evidence in this paper also highlights how the sphere of a city's influence can go beyond its borders. Innovative thinking in a particular city can directly influence the behavior of firms within the city, and then migrate more broadly through the social connections of its residents. Future research should explore whether these cross-city social connections have become more important in the post-COVID environment, where individuals work from home and have technology that facilitates cross-city communication.

Given the importance of social ties, their determinants warrant future research. It is especially important to better understand if it is the communication between managers that cause their strategies to converge or whether it's just that people with similar views and beliefs tend to communicate. If we believe that it is the communication that influences strategy, then one might want to think more carefully about the extent to which some regions are more influential than others. For example, we might explore whether or not firms in locations with greater social ties tend to be more influential.

References

Abdi, Hervé, 2010, Partial least squares regression and projection on latent structure regression (pls regression), *Wiley interdisciplinary reviews: computational statistics* 2, 97–106.

Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2012, The network origins of aggregate fluctuations, *Econometrica* 80, 1977–2016.

Ahern, Kenneth R, 2017, Information networks: Evidence from illegal insider trading tips, *Journal of Financial Economics* 125, 26–47.

Ali, Usman, and David Hirshleifer, 2020, Shared analyst coverage: Unifying momentum spillover effects, *Journal of Financial Economics* 136, 649–675.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Au, Shiu-Yik, Ming Dong, and Xinyao Zhou, 2023, Does social interaction spread fear among institutional investors? evidence from covid-19, *Management Science* forthcoming.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, 2018a, Social connectedness: Measurements, determinants, and effects, *Journal of Economic Perspectives* 32, 259–80.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2018b, The economic effects of social networks: Evidence from the housing market, *Journal of Political Economy* 126, 2224–2276.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2018c, The economic effects of social networks: Evidence from the housing market, *Journal of Political Economy* 126, 2224–2276.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel, 2019a, House price beliefs and mortgage leverage choice, *The Review of Economic Studies* 86, 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel, 2019b, Social connectedness in urban areas, Technical Report w26029, NBER Working Paper.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel, 2021, International trade and social connectedness, *Journal of International Economics* 129, 103418.

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, 2019c, Peer effects in product adoption, Technical report, National Bureau of Economic Research.

Bali, Turan, David Hirshleifer, Lin Peng, and Yi Tang, 2019, Attention, social interaction, and investor attraction to lottery stocks, Technical report, Working Paper.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *The Journal of Financial Economics* 99, 427–446.

Bali, Turan G., and Armen Hovakimian, 2009, Volatility spreads and expected returns, *Management Science* 55, 1797–1812.

Bernard, Victor L, and Jacob K Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305–340.

Bernile, Gennaro, Alok Kumar, and Johan Sulaeman, 2015, Home away from home: Geography of information and local investors, *The Review of Financial Studies* 28, 2009–2049.

Beshears, John, James J Choi, David Laibson, Brigitte C Madrian, and Katherine L Milkman, 2015, The effect of providing peer information on retirement savings decisions, *The Journal of finance* 70, 1161–1201.

Breschi, Stefano, and Camilla Lenzi, 2016, Co-invention networks and inventive productivity in us cities, *Journal of Urban Economics* 92, 66–75.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt, 2022, Social capital I: Measurement and associations with economic mobility, *Nature* 608, 108–121.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *The Journal of Finance* 63, 1977–2011.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, 2008, The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116, 951–979.

Cohen, Lauren, Umit G Gurun, and Christopher Malloy, 2017, Resident networks and corporate connections: Evidence from World War II internment camps, *The Journal of Finance* 72, 207–248.

Cohen, Lauren, and Dong Lou, 2012, Complicated firms, *Journal of Financial Economics* 104, 383–400.

DellaVigna, Stefano, and Joshua Pollet, 2009, Investor inattention and Friday earnings announcements, *Journal of Finance* 64, 709–749.

Duggan, Maeve, Shannon Greenwood, and Andrew Perrin, 2016, Social media update 2016. Pew Research Center.

Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence, 2017, The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation, *Journal of Accounting and Economics* 64, 221–245.

Engelberg, Joseph, Arzu Ozoguz, and Sean Wang, 2018, Know thy neighbor: Industry clusters, information spillovers, and market efficiency, *Journal of Financial and Quantitative Analysis* 1937–1961.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics non-parametrically, *The Review of Financial Studies* 33, 2326–2377.

Garcia, Diego, and Øyvind Norli, 2012, Geographic dispersion and stock returns, *Journal of Financial Economics* 106, 547–565.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Harvey, Campbell R, 2017, Presidential address: The scientific outlook in financial economics, *The Journal of Finance* 72, 1399–1440.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Heimer, Rawley Z, 2016, Peer pressure: Social interaction and the disposition effect, *The Review of Financial Studies* 29, 3177–3209.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.

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

Hong, Harrison, and Jeremy C Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of Finance* 54, 2143–2184.

Hong, Harrison, and Jiangmin Xu, 2019, Inferring latent social networks from stock holdings, *Journal of Financial Economics* 131, 323–344.

Hvide, Hans K., and Östberg, 2015, Social interaction at work, *Journal of Financial Economics* 117, 628–652.

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

Jensen, Theis Ingerslev, Bryan T. Kelly, and Lasse Heje Pedersen, 2021, Is there a replication crisis in finance?, Technical report, National Bureau of Economic Research.

Jin, Zuben, and Frank Weikai Li, 2020, Geographic links and predictable returns, *Available at SSRN 3617417* .

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kothari, S.P., Jonathan Lewellen, and Jerold B. Warner, 2006, Stock returns, aggregate earnings surprises, and behavioral finance, *Journal of Financial Economics* 79, 537–568.

Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou, 2021, Social proximity to capital: Implications for investors and firms, *The Review of Financial Studies* .

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel, 2020, The geographic spread of Covid-19 correlates with structure of social networks as measured by Facebook, Technical report, National Bureau of Economic Research.

Lee, Charles MC, Stephen Teng Sun, Rongfei Wang, and Ran Zhang, 2019, Technological links and predictable returns, *Journal of Financial Economics* 132, 76–96.

Lou, Dong, and Christopher Polk, 2022, Comomentum: Inferring arbitrage activity from return correlations, *The Review of Financial Studies* 35, 3272–3302.

Loughran, Tim, and Jay W. Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Finance* 46, 1629–1650.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

McNerney, James, Charles Savoie, Francesco Caravelli, Vasco M. Carvalho, and J. Doyne Farmer, 2022, How production networks amplify economic growth, *Proceedings of the National Academy of Sciences* 119, e2106031118.

Mendenhall, Richard R, 1991, Evidence on the possible underweighting of earnings-related information, *Journal of Accounting Research* 29, 170–179.

Moskowitz, Tobias J, and Mark Grinblatt, 1999, Do industries explain momentum?, *The Journal of Finance* 54, 1249–1290.

Newey, Whitney K, and Kenneth D West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Parsons, Christopher A, Riccardo Sabbatucci, and Sheridan Titman, 2020, Geographic lead-lag effects, *The Review of Financial Studies* 33, 4721–4770.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Pirinsky, Christo, and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *The Journal of Finance* 61, 1991–2015.

Pool, Veronika K., Noah Stoffman, and Scott E. Yonker, 2012, No Place Like Home: Familiarity in Mutual Fund Manager Portfolio Choice, *The Review of Financial Studies* 25, 2563–2599.

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

Shiller, Robert J., 2010, *Irrational Exuberance* (Princeton University Press, Princeton, N.J.).

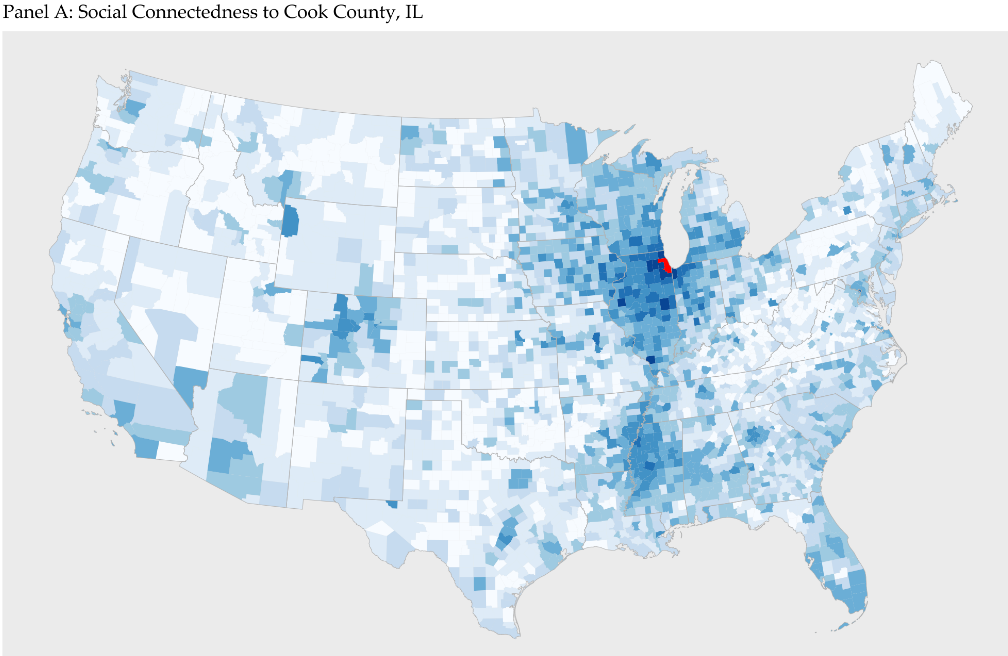
Shue, Kelly, 2013, Executive networks and firm policies: Evidence from the random assignment of mba peers, *The Review of Financial Studies* 26, 1401–1442.

Shumway, Tyler, 1997, The delisting bias in crsp data, *Journal of Finance* 52, 327–340.

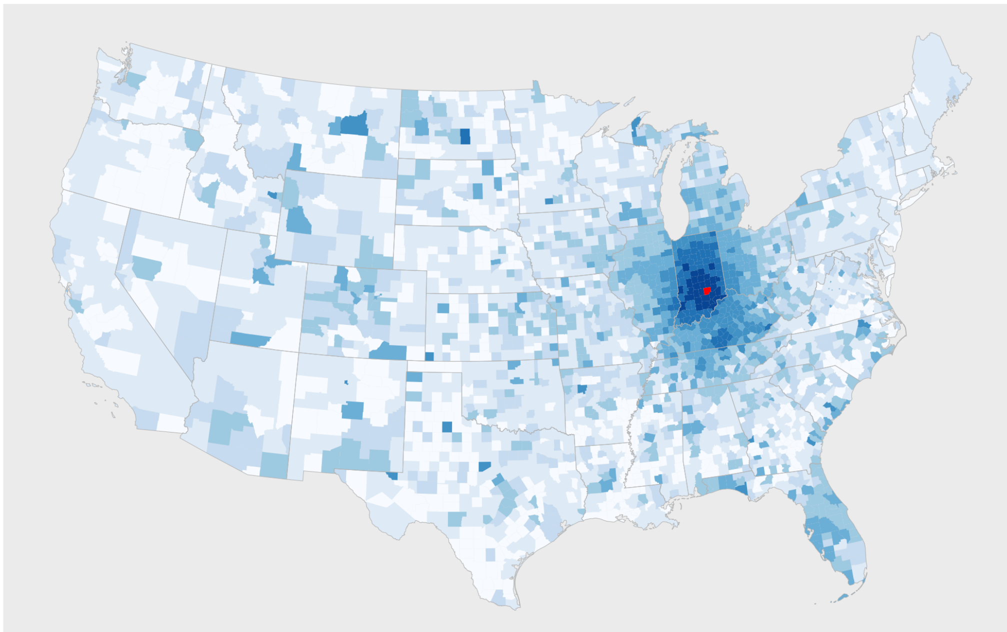
Figure 1: Examples of Social Connectedness

This figure shows county-level heat maps of the social connectedness to Cook County, IL, in Panel A, and Bartholomew County, IN, in Panel B. The focal counties are in red and darker colors indicate higher social connectedness to the focal counties. Panel C presents CMI's headquarters county (Bartholomew County, IN), and the locations of CMI's industry peers, with dark blue indicating higher social connectedness to Bartholomew County. Examples of the industry peer firms include Caterpillar (ticker: CAT), Clarcor (ticker: CLC), and Stanley Black & Decker (ticker: SWK), located in Peoria, IL, Williamson, TN, and Hartford, CT, respectively. Counties without any peer firms' presence are presented in dark grey.

Panel A: Social Connectedness to Cook County, IL



Panel B: Social Connectedness to Bartholomew County, IN



Panel C: Social Connectedness to Bartholomew County, IN, and the Presence of "Machinery" Firms

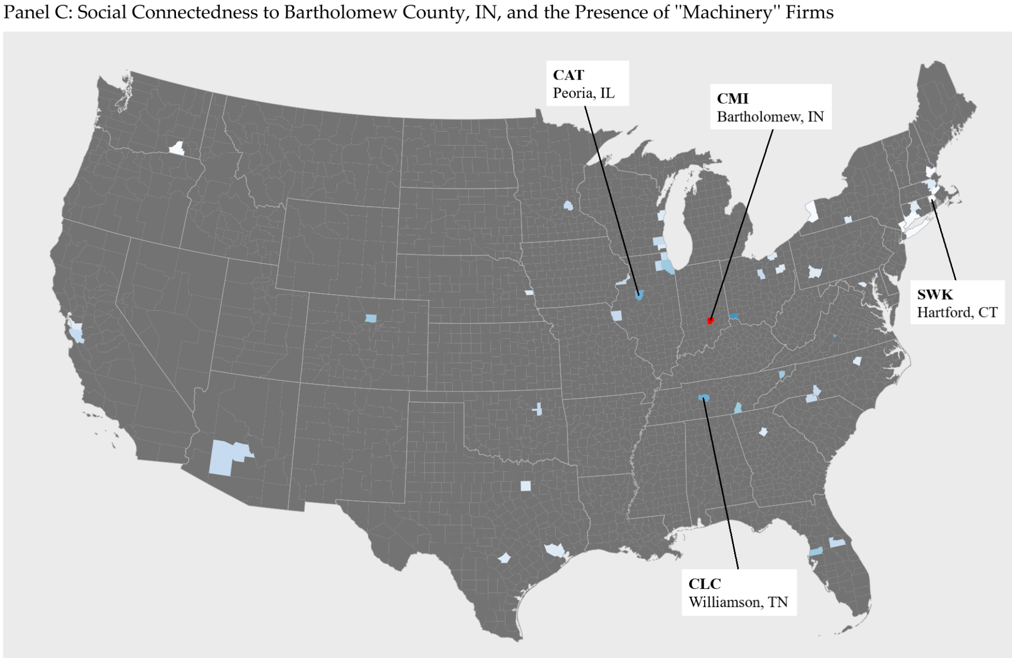


Figure 2: *SPFRET* Strategy Returns

The figure depicts the monthly Fama-French five-factor (FF5)-adjusted abnormal returns to a long-short (LS) portfolio based on *SPFRET* (Panel A) and $SPFRET_{\perp}$ (Panel B). The reported returns are obtained as the intercept plus the residuals from a regression of monthly returns on the FF5 factor model and averaged for each year. *SPFRET* is the SCI-weighted average return of stocks from the same Fama-French 48-industry as the focal stock but from a different state. $SPFRET_{\perp}$ is the abnormal return generated from a panel regression of *SPFRET* on the equal-weighted industry return, using all available stock-month observations prior to the portfolio formation month (i.e., month *t*) and month fixed effects. All returns are in percentages.

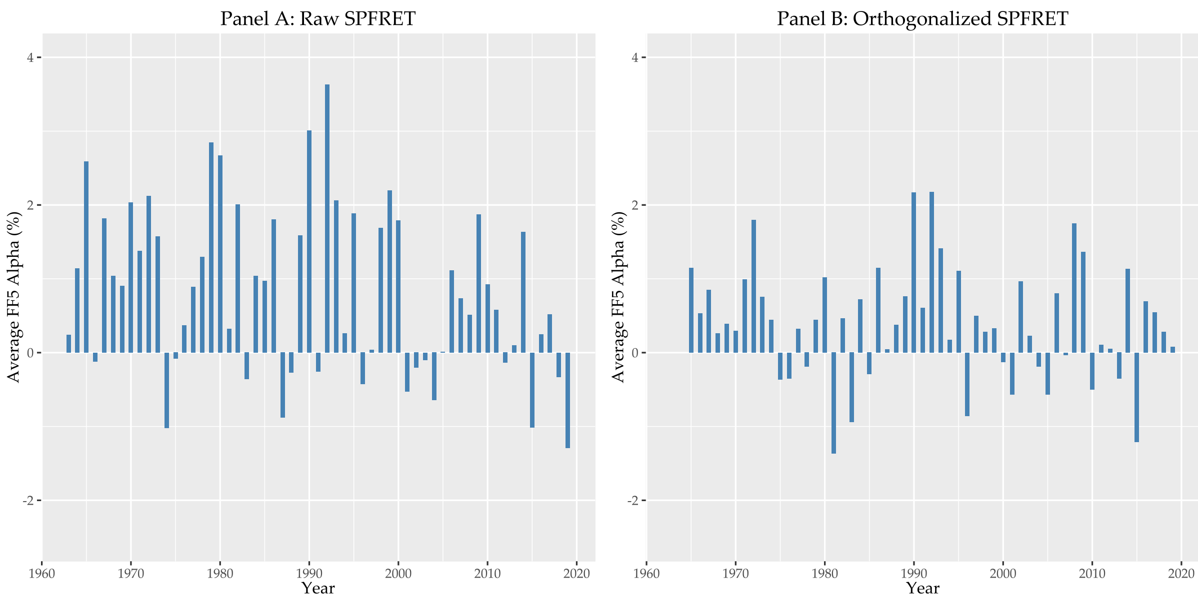


Figure 3: Lead-lag Strategy Performance (1989–2019)

The figure shows the Fama-French five-factor portfolio alphas of long-short strategies based on the lead-lag predictors (*SPFRET*, *INDRET*, *GEORET*, *CFRET*, *CRET*, *TECHRET*, *TNICRET*, and *CONGRET*). The orange bars show the 90% confidence intervals. The portfolios are value-weighted. *SPFRET* refers to social peer firm returns. *INDRET* is the industry return. *CRET* is the customer return. *GEORET* is the geographic momentum. *CFRET* refers to shared analyst coverage returns. *TECHRET* is the technology-linked firm return. *TNICRET* is the text-based industry return. *CONGRET* is the pseudo-conglomerate return. The sample period runs from July 1989 to November 2019, due to the availability of *TNICRET*.

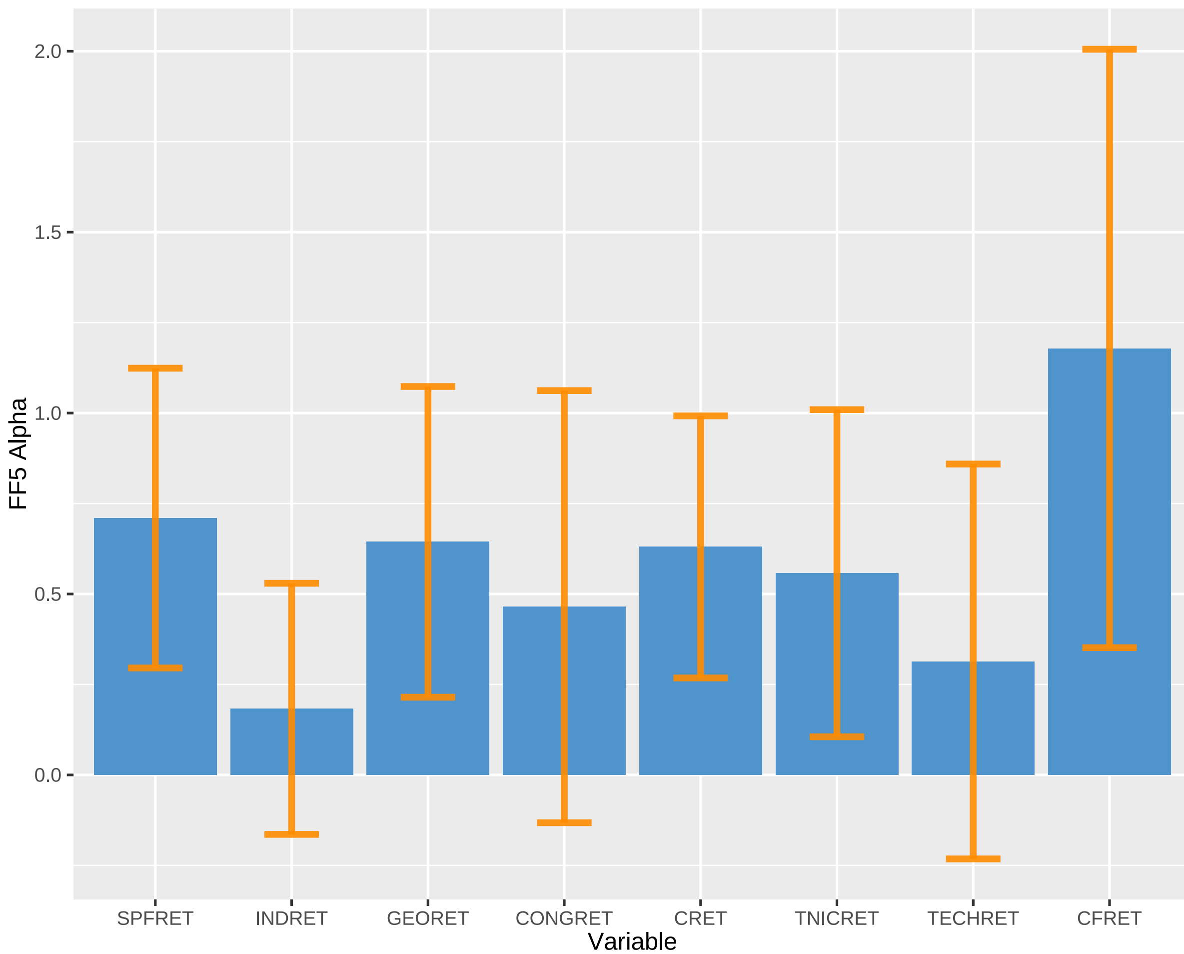


Figure 4: Decomposing *SPFRET* Portfolio Returns

The figure decomposes the average return of the *SPFRET*-sorted value-weighted long-short portfolio into its various components. Each bar shows the amount of the average return that is explained by the variable(s) indicated below it. *INDRET* represents the contribution of *INDRET*, the equal-weighted Fama-French 48-industry return. *GEORET* represents the contribution of geographic momentum. *CONGRET* represents the contribution from the complicated firm effect. *CRET* represents the contribution from the customer return. *TNICRET* represents the contribution of the text-based industry return of Hoberg and Phillips (2018). *TECHRET* represents the contribution of the technological spillover effect. *CFRET* represents the contribution of the analyst-linked firm return. *STRATRET* represents the contribution of the strategy similarity returns. *FF5* represents the combined contribution of the Fama-French five factors. The last bar shows the alpha, the portion of the average return not explained by variables shown in the plot. Red indicates negative values. The sample period is July 1994 to November 2019, due to the availability of *STRATRET*.

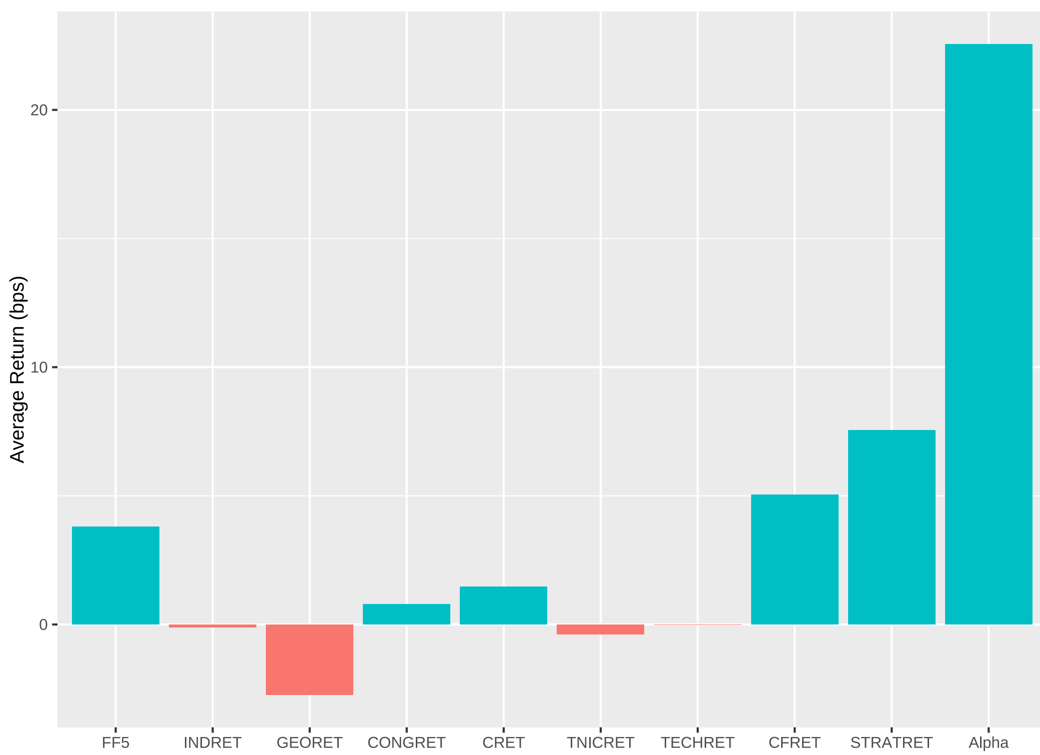


Figure 5: Relative Variable Importance

The figure depicts the relative importance of lead-lag predictors (*SPFRET*, *INDRET*, *GEORET*, *CFRET*, *CRET*, *TECHRET*, *TNICRET*, and *CONGRET*) in a partial least squares (PLS) model that uses these predictors along with all the other controls (i.e., *INDMOM*, *RET*, *MOM*, *SIZE*, *BM*, *BMKT*, *ILLIQ*, *SKEW*, *COSKEW* and *MAX*). Starting from July 2000, a new PLS model is trained each month using all the data available up to that point. For each training set, the difference between the full-model R^2 and the R^2 is obtained by dropping one of the eight lead-lag predictors while keeping the coefficient estimates for the rest of the predictors fixed. These marginal changes in R^2 are then normalized so that their sum equals one. The relative variable importance in each panel is then calculated by taking the average over the normalized values over all the training sets. *CFRET* refers to shared analyst coverage returns. *SPFRET* refers to social peer firm returns. *INDRET* is the industry return. *CRET* is the customer return. *GEORET* is the geographic momentum. *TECHRET* is the technology-linked firm return. *TNICRET* is the text-based industry return. *CONGRET* is the pseudo-conglomerate return.

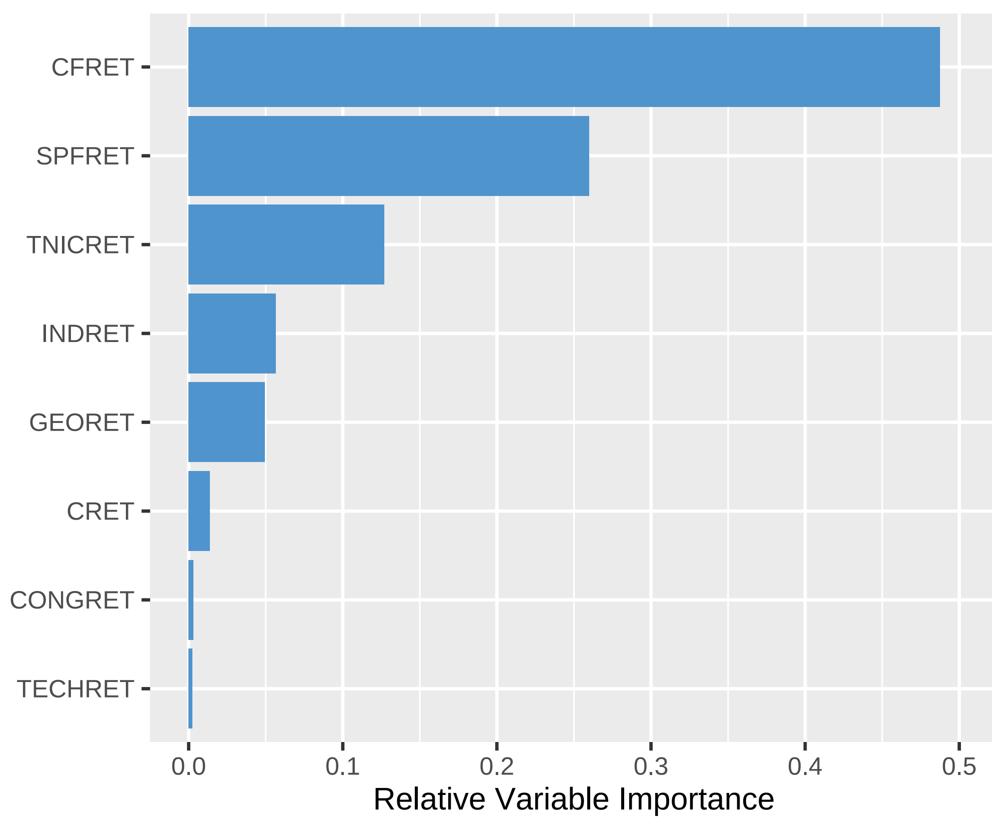


Figure 6: PLS-Based Long-Short Portfolio Returns

The figure shows the cumulative log returns of decile-10 minus decile-1 long-short portfolios that are sorted based on the out-of-sample PLS predictor with and without *SPFRET* from August 2000 through December 2019. Portfolios are value-weighted and exclude the largest 100 stocks. Shaded areas indicate the NBER recession periods.

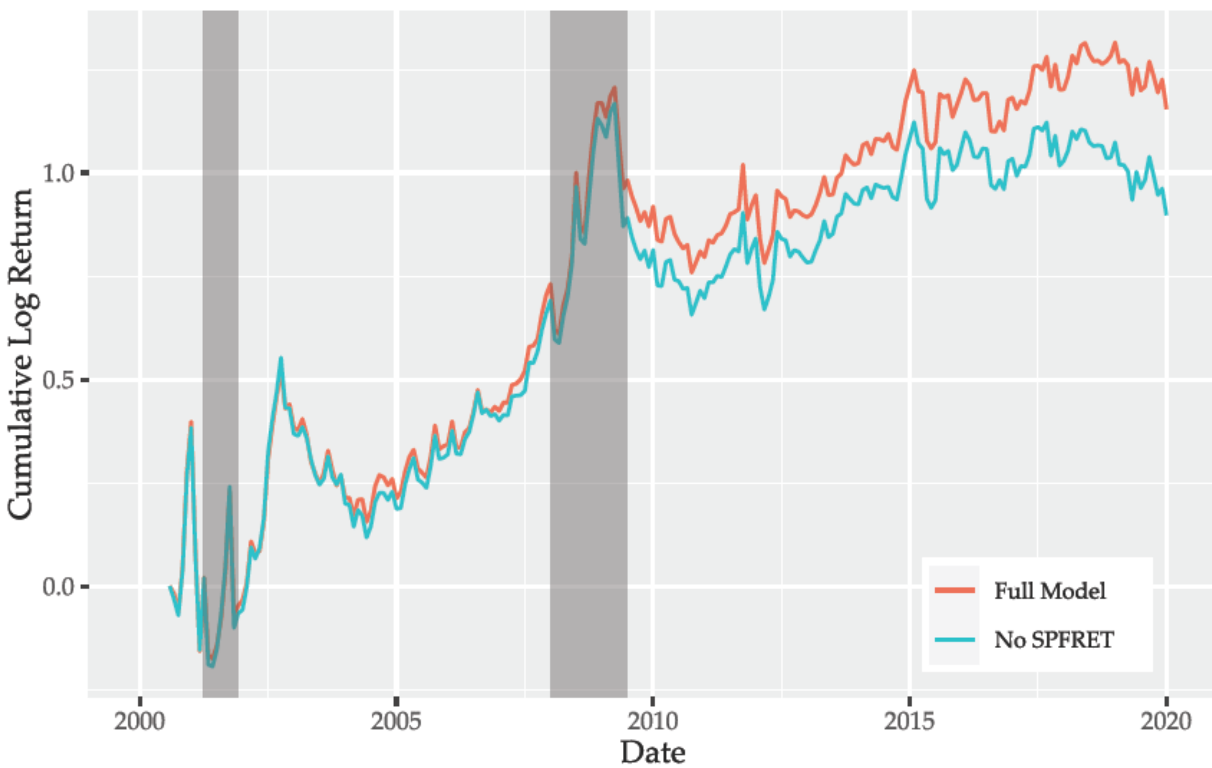


Table 1: Summary Statistics and Correlations

The table provides summary statistics and correlation tables for the pairwise similarity measures used in our analyses. SCI is the social connectedness between the firms' headquarters locations. *STRATSIM* is the 10-K based strategy similarity between the two firms. $STRATSIM^{PROD}$ and $STRATSIM^{TECH}$ are strategy similarity measures based on product and technology similarities, respectively. *ANALYST* and *CSTMR* are firm similarities based on the shared analysts and shared customers. *Same County* is an indicator that equals one if the two firms are headquartered in the same county, and zero otherwise.

Panel A: Summary Statistics of Similarity Measures

	N	Mean	Std. Dev.	Min.	25%	75%	Max.
SCI	632,427	0.000	0.003	0.000	0.000	0.000	0.819
STRATSIM	196,011	0.080	0.067	0.000	0.031	0.114	0.940
STRATSIM^{PROD}	207,367	0.091	0.064	0.000	0.040	0.129	0.815
STRATSIM^{TECH}	43,083	0.325	0.302	0.000	0.065	0.539	1.000
ANALYST	231,698	0.019	0.032	0.002	0.011	0.025	0.989
CSTMR	27,836	0.016	0.122	0.000	0.000	0.000	1.000
Same County	632,427	0.025	0.154	0.000	0.000	0.000	1.000

Panel B: Correlations of Similarity Measures

	SCI	STRATSIM	STRATSIM^{PROD}	STRATSIM^{TECH}	ANALYST	CSTMR
STRATSIM	0.065					
STRATSIM^{PROD}	0.048	0.133				
STRATSIM^{TECH}	0.066	0.036	0.135			
ANALYST	0.018	0.013	0.040	0.031		
CSTMR	-0.001	0.013	0.035	0.042	0.028	
Same County	0.232	0.052	0.068	0.043	0.008	0.010

Table 2: Strategy Similarity and Social Connectedness

The table reports the results of the panel regressions of strategy similarities between firm pairs on the SCI between the firm headquarters counties. *STRATSIM* is the cosine similarity based on the overlap of strategy-related words in the 10-K's. $STRATSIM^{TECH}$ is the similarity based on the fraction of patent filings in each technology category (Lee et al., 2019). $STRATSIM^{PROD}$ is the cosine similarity based on the overlap of product-related words in the 10-K's (Hoberg and Phillips, 2016, 2018). *Same County* is a dummy variable indicating whether the two firms are headquartered in the same county. Similarities are calculated only for firms that are in the same Fama-French 48 industry. The sample period is 1993-2019 for *STRATSIM* and $STRATSIM^{PROD}$ regressions, while it is 1993-2010 for the $STRATSIM^{TECH}$ regression. All regressions include year, industry, and county fixed effects. Standard errors are clustered on both year and industry levels and the corresponding *t*-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Univariate Regressions

	STRATSIM	$STRATSIM^{TECH}$	$STRATSIM^{PROD}$
	(1)	(2)	(3)
SCI	0.046***	0.047***	0.049***
	(6.368)	(2.844)	(12.466)
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Observations	10,415,430	1,505,475	11,052,778
R^2	0.092	0.059	0.147

Panel B: Controlling for Same County

	STRATSIM	$STRATSIM^{TECH}$	$STRATSIM^{PROD}$
	(1)	(2)	(3)
SCI	0.044***	0.040***	0.047***
	(6.005)	(3.112)	(10.251)
Same County	0.017***	0.039**	0.039***
	(3.647)	(2.432)	(5.388)
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Observations	10,415,430	1,505,475	11,052,778
R^2	0.092	0.059	0.147

Table 3: Comovements in Firm Fundamentals and Stock Returns

The table presents the comovements between the focal firm and its industry peers using panel regression analysis. In columns 1-4, the dependent variables are firms' fundamentals, measured annually: ΔEPS is the change in EPS scaled by lagged stock price, $\Delta Sales$ is the percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value. Only the firms with the same fiscal year-end as the focal firm are included. In column 5, the dependent variable is monthly stock returns. In Panel A, the independent variable is the corresponding fundamental of the contemporaneous SCI-weighted industry portfolio (excluding firms from the same state), while the independent variables in Panel B and C also include the contemporaneous equal-weighted and value-weighed industry portfolios (excluding the focal firm) respectively. We follow Gu et al. (2020) and cross-sectionally rank the independent variables and scale them into the [0,1] interval. All regressions include time fixed effects. All dependent variables are scaled up by 100. Standard errors are clustered by both time and firm and the corresponding t-statistics are reported in parentheses. For all panels, *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Univariate Analysis of SCI-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
SCI-Weighted	2.445***	30.848***	13.374***	35.305***	7.912***
	(6.024)	(9.735)	(10.031)	(11.522)	(27.047)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,035	124,232	121,090	121,124	1,711,696
R^2	0.018	0.051	0.047	0.041	0.147

Panel B: Multivariate Analysis with Equal-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
SCI-Weighted	3.463***	26.496***	13.526***	36.495***	7.718***
	(7.260)	(7.096)	(10.528)	(8.409)	(23.620)
Equal-Weighted	−1.283**	5.108**	−0.183	−1.371	0.226
	(−2.517)	(2.387)	(−0.179)	(−0.469)	(1.284)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,035	124,232	121,090	121,124	1,711,696
R^2	0.018	0.052	0.047	0.041	0.147

Panel C: Multivariate Analysis with Value-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
SCI-Weighted	2.426***	29.034***	14.139***	34.255***	7.201***
	(6.255)	(10.831)	(10.477)	(10.710)	(25.469)
Value-Weighted	0.148	0.497	−0.949	3.279	0.923***
	(0.581)	(0.435)	(−1.577)	(1.420)	(9.322)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	114,931	120,009	119,003	120,140	1,664,695
R^2	0.018	0.048	0.047	0.044	0.149

Table 4: Return Predictability of Social Peer Firm Returns: Portfolio Analysis

The table reports the results of the univariate and bivariate portfolio sort based on *SPFRET*, the SCI-weighted average return of a firm's industry peers (excluding those from the same state). In Panel A, for each month, we sort all common stocks into deciles based on *SPFRET* and calculate both the equal-weighted and value-weighted one-month-ahead returns for the decile portfolios, as well as the return of the portfolio that long the decile-10 portfolio and short the decile-1 portfolio. We present the raw returns and the FF5 alphas for the equal- and value-weighted portfolios respectively. In Panel B, we first sort stocks into quintiles based on *INDRET*, and then, within each *INDRET* quintile, we further sort stocks into quintiles based on *SPFRET*. We then report the one-month-ahead FF5 alphas for the 25 equal-weighted and value-weighted portfolios and the portfolios with long positions in stocks in the *SPFRET* quintile 5 and short positions in the *SPFRET* quintile 1 stocks. Final rows in both panels report the average alphas for the *SPFRET* quintiles. All returns are in percentages. The corresponding t-statistics based on Newey and West (1994) standard errors are reported in parentheses below the alphas. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Univariate Sort

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.277	0.534**	0.799***	0.970***	1.091***	1.219***	1.277***	1.438***	1.630***	1.766***	1.489***
	(1.146)	(2.473)	(3.967)	(4.436)	(4.933)	(5.838)	(5.898)	(6.832)	(7.131)	(7.233)	(8.569)
FF5 Alpha EW	−0.846***	−0.525***	−0.234**	−0.122	−0.036	0.083	0.142**	0.358***	0.610***	0.723***	1.568***
	(−7.043)	(−4.233)	(−2.050)	(−1.403)	(−0.596)	(1.481)	(2.304)	(5.169)	(6.341)	(6.135)	(7.058)
Raw Return VW	0.443**	0.655***	0.935***	0.834***	1.103***	1.032***	1.058***	1.168***	1.098***	1.208***	0.765***
	(2.122)	(3.360)	(5.049)	(4.410)	(5.771)	(5.533)	(5.650)	(7.160)	(5.275)	(6.188)	(4.781)
FF5 Alpha VW	−0.554***	−0.251**	0.007	−0.065	0.146	0.007	0.057	0.199***	0.255**	0.286***	0.840***
	(−5.516)	(−2.546)	(0.067)	(−0.711)	(1.415)	(0.082)	(0.705)	(2.706)	(1.985)	(3.177)	(5.216)

Panel B: Bivariate Sort

	SPFRET					
	Equal-Weighted			Value-Weighted		
	(1 Low)	(5 High)	(5-1)	(1 Low)	(5 High)	(5-1)
1 (Low INDRET)	-1.113^{***}	-0.537^{***}	0.576^{***}	-0.725^{***}	-0.240	0.485^{***}
	(-7.309)	(-3.773)	(4.184)	(-5.289)	(-1.457)	(2.794)
2	-0.572^{***}	0.088	0.660^{***}	-0.618^{***}	0.146	0.764^{***}
	(-4.813)	(0.908)	(5.043)	(-5.214)	(1.218)	(4.875)
3	-0.250^{***}	0.218^{**}	0.468^{***}	-0.057	0.111	0.168
	(-2.684)	(2.345)	(4.517)	(-0.529)	(0.938)	(1.212)
4	-0.010	0.472^{***}	0.482^{***}	0.004	0.186^{*}	0.182
	(-0.126)	(4.249)	(3.318)	(0.042)	(1.648)	(1.108)
5 (High INDRET)	0.373^{***}	1.002^{***}	0.630^{***}	0.108	0.440^{***}	0.332^{**}
	(3.446)	(6.219)	(4.599)	(0.875)	(3.487)	(2.082)
Average	-0.314^{***}	0.249^{***}	0.563^{***}	-0.258^{***}	0.129^{**}	0.386^{***}
	(-5.334)	(5.278)	(7.384)	(-4.931)	(2.185)	(4.844)

Table 5: Understanding *SPFRET* Portfolio Alphas: Spanning Regressions

The table presents the results of spanning regressions where the value-weighted long-short (LS) portfolio return of *SPFRET* is regressed against the value-weighted long-shortportfolio returns obtained using other variables that capture economic linkages between firms. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). The other variables that are used to form long-shortportfolios include industry momentum (*INDRET*), geographic momentum (*GEORET*), customer return (*CRET*), the technology-linked firm return (*TECHRET*), the pseudo-conglomerate return (*CONGRET*), the text-based industry return (*TNICRET*), the technology-linked firm return (*TECHRET*), the analyst-linked firm return (*CFRET*), and the strategy similarity based peer return (*STRATPRET*). We control for the Fama-French five-factor model but do not report those coefficients for brevity. The sample period is July 1994 to November 2019, due to the availability of *STRATRET*. All returns are reported in percentages. Newey-West heteroskedasticity and autocorrelation-corrected *t*-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	SPFRET								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.416*	0.436***	0.412**	0.381**	0.362**	0.342**	0.366**	0.308*	0.226
	(1.754)	(2.689)	(2.559)	(2.346)	(2.247)	(2.076)	(2.353)	(1.818)	(1.384)
INDRET		0.775***	0.763***	0.679***	0.670***	0.607***	0.582***	0.540***	0.407***
		(13.998)	(13.831)	(12.032)	(12.365)	(9.508)	(8.804)	(7.674)	(4.835)
GEORET			0.034	0.023	0.022	-0.015	-0.033	-0.069	-0.048
			(0.551)	(0.409)	(0.387)	(-0.255)	(-0.629)	(-1.248)	(-0.984)
CONGRET				0.145***	0.135***	0.110***	0.096**	0.072*	0.062**
				(4.190)	(3.406)	(2.885)	(2.476)	(1.925)	(2.093)
CRET					0.043	0.033	0.021	0.034	0.038
					(0.908)	(0.648)	(0.387)	(0.603)	(0.633)
TNICRET						0.129*	0.109	0.036	-0.011
						(1.666)	(1.546)	(0.494)	(-0.149)
TECHRET							0.083	0.063	0.007
							(1.330)	(1.086)	(0.147)
CFRET								0.129**	0.086**
								(2.466)	(1.971)
STRATRET									0.312***
									(7.112)
Observations	305	305	305	305	305	305	305	305	305
R^2	0.013	0.709	0.710	0.730	0.732	0.737	0.741	0.750	0.786

Table 6: *SPFRET* and Cross-sectional Return Predictions

The table reports the results of Fama-MacBeth regressions where stock returns are regressed on lagged social peer firm returns. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the focal firms' headquarters states). *INDRET* is the short-term industry momentum. *GEORET* is the geographic momentum, *CFRET* is the analyst-connected firm return, *CRET* is the customer return, *TECHRET* is the technology-linked firm return. *CONGRET* is the pseudo-conglomerate return. *TNICRET* is the text-based industry return. *TECHRET* is the technology-linked firm return. *STRATRET* is the strategy similarity-based peer return. We also include the following control variables: *RET*, *SIZE*, *BMKT*, *BM*, *MOM*, *IVOL*, *ILLIQ*, *MAX*, *SKEW*, and *COSKEW*. Section 2 provides detailed descriptions. The sample period is July 1994 to November 2019. All returns are reported in percentages. Missing values of independent variables are imputed with the monthly medians. In order to standardize the independent variables, we cross-sectionally rank them and map the rankings into the $[0, 1]$ interval (Gu et al. (2020)). t-statistics are computed with Newey and West (1994) standard errors and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

					RET_{t+1}				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SPFRET	1.276***	0.710***	0.703***	0.697***	0.686***	0.601***	0.602***	0.387***	0.431***
	(5.222)	(3.808)	(3.737)	(3.710)	(3.732)	(3.637)	(3.794)	(2.689)	(2.900)
INDRET		0.829***	0.830***	0.834***	0.827***	0.692***	0.695***	0.514***	0.535***
		(3.873)	(3.914)	(3.884)	(3.837)	(3.530)	(3.555)	(2.807)	(2.932)
GEORET			0.304***	0.305***	0.297***	0.268**	0.271**	0.230**	0.229**
			(2.706)	(2.710)	(2.657)	(2.477)	(2.574)	(2.215)	(2.212)
CONGRET				−0.028	−0.037	−0.083	−0.076	−0.199	−0.182
				(−0.232)	(−0.312)	(−0.682)	(−0.598)	(−1.526)	(−1.383)
CRET					0.433***	0.385***	0.398***	0.314***	0.312***
					(3.327)	(3.075)	(3.250)	(2.633)	(2.641)
TNICRET						0.687***	0.692***	0.474***	0.473***
						(3.801)	(3.878)	(3.328)	(3.311)
TECHRET							−0.077	−0.229	−0.218
							(−0.437)	(−1.297)	(−1.219)
CFRET								1.187***	1.200***
								(6.002)	(6.094)
STRATRET									−0.187
									(−1.623)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	305	305	305	305	305	305	305	305	305
# Stocks	3,689	3,689	3,689	3,689	3,689	3,689	3,689	3,689	3,689
R^2	0.059	0.061	0.062	0.062	0.063	0.064	0.065	0.067	0.068

Table 7: Information Environment and Return Predictability

The table examines how firm characteristics affect the return predictability of social peer firm returns by using bivariate portfolio sorting. In Panel A, each month, we first sort firms into deciles based on size, institutional ownership, and analyst coverage. In Panel B, we first sort firms based on whether they are headquartered in an industry cluster. For a given Fama-French 48 industry and month, we measure the total industry market capitalization at the county level and define the top 20% (columns 1 to 3) or top 10% (columns 4 and 6) of counties as the industry cluster. After sorting based on firms' characteristics, we further sort firms based on *SPFRET*. We report the value-weighted FF5 alphas for decile 1, decile 10, and 10-1 long-short portfolios. All returns are reported in percentages. We report *t*-statistics based on Newey-West adjusted standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Size, Institutional Ownership, and Analyst Coverage

	Size			Inst. Own.			Analyst Cov.		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Low	−0.930***	0.882***	1.811***	−1.037***	0.360	1.397***	−0.635***	0.582**	1.218***
	(−6.035)	(5.776)	(6.815)	(−3.593)	(1.259)	(3.727)	(−3.448)	(2.382)	(3.827)
High	−0.340***	0.231*	0.571***	−0.674***	0.086	0.760***	−0.418***	0.252	0.670***
	(−3.404)	(1.878)	(3.773)	(−2.922)	(0.463)	(2.649)	(−2.851)	(1.571)	(2.836)

Panel B: Out-Cluster and In-Cluster

	Top 20%			Top 10%		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Out-Cluster	−0.732***	0.531***	1.263***	−0.734***	0.417***	1.150***
	(−5.862)	(4.516)	(6.150)	(−6.684)	(3.446)	(5.806)
In-Cluster	−0.437***	0.209**	0.646***	−0.405***	0.263**	0.669***
	(−4.119)	(2.130)	(4.011)	(−3.436)	(2.164)	(3.599)

Table 8: Return Predictability in the Long Run

The table reports the long-run performance of calendar-time portfolios sorted by social peer firms' returns. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). *SPFMOM* is the cumulative *SPFRET* over months $t-11$ through $t-1$. Panel A (B) reports the FF5 alpha of the equal-weighted (value-weighted) portfolio return for deciles 1 and 10, and the 10-1 return differences for the following months relative to month t: 1–3, 4–6, 7–12, and 13–24. t-statistics based on Newey-West standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: FF5 Alphas (Equal-Weighted)

	SPFRET			SPFMOM		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Month 1	−0.846***	0.722***	1.568***	−0.340**	0.354***	0.694***
	(−7.043)	(6.123)	(7.052)	(−2.322)	(3.048)	(2.907)
Months 2–3	−0.171	0.065	0.236	−0.315**	0.297***	0.611***
	(−1.374)	(0.720)	(1.240)	(−2.317)	(2.768)	(2.867)
Months 4–6	−0.192*	0.061	0.254*	−0.330***	0.207*	0.536***
	(−1.955)	(0.814)	(1.677)	(−2.737)	(1.934)	(2.728)
Months 7–12	−0.184**	0.105*	0.289***	−0.226***	0.048	0.274*
	(−2.302)	(1.914)	(2.893)	(−2.726)	(0.518)	(1.877)
Months 13–24	−0.054	0.023	0.077	−0.093	−0.024	0.069
	(−1.275)	(0.413)	(1.477)	(−1.282)	(−0.232)	(0.467)

Panel B: FF5 Alphas (Value-Weighted)

	SPFRET			SPFMOM		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Month 1	−0.554***	0.287***	0.840***	−0.098	0.201*	0.299
	(−5.526)	(3.162)	(5.203)	(−0.671)	(1.778)	(1.353)
Months 2–3	0.057	−0.012	−0.069	−0.163	0.243**	0.405**
	(0.486)	(−0.115)	(−0.351)	(−1.244)	(2.485)	(2.069)
Months 4–6	−0.052	−0.007	0.045	−0.230**	0.151	0.382**
	(−0.559)	(−0.090)	(0.300)	(−2.289)	(1.589)	(2.272)
Months 7–12	−0.140**	0.072	0.213**	−0.216***	0.009	0.225*
	(−1.968)	(1.355)	(2.161)	(−2.701)	(0.120)	(1.810)
Months 13–24	−0.024	0.052	0.076	−0.038	0.014	0.051
	(−0.664)	(0.924)	(1.279)	(−0.575)	(0.156)	(0.425)

Table 9: Predicting Future Earnings Surprises, Analyst Forecast Errors, and Announcement Returns

The table presents the panel regression results of using social peer firm returns to predict a firm's future earnings surprises, analyst forecast errors, and cumulative abnormal returns around earnings announcements. The dependent variables are the focal firm's standardized unexpected earnings (SUE) (columns 1 to 4), analyst forecast errors (FE) (columns 5 to 8), and the three-day abnormal returns around the earnings announcements (columns 9 to 12) for the next four quarters. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). *SPFMOM* is the cumulative *SPFRET* over months $t-11$ through $t-1$. Panel B includes short- and long-term industry momentum (*INDRET* and *INDMOM*). All dependent variables are scaled up by 100. Missing values of independent variables are imputed with the monthly medians. We cross-sectionally standardize all independent variables by mapping them into the $[0,1]$ interval by scaling the monthly ranks by the number of observations (Gu et al. (2020)). All regressions include time and firm fixed effects. t-statistics are computed with standard errors clustered by month and stock and are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Predictability of Social Peer Firm Returns

	Earnings Growth				Forecast Error				Earnings Return			
	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q1	(6) Q2	(7) Q3	(8) Q4	(9) Q1	(10) Q2	(11) Q3	(12) Q4
SPFMOM	34.067***	25.649***	18.217***	11.812***	14.655***	14.286***	14.327***	11.194***	0.141***	0.096*	−0.003	−0.018
	(21.318)	(16.372)	(11.822)	(7.753)	(7.151)	(6.212)	(6.076)	(5.109)	(2.819)	(1.850)	(−0.053)	(−0.338)
SPFRET	10.925***	12.916***	11.340***	9.178***	4.830***	3.588**	5.874***	5.741***	0.140***	0.013	0.046	−0.0004
	(9.823)	(11.968)	(9.986)	(8.031)	(3.412)	(2.296)	(3.486)	(3.504)	(3.315)	(0.325)	(1.077)	(−0.009)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,457,322	1,462,274	1,467,432	1,472,775	817,742	737,599	673,727	621,535	1,490,579	1,463,703	1,438,009	1,412,416
R^2	0.174	0.175	0.175	0.174	0.132	0.151	0.161	0.169	0.052	0.056	0.059	0.058

Panel B: Controlling for Industry Momentum

	Earnings Growth				Forecast Error				Earnings Return			
	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q1	(6) Q2	(7) Q3	(8) Q4	(9) Q1	(10) Q2	(11) Q3	(12) Q4
SPFMOM	16.973***	12.622***	9.360***	7.949***	6.214***	7.624***	9.788***	6.425***	0.138**	0.155**	0.119**	0.103*
	(9.837)	(7.431)	(5.357)	(4.539)	(2.898)	(3.118)	(4.040)	(2.639)	(2.359)	(2.538)	(2.033)	(1.671)
SPFRET	2.991***	4.754***	4.548***	4.121***	1.742	−0.041	3.696**	4.079**	0.127***	0.027	0.065*	−0.003
	(3.252)	(5.178)	(4.670)	(4.511)	(1.294)	(−0.030)	(2.272)	(2.369)	(3.334)	(0.735)	(1.664)	(−0.082)
INDMOM	22.306***	16.767***	11.266***	4.678**	10.834***	8.425***	5.765**	6.116**	0.003	−0.077	−0.162**	−0.164**
	(12.367)	(9.192)	(6.024)	(2.464)	(4.989)	(3.544)	(2.408)	(2.211)	(0.041)	(−1.224)	(−2.441)	(−2.501)
INDRET	11.175***	11.647***	9.762***	7.353***	4.145***	4.932***	2.945	2.217	0.019	−0.018	−0.024	0.008
	(9.674)	(9.711)	(7.661)	(6.260)	(2.760)	(2.850)	(1.569)	(1.108)	(0.436)	(−0.410)	(−0.526)	(0.162)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,457,322	1,462,274	1,467,432	1,472,775	817,742	737,599	673,727	621,535	1,490,579	1,463,703	1,438,009	1,412,416
R^2	0.175	0.176	0.176	0.174	0.132	0.151	0.161	0.169	0.052	0.056	0.059	0.058

Internet Appendix

Appendix I: Examples of Strategy Similarity Constructions

This appendix provides examples of strategy similarity based on firms' fiscal 2010 10-K filings. In each example, we present the similarity scores and excerpts from strategy-related discussions for firm pairs in the same industry and the raw similarity scores calculated based on 10-K filings. The similarity scores are calculated based on all paragraphs focusing on strategy discussion from the 10-K files. We use the method of Dyer et al. (2017) to find the strategy-related paragraphs in each 10-K. The excerpts are intended to demonstrate how the similarity measure works and do not represent the full strategy discussion of the 10-K filings that we use to compute the similarity scores. We discard generic words (i.e., any word that appears in more than 25% of 10-K filings) when computing strategy similarity. The generic words are in grey. Overlapping non-generic words are in blue and are shown in boldface.

Firm Pair 1: GRANITE CONSTRUCTION and STERLING CONSTRUCTION

- FF-48 Industry: Construction
- Similarity Score: 0.407
- Number of Unique Words:
  - GRANITE CONSTRUCTION: 175
  - STERLING CONSTRUCTION: 174
- Number of Overlapping Words: 71

Sample Excerpts:

GRANITE CONSTRUCTION INC (Accession Number: 0000861459-11-000014)

*We **participate** in joint **ventures** with other **construction** companies mainly on **projects** in our **Large Project Construction** segment. Joint **ventures** are **typically** used for **large**, **technically complex projects**, including design/**build projects**, where it is **desirable** to **share risk** and **resources**. Joint **venture partners typically** provide **independently prepared estimates**, shared financing and equipment and **often bring local knowledge** and **expertise** (see Joint **ventures**; Off-Balance-Sheet Arrangements under Item 7. Management's Discussion and **Analysis** of Financial Condition and Results of Operations).*

STERLING CONSTRUCTION CO INC (Accession Number: 0000874238-11-000006)
*We **participate** in joint **ventures** with other **large construction** companies and other **partners**, **typically** for **large**, **technically complex projects**, including design-**build projects**, when it is **desirable** to **share risk** and **resources** in order to seek a competitive advantage or when the **project** is too **large** for us to obtain sufficient bonding. Joint **venture partners typically** provide **independently prepared estimates**, furnish employees and equipment enhance bonding capacity and **often** also **bring local knowledge** and **expertise**. We select our joint **venture partners** based on our **analysis** of their **construction** and financial capabilities, **expertise** in the type of work to be performed and past working relationships with us, among other criteria.*

<u>Firm Pair 2: NATIONAL BEVERAGE CORP and COCA COLA</u>

- FF-48 Industry: Candy and Soda
- Similarity Score: 0.131
- Number of Unique Words:
  - NATIONAL BEVERAGE: 77
  - COCA COLA: 128
- Number of Overlapping Words: 13

Sample Excerpts:

NATIONAL BEVERAGE CORP (Accession Number: 0000950123-10-065795)
*Our fantasy of flavors **strategy** emphasizes our distinctive flavored soft drinks, energy drinks, juices and other specialty beverages. Although cola drinks account for approximately 50 of the soft drink industry's domestic grocery channel **volume**, colas account for less than 20 of our total **volume**. We continue to emphasize expanding our **beverage** portfolio beyond traditional carbonated soft drinks through new product development inspired by lifestyle enhancement trends, innovative package enhancements and the development of products **designed** to provide functional benefits to the **consumer**. Such products include our lines of energy drinks and vitamin-enhanced waters. We intend to expand our product offerings through in-house development and acquisitions, to further our **strategy** within the evolving functional category geared toward **consumer** health and wellness.*

COCA COLA CO (Accession Number: 0001047469-11-001506)
*Marketing investments are **designed** to enhance **consumer** awareness of and increase consumer-preference for our brands. This produces long-term growth in unit case **volume**, per capita consumption and our shareof worldwide nonalcoholic **beverage** sales. Through our relationships with our bottling partners and those who sellour products in the marketplace, we create and implement integrated marketing programs, both globally and locally,that are **designed** to heighten **consumer** awareness of and product appeal for our brands. In developing a **strategy** for a Company brand, we conduct product and packaging research, establish brand positioning, develop precise consumer communications and solicit **consumer** feedback. Our integrated marketing activities include, but*

are not limited to, advertising, point-of-sale merchandising and sales promotions.

Firm Pair 3: ACI WORLDWIDE and HEALTHCARE SERVICES GROUP

- FF-48 Industry: Business Services
- Similarity Score: 0.029
- Number of Unique Words:
  - ACI WORLDWIDE: 142
  - HEALTHCARE SERVICES GROUP: 74
- Number of Overlapping Words: 3

Sample Excerpts:

ACI WORLDWIDE INC. (Accession Number: 0000950123-11-015619)
*Our product development efforts focus on new products and improved versions of existing products. We facilitate user group meetings. The user groups are generally organized geographically or by product lines.The groups help us determine our product strategy, development plans and aspects of customer support. We believe that the timely development of new applications and enhancements is essential to maintain our competitive positioning the market. To compete successfully, we need to maintain a successful research and development effort. If we fail to enhance our current products and develop new products in response to changes in technology and industrystandards, bring product enhancements or new product developments to market quickly enough, or accurately predictfuture changes in our customers **needs** and our competitors develop new technologies or products, our products could become less competitive or obsolete. Adoption of open systems technology. In an effort to leverage lower-cost computing technologies and current technology staffing and resources, many financial institutions, retailers and electronicpayment processors are seeking transition their systems from proprietary technologies to open technologies. Our continued investment in open systems technologies is, in part, designed to address this demand.*

HEALTHCARE SERVICES GROUP INC (Accession Number: 0000950123-11-015770)
*Dietary consists of managing the client's dietary department which is principally responsible forfood purchasing, meal preparation and providing dietician consulting professional services, which includes the development of a menu that meets the patient's dietary **needs**. We began Dietary operations in 1997.*

Firm Pair 4: LANDAUER and ArcSight

- FF-48 Industry: Business Services
- Similarity Score: 0.029
- Number of Unique Words:
  - LANDAUER: 127
  - ArcSight: 147
- Number of Overlapping Words: 4

Sample Excerpts:

LANDAUER INC (Accession Number: 0000892626-10-000090)
Sample Excerpt: *The Company's domestic radiation monitoring services are largely based on the Luxel+ dosimetersystem in which all analyses are performed at the Company's laboratories in Glenwood, Illinois. Luxel+ employs the Company's proprietary OSL technology. The Company's InLight dosimetry system enables certain customers to maketheir own measurements using OSL technology. InLight is marketed to domestic and international radiation measurement laboratories found at nuclear power plants, military installations, the Department of Homeland Security,national research laboratories, and commercial services. Landauer has positioned the InLight system as both a productline and a radiation monitoring service in ways that others can benefit directly from the technical and operationaladvantage of OSL technology.*

ArcSight Inc (Accession Number: 0000950123-10-064592)
Sample Excerpt: *Enterprise-Class Technology and Architecture. We design our solutions to serve the needs of thelargest organizations, which typically have highly complex, geographically dispersed and heterogeneous business and technology infrastructures. We deliver enterprise-class solutions by providing interoperability, flexibility, scalabilityand efficient archiving.*

Appendix II

- Figure A.1: Strategy Alphas for Post-2000
- Table A.1: Variable Descriptions
- Table A.2: Summary Statistics and Correlations for Return Variables
- Table A.3: Comovements in Firm Fundamentals and Stock Returns with Industry Portfolio
- Table A.4: Return Predictability of Social Peer Firm Returns: Within Industry Univariate Sort
- Table A.5: Long-Run Predictability, Earnings Surprises, Analyst Forecast Errors, and Earnings Returns using Orthogonalized Social Peer Firm Return Measures
- Table A.6: Return Predictability of Peer Firms' Returns: Robustness Checks
- Table A.7: Return Predictability of In-State Peer Firms and Non-Peer Firms

Figure A.1: Strategy Alphas for Post-2000

The figure presents the Long-Short FF5 alphas for the 2000/01-2019/11 sample. In Panel A, we report the long-short alphas for the lead-lag predictors (*SPFRET*, *INDRET*, *GEORET*, *CFRET*, *CRET*, *TECHRET*, *TNICRET*, and *CONGRET*, and *STRATRET*). The orange bars show the 90% confidence intervals. The portfolios are value-weighted and exclude the largest 100 firms.

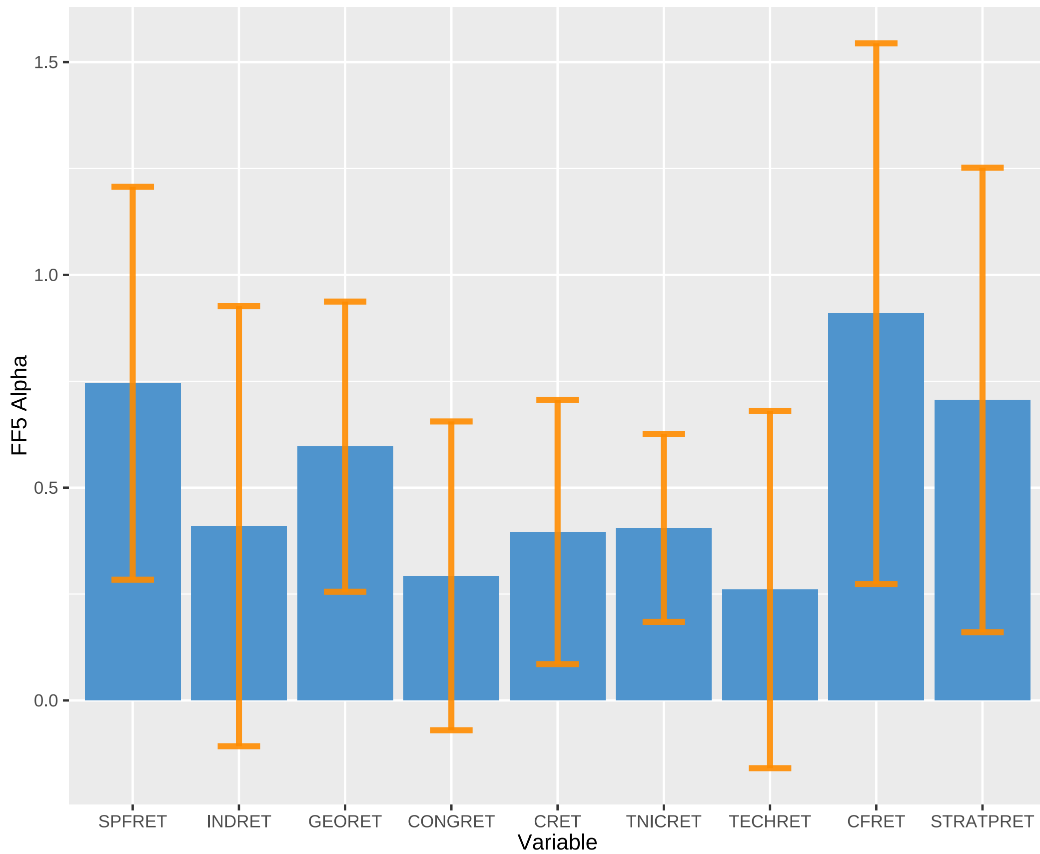


Table A.1: Variable Descriptions

Variable	**Definition**
Social Connectedness Index (SCI)	Number of Facebook friends links between firms' headquarters counties, scaled by the product of populations of the two counties
Strategy Similarity (STRATSIM)	Cosine similarity of business strategy word vectors of two firms. The word vectors are derived from 10-K documents filed in the past five years, using only the paragraphs that discuss firm strategies. The strategy paragraphs are identified using LDA methodology in Dyer et al. (2017).
Product Similarity (STRATSIM^{PROD})	Cosine similarity between product-related 10-K word vectors of two firms, calculated using the methodology in Hoberg and Phillips (2016).
Technology Similarity (STRATSIM^{TECH})	Cosine similarity between patent distribution vectors of two firms, calculated using the methodology in Lee et al. (2019).
Analyst Similarity (ANALYST)	Cosine similarity between the analyst vectors of two firms.
Customer Similarity (CSTMR)	Cosine similarity between the customer vectors of two firms derived from Compustat Segment data.
Same County	Dummy variable that indicates whether two firms are headquartered in the same county.
SCI-Weighted Fundamentals	Fundamentals (ΔEPS, $\Delta Sales$, $\Delta Employees$, $NewCapital$) of an SCI-weighted portfolio composed of all firms in the focal firm's industry, excluding same-state firms. ΔEPS is the change in EPS scaled by lagged stock price, $\Delta Sales$ is the percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value.
Social Peer Firm Return (SPFRET)	SCI-weighted returns based on all firms in the focal firm's industry, excluding same-state firms. $SPFRET_{ALL}$ is the SCI-weighted returns based on all the firms in the focal firm's industry except for the focal firm itself. SPFRET$_{DIST}$ is an alternative social peer firm return measure for which social ties are measured with the inverse distance between firms' headquarters locations. $SPFRET_{\perp}$ is $SPFRET$ orthogonalized against industry momentum and potentially other control variables.
Strategy Peer Firm Return (STRATRET)	*STRATSIM*-weighted returns based on all firms in the focal firm's industry, excluding same-state firms.
Social Peer Firm Momentum (*SPFMOM*)	The compounded *SPFRET* from months $t-11$ to $t-1$.
Industry Return (INDRET)	The equal-weighted average return of stocks with the same Fama-French 48 industry classification as the focal stock, *INDMOM* is obtained by compounding *INDRET* from month $t-11$ to $t-1$.
Geographic Return (GEORET)	The equal-weighted average return of all stocks from the same economic area (EA) as the focal stock but from a different FF48 industry, constructed following Parsons et al. (2020). *GEOMOM* is obtained by compounding *GEORET* from month $t-11$ to $t-1$.
Analyst Momentum (CFRET)	The weighted average return of stocks that share at least one analyst with the focal stock over the previous 12 months, where weights are the number of shared analysts between stocks. The variable is constructed following Ali and Hirshleifer (2020).
Customer Return (CRET)	The equal-weighted average stock return of the main customers of the focal firm, where a six-month gap is required between the fiscal year-end of the supplier and stock returns. Construction of the variable follows Cohen and Frazzini (2008).
Technology-Linked Firm Return (TECHRET)	The weighted average stock return of technology-linked peer firms, where the weights are the technology closeness between the peer firm and the focal firm, determined by the similarities between patent distributions across different technology categories. The variable is constructed following Lee et al. (2019).
Pseudo-Conglomerate Return (CONGRET)	The sales-weighted return of value-weighted, single-segment firm portfolios, formed for each segment that a conglomerate firm operates in. Construction of the variable follows Cohen and Lou (2012).

Variable	Definition
Text-Based Industry Momentum (TNICRET)	Equal-weighted stock return of peer firms identified through 10-K product text, constructed following Hoberg and Phillips (2016).
Monthly Return (RET)	The monthly stock return. Following Shumway (1997), we adjust stock returns for delisting to avoid survivorship bias.
Firm Size (SIZE)	The logarithm of the market capitalization (in million dollars) as measured at the end of the previous June.
Market Beta (BMKT)	The CAPM beta are computed using a 60-month window with a minimum window of 24 months using a one-factor market model.
Book-to-market (BM)	Computed as the book value of stockholders' equity, plus deferred taxes and investment tax credit (if available), minus the book value of the preferred stock for the last fiscal year, scaled by the market value of equity at the end of December of $T-1$. Depending on availability, the redemption, liquidation, or par value (in that order) is used to estimate the book value of the preferred stock.
Momentum (MOM)	Obtained by compounding RET from month $t-11$ to $t-1$.
Idiosyncratic Volatility (IVOL)	Computed as the standard deviation of the daily residuals obtained by regressing the daily excess stock returns on the daily market excess return, small-minus-big (SMB) and high-minus-low (HML) factors over the previous month.
Illiquidity (ILLIQ)	Amihud's illiquidity (Amihud, 2002), defined as the average daily ratio of the absolute stock return to the dollar trading volume within the previous month.
Maximum Return (MAX)	The maximum daily stock return realized over the previous month (Bali et al., 2019).
Skewness (SKEW)	The sample skewness of the daily stock returns from the previous month.
Coskewness (COSKEW)	The stock's monthly coskewness constructed following Harvey and Siddique (2000).
Cumulative Abnormal Return (CAR)	Market-adjusted returns cumulated over a three-day window around earnings announcements.
Standardized Unexpected Earnings (SUE)	Calculated as the difference between a stock's quarterly earnings minus the same-quarter value from the previous year, divided by its standard deviation over the past eight quarters.
Analyst Forecast Errors (FE)	Calculated as the difference between the announced earnings and analysts' consensus forecast, scaled by the stock price five trading days before the earnings announcement date (DellaVigna and Pollet, 2009).

Table A.2: Summary Statistics and Correlations for Return Variables

The table provide summary statistics and correlation tables for the return variables used in our analyses. *RET* is the contemporaneous return. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). *SPFMOM* is the cumulative *SPFRET* over months $t-11$ through $t-1$. *INDRET* is the equal-weighted average return of stocks for a given industry. *INDMOM* is the cumulative *INDRET* over months t-11 through t-1. *GEORET* is the equal-weighted average return of peer firms that are in the same economic area as a stock (excluding those in the same industry). *CONGRET* is the pseudo-conglomerate return. *CRET* is the equal-weighted average stock return of a firm's main customers. *TNICRET* is the text-based industry return of Hoberg and Phillips (2018). *TECHRET* is the weighted average stock return of technology-linked peer firms, where the weights are the technology closeness between the peer firm and the focal firm. *CFRET* is the average return of peer stocks that share at least one analyst with the focal stock over the previous 12 months, weighted by the number of shared analysts between stocks. *STRATRET* is the strategy-similarity-weighted return of peer firms that are in the same Fama-French 48 industry as the focal firm. All returns are reported in percentages. Before calculating the correlations, all variables except *RET* are cross-sectionally standardized by first ranking and then mapping them into the $[0,1]$ interval (Gu et al. (2020)). All of the reported statistics and correlations are averages over monthly cross sections.

Panel A: Summary statistics of Return Variables

	N	Mean	Std. Dev.	Min.	25%	75%	Max.
RET	2608	1.741	11.713	-46.511	-4.386	6.569	157.254
SPFRET	2525	1.722	3.837	-14.521	-0.643	3.837	28.001
SPFMOM	2525	19.878	18.671	-32.533	7.887	29.027	164.051
INDRET	2543	1.187	3.215	-8.614	-0.884	3.145	13.377
INDMOM	2543	15.027	15.711	-24.081	4.651	24.479	71.097
GEORET	1612	1.243	1.929	-4.266	0.136	2.270	7.572
CONGRET	731	1.255	3.942	-12.948	-1.055	3.459	18.951
CRET	882	1.221	6.995	-26.711	-2.616	4.757	45.661
TNICRET	2885	1.767	6.163	-33.748	-1.310	4.534	74.585
TECHRET	805	1.631	2.401	-9.251	0.339	2.811	17.265
CFRET	2682	1.481	3.995	-18.192	-0.820	3.647	30.652
STRATRET	1869	1.840	4.398	-16.094	-0.878	4.266	28.360

Panel B: Correlations of Return Variables

	RET	SPFRET	INDRET	GEORET	CONGRET	CRET	TNICRET	TECHRET	CFRET
SPFRET	0.166								
INDRET	0.204	0.642							
GEORET	0.009	-0.013	-0.019						
CONGRET	0.145	0.359	0.358	0.021					
CRET	0.103	0.141	0.144	0.054	0.150				
TNICRET	0.167	0.367	0.380	0.062	0.254	0.179			
TECHRET	0.122	0.250	0.271	0.049	0.205	0.134	0.278		
CFRET	0.266	0.494	0.495	0.061	0.412	0.221	0.441	0.370	
STRATRET	0.177	0.749	0.703	-0.012	0.391	0.155	0.350	0.353	0.514

Table A.3: Comovements in Firm Fundamentals and Stock Returns with Industry Portfolio

The table presents the comovements between the focal firm and its industry peers using panel regression analysis. The key independent variable is the corresponding fundamentals of the contemporaneous equal- and value-weighted industry portfolios (excluding the focal firm). In columns 1-4, the dependent variables are firms' fundamentals, measured annually: ΔEPS is the change in EPS scaled by lagged stock price, $\Delta Sales$ is the percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value. Only the firms with the same fiscal year-end as the focal firm are included. In column 5, the dependent variable is monthly stock returns. All regressions include time fixed effects and we present the coefficient estimates in percentages. Standard errors are clustered by both time and firm and the corresponding t-statistics are reported in parentheses. For all panels, *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Equal-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
Equal-Weighted	1.470***	28.165***	11.076***	28.904***	6.822***
	(3.455)	(9.826)	(8.385)	(10.892)	(25.372)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,732	124,983	121,845	121,848	1,711,696
R^2	0.017	0.048	0.043	0.034	0.141

Panel B: Value-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
Value-Weighted	1.312***	18.083***	6.926***	23.947***	5.021***
	(4.168)	(9.746)	(6.845)	(8.354)	(24.613)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,626	120,737	119,748	120,858	1,667,918
R^2	0.017	0.035	0.037	0.029	0.134

Table A.4: Return Predictability of Social Peer Firm Returns: Within Industry Univariate Sort

The table reports the results of the univariate portfolio sorts based on *SPFRET*, the SCI-weighted average return of a firm's industry peers (excluding those from the same state), averaged over the FF48 industries. Each month, for each FF48 industry, we sort all common stocks into deciles based on *SPFRET* and calculate both the equal-weighted and value-weighted one-month-ahead returns for the decile portfolios, as well as the return of the portfolio that long the decile 10 portfolio and short the decile 1 portfolio. We then take the average of these portfolio returns over all industries and present both the average (raw) returns and the FF5 alphas. All returns are in percentages. The corresponding *t*-statistics based on Newey and West (1994) standard errors are reported in parentheses below the alphas. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.935***	1.042***	0.990***	1.104***	1.121***	1.184***	1.177***	1.199***	1.203***	1.286***	0.351***
	(4.387)	(4.724)	(4.435)	(4.942)	(4.888)	(5.345)	(5.414)	(5.565)	(5.643)	(6.246)	(5.217)
FF5 Alpha EW	−0.293***	−0.151**	−0.233***	−0.137*	−0.109	−0.022	−0.027	−0.059	−0.028	0.058	0.351***
	(−4.577)	(−2.085)	(−3.545)	(−1.912)	(−1.543)	(−0.308)	(−0.428)	(−0.818)	(−0.428)	(0.886)	(4.707)
Raw Return VW	0.914***	1.041***	0.979***	1.066***	1.084***	1.160***	1.106***	1.165***	1.144***	1.229***	0.314***
	(4.510)	(4.901)	(4.686)	(5.121)	(4.973)	(5.744)	(5.392)	(5.743)	(5.742)	(6.499)	(4.607)
FF5 Alpha VW	−0.310***	−0.120	−0.240***	−0.158**	−0.135*	−0.045	−0.086	−0.078	−0.067	0.025	0.335***
	(−4.625)	(−1.617)	(−3.386)	(−2.066)	(−1.713)	(−0.655)	(−1.193)	(−0.908)	(−1.020)	(0.345)	(4.529)

Table A.5: Long-Run Predictability of the Orthogonalized Social Peer Firm Returns

The table presents robustness checks of Tables 8 and 9, replacing *SPFRET* and *SPFMOM* with the orthogonalized versions of the corresponding variables, $SPFRET_{\perp}$ and $SPFMOM_{\perp}$, respectively. Panel A replicates Table 8, and Panel B replicates Table 9 Panel A. See Tables 8 and 9 for detailed descriptions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Calendar-Time Portfolio Returns based on $SPFRET_{\perp}$ or $SPFMOM_{\perp}$ (FF5 alphas)

	$SPFRET_{\perp}$			$SPFMOM_{\perp}$		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Months 1–3	-0.204^{***}	0.124^{**}	0.329^{***}	-0.172^{**}	0.215^{***}	0.387^{***}
	(-3.229)	(2.357)	(4.177)	(-2.484)	(2.959)	(3.056)
Months 4–6	-0.104^{**}	0.031	0.136^{**}	-0.095	0.122	0.216
	(-2.080)	(0.714)	(2.437)	(-0.872)	(1.628)	(1.488)
Months 7–12	-0.139^{**}	0.010	0.149^{***}	-0.134	0.016	0.150
	(-2.207)	(0.226)	(2.966)	(-1.513)	(0.263)	(1.529)
Months 13–24	-0.128^{**}	0.004	0.131^{***}	-0.090	0.021	0.110
	(-2.140)	(0.083)	(3.631)	(-1.223)	(0.359)	(1.409)

Panel B: Earnings Growth, Analyst Forecast Errors, and Earnings Announcement Returns

	Earnings Growth				Forecast Error				Earnings Return			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$SPFMOM_{\perp}$	7.081***	5.605***	5.478***	3.798**	3.465**	4.058*	4.924**	2.085	0.153***	0.225***	0.142**	0.140**
	(4.641)	(3.694)	(3.569)	(2.483)	(2.004)	(1.952)	(2.512)	(1.106)	(2.681)	(3.857)	(2.547)	(2.349)
$SPFRET_{\perp}$	0.565	1.439	1.414	1.977**	0.490	0.950	1.756	1.433	0.120***	0.043	0.063	0.028
	(0.646)	(1.604)	(1.459)	(2.282)	(0.487)	(0.939)	(1.459)	(0.940)	(3.180)	(1.227)	(1.552)	(0.717)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	812,690	817,056	821,603	826,164	549,509	492,687	447,513	413,251	921,282	906,370	891,737	876,953
R^2	0.197	0.201	0.202	0.200	0.148	0.169	0.176	0.185	0.064	0.070	0.071	0.068

Table A.6: Return Predictability of Peer Firms' Returns: Robustness Checks

The table presents robustness checks for Table 6, column 9 with alternative specifications. Section 2 provides detailed variable descriptions. Column 1 is the same as Table 6, column 9. Column 2 replaces the equal-weighted industry return with its value-weighted version. In column 3, we replace the Fama-French 48 classification with the TNIC-based industry classification while calculating *SPFRET*. Column 4 considers the geographical proximity-based social peer firm return, $SPFRET_{DIST}$. In column 5, *SPFRET* is calculated excluding peer firms headquartered in states where the focal firm has economic presences. In column 6, we add *SIMRET* to the controls, which is the socioeconomicsimilarity-weighted peer firm return. In column 7, we construct *SPFRET* by excluding peer firms within 100 miles of the focal firm's headquarters. Column 8 replaces county-based *SPFRET* with its ZIP code-based version. In column 9, we use an orthogonalized *SPFRET*, defined as a regression residual of *SPFRET* on all the other independent variables. Column 10 uses an alternative standardization method and transform all the explanatory variables by subtracting the sample mean and dividing by its standard deviation. In column 11, we use a panel regression with month fixed effects and standard errors double-clustered by month and by stock. All returns are reported in percentages. Missing values of independent variables are imputed with the monthly medians. In all columns except column 11, standard errors are adjusted based on Newey and West (1994). *t*-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	RET_{t+1}										
	(1)	(2)$_{VW}$	(3)$_{TNIC}$	(4)$_{DIST}$	(5)$_{No\ Overlap}$	(6)$_{SIM}$	(7)$_{100}$	(8)$_{ZIP}$	(9)$_{\perp}$	(10)$_{SD}$	(11)$_{Panel}$
SPFRET	0.560***	0.993***	0.878***	0.431***	0.275***	0.402***	0.553***	0.526***	0.463***	0.140***	0.767***
	(3.729)	(5.481)	(6.287)	(5.456)	(2.665)	(3.355)	(3.715)	(4.018)	(3.972)	(3.014)	(3.787)
INDRET	0.747***	0.090	1.007***	0.917***	0.952***	0.694***	0.751***	0.812***	1.041***	0.236***	0.762**
	(4.310)	(0.571)	(5.907)	(5.447)	(5.424)	(3.706)	(4.031)	(5.812)	(5.821)	(3.864)	(1.962)
GEORET	0.248**	0.229**	0.249**	0.230***	0.251***	0.254***	0.246**	0.246**	0.205*	0.066***	0.319*
	(2.562)	(2.353)	(2.561)	(2.377)	(2.577)	(2.617)	(2.551)	(2.513)	(1.902)	(2.690)	(1.951)
CONGRET	−0.050	−0.013	−0.035	−0.032	−0.012	−0.064	−0.046	−0.033	−0.060	−0.015	−0.200
	(−0.363)	(−0.100)	(−0.249)	(−0.226)	(−0.085)	(−0.483)	(−0.334)	(−0.237)	(−0.401)	(−0.822)	(−1.021)
CRET	0.448***	0.465***	0.443***	0.445***	0.455***	0.444***	0.444***	0.447***	0.422***	0.068***	0.397**
	(3.364)	(3.577)	(3.349)	(3.373)	(3.425)	(3.337)	(3.345)	(3.357)	(3.120)	(3.638)	(2.327)
TNICRET	0.654***	0.694***	0.139	0.669***	0.681***	0.653***	0.657***	0.669***	0.706***	0.153***	0.834***
	(4.726)	(4.968)	(1.220)	(3.373)	(4.824)	(4.753)	(4.741)	(4.752)	(4.745)	(4.274)	(3.621)
TECHRET	−0.015	0.015	−0.024	−0.008	0.005	−0.017	−0.023	−0.015	−0.047	−0.013	0.074
	(−0.079)	(0.080)	(−0.119)	(−0.042)	(0.027)	(−0.089)	(−0.118)	(−0.077)	(−0.240)	(−0.399)	(0.292)
CFRET	1.563***	1.614***	1.532***	1.587***	1.628***	1.555***	1.561***	1.590***	1.644***	0.428***	1.835***
	(8.301)	(8.215)	(8.216)	(8.258)	(8.359)	(8.257)	(8.188)	(8.355)	(7.817)	(7.406)	(8.288)
SIMRET					0.229						
					(1.478)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	276	276	276	276	276	276	276	276	253	276	276
# Stocks	3,867	3,867	3,867	3,867	3,867	3,867	3,867	3,867	3,966	3,867	3,867
Time FE	–	–	–	–	–	–	–	–	–	–	Yes
Observations	–	–	–	–	–	–	–	–	–	–	1,067,332
R^2	0.070	0.070	0.069	0.070	0.069	0.070	0.069	0.069	0.070	0.071	0.126

Table A.7: Return Predictability of In-State Peer Firms and Non-Peer Firms

The table presents robustness checks for Table 6, column 9 with two additional controls. $SPFRET_{STATE}$ is the SCI-weighted return of same-state peer firms and $NPFRET$ is the SCI-weighted return of firms from other industries (excluding same-state firms). Section 2 provides detailed descriptions of the other variables. All returns are reported in percentages. Missing values of independent variables are imputed with the monthly medians. We cross-sectionally standardize all independent variables by mapping their ranks into the $[0,1]$ interval, similar to (Gu et al. (2020)). We report t-statistics with standard errors based on Newey and West (1994) adjustments in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	RET_{t+1}	
	(1)	(2)
SPFRET	0.532***	0.543***
	(3.607)	(3.639)
SPRET_STATE	0.361***	
	(4.578)	
NPFRET		−0.140
		(−1.609)
INDRET	0.672***	0.706***
	(3.824)	(4.376)
GEORET	0.239**	0.251***
	(2.502)	(2.754)
CONGRET	−0.055	−0.044
	(−0.402)	(−0.333)
CRET	0.441***	0.450***
	(3.331)	(3.399)
TNICRET	0.642***	0.651***
	(4.699)	(4.739)
TECHRET	−0.028	−0.008
	(−0.145)	(−0.043)
CFRET	1.528***	1.556***
	(8.334)	(8.325)
Controls	Yes	Yes
# Periods	276	276
# Stocks	3,867	3,867
R^2	0.070	0.070