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

Lin Peng Sheridan Titman Muhammed Yönaç Dexin Zhou

March 17, 2026

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), Gerard Hoberg (discussant), Xing Huang (discussant), Ted Joyce, Xin Li (discussant), Dong Lou, Sebastiano Manzan, Pierre Magontier (discussant), Javier Mella (discussant), Ramana Nanda, Quoc Nguyen (discussant), Ali Sharifkhani (discussant), Jun Wang, Junbo Wang (discussant), Wei Xiong, Jing Zhao (discussant), and conference and seminar participants at AsianFA Meeting, CFEA, CICF, EFA, E(astern)FA, FMA, FMA Consortium on Asset Management, FSU SunTrust Beach Conference, MFA, NBER BF Working Group, SFW, SFS Cavalcade Asia-Pacific, UEA, 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

For a variety of reasons, economists group firms into clusters that characterize how the firms do business. Financial economists, for example, seek to identify comparable firms to benchmark both returns and financial decisions, and economists studying industrial organizations seek to identify sets of firms that compete for customers and labor. As a first step, economists use industry codes that are assigned to the firms, but given the heterogeneity within these codes, a literature has evolved that studies refinements of these industry classifications (see, for example, Cohen and Frazzini 2008; Hoberg and Phillips 2018; Lee et al. 2019). For financial economists, an important insight from this literature is that firms may be in similar lines of business, but because they implement different strategies, their financial structure and return patterns may be very different.

The premise of this study is that the strategies implemented by corporate executives are determined in part by their opinions and views of the world, and that these opinions and views depend in part on where the executives live. We conjecture that not only do people living in the same place tend to have similar views, but that individuals in different, but socially connected locations are also likely to have similar views. This could arise because interactions between individuals in socially connected places shape their views, or it could be that residents of these places share similar characteristics, which are the basis for their views. Regardless of the cause, we conjecture that because their views are likely to be similar, corporate executives in socially-connected locations will tend to implement similar strategies. Hence, the within-industry differences in various aspects of corporate strategy are likely to be smaller among firms located in more socially connected cities.

We find that this is indeed the case: industry peer firms headquartered in more socially-connected counties, even after excluding same-state peers, do tend to be more similar along a number of dimensions. The fundamentals and stock returns of such peer firms also tend to exhibit more co-movements than those from less connected places. However, our evidence suggests the valuation implications of these social connectedness-based similarities are not fully appreciated by stock market participants. Specifically, we find that a lead-lag strategy that exploits the underreaction of stock returns to the returns of their socially connected industry peers generates significant excess returns.

Our analysis captures the multi-faceted concept of social connectedness by using three different measures. The first one is Facebook's Social Connectedness Index (SCI) (see Bailey et al., 2018a), which measures the relative probability that individuals in two U.S. counties are friends on Facebook..[1] The second measure is the geographic proximity between two locations (*PROX*). It has long been shown that proximity can facilitate the flow of ideas (see Hong et al., 2004; Pool et al., 2015, for example). Our third measure is based on job-to-job (*J2J*) transition data provided by the United States Census Bureau, which tracks the volume of worker flows between states within a given industry.[2]

We measure the similarity of firm strategies along three dimensions. The first dimension of strategy similarity, *STRATSIM*, is measured by the similarity of text from a firm's 10-K filings. The measure is constructed by applying the Latent Dirichlet Allocation model of Dyer et al. (2017) to identify paragraphs in the filings that pertain to business operations and strategy, and then computing the cosine similarity of the resulting word vectors across firm pairs. The second measure, $STRATSIM^{PROD}$, is the product market similarity score of Hoberg and Phillips (2016), which quantifies the overlap between firms' business descriptions in their 10-K filings using cosine similarity of product-related word vectors. The third measure, $STRATSIM^{TECH}$, is the technology similarity score of Lee et al. (2019), which measures the proximity of firms' patent portfolios using the cosine similarity of vectors constructed from the USPTO technology class distributions of each firm's patents.

We find that each of these pairwise similarity measures exhibits a significant positive correlation with the social connectedness of the firms' headquarters. We find that among the three social connectedness measures, *SCI* shows the strongest explanatory power for strategy similarities. For example, the 10-K-based strategic similarity is 9.0% standard deviation higher for firm pairs with *SCI* in the top quartile compared to those in the bottom quartile. Similarly, the same change in *SCI* corresponds to 19.9% and 9.4% of the

[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 representativeness of its user base makes SCI a unique measure of the real-world geographic structure of U.S. social networks at a population scale. See, Bailey et al. (2018a) and Kuchler et al. (2020) for further discussion on these points.

[2] This measure may simply reflect the fact that people tend to migrate to locations that are culturally similar to what they previously experienced, or it could be causal, i.e., locations become culturally similar due to migration. For example, Stoyanov and Zubanov (2012) show that workers can carry productivity from one firm to another.

standard deviation of product-based and technology similarity, respectively.

The labor flow-based measure is the second most economically important social connectedness measure. Going from the bottom to the top quartile, the change in the labor-based measure explains 6.6%, 15.9%, and 6.1% standard deviations in the 10-K, product-based, and technology-based strategy similarity, respectively.

The geographic proximity-based measure is the weakest among the three social connectedness measures. This is primarily a consequence of our methodological choice rather than a limitation of geographic proximity itself. While geographic proximity naturally implies stronger interactions for nearby county pairs, distance is also correlated with other forms of local similarity that are conceptually distinct from social connectedness.[3] Since same-state firm pairs are disproportionately geographically close, their exclusion removes precisely the pairs for which the proximity effect would be strongest, while leaving the SCI and J2J measures largely unaffected. Excluding the same-state peers therefore mechanically attenuates the explanatory power of geographic proximity, but helps isolate social connectedness from institutional and geographic confounds, as emphasized in Pirinsky and Wang 2006, Parsons et al. 2020, and Jin and Li 2020.

Motivated by these pairwise results, we next create individually tailored tracking portfolios for each focal firm that is designed to best capture their strategies. These firm specific tracking portfolios include each of the focal firm's industry peers, weighted by either our measure of social connectedness of the two firms' locations or by measures of the similarities of the firms' strategies. Relative to equally-weighted and value-weighted industry portfolios, these specially tailored tracking portfolios should more accurately predict the behavior and performance of the focal stocks.

Our tests confirm this conjecture. We find that movements in both firm fundamentals and stock returns tend to be better explained by the social connectedness-weighted indexes, even after controlling for equal- and value-weighted industry portfolios and strategy-linked peer portfolios based on product market, technology, and 10-K-based similarity.

[3] These confounding channels include shared regulatory regimes, tax and labor policies, political environments, exposure to statewide shocks, and access to overlapping labor and product markets, all of which are highly correlated with geographic proximity but conceptually distinct from social connectedness. In additional results that we present later, we include same-state peers in our analysis and find that they do not subsume the effect of socially-connected distant peers.

We then ask whether market prices fully reflect the importance of these social ties. Under the null hypothesis that prices fully reflect the importance of social ties, the correlation between the date t return of the focal firm and the date t-j return of its tracking portfolio should be zero. However, under the alternative investor inattention hypothesis, the return of the focal firm underreacts to the return of its industry tracking portfolio, implying that the correlation between its current return and the lagged tracking portfolio return is positive. We do this by calculating three socially connectedness-weighted returns of each focal firm's industry peers based on *SCI*, *PROX*, and *J2J* (henceforth *SCIRET*, *PROXRET*, and *J2JRET*). We also construct a composite social firm peer return signal (*SPFRET*) by combining the three individual signals. We examine the extent to which these signals predict the focal firm's future returns. If market prices fail to fully reflect the similarity of socially connected industry peers, the social connectedness-weighted returns should lead the returns of the focal firms. We find strong evidence that supports the conjecture. For example, we show that a trading strategy that exploits these lead-lag relationships using SPFRET earns significant abnormal returns of 73 basis points (bps) per month.

Moreover, *SPFRET* continues to predict a firm's future returns for up to the next 12 months. The one-year cumulative long-short portfolio alpha is 3.91% and 3.95% 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% and 2.84%, but still significantly different from 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. A parallel lead-lag pattern arises for strategy-similarity-based tracking portfolios. Importantly, the return predictability associated with social connectedness and with strategy similarity reflects complementary information, with neither signal subsuming the other.

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 *SPFRET* exhibit 11 percentage points higher *SUE* in the following quarter. Moreover, firms with the highest past cumulative social peer firm returns (*SPFMOM*) have, on average, standardized unexpected earnings (*SUE*) that are 33.8 percentage points higher in the next quarter than those with the lowest *SPFMOM* (or 25.3% of the standard deviation of *SUE*). 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.

Consistent with the documented attenuation of many stock return predictors in recent decades (e.g., McLean and Pontiff, 2016), the abnormal returns to the *SPFRET* strategy are weaker in the post-2000 period. One contributing factor is that industry momentum (*INDRET*) itself declines sharply and becomes statistically insignificant after 2000 (Ali and Hirshleifer, 2020). Motivated by the approach in Jensen et al. (2021), we therefore refine our strategy by excluding the 100 largest firms by market capitalization, which tend to be efficiently priced and carry disproportionate weight in value-weighted portfolios. Under this refinement, the *SPFRET* portfolio continues to generate statistically significant abnormal returns of 63 basis points per month in the post-2000 period, only modestly below the full-sample alpha of 73 basis points.

Finally, we demonstrate that the usefulness of social connectedness for capturing valuable information, as established above, can be extended more broadly to non-industry-based tracking portfolios. We find that peer returns from other industries in connected counties also exhibit strong predictability, resulting in a significant long-short portfolio alpha of 33.5 bps per month. Furthermore, applying social connectedness as weights to strategies based on product market overlap, shared analyst coverage, and industry input-output links generates significant incremental return predictability beyond the original strategies.

A natural concern is that the correlation between social connectedness and strategic similarity may reflect common unobserved factors rather than a causal effect of social ties.

We mitigate this by excluding same-state firm pairs, which removes variation driven by shared regulatory environments, tax regimes, and local labor markets, and by controlling for a rich set of firm-pair characteristics including geographic proximity, firm size, and industry composition. We cannot, however, fully rule out omitted variables. The strategic similarity we document could arise because social interactions between executives in connected locations directly shape their views, or because residents of socially connected places share similar characteristics that independently lead to similar strategies. Our analysis does not distinguish between these channels. Rather, our contribution is to document that social connectedness captures meaningful variation in strategic similarity beyond what existing industry classifications, geographic proximity, or standard firm-level controls explain, and that market participants do not fully incorporate this information into prices.

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 the performance of professional investors.[4] Several more recent papers focus on 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. Similarly, Fracassi (2017) finds that executives sharing education and past employment histories tend to adopt more similar corporate policies. Furthermore, Kuchler et al. (2021) provide evidence that a firm's social proximity to institutional capital influ-

[4]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 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).

ences 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. More broadly, there is a long literature arguing that firms in similar industries tend to geographically colocate (e.g., Ellison and Glaeser, 1997, 1999) and such colocation could lead to increased productivity through knowledge or infrastructure sharing (e.g., Krugman, 1991; Storper and Venables, 2004). Our results imply that knowledge sharing may extend beyond the confines of cities and that social ties and labor flows can act as potential channels.

Lastly, this paper adds to the extensive literature that studies return predictability in financial markets.[5] 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).[6] 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 show 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 connectedness 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

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

[6]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).

focal firms' returns. In Section 5, we show that the returns of socially connected peer firms capture earnings-related information about the focal firm. In Section 6, we present robustness tests of the earlier empirical results. Section 7 concludes.

2 Data and Variable Definitions

Our main sample includes all common stocks traded on the NYSE, AMEX, and NASDAQ exchanges, covering the period from July 1963 through December 2023. The Center for Research in Security Prices (CRSP) provides daily and monthly return and volume data. To obtain some of the return variables, we also use a larger CRSP sample from June 1953 to December 2023. 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 the 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

2.1.1 Social Connectedness Measures

We consider three measures of social connectedness between two regions. First, we follow Bailey et al. (2018a) and measure the social connectedness between individuals from two U.S. counties based on the number of Facebook friendship links. Specifically, the *Social Connectedness Index* (*SCI*) is defined as the total number of Facebook friendship links between two U.S. counties divided by the product of the populations of the two counties, based on data obtained in April 2016. As the world's largest online social networking service, Facebook's scale and the relative representativeness of its user body make the SCI a comprehensive measure for real-world social connections.[7] In Internet Appendix I, we

[7] 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

provide a detailed illustration of the SCI measure.

Our second measure of social connectedness is geographic proximity (*PROX*), defined as the inverse of the geodesic distance between two counties. This measure is motivated by prior studies showing that social connections are more likely to be formed in geographically close areas (e.g., Hong et al., 2004; Pool et al., 2012; Bailey et al., 2018b).

The third measure of social connectedness is based on the intensity of labor flows, which are the job-to-job movements at the state-industry level. Specifically, we obtain the quarterly job-to-job employee flows data from the U.S. Census Bureau's Longitudinal Household-Employer Dynamics (LEHD), which is available starting in 2000, and we aggregate the employee counts to an annual frequency to remove seasonal effects. The employee flows data are at the state-industry level, where industries are given by the North American Industry Classification System (NAICS) sectors. We define the variable J2J, the scaled job-to-job employee flow, as the total number of employee movements between two counties divided by the total (unique) employee flows into or out of the two counties.[8]

2.1.2 Social Connection-based Return Predictors

Based on the three social connectedness measures, we first associate peer firms with a focal firm based on the social connectedness of the firms' headquarters locations.[9] We then

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); 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.

[8]Formally, J2J between firm i and j in year y is given by

$$\begin{aligned}\mathrm{J2J}_{y,i,j} \equiv \left(\mathrm{Flow}_{y,i,j} + \mathrm{Flow}_{y,j,i}\right) / (&\mathrm{Outflow}_{y,i} + \mathrm{Inflow}_{y,i} + \mathrm{Outflow}_{y,j} + \mathrm{Inflow}_{y,j} - \\ &\mathrm{Flow}_{y,i,i} - \mathrm{Flow}_{y,j,j} - \mathrm{Flow}_{y,i,j} - \mathrm{Flow}_{y,j,i}),\end{aligned}$$

where $Flow_{y,i,j}$ is the total count of employees that leave the state-NAICS of firm i to join the state-NAICS of firm j in year q, whereas $Inflow_{y,i}$ and $Outflow_{y,i}$ are the total inflow to and outflow from the state-NAICS of firm i in year y. Mathematically, $J2J_{y,i,j}$ is the probability of an employee who has moved into or out of either i or j in year y having moved between i and j.

[9]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 the SEC's Edgar database for 2023. 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

construct three versions of social connectedness-based industry peer returns: *SCIRET*, *PROXRET*, and *J2JRET*. *SCIRET* is the *SCI*-weighted same-industry peer firm returns, where industry is defined based on the Fama-French 48-industry definition. We exclude peer firms from the same state as the focal firm to alleviate the concerns that our results are driven by firms in close geographic proximity (e.g., Pirinsky and Wang, 2006; Parsons et al., 2020). *PROXRET* and *J2JRET* are defined similarly, with the SCI weights being replaced by *PROX* and *J2J*, respectively.

Formally, for a given (focal) firm i and for a given month t, component k of *SPFRET* is defined as

$$\text{Component}_{i,t}^{k} \equiv \frac{\sum_{j \in I_i} w_{i,j}^{k} RET_{j,t}}{\sum_{j \in I_i} w_{i,j}^{k}}, \tag{1}$$

where $w_{i,j}^{k}$ is the measure of connectedness between firm i and firm j associated with component k and I_i is the set of firms in the Fama-French 48-industry (FF48) to which firm i belongs, excluding firms headquartered in the same state as firm i. We consider three weights: *SCI*, *PROX* and *J2J*.

We provide two maps to illustrate how social connectedness affects portfolio weighting based on *SCI* and *J2J*. We first illustrate the construction of *SCIRET* 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 A.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.[10] Counties without the pres-

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

[10]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.

ence of such firms are presented in dark grey. When calculating *SCIRET* for CMI, peer firms headquartered in counties with higher SCI with Bartholomew are assigned higher weights. Consequently, CAT and CLC have corresponding weights of 3.27% and 4.33%, respectively, while the weight on SWK is only 0.54%. 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.[11]

In Figure A.1 Panel D, we present the heat map of the J2J between the machinery firms in Bartholomew County, IN, and other counties in the continental U.S. with a machinery industry presence, created using J2J data from 2014. Since J2J is on the state-industry level, all counties with a machinery firm presence in the same state have the same color in the heat map. We can see that while there has been substantial flow between machinery firms in Bartholomew and Williamson in 2014, there has been no flow between either Bartholomew and Peoria or Bartholomew and Hartford in the same year. Consequently, if we recalculate the weights of CAT, CLC, and SWK in CMI's peer portfolio from April 2016 using J2J, the weights of CAT and SWK are zero, while CLC ends up having a weight of 2.08%.[12]

Composite Social Peer Firm Return We define the social peer firm return (*SPFRET*) as a composite social-connectedness-related peer firm return using information from the three social peer firm returns, *SCIRET*, *PROXRET*, and *J2JRET*.

SPFRET is computed in a two-step process, starting in June 1963 and iterating forward in annual increments. First, using data available as of 1963, we regress the contemporaneous focal stock return on *SCIRET*, *PROXRET*, and *J2JRET*, with the three peer returns rank-normalized following Gu et al. (2020).[13] Second, we use the coefficients obtained from the first step as weights and interact the weights with the realized and

[11] By comparison, the PROX-based weights assigned to CAT, CLC, and SWK are 2.05%, 2.03%, and 0.67%, respectively.

[12] We also try averaging J2J over all years up to the portfolio formation. The results are almost indistinguishable from using only the most recent J2J, which suggests that J2J is highly persistent.

[13] We extend our main sample back to July 1953 to obtain an initial training window of ten years. *PROXRET* is available throughout our sample period. We backfill SCI to also obtain *SCIRET* for the whole sample. *J2J* data first became available at the end of 2000. We apply a standard six-month gap between the data availability and return calculations. Hence, *J2JRET* first enters our model in the training period that ends in June 2001. We backfill *J2JRET* for trainings after this date.

rank-normalized monthly peer returns in the following twelve-month period to compute the corresponding *SPFRET* for a focal stock. We repeat this process at the end of each June to obtain *SPFRET* for the period of July 1963 through December 2023.

For our long-run prediction analysis, we also consider *SPFMOM*, the long-run version of *SPFRET*, by interacting the weights above with the rank-normalized *SCIMOM*, *PROXMOM*, and *J2JMOM*, the long-run versions of *SCIRET*, *PROXRET*, and *J2JRET*, obtained by compounding the corresponding variable from month $t-11$ to $t-1$.

Composite Strategy Peer Firm Return We define the strategy peer firm return (*STRATRET*) as a composite strategy-related peer firm return using information from the three strategy peer firm returns, *TECHRET*, *TNICRET*, and *10KRET*.

STRATRET is computed using the same two-step process in *SPFRET*. That is, we recursively train OLS models using *TECHRET*, *TNICRET*, and *10KRET* as features (where they are available) and the contemporaneous focal stock return as the target. We then use the coefficients of these regressions as weights to form a composite strategy-linked peer return *STRATRET*.[14]

2.1.3 Firm Fundamentals

Following Parsons et al. (2020), we examine the comovement between focal firms and their peer firms' fundamental changes. We examine five firm fundamental variables, including ΔEPS, $\Delta Sales$, $\Delta Employees$, $NewCapital$, and $Capx$, 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 percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value.[15] $Capx$ is the capital expenditures scaled by sales.

[14] *TECHRET* is available throughout our sample (July 1963 to December 2023), and we backfill *TNICRET* and *10KRET* paying attention to when these two variables first become available. As *TNICRET* first becomes available in July 1989, our first model is trained over the data between July 1963 and June 1989 using *TECHRET* and the back-filled *TNICRET*. We then use the coefficients of this model to compose *STRATRET* over the following year. We retrain our model annually, expanding the training data set by one year in each training. Starting in June 1994, we add *10KRET* to the model, again with back-filling.

[15] 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).

2.2 Measuring Firm Strategy Similarities

As firm strategy is a complex concept that can influence many aspects of the firm's decisions, we use multiple measures to capture different facets of its 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).[16] Next, we multiply each word's topic-specific probability by its frequency in the paragraph, summing these products to obtain a paragraph-level score for each topic. 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).[17] 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.

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

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

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

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 using 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).[18] 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.

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

2.3 Controls

Lead-Lag Controls We create 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 2023 unless otherwise mentioned.

For a given stock and 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).[19] *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).[20] *CONGRET* is the pseudo-conglomerate 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 start from July 1977.

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

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

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

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).

2.4 Summary Statistics

Table **1** provides summary statistics (Panel A) and correlations (Panel B) for the pairwise variables used in our analyses, including *SCI, PROX, J2J* 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 all three county-pair social connectedness measures are positively related to measures of firm-pair similarity, such as strategy, product, and technology similarities. This provides preliminary evidence that social connectedness between firms' headquarters locations is associated with the strength of social ties between the locations.

Appendix Table A.2 presents the summary statistics of *SCIRET, PROXRET, J2JRET,* and *SPFRET*, together with stock returns and return predictors established in the previous studies. The distributions of 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 0.168 and 0.208, respectively. *SPFRET* is highly correlated with *CFRET* of Ali and Hirshleifer (2020), with a coefficient of 0.469. Furthermore, the average cross-sectional correlation between *SPFRET* and *INDRET* is considerable, at 0.640. This is likely due to the construction of

SPFRET, which relies on the SCI-weighted, same-industry firm returns. In subsequent analyses, we control for industry effects using measures constructed with both the Fama-French 48 industries and the TNIC-based industry classifications (Hoberg and Phillips 2010), respectively.

[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 stock returns. In this section, we empirically test these predictions.

3.1 Social Ties and Strategy Similarity

We formally examine the extent to which social connectedness between firms' headquarters relates to firms' strategy similarity using the following pairwise regression analyses:

$$\begin{aligned} Similarity_{i,j,t} = {} & \beta_1 SCI_{i,j,t} + \beta_2 PROX_{i,j,t} + \beta_3 J2J_{i,j,t} + \beta_4 Similar\ Size_{i,j,t} \\ & \alpha_t + \delta_{Ind} + \zeta_{County,i} + \eta_{County,j} + \epsilon_{i,j,t}, \end{aligned} \tag{2}$$

where *Similarity* is one of the three strategy similarity measures between the two firms, $STRATSIM$, $STRATSIM^{TECH}$, or $STRATSIM^{PROD}$. *SCI*, *PROX*, and *J2J* are proxies for the social connectedness between the firms' headquarters counties. We control for size-based homophily using an indicator variable that equals one if the sizes of the companies are within a ten-fold range (e.g., Hoberg and Neretina, 2023). We additionally control for the industry affiliation of the two companies, the headquarters counties, and time fixed effects. We limit our firm pairs so that they are in the same industry.[21]

[21] 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 (ranking is done within the industry in Eq.2). This non-parametric transformation allows the analysis to

This set of results is reported in Table **2**. We focus on the relationship between proxies for social connectedness and similarity measures. In columns 1-4, our dependent variable is the *STRATSIM* measure based on the 10-K filing. In column 1, we focus on the Facebook-based SCI measure. We find that *SCI* is strongly and positively related to strategy similarity. We find a strong positive relationship between *SCI* and the 10-K-based strategy similarity measure. Specifically, comparing firm pairs with *SCI* in the top quartile with those in the bottom quartile, the strategy similarity is 0.60 percentage points ($= 0.8\% \times 0.75$) higher, which corresponds to 9.0% of the standard deviation of the dependent variable.[22] This result confirms that stronger connections between firms' headquarters locations are associated with a greater level of overall firm strategy. In columns 2 and 3, we replace the *SCI* measure with *PROX* and *J2J*, respectively. We obtain very similar results. In column 4, the key independent variable is the average of the three connectedness measures. As expected, we find that this measure positively associates with the 10-K-based similarity measure. We also note that the coefficient of *Similar Size* is positive and highly significant, indicating that firms with similar sizes are likely to adopt similar corporate strategies, consistent with Hoberg and Neretina (2023).

Next, we focus on the two specific dimensions of firm strategy similarity, based on technology and product market, respectively. We consider companies' product market similarities (i.e., $STRATSIM^{PROD}$) and report this set of results in columns 5 to 8. As reported in column 5, 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 product market similarity is 1.3 percentage points higher, which corresponds to 19.9% of the dependent variable's standard deviation. When we consider the other two social connectedness measures, we find very similar results. In column 8, we find a strong positive relationship between the average connectedness measure and the product similarity measures.

In columns 9 to 12, we report the analyses 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.

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).

[22] Since *SCI* is rank-normalized between 0 and 1, we multiply the coefficient by 0.75 to assess the corresponding relationship moving from the bottom to top quartile.

Specifically, comparing firm pairs with *SCI* in the top quartile with those in the bottom quartile, the technology similarity is 2.9 percentage points higher, which corresponds to 9.4% of the dependent variable's standard deviation.

In columns 10 and 11, we repeat this analysis with *PROX* and *J2J*, and our conclusion remains unchanged. In column 12, we include the average connectedness measure as the independent variable. We show a strong positive relationship between this measure and the similarity measure.

Finally, to summarize the key results for this set of analyses, we create a new dependent variable by taking the average of all three strategy similarity measures. We find that the average connectedness measure, again, strongly positively correlates with the average strategy similarity measure.[23]

In summary, our results suggest that social connectedness is associated with higher strategy similarities, across all proxies of social connectedness. This relationship is not entirely driven by size-based homophily.

[Insert Table **2** here]

3.2 The Comovements of Firm Fundamentals and Stock Returns

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 headquartered in socially connected areas tend to comove more strongly in their fundamentals and returns. Following Parsons et al. (2020),

[23] Another way to view product market similarity is a more refined alternative to the Fama-French industry classification approach. We report regressions with 10-K-based similarity and technology similarity on as dependent variables and include the product market similarity measure as a control. We find that our results that social connectedness positively relates to firm similarity measures remain robust.

we conduct the panel regressions:

$$
\begin{aligned}
Fundamental_{i,j,t} = {} & \beta_1 Fundamental_{SCI,i,j,t} + \beta_2 Fundamental_{J2J,i,j,t} \\
& + \beta_3 Fundamental_{PROX,i,j,t} + \beta_4 Fundamental_{j,t} \\
& + \beta_5 Fundamental_{TNIC,i,t} + \beta_6 Fundamental_{TECH,i,t} + \beta_7 Fundamental_{10K,i,t} \\
& + \alpha_t + \epsilon_{i,j,t} \qquad (3) \\
RET_{i,j,t} = {} & \beta_1 RET_{SCI,i,j,t} + \beta_2 RET_{J2J,i,j,t} \\
& + \beta_3 RET_{PROX,i,j,t} + \beta_4 RET_{j,t} \\
& + \beta_5 RET_{TNIC,i,t} + \beta_6 RET_{TECH,i,t} + \beta_7 RET_{10K,i,t} + \alpha_t + \epsilon_{i,j,t} \qquad (4)
\end{aligned}
$$

The dependent variable $Fundamental_{i,j,t}$ is a fundamental measure for a focal firm i, in the Fama-French 48 (FF48) industry j, and at time t. We consider five fundamental measures, ΔEPS, $\Delta Sales$, $\Delta Employees$, $NewCapital$, and $Capx$ measured at the annual frequency. We also consider monthly firm returns as a key dependent variable.[24]

Our key independent variables measure the changes in the fundamentals and returns of FF48 industry peers, weighted by their social and economic connectedness to the focal firm's headquarters. Specifically, $Fundamental_{SCI}$ (RET_{SCI}), $Fundamental_{J2J}$ (RET_{J2J}), and $Fundamental_{PROX}$ (RET_{PROX}) represent these changes weighted by SCI, J2J, and PROX, respectively. To isolate the effect of distant network connections and mitigate the influence of geographical co-location (Pirinsky and Wang (2006); Parsons et al. (2020)), we intentionally exclude firms located in the focal firm's headquarters state.

To isolate the unique impact of connectedness, we implement several controls for alternative sources of comovement. To account for standard industry affiliation, we control for the average change in industry fundamentals ($Fundamental_{j,t}$) and returns ($RET_{j,t}$), considering both equal- and value-weighted portfolios.[25] Furthermore, we control for contemporaneous strategy-linked peer fundamentals and returns to ensure that social ties provide additional expplanatory power beyond strategy similarity. This includes $Fundamental_{TNIC}$ (RET_{TNIC}) for peers within the same TNIC category, $Fundamental_{TECH}$

[24] In columns 1-5, 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.

[25] We exclude the focal firm when forming the equal- and value-weighted industry peers.

(RET_{TECH}) for *TECH*-weighted peers, and $Fundamental_{10K}$ (RET_{10K}) for peers weighted by 10-K text strategy similarities. Finally, we include time fixed effects (α_t) to absorb unobserved aggregate shocks.

In Table 3 Panel A, we examine the relationship between social connectedness-weighted portfolios and focal firms across multiple financial and operational metrics, while controlling for equal-weighted FF48 industry portfolios. Column 1 reveals significant comovement in EPS growth between focal firms and their social peer firm portfolios based on all three types of social proximity weighting. The economic impact is substantial: focal firms with the highest *SCI*-weighted firm EPS growth demonstrate 1.81 percentage points higher ΔEPS than those with low social peer firm ΔEPS. The effects of *PROX* and *J2J* weighting are also highly significant both economically and statistically. Focal firms with the highest *PROX*-weighted (*J2J*-weighted) firm EPS growth demonstrate 1.89 (1.52) percentage points higher ΔEPS than those with low social peer firm ΔEPS. This earnings comovement could be driven by both demand-side (i.e., customer demand) and supply-side (i.e., firm investment and production) factors. On the demand side, we rely on sales as a proxy for customer demand for the firm's product. As shown in column 2, firms with the highest *SCI*-weighted $\Delta Sales$ exceed those with the lowest social peer firm sales growth by 11.06 percentage points, a pattern that remains robust across *J2J* and *PROX*-weighted industry peer portfolios. These results may indicate that firms located in socially connected areas may have similar product market overlaps, and thus the demand from their clientele becomes more correlated.

On the supply side, changes in employment, new capital, and capital expenditure indicate firms' strategic choices in expanding their production capacity. We observe significant comovement in firms' strategic variables, including changes in employee numbers, newly raised capital, and firm investment (measured by capital expenditure), as reported in columns 3 through 5.

Given strong comovements between focal firm fundamentals and peer firms weighted by social connectedness, we expect that focal firms' returns comove strongly with those portfolios. To investigate this hypothesis, we conduct the regression 3 with monthly returns as the dependent variable of interest. We report this result in column 6. We find that across the board, social ties-weighted portfolios exhibit strong comovements with focal firms. For example, firms with the highest SCI-weighted firm returns outperform those with the lowest social peer firm returns by 5.43%. 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.

In regressions involving firm fundamentals (Columns 1-5), equal-weighted industry portfolio has the smallest coefficient among all independent variables. In Panel B, we conduct the same analyses by replacing the equal-weighted industry portfolio with its value-weighted version, and our results remain robust. These findings suggest that the social connectedness-weighted peer portfolios are more tightly connected with focal firms in their fundamentals than equal- or value-weighted industry peers.[26]

[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 connectedness peer portfolios. 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 socially connected peers, it is possible that their returns could predict the future performance of focal firms. In this section, we conduct a formal analysis to explore whether the returns of socially connected peers can serve as a predictor of focal firms' returns. We begin by presenting evidence that a long-short portfolio sorted by lagged socially connectedness-weighted peer firm returns yields significant abnormal returns. We then examine the extent to which any abnormal returns can be explained by the existing predictive relationships.

[26] In Panel A and B of Appendix Table IA2, we report the univariate relationship between equal- and value-weighted industry portfolio accounting and return metrics and that of the focal firm. We find that the relationship between industry portfolios (equal or value-weighted) and focal firms is positive and strong. Together with the main results presented in Table **3**, this suggests that such positive relationships are likely subsumed by industry portfolios weighted by SCI, PROX, or J2J measures. To alleviate concerns of multicollinearity, in Panel C of Appendix Table IA2, we further report the results of the regressions in Table **3** without equal- or value-weighted industry controls. SCI-weighted variables remain significant except ΔEPS, which has a coefficient of 0.64 with t-statistics of 1.54. All other SCI-weighted metrics remain statistically and economically significant.

4.1 Portfolio Analysis

We first perform portfolio analysis and investigate the returns to a long-short portfolio sorted by *SPFRET* and its components. Specifically, for each month, we sort our sample firms into deciles based on *SCIRET*, *PROXRET*, and *J2JRET*. These three measures represent the lag industry peer returns weighted by *SCI*, *PROX*, and *J2J*, respectively. Additionally, these three measures may provide important information about future returns. Therefore, we also construct a composite return measure based on all three social connectedness-weighted return variables. We name this variable social peer firm returns *SPFRET*. We then hold the portfolios for a month and calculate the returns for each portfolio and the return differential between the portfolios with the highest and the lowest values of the sorting variable.

Table 4 reports the results of the univariate portfolio analyses. The first row in each panel 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 and French (2015) five-factor alphas (FF5) for equal-weighted portfolios in the second row. The third and fourth rows report the raw returns and FF5 alphas for value-weighted portfolios, respectively.

In Panels A, B, and C, we report the portfolio returns based on *SCIRET*, *PROXRET*, and *J2JRET*, respectively. We find that the social connectedness-weighted peer firm returns offer significant power in predicting future focal firm returns. For example, the *SCIRET*-sorted long-short portfolio produces an FF5 alpha of 161 basis points per month (t = 7.71) for the equal-weighted portfolio and 67 basis points (t = 4.14) for the value-weighted portfolio. The *PROXRET*-sorted long-short portfolio generates 154 (69.9) bps FF5 alpha when we apply equal (value) weighting. The *J2JRET*-sorted portfolio generates significant alpha when we use equal weighting, but the economic and statistical significance of the outperformance weakens once we consider value-weighted strategies.[27]

In Panel D, we use the combined social peer firm return measure (*SPFRET*) as the sorting variable.[28] We show that the *SPFRET*-sorted long-short portfolio produces eco-

[27]The weaker performance of *J2JRET* is mostly due to its being available after July 2001. The other predictors are available from July 1963.

[28]textitSPFRET is constructed as an OLS-weighted combination of *SCIRET*, *PROXRET*, and *J2JRET*, where the weights are estimated by regressing the contemporaneous focal stock return on the three (rank-normalized) peer returns and normalizing the coefficients to sum to one. The model is re-estimated annually using all data available up to that point. Figure 1 plots these time-varying weights and shows that

nomically large and highly significant excess returns. For example, the FF5 alpha is 164 basis points per month (t = 7.67) for the equal-weighted portfolio and 73 basis points (t = 4.33) for the value-weighted portfolio.

[Insert Table 4 here]

To provide a temporal perspective of the strategy returns, we present Figure 2, which exhibits the average monthly long-short portfolio alpha over time for a *SPFRET* strategy. 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 also examine the incremental return predictability of *SPFRET* by cross-sectionally ranking *SPFRET* within an industry and forming long-short portfolios using this within-industry version of *SPFRET*. We find that returns are positive for most of the sample period. Given the skewed distribution of firm sizes, we also remove the largest 100 firms to reduce the influence of the largest firms, similar to Jensen et al. (2021). We then present the time series of portfolio returns in Panel B. This removal enhances the performance of our strategy in the post-2000 period. Overall, the strategy based on *SPFRET* shows consistent performance throughout the sample.[29]

[Insert Figure 2 here]

Our *SPFRET* portfolio is constructed using peer firms in the focal firm's industry. 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* and its components 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 *SCIRET*, *PROXRET*, *J2JRET*, and *SPFRET*. We calculate the returns for each of the 25

SCIRET receives the largest and most stable weight throughout the sample, consistent with SCI being the most informative of the three social connectedness measures.

[29] The overall decline in SPFRET profitability toward the end of our sample period is possibly attributable to learning in markets (e.g., McLean and Pontiff, 2016). Figure IA1 plots the alphas of key cross-firm return prediction anomalies. As reported in Panel A, we find that none of the value-weighted anomaly portfolios is significant post year 2000. In Panel B, we exclude the top 100 stocks in market capitalization, and only *SPFRET* shows a statistically significant profit post-2000.

portfolios in the month that follows, along with the returns of five high-minus-low portfolios based on the sorting variable for each *INDRET* quintile.

Table 5 reports the results of the bivariate portfolio sort with equal-weighted and value-weighted abnormal returns. For each of the sorting variables, we report in a separate row the FF5 alphas of the lowest and highest quintile portfolios, as well as the high-minus-low portfolios, averaged across all five *INDRET* quintiles. We find that equal-weighted strategies based on all three social connectedness measures continue to generate significant alphas after controlling for industry momentum. Value-weighted strategies continue to generate significant alphas in both *SCIRET* and *PROXRET*-based strategies.

Finally, when we combine the three predictive variables into *SPFRET*, we find that its predictability remains robust after controlling for industry momentum. The high-minus-low average alpha of the equal-weighted portfolio for *SPFRET* is 55 basis points per month with a *t*-statistic of 8.15 while the economic magnitude under value-weight is 32 basis points with a *t*-statistic of 4.52. 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).[30]

Given that *SPFRET* encompasses the information from all three socially connected peer firm return measures, we will focus on this variable in the remainder of our analyses.

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

[30] Another way to control for industry momentum is to sort *SPFRET* within each industry. We report these results in Table IA3. Consistent with the double-sort results presented in Table 4, *SPFRET*-based strategy still generates significant abnormal returns.

contribution of *SPFRET* to the predictive power in a machine learning-based composite predictor that encompasses all individual predictors.

4.2.1 Spanning Tests and Portfolio Comparisons

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. We then directly compare the composition and the performances of these portfolios.

Spanning Tests We follow Ali and Hirshleifer (2020) and estimate the following time-series regression:

$$RET_{SPFRET,t} = \alpha + \beta F_t + \epsilon_t, \tag{5}$$

where RET_{SPFRET} is the size-controlled 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. Specifically, we consider the following long-short portfolios.

We first consider the industry momentum strategy based on the FF 48 industry (i.e., *INDRET*). Industry momentum is highly relevant as *SPFRET* is constructed using firms from the same industry. Second, we include the geographic momentum documented in Parsons et al. (2020). Third, 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. Next, we include the lead-lag effect due to the customer-supplier relationship (Cohen and Frazzini, 2008). 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. We also consider 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. Finally, we include 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.

To ensure that the coefficients are comparable across different variables, we focus on a sample period of July 1989 through November 2023, so that return information exists for all the portfolios considered in our test. We report the alphas of the lead-lag portfolios in Figure 3.[31] 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]

Figure 4 further illustrates the relative contributions of the economic linkage-based variables.[32] The figure shows that the lead-lag effects documented in the existing literature explain 66% of SPFRET-based returns, and *CFRET* is the most prominent variable in this comparison, which explains 34%. *TNICRET* and *INDRET* are the two other prominent variables, explaining 15% and 21% of the *SPFRET*-based portfolio return respectively.

[Insert Figure 4 here]

In sum, the tests show that social connectedness between firm locations corresponds to important economic ties across firms, such as industry or product similarity, customer-supplier links, and technology similarity. But our social tie-based peer firm measure also captures a substantial amount of cross-firm linkages that existing measures do not account for. We further explore this next.

[31] These returns are constructed using only the lead-lag variable, without considering firm sizes.

[32] In Table IA4, we report the full spanning results, by adding the control portfolio returns gradually into our regressions. Overall, we show that our *SPFRET*-based strategy generates positive alpha regardless of controls. In the bar plot, we multiply the coefficient estimates from the last column of Table IA4 and the average returns of the corresponding long-short portfolios. We then scale this product by the average return of the *SPFRET* portfolio to gauge the contribution of the relevant factor to the *SPFRET*-based strategy.

Portfolio comparison We analyze the degree to which peer portfolios formed using *SPFRET* differ from those constructed using alternative strategy-related predictors. We conduct a direct comparison of value-weighted long-short portfolios sorted on *SPFRET* against those sorted on text-based network industry returns (*TNICRET*), technology similarity returns (*TECHRET*), 10-K text similarity returns (*10KRET*), strategic similarity returns (*STRATRET*), and standard industry returns (*INDRET*). Due to data availability for *10KRET*, our sample period for this analysis is July 1994 to December 2023.

First, we assess the similarity in portfolio composition by calculating the average monthly cosine similarity of portfolio weights. The analysis reveals only a moderate to low degree of overlap. The cosine similarity between the *SPFRET* portfolio and the portfolios based on *INDRET*, *STRATRET*, *TNICRET*, T*ECHRET*, and *10KRET* is 0.53, 0.29, 0.25, 0.17, and 0.14, respectively. This confirms that the various signals lead to the construction of markedly different portfolios.

Next, we compare their performance. While the correlation between the Fama-French five-factor abnormal returns of the *SPFRET* portfolio and those based on *INDRET*, *TNICRET*, *TECHRET*, and *10KRET* is relatively high (0.82, 0.81, 0.67, and 0.62, respectively), a comparison of risk-adjusted returns tells a more complete story. The trading strategy based on *SPFRET* delivers the highest Sharpe ratio of 0.19. This is followed by the strategies based on *TNICRET* (0.17), *STRATRET* (0.16), and *10KRET* (0.10). In contrast, the *TECHRET* strategy yields a much lower Sharpe ratio of 0.04, and the *INDRET* strategy results in a negative Sharpe ratio of -0.10. Taken together, these results indicate that while there is a common component across the predictors, the portfolio derived from social peer returns is distinct and generates superior risk-adjusted performance.[33] In the next section, we provide a formal test of this incremental return predictability in a multivariate setting.

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

[33] Explanations for the superior predictive power arise because *SPFRET* either measures these similarities with less error or captures subtle strategic aspects that are more difficult for market participants to observe through conventional data.

stock level while accounting for a comprehensive set of controls.

We estimate the following regression:

$$RET_{i,t+1} = \alpha + \beta Predictor_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1}, \tag{6}$$

where $RET_{i,t+1}$ is the one-month-ahead firm return. *Predictor* is a predictive variable. The key predictor of interest is *SPFRET*, or the social peer firm returns. We compare the results based on the *SPFRET* with other firm similarity-based predictive variables, including the technological spillover effect (*TECHRET*), the text-based industry momentum (*TNICRET*), the lagged return based on 10-K similarity (*(10KRET)*) and the composite measure based on *TECHRET, TNICRET*, and *10KRET*, which we call *STRATRET*. *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*), and the complicated firm effect (*CONGRET*). 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 1.16 percentage point higher average return in the following month.

In columns 2 to 4, we include *TECHRET*, *TNICRET*, and *10KRET* as predictors, respectively. We find that all three predictors exhibit a positive relationship with next month's return, with *TNICRET* and *10KRET* showing statistically significant coefficients. In column 5, we include the composite strategy return measure (*STRATRET*) as the predictor. We find that combining the three strategy similarity-based return predictors exhibits strong power in predicting the following month's return.

In columns 6 through 9, we compare the predictive power of *SPFRET* and another strategy-based return measure. We find that *SPFRET* continues to exhibit strong power in predicting the following month's return. In column 9, we include both *SPFRET* and the composite *STRATRET* measure, and we find that *SPFRET* coefficient drops in magnitude to 0.80, but is still significant at the 1% level.

In column 10, we further include other lead-lag measures in the regression. We show that *SPFRET* robustly predicts the following month's return. In terms of the economic magnitude, moving from firms with the lowest *SPFRET* to the highest, the next month's return is expected to increase by 29 bps.

In sum, the comprehensive set of strategy similarity-based return predictors and predictors based on other economic linkages 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 (6) through partial least squares (PLS).[34]

To obtain our results, starting in July 1999, we train new partial least squares (PLS) models every twelve months, using all the data available up to that point. 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-

[34]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).

weighted returns of the corresponding portfolios. We then compute the cumulative log returns over the period of August 1973 through December 2023.[35]

Figure 5 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. Over a sample period of 46 years from August 1973 to December 2023, the value generated by the former portfolio is almost twice the one generated by the latter, which corresponds to a 1.39% per year outperformance in terms of log returns. In terms of raw returns, the LS portfolio based on the full model outperforms by 1.61% per year.

[Insert Figure 5 here]

5 Drivers of Return Predictability: Underreaction vs. 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.[36] 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.

[35] We use a 10-year initial training window for PLS. Therefore, the composite predictor is first available for July 1973.

[36] 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.

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 116 basis points per month for small firms, which is over twice as large 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, the 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 outside 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 three 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 (150 basis points based on *SPFRET* and 68 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 22 and 53 basis points, respectively. The effect of *SPFRET* remains positively associated with returns for months 4–6 and 7–12, and we see no reversals after the first twelve months.[37]

The 12-month cumulative return for the long-short portfolio sorted by *SPFRET* is 3.91%. The long-short strategy based on *SPFRET* corresponds to a contemporaneous return of 7.61%. Compared to the contemporaneous effects, the cumulative return indicates that 34% of the information in *SPFRET* is associated with a delayed reaction.[38] Panel B presents the Fama and French (2015) five-factor alpha for value-weighted portfolios. The

[37]In Panel A of Table IA5, we conduct these analyses after orthogonalizing *SPFRET* and *SPFMOM* with respect to all return variables and other controls in Table **6**. In case of *SPFMOM*, we also orthogonalize with respect to the long-run versions of the return variables. Although the results are weaker than those for the raw versions, there is still no sign of return reversal at least 12 months after portfolio formation.

[38]The delayed reaction is calculated as $3.91\%/(7.61\% + 3.91\%)$.

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 firms' future fundamental performance. 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:

$$Fundamental_{i,t+s} = \alpha + \beta_1 SPFRET_{i,t} + \beta_2 SPFMOM_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+s}, \tag{7}$$

where *Fundamental* represents a fundamental-related variable. *SPFRET* and *SPFMOM* are social peer firm returns and momentum, respectively.[39] X is a vector of control variables, including long-term and short-term industry momentum, *INDMOM* and *INDRET*, as well as time and firm fixed effects. We report these results in Table **9**. 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.[40] Columns 1-4 of Table **9** Panel A

[39] 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. Panel B of Table IA5 presents the results and shows that our findings are robust.

[40] 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

reveal that *SPFMOM* and *SPFRET* strongly predict future SUEs. Specifically, an increase from the lowest *SPFMOM* to the highest leads to a 33.8 percentage points increase in the following quarter's SUE, which corresponds to 25.3% 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. Table 9 Panel B shows that the results are robust to controlling for short- and long-term industry momentum.

[Insert Table 9 here]

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 errors in quarters $t+1$ to $t+4$ as the dependent variables.[41] As reported in columns 5-8 of Table 9 Panel A, both *SPFMOM* and *SPFRET* predict analyst forecast errors over the following year. Table 9 Panel B shows that the *SPFMOM* still robustly predicts analyst forecast errors even after including industry momentum controls. While the coefficients of *SPFRET* remain positive, they become insignificant. 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

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.

[41] 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.

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 columns 9-12 of Table **9** Panel A, we find that both *SPFRET* and *SPFMOM* are strongly related to earnings announcement returns in the following quarter.In terms of economic magnitude, companies with the highest *SPFRET* realize a following-quarter CAR that is 11.1 basis points higher than those with the lowest. Similarly, companies with the highest *SPFMOM* outperform those with the lowest by 4.3 basis points. As indicated in Panel B, while these coefficients have become insignificant after controlling for industry momentum variables, both variables still exhibit positive relationships with future earnings announcement returns.

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 analyses using alternative specifications.

6.1 Predictability Post 2000

Consistent with the 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. One reason for this is that industry momentum (*INDRET*) is substantially weaker and becomes insignificant post-2000 (Ali and Hirshleifer, 2020). Indeed, as shown in Panel A of Figure IA1, this is the case for all of the lead-lag predictors we studied. In Panel B, 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.

Panel B shows that, in contrast to the value-weighted industry momentum portfolio, the *SPFRET* portfolio still exhibits significant abnormal returns of 63 basis points, which is only slightly smaller than the full sample alpha of 73 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 IA6 presents the results, with column 1 the same as column 9 of Table **6** to facilitate the comparison. Since we are now using the composite variable *SPFRET* rather than its individual components, and *SPFRET* in a given month is constructed using the available components, we can conduct our analyses using a longer sample period. Specifically, the sample we use for the analyses below runs from January 1982 to November 2023 and is only constrained by the availability of *CFRET*, one of the peer return controls used in the regressions.

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 64.0%. The correlation between *SPFRET* and the value-weighted *INDRET*, on the other hand, is 54.1%. Hence, the equal-weighted *INDRET* provides a more stringent control and as such, presumably provides more conservative results. Column 2 presents the results when value-weighted *INDRET* is used instead. Our result remains robust.

Constructing *SPFRET* without SCI One concern regarding the use of *SPFRET* is that one of its components, *SCIRET*, 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 drop *SCIRET* and retrain an OLS with only *PROXRET* and *J2JRET* to obtain another version of the social peer firm return, $SPFRET_{No\ SCI}$. Column 3 shows that $SPFRET_{No\ SCI}$ significantly predicts future returns, with a comparable coefficient of 0.532. This suggests that one can use geographic proximity to proxy for social ties and generate significant abnormal profits in real-time portfolios.[42]

Firms' Economic Presence Beyond Headquarters In column 4, 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 and reconstruct *SPFRET*. We also still exclude peers headquartered in the same state as the focal firm. This exercise is highly conservative, as it eliminates 20% of out-of-state 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. Accordingly, we control for lagged portfolio returns that are weighted based on the similarity between the focal county and peer firms' counties with respect to five county characteristics: population, education, political inclination, income, and unemployment rate.[43] In column 5, 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

[42] The average cross-sectional correlation between *SPFRET* and $\mathrm{SPFRET}_{\mathrm{No\ SCI}}$ is 97.2%. Despite this relatively high correlation, when we include both *SPFRET* and $\mathrm{SPFRET}_{\mathrm{No\ SCI}}$ in the same regression, *SPFRET* remains highly significant at 1.075 with t-statistics of 4.325. In comparison, the coefficient of $\mathrm{SPFRET}_{\mathrm{No\ SCI}}$ is much smaller, at 0.401, and is also only marginally significant with a t-statistic of 1.712. 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.

[43] For each county, we form a five-dimensional vector based on these five socioeconomic variables. Each dimension is normalized to a value between 0 and 1 following Gu et al. (2020). *SIMRET* is then constructed as the weighted out-of-state industry peer firm returns, where the weight assigned to a peer firm is one minus the Euclidean distance between focal and peer counties in this five-dimensional space.

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 headquarters state in constructing *SPFRET*. In column 6, we further exclude peer firms that are located within a 100-mile radius of the focal firm's headquarters to construct *SPFRET*. Column 7 then examines 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 8 uses 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 9, we use an alternative standardization procedure in which independent variables are demeaned and then divided by the variable's standard deviation, and find the results to be robust.[44] In column 10, 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 "same-state" version of *SPFRET*, which is constructed by using only the industry peers from the same state as the focal firm, excluding the focal firm itself. Appendix Table IA7 presents the results. We find that, when used together in a Fama-MacBeth regression with other controls, 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 significantly smaller than *SPFRET*. This result further confirms that our results are not driven solely by the return predictability of geographically proximate industry firms.

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

6.3 Enhancement Effects for Other Predictors

Next, we examine whether social ties can enhance the effectiveness of other lead-lag strategies. Building on previous literature, we construct predictive signals using return information from product market overlaps (Hoberg and Phillips, 2010), shared analyst coverage (Ali and Hirshleifer, 2020), and industry input-output links (Menzly and Ozbas, 2010). We then augment these baseline predictors by weighting the lagged returns of connected firms according to the strength of their social ties to the focal firm.

We first introduce $TNICRET_{SPF}$. This measure is derived by first selecting peer firms that share the same TNIC category as the focal firm. We compute $TNICRET_{SCI}$, $TNICRET_{PROX}$, and $TNICRET_{J2J}$, which use *SCI*, *PROX*, and *J2J* as weights. These variables are then combined to obtain $TNICRET_{SPF}$ using the same weights that are applied to the components of $SPFRET$.[45]

Similarly, $CFRET_{SPF}$ is formulated by first forming measures based on three social connectedness measures as weights to firms share analyst coverage with the focal firm. We then combine the three measures into $CFRET_{SPF}$ using the $SPFRET$ weights.

Finally, $IORET_{SPF}$ is the composite of SCI-, $PROX$-, and $J2J$-weighted $IORET$, which is a weighted average peer return where the weights are obtained using the industry links from the Historical Input-Output Accounts database of the Bureau of Economic Analysis (BEA).

We posit that these social connectedness-enhanced measures, $TNIC_{SPF}$, $CFRET_{SPF}$, and $IORET_{SPF}$, could augment the predictive capability of trading strategies, even in the presence of existing predictors. Table A.3 reports the results of the dependent bivariate portfolio sorting based on these two sets of variables, where stocks are first sorted into quintiles on the original (non-social connectedness-weighted) variable and then on the socially connectedness-weighted version within each of these quintiles. We then calculate the value-weighted returns for each of the resultant 25 portfolios (per variable set) over the month following the portfolio formation and report the average FF5 alphas for the lowest and highest SPF-weighted quintiles, as well as the 5-1 LS portfolios. Results show that controlling for *TNICRET*, $TNICRET_{SPF}$-sorted 5-1 portfolio commands an FF5 alpha

[45]To be consistent with the methodology that we use to form *SPFRET*, we also exclude peers that are headquartered in the same state as the focal firm. This approach is aimed at capturing the nuanced interactions based on geographical and product market proximity.

of 19 bps per month, with the return differential most significant for the lowest *TNICRET* quintile. Turning to *CFRET* as the return predictor, its $CFRET_{SPF}$-sorted counterpart earns an alpha of 38 bps per month controlling for *CFRET*, with the strongest result coming from the lowest and highest *CFRET* quintiles. Finally, $IORET_{SPF}$-sorted portfolio earns an alpha of 21 bps per month controlling for *IORET*. Together, the results suggest that augmenting the product market, shared analyst, or industry input-output related return predictors with social connectedness enhances the power of these predictors.

Panel B of Table A.3 uses Fama-MacBeth regressions to further control for the other return predictors. In columns 1, 2, and 3, we regress next month returns of focal stock on the original *TNICRET*, *CFRET*, and *IORET* as defined earlier in Section 2.3. In columns 4, 5, and 6, we replace the original variables with their social connectedness-weighted counterparts, $TNICRET_{SPF}$, $CFRET_{SPF}$, and $IORET_{SPF}$. We find that both the original strategies and the social connectedness-weighted strategies significantly predict the next month's returns. In columns 7, 8, and 9, we include both the original and the social connectedness-weighted versions of the two strategies. Column 7 of Table A.3 shows that, when used together with *TNICRET*, the lowest-to-highest increase in this month's $TNICRET_{SPF}$ gives a 6.7 bps increase in next month's focal firm return. However, the effect is not statistically significant. In the case of shared-analysts-based strategy, column 8 shows that $CFRET_{SPF}$ remains economically and statistically highly significant in the presence of *CFRET*, with an economic magnitude of 22% of that of *CFRET*. Finally, column 9 shows that $IORET_{SPF}$ is stronger than the original variable, *IORET*, both economically and statistically. The results suggest that, in the case of *CFRET* and *IORET*, the social connectedness-weighted variables have incremental information that is not included in the original version.

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 adopt three social connectedness measures based on online social networks, geographic proximity, and labor flows. 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 social connectedness 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.

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 connectedness, 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 underreact 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 socially connected industry peers are 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. Importantly, applying social connectedness as weights to strategies based on product market overlap, shared analyst coverage, and industry input-output links generates significant incremental return predictability beyond the original strategies.

Our findings raise a number of issues that suggest future avenues of continuing research. Most importantly, it illustrates that the nature of the individuals who 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 whether it is the communication between managers that causes 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 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.

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.

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.

Ellison, Glenn, and Edward L Glaeser, 1997, Geographic concentration in us manufacturing industries: a dartboard approach, *Journal of political economy* 105, 889–927.

Ellison, Glenn, and Edward L Glaeser, 1999, The geographic concentration of industry: does natural advantage explain agglomeration?, *American Economic Review* 89, 311–316.

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.

Fracassi, Cesare, 2017, Corporate finance policies and social networks, *Management Science* 63, 2420–2438.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *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 Ekaterina Neretina, 2023, Do trade associations matter to corporate strategies?, *Available at SSRN 4575314* .

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *The Review of Financial Studies* 23, 3773–3811.

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, Jeffrey D Kubik, and Jeremy C Stein, 2004, Social interaction and stock-market participation, *The journal of finance* 59, 137–163.

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.

Krugman, Paul, 1991, Increasing returns and economic geography, *Journal of political economy* 99, 483–499.

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.

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

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

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.

Storper, Michael, and Anthony J Venables, 2004, Buzz: face-to-face contact and the urban economy, *Journal of economic geography* 4, 351–370.

Stoyanov, Andrey, and Nikolay Zubanov, 2012, Productivity spillovers across firms through worker mobility, *American Economic Journal: Applied Economics* 4, 168–198.

Figure 1: Composition of *SPFRET*

The figure depicts the weights of *SCIRET*, *PROXRET*, and *J2JRET* in *SPFRET* as a function of time. The weights are obtained by regressing the next month's focal stock return on this month's *SCIRET*, *PROXRET*, and *J2JRET* using OLS and normalizing the sum of the regression coefficients to one. The initial training period runs from July, 1953 to June, 1963. The model is re-trained at the end of each June thereafter, using all available data from July 1953 to that point. Due to the availability of data on job-to-job employee flows, *J2JRET* first enters the model in the training period that ends in June, 2001. We backfill *J2JRET* by assigning the job-to-job employee flows data from year 2000 to all months prior to June, 2001.

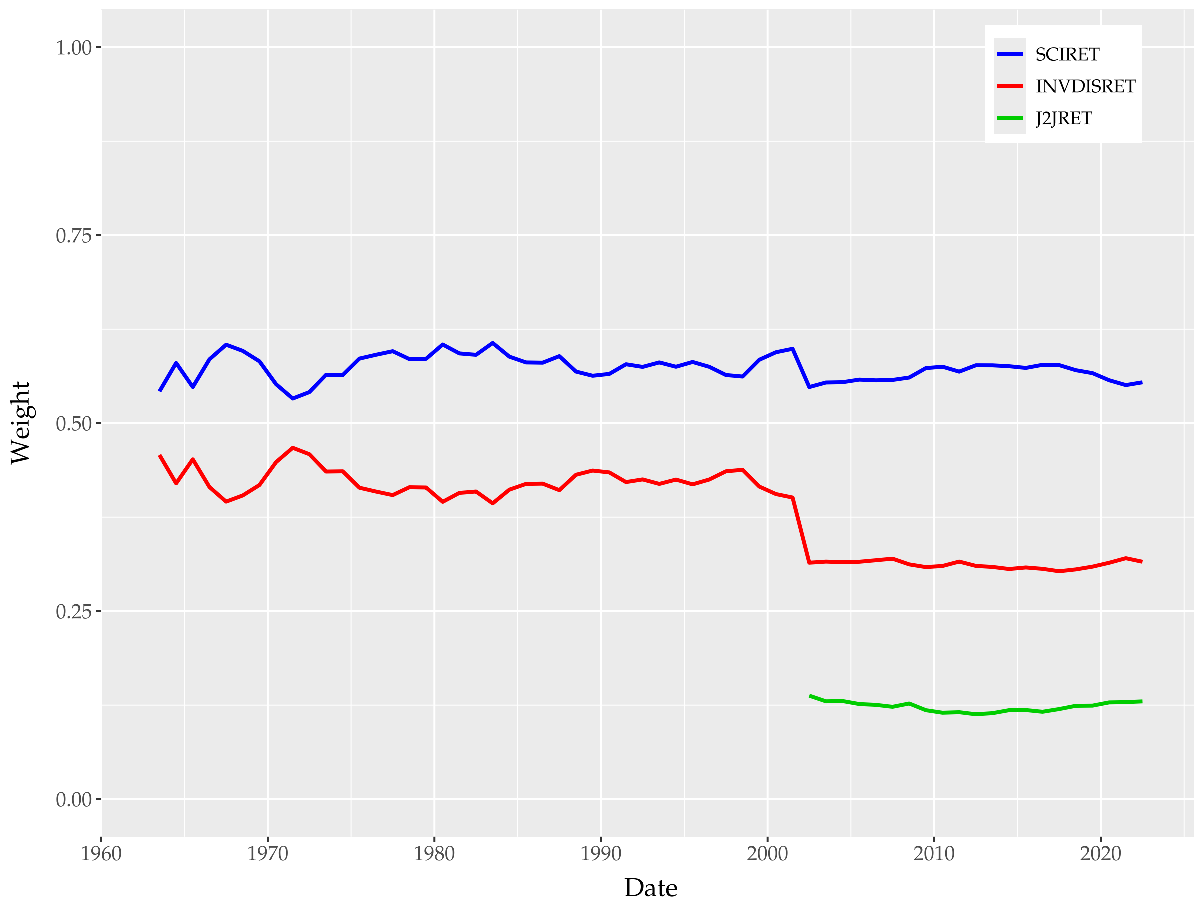


Figure 2: *SPFRET* Strategy Returns

The figure depicts for each year in our sample the average Fama-French five-factor (FF5)-adjusted abnormal returns to a long-short (LS) portfolio based on *SPFRET*.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. Panel A uses all stocks in a given month after the usual price, exchange code, and share code filters, while Panel B excludes the largest 100 stocks. All returns are in percentages.

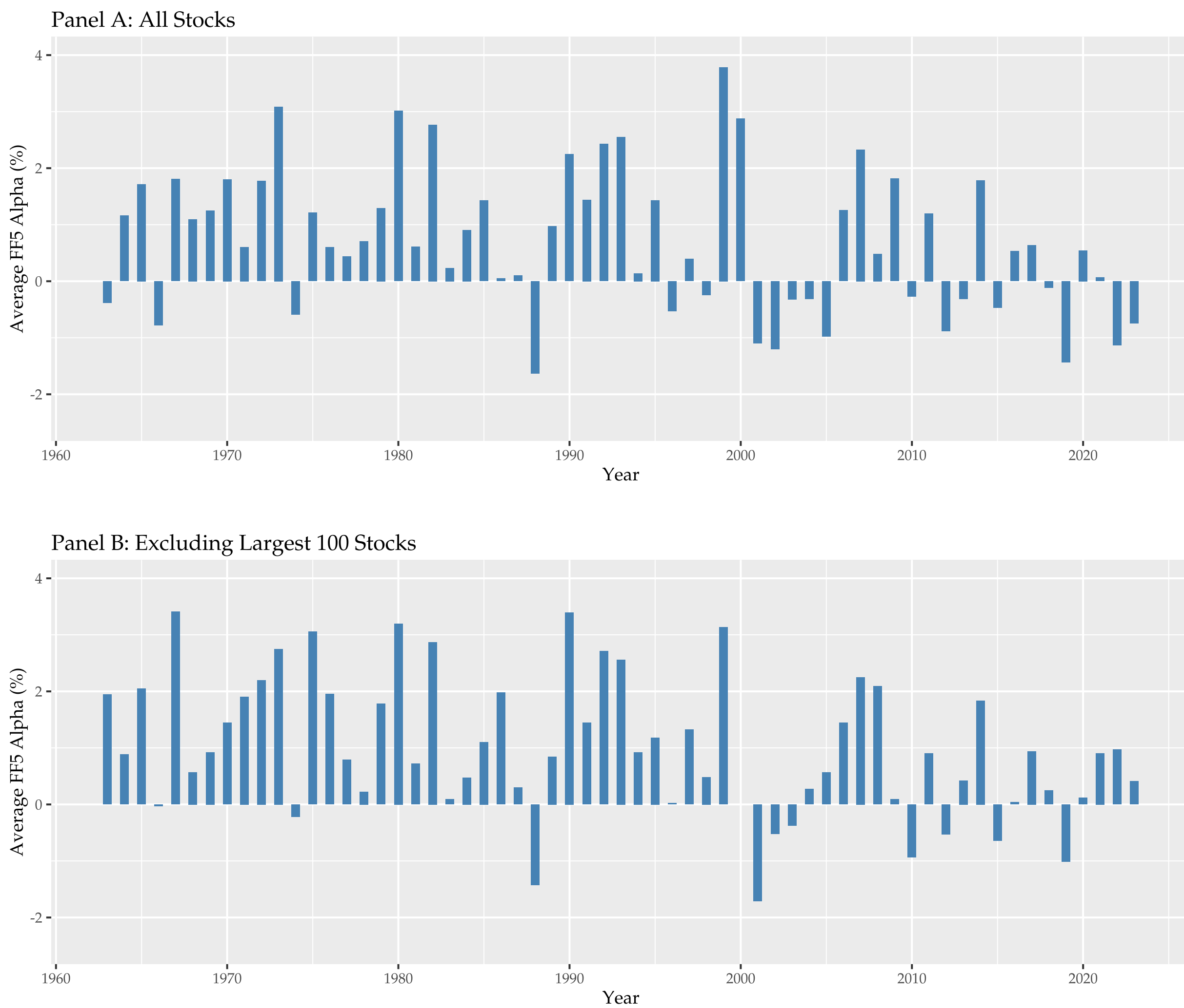


Figure 3: Lead-lag Strategy Performance (1982–2023)

The figure shows the Fama-French five-factor portfolio alphas of long-short strategies based on the lead-lag predictors (*SPFRET*, *INDRET*, *GEORET*, *CRET*, *CONGRET*), *STRATRET*, and *CFRET*. The orange bars show the 90% confidence intervals. The portfolios are value-weighted. *SPFRET* refers to the composite peer firm return, obtained by optimally combining *SCIRET*, *PROXRET*, and *J2JRET* to predict next month's focal firm return. *INDRET* is the industry return. *CRET* is the customer return. *GEORET* is the geographic momentum. *CFRET* refers to shared analyst coverage returns. *STRATRET* refers to the composite strategy return with *TECHRET*, *TNICRET*, and *10KRET* as the components. *CONGRET* is the pseudo-conglomerate return. The sample period runs from March 1982 to December 2023, due to the availability of *CFRET*.

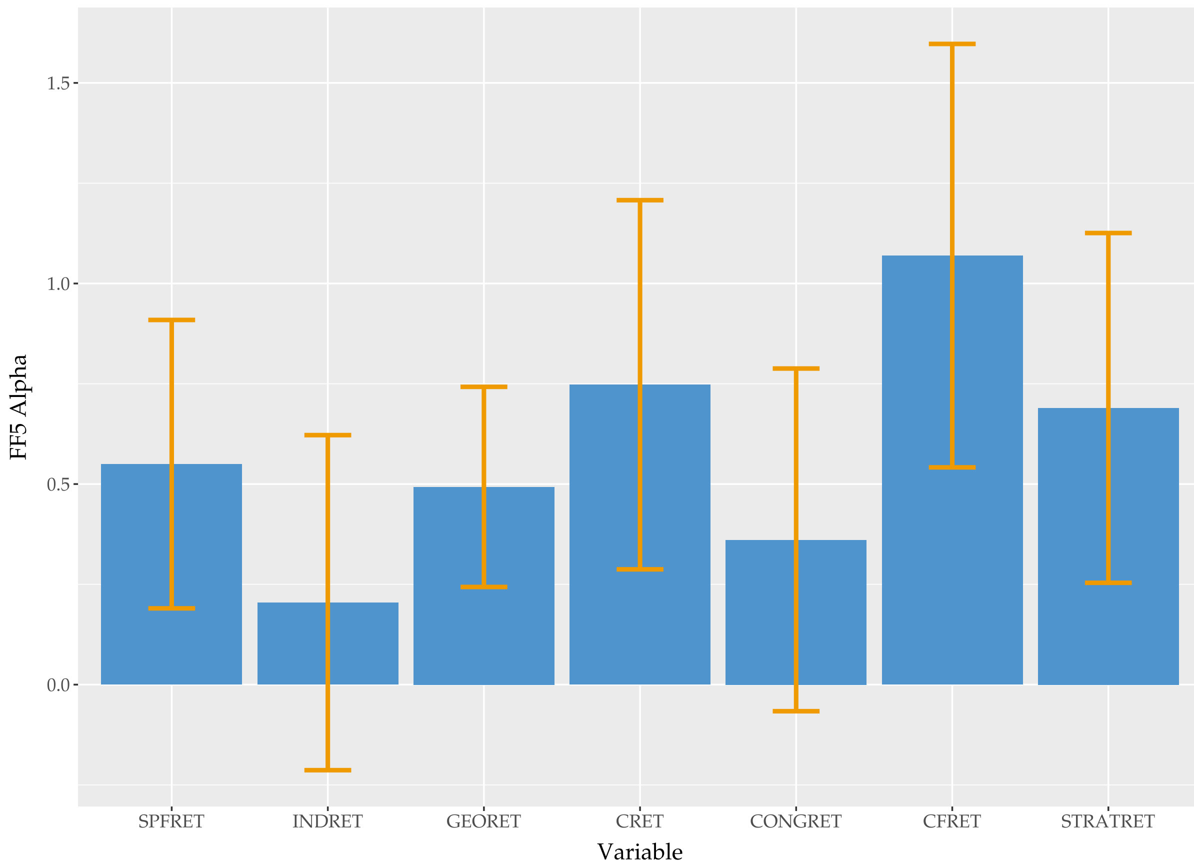


Figure 4: Decomposing *SPFRET* Portfolio Returns

The figure decomposes the average return of the *SPFRET*-sorted long-short portfolio into its various components. Each bar shows the amount of the average portfolio 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. *CFRET* represents the contribution of the analyst-linked firm return. *TECHRET* represents the contribution of the technology-linked firm return. *TNICRET* represents the contribution of the product-linked firm return.*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 August 1989 to December 2023, due to the availability of *TNICRET*.

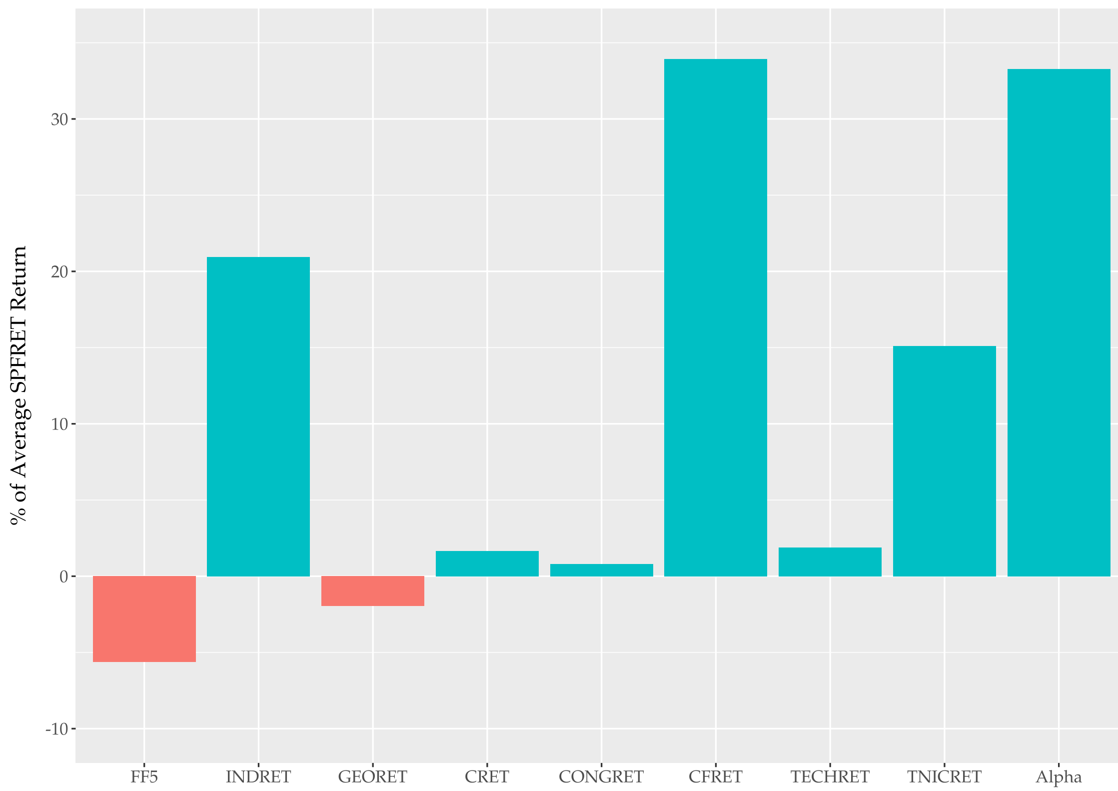


Figure 5: PLS-Based Long-Short Portfolio Returns

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 1971 through December 2023. Portfolios are value-weighted. Shaded areas indicate the NBER recession periods.

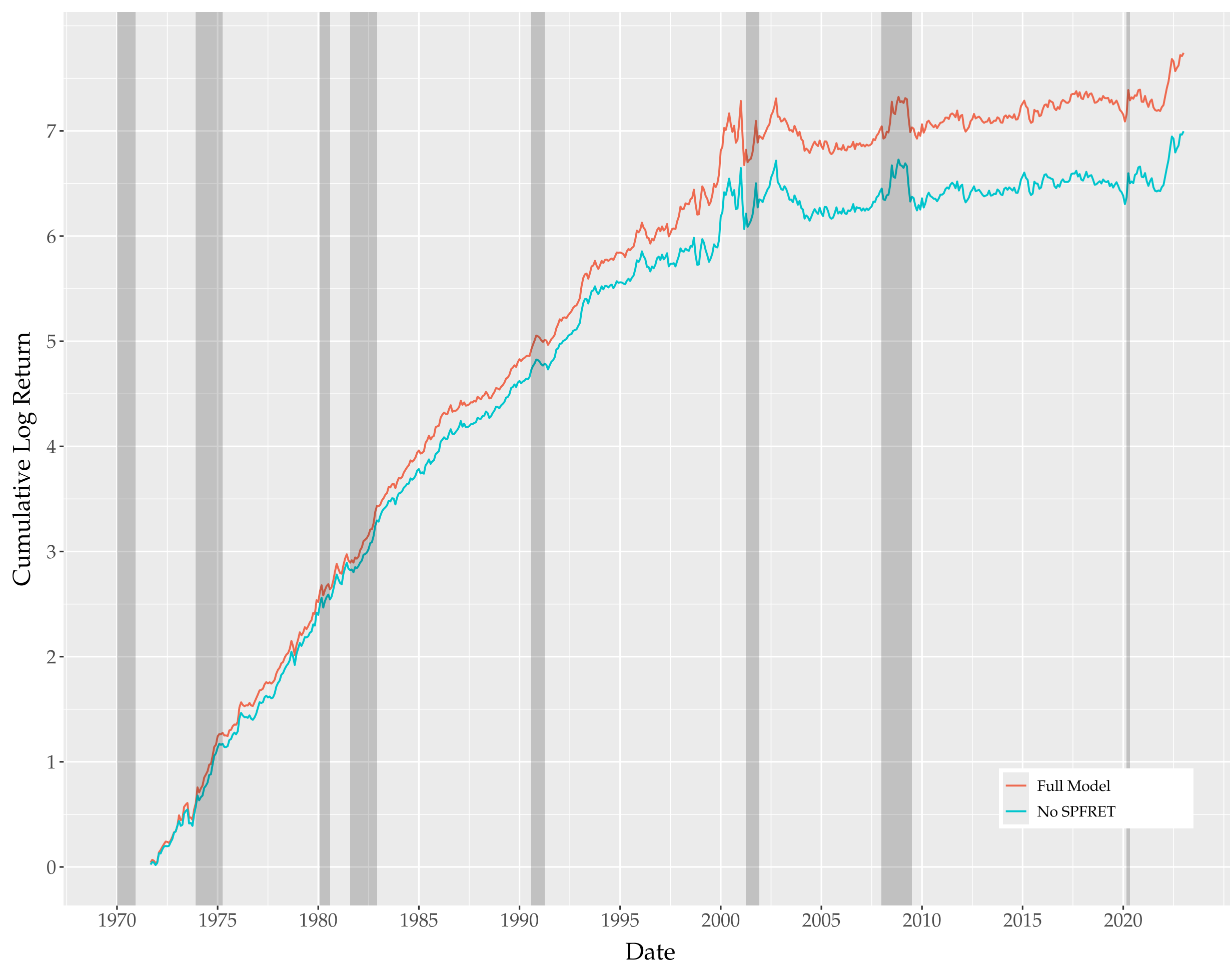


Table 1: Summary Statistics and Correlations

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	632,443	0.000	0.003	0.000	0.000	0.000	0.819
PROX	632,443	0.027	0.154	0.000	0.001	0.002	1.000
J2J	632,101	0.013	0.044	0.000	0.000	0.001	0.486
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	235,861	0.048	0.123	0.000	0.000	0.002	1.000
CSTMR	32,469	0.018	0.129	0.000	0.000	0.000	1.000
Same County	632,443	0.025	0.154	0.000	0.000	0.000	1.000

Panel B: Correlations of Similarity Measures

	SCI	PROX	J2J	STRATSIM	$STRATSIM^{PROD}$	$STRATSIM^{TECH}$	ANALYST	CSTMR
PROX	0.491							
J2J	0.268	0.251						
STRATSIM	0.065	0.012	0.050					
$STRATSIM^{PROD}$	0.048	0.050	0.057	0.133				
$STRATSIM^{TECH}$	0.066	0.026	0.034	0.036	0.135			
ANALYST	0.107	0.129	0.074	0.116	0.255	0.212		
CSTMR	0.002	0.030	0.001	0.014	0.034	0.042	0.171	
Same County	0.232	0.259	0.213	0.052	0.068	0.043	0.068	0.005

Table 2: Strategy Similarity and Social Connectedness

The table reports the results of the panel regressions of strategy similarities between firm pairs on the three measures of social connectedness between firms. *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). *Avg. SIM* is the average of $STRATSIM^{10K}$, $STRATSIM^{TECH}$, and $STRATSIM^{PROD}$, whichever is available. Similarities are calculated only for firms that are in the same Fama-French 48 industry. The sample period is 1993-2023 for $STRATSIM^{10K}$ and $STRATSIM^{PROD}$ regressions, while it is 1993-2010 for the $STRATSIM^{TECH}$ regression. *SCI* is the social connectedness index between the firms' headquarters counties. *J2J* is is the scaled employee flow between the state-industry pairs that the firms belong to. *PROX* is geographic proximity between the firms' headquarters counties, measured as the inverse of the geodesic distance. All three measures are rank normalized within each industry each year. *Avg. Conn.* is the average of *SCI*, *PROX*, and *J2J*, whichever is available. *Similar Size* is a dummy variable that equals one if the market capitalizations of the two firms are within 10 times of each other. 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.

	STRATSIM				$STRATSIM^{PROD}$				$STRATSIM^{TECH}$				Avg. SIM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
SCI	0.008***				0.017***				0.038***				
	(7.370)				(6.536)				(4.034)				
PROX		0.005***				0.011**				0.022**			
		(5.779)				(2.428)				(2.485)			
J2J			0.006**				0.014***				0.025***		
			(2.284)				(24.006)				(3.102)		
Avg. Conn				0.009***				0.020***				0.035***	0.015***
				(4.751)				(4.480)				(2.955)	(4.495)
Similar Size	0.003***	0.003***	0.003***	0.003***	0.011***	0.011***	0.011***	0.011***	0.036***	0.036***	0.036***	0.036***	0.005*
	(3.173)	(3.220)	(3.237)	(3.169)	(12.742)	(12.515)	(12.283)	(12.630)	(12.404)	(12.468)	(12.482)	(12.476)	(1.785)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,309,349	5,309,349	5,307,985	5,309,349	5,436,299	5,436,299	5,435,983	5,436,299	694,631	694,631	694,631	694,631	9,088,967
R^2	0.236	0.236	0.236	0.236	0.323	0.322	0.323	0.324	0.244	0.244	0.243	0.244	0.130

Table 3: Comovements in Firm Fundamentals and Stock Returns

The table presents the comovements in firm fundamentals and stock returns between the focal firm and its peers using panel regression analysis. In columns 1-5, 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. *Capx* is capital expenditures scaled by sales. In column 6, the dependent variable is monthly stock returns. The main independent variables are the corresponding fundamental of the contemporaneous *SCI*-weighted, *J2J*-weighted, and *PROX*-weighted industry portfolio (excluding firms from the same state). Independent variables in Panel A (B) also include the contemporaneous equal-weighted (value-weighted) industry portfolios (excluding the focal firm), equal-weighted portfolio of peers from the same *TNIC* classification, *TECH*-weighted peer portfolio, and 10-K (strategy) similarity weighted peer portfolio. *TNIC*, *TECH*, and *10K* portfolios exclude same-state peers. 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: Multivariate Analysis with Equal-Weighted Industry Portfolio

Peer Type	ΔEPS (1)	ΔSales (2)	ΔEmployees (3)	NewCapital (4)	Capx (5)	Returns (6)
SCI	1.814***	11.061***	6.495***	3.891***	4.379***	5.430***
	(3.616)	(4.690)	(5.321)	(5.889)	(4.483)	(17.637)
PROX	1.887***	13.121***	6.876***	1.720***	4.743***	3.592***
	(3.910)	(6.490)	(5.983)	(3.483)	(5.439)	(14.646)
J2J	1.519***	5.679***	3.842***	2.309***	7.431***	2.351***
	(6.418)	(5.932)	(5.908)	(6.635)	(8.829)	(15.726)
Industry EW	−2.917***	−14.513***	−10.302***	−3.874***	−1.464*	−5.499***
	(−5.462)	(−6.294)	(−9.151)	(−5.624)	(−1.750)	(−16.760)
TNIC	4.571***	34.690***	17.492***	8.645***	18.622***	8.066***
	(10.624)	(13.719)	(12.350)	(13.948)	(19.803)	(27.134)
TECH	−0.068	−1.617	−2.233**	−0.305	−2.493***	2.608***
	(−0.254)	(−1.200)	(−2.380)	(−0.608)	(−4.889)	(9.632)
10K	−1.403***	3.576**	2.623**	1.902**	5.034***	1.580***
	(−3.045)	(2.011)	(2.015)	(2.344)	(5.734)	(12.771)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76,086	75,089	96,507	76,074	76,044	1,305,855
R^2	0.044	0.099	0.067	0.066	0.197	0.144

Panel B: Multivariate Analysis with Value-Weighted Industry Portfolio

Peer Type	ΔEPS (1)	ΔSales (2)	ΔEmployees (3)	NewCapital (4)	Capx (5)	Returns (6)
SCI	0.717*	5.952***	3.519***	2.671***	2.677***	3.325***
	(1.698)	(3.195)	(3.355)	(5.112)	(2.897)	(14.993)
PROX	0.853**	8.433***	3.486***	0.831	3.269***	1.534***
	(2.137)	(4.516)	(3.419)	(1.635)	(3.872)	(8.621)
J2J	1.345***	5.275***	3.554***	2.230***	7.241***	1.895***
	(6.173)	(6.017)	(5.843)	(6.801)	(8.622)	(13.738)
Industry VW	−0.341	−3.674***	−3.335***	−2.269***	2.454***	−0.450***
	(−1.412)	(−4.074)	(−7.743)	(−8.242)	(3.600)	(−3.462)
TNIC	4.311***	33.582***	16.511***	8.342***	18.441***	7.623***
	(10.330)	(13.581)	(11.662)	(13.539)	(19.583)	(26.696)
TECH	−0.154	−2.280*	−2.576***	−0.472	−2.549***	2.420***
	(−0.576)	(−1.671)	(−2.766)	(−0.930)	(−5.010)	(8.936)
10K	−1.519***	3.024*	2.138*	1.698**	5.040***	1.333***
	(−3.324)	(1.733)	(1.659)	(2.103)	(5.747)	(10.864)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76,086	75,089	90,507	76,074	76,044	1,305,855
R^2	0.042	0.097	0.065	0.066	0.198	0.142

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

The table reports the results of the univariate portfolio sort based on different versions of the peer return. In Panel A, the sorting variable is *SCIRET*, the SCI-weighted average return of a firm's industry peers. In Panel B, the sorting is done on *PROXRET*, the PROX-weighted average return of industry peers. In Panel C, stocks are sorted on *J2JRET*, the job-to-job employee flow weighted return of industry peers. In Panel D, the sorting is done on *SPFRET*, the composite peer return obtained by finding the optimal combination between *SCIRET*, *PROXRET*, and *J2JRET* to predict next month's focal firm return. While forming the peer returns, peers headquartered in the same state as the focal firm are excluded. In all panels, for each month, we sort all common stocks into deciles based on the corresponding sorting variable 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. 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: SCIRET

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.218	0.523**	0.759***	0.983***	1.112***	1.191***	1.168***	1.386***	1.585***	1.759***	1.541***
	(0.916)	(2.399)	(3.604)	(4.550)	(5.565)	(5.685)	(5.546)	(6.220)	(7.240)	(7.413)	(8.793)
FF5 Alpha EW	-0.876***	-0.523***	-0.280***	-0.062	-0.001	0.066	0.033	0.286***	0.536***	0.731***	1.607***
	(-7.868)	(-4.649)	(-2.813)	(-0.712)	(-0.013)	(1.134)	(0.568)	(3.811)	(5.117)	(6.583)	(7.710)
Raw Return VW	0.506***	0.760***	0.951***	0.894***	1.080***	1.187***	0.922***	1.018***	1.143***	1.159***	0.652***
	(2.587)	(4.130)	(5.080)	(4.829)	(5.919)	(6.697)	(5.023)	(5.648)	(6.143)	(6.131)	(4.374)
FF5 Alpha VW	-0.447***	-0.168*	0.014	-0.024	0.100	0.207**	-0.087	0.032	0.229**	0.227**	0.673***
	(-4.617)	(-1.782)	(0.138)	(-0.291)	(1.189)	(2.483)	(-1.106)	(0.346)	(2.171)	(2.382)	(4.141)

Panel B: PROXRET

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.266	0.539**	0.719***	1.008***	1.076***	1.146***	1.225***	1.364***	1.615***	1.727***	1.461***
	(1.132)	(2.479)	(3.423)	(4.689)	(4.904)	(5.293)	(5.686)	(6.307)	(7.172)	(7.185)	(8.052)
FF5 Alpha EW	-0.847***	-0.515***	-0.319***	-0.060	-0.027	0.034	0.111*	0.268***	0.571***	0.694***	1.542***
	(-7.595)	(-4.789)	(-3.477)	(-0.697)	(-0.434)	(0.587)	(1.846)	(3.310)	(6.527)	(5.382)	(7.065)
Raw Return VW	0.434**	0.840***	0.888***	1.048***	0.997***	1.021***	1.018***	1.105***	1.166***	1.107***	0.673***
	(2.171)	(4.635)	(4.901)	(5.469)	(5.639)	(5.733)	(5.742)	(6.057)	(6.278)	(5.669)	(4.397)
FF5 Alpha VW	-0.538***	-0.090	-0.033	0.125	0.051	0.049	-0.036	0.091	0.257**	0.161	0.699***
	(-5.571)	(-0.899)	(-0.361)	(1.429)	(0.643)	(0.588)	(-0.549)	(1.156)	(2.330)	(1.454)	(3.960)

Panel C: J2JRET (Post-2001)

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.501	0.590	0.626*	0.877**	0.883**	0.826**	0.936***	1.122***	1.057***	1.131***	0.630***
	(1.227)	(1.548)	(1.684)	(2.464)	(2.504)	(2.351)	(2.751)	(3.313)	(3.154)	(3.311)	(3.010)
FF5 Alpha EW	-0.443***	-0.248*	-0.201*	0.104	0.078	-0.023	0.093	0.240*	0.161	0.295**	0.738***
	(-3.465)	(-1.965)	(-1.968)	(1.155)	(0.982)	(-0.278)	(0.998)	(1.839)	(1.528)	(2.386)	(3.616)
Raw Return VW	0.655*	0.915***	0.972***	0.871***	0.840***	0.782**	0.918***	0.773***	0.727**	0.725**	0.070
	(1.968)	(2.983)	(2.861)	(2.688)	(2.639)	(2.381)	(2.930)	(2.629)	(2.376)	(2.423)	(0.314)
FF5 Alpha VW	-0.173	0.098	0.158	0.147	0.001	-0.084	0.081	-0.042	-0.109	-0.034	0.139
	(-1.207)	(0.905)	(1.239)	(1.103)	(0.007)	(-0.773)	(0.671)	(-0.328)	(-0.818)	(-0.235)	(0.595)

Panel D: SPFRET

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.201	0.553**	0.778***	0.947***	1.108***	1.141***	1.200***	1.410***	1.574***	1.773***	1.572***
	(0.851)	(2.542)	(3.670)	(4.377)	(5.541)	(5.297)	(6.083)	(6.495)	(7.003)	(7.518)	(8.851)
FF5 Alpha EW	-0.898***	-0.492***	-0.256***	-0.108	-0.010	0.025	0.054	0.317***	0.538***	0.741***	1.639***
	(-7.805)	(-4.716)	(-2.730)	(-1.300)	(-0.168)	(0.442)	(0.893)	(4.192)	(5.353)	(6.320)	(7.672)
Raw Return VW	0.453**	0.811***	0.912***	0.926***	1.126***	1.022***	0.955***	1.140***	1.155***	1.138***	0.685***
	(2.300)	(4.572)	(5.017)	(4.821)	(6.241)	(6.017)	(5.258)	(6.103)	(6.288)	(6.063)	(4.454)
FF5 Alpha VW	-0.529***	-0.097	-0.015	-0.015	0.180**	0.029	-0.072	0.152*	0.233**	0.204**	0.732***
	(-5.345)	(-0.920)	(-0.162)	(-0.186)	(2.122)	(0.445)	(-0.923)	(1.665)	(2.215)	(2.165)	(4.327)

Table 5: Portfolio Analysis II: Bivariate Sort

The table reports the results of the bivariate portfolio sort based on *INDRET* and the sorting variables used in Table 4 in separate rows. Each month, we first sort all stocks into quintiles based on *INDRET*, and then, within each *INDRET* quintile, we further sort them into quintiles based on one of the variables given in Table 4 and compute both the equal-weighted and value-weighted average return of all portfolios over the next month. For each variable quintile, we then compute the average of the five portfolio returns across the *INDRET* quintiles. We report the FF5 alphas for these return series for the lowest and highest quintiles, as well as the FF5 alphas for the long-short portfolios. 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.

	Equal-Weighted			Value-Weighted		
	(1 Low)	(5 High)	(5-1)	(1 Low)	(5 High)	(5-1)
SCIRET	-0.311***	0.232***	0.543***	-0.225***	0.159***	0.384***
	(-6.001)	(5.114)	(7.964)	(-4.800)	(2.844)	(4.958)
PROXRET	-0.310***	0.194***	0.504***	-0.229***	0.083	0.312***
	(-6.347)	(4.342)	(7.395)	(-4.456)	(1.465)	(4.391)
J2JRET (Post-2001)	-0.160**	0.092	0.252***	-0.066	0.019	0.085
	(-2.423)	(1.502)	(3.088)	(-1.022)	(0.265)	(0.865)
SPFRET	-0.332***	0.219***	0.551***	-0.239***	0.080	0.319***
	(-6.223)	(4.929)	(8.154)	(-5.072)	(1.374)	(4.517)

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 composite social-peer return, obtained by regressing next month's focal stock return on *SCIRET*, *PROXRET*, and *J2JRET*, and using the regression coefficients as weights. *STRATRET* is the composite strategy-linked peer return, obtained by regressing next month's focal stock return on *TECHRET*, *TNICRET*, and *10KRET*, according to their availability, and using the regression coefficients as weights. *INDRET* is the short-term industry momentum. *GEORET* is the geographic momentum. *CFRET* is the analyst-connected firm return. *CRET* is the customer return. *CONGRET* is the pseudo-conglomerate 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 December 2023 due to the availability of *10KRET*. 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)	(10)
SPFRET	1.159***					1.156***	0.818***	1.132***	0.802***	0.292**
	(5.422)					(5.549)	(5.382)	(5.607)	(5.392)	(2.083)
TECHRET		0.239				−0.020				
		(1.037)				(−0.102)				
TNICRET			1.348***				1.028***			
			(4.894)				(4.690)			
10KRET				0.527***				0.302**		
				(2.871)				(2.006)		
STRATRET					1.353***				1.026***	0.676***
					(4.893)				(4.618)	(4.113)
INDRET										0.358*
										(1.830)
GEORET										0.234***
										(2.880)
CRET										0.356**
										(2.224)
CONGRET										−0.289**
										(−2.358)
CFRET										1.052***
										(6.018)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	353	353	353	353	353	353	353	353	353	353
# Stocks	3,682	3,682	3,682	3,682	3,682	3,682	3,682	3,682	3,682	3,682
R^2	0.062	0.060	0.062	0.061	0.062	0.063	0.064	0.064	0.064	0.069

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	−0.477*	0.685**	1.162***	−1.443***	0.065	1.508***	−0.973***	0.448	1.421***
	(−1.895)	(2.371)	(3.908)	(−4.588)	(0.184)	(4.074)	(−3.392)	(1.378)	(4.253)
High	−0.487***	0.041	0.528***	−0.816***	−0.142	0.697**	−0.778***	−0.153	0.625**
	(−3.114)	(0.270)	(3.104)	(−3.066)	(−0.593)	(2.523)	(−2.812)	(−0.616)	(2.416)

Panel B: Out-Cluster and In-Cluster

	Top 20%			Top 10%		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Out-Cluster	−0.916***	0.314**	1.230***	−0.946***	0.316*	1.262***
	(−5.548)	(1.966)	(6.177)	(−5.769)	(1.885)	(5.971)
In-Cluster	−0.584***	−0.006	0.577***	−0.537***	−0.078	0.459**
	(−3.651)	(−0.040)	(3.295)	(−3.313)	(−0.466)	(2.525)

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 composite social-peer return obtained by optimally combining *SCIRET*, *PROXRET*, and *J2JRET* to predict next month's focal stock return. *SPFMOM* is the composite social-peer return over months $t-11$ through $t-1$ obtained by combining the compound *SCIMOM*, *PROXMOM*, and *J2JMOM* using the same weights in *SPFRET*. 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, 2–3, 4–6, 7–12, 13–24, and 25–36. *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	-0.502***	0.995***	1.498***	0.046	0.721***	0.675***
	(-4.694)	(8.501)	(7.313)	(0.317)	(5.612)	(2.828)
Months 2–3	0.094	0.309***	0.215	0.017	0.543***	0.526**
	(0.952)	(3.764)	(1.274)	(0.127)	(4.660)	(2.429)
Months 4–6	0.076	0.223***	0.147	-0.027	0.416***	0.443**
	(0.832)	(3.080)	(1.031)	(-0.208)	(3.754)	(2.119)
Months 7–12	0.022	0.278***	0.257***	0.070	0.219**	0.149
	(0.354)	(5.428)	(2.978)	(0.823)	(2.312)	(0.976)
Months 13–24	0.109***	0.121**	0.013	0.088	0.145	0.057
	(2.652)	(2.197)	(0.217)	(1.427)	(1.552)	(0.471)
Months 25–36	0.015	0.095*	0.080*	-0.094	0.186**	0.280***
	(0.391)	(1.927)	(1.653)	(-1.498)	(2.356)	(2.754)

Panel B: FF5 Alphas (Value-Weighted)

	SPFRET			SPFMOM		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Month 1	-0.436***	0.273***	0.709***	0.083	0.290***	0.207
	(-4.596)	(2.878)	(4.245)	(0.567)	(2.587)	(0.972)
Months 2–3	0.178	0.059	-0.119	-0.068	0.348***	0.416**
	(1.518)	(0.663)	(-0.674)	(-0.543)	(3.510)	(2.217)
Months 4–6	-0.008	0.014	0.022	-0.175	0.180**	0.355**
	(-0.091)	(0.207)	(0.170)	(-1.539)	(1.992)	(2.094)
Months 7–12	-0.065	0.178***	0.244***	-0.020	0.102	0.122
	(-1.127)	(3.207)	(2.711)	(-0.230)	(1.222)	(0.855)
Months 13–24	0.053	0.049	-0.004	0.019	0.030	0.011
	(1.324)	(0.799)	(-0.062)	(0.282)	(0.324)	(0.092)
Months 25–36	0.013	0.057	0.043	-0.100	0.159*	0.259**
	(0.307)	(1.039)	(0.763)	(-1.449)	(1.787)	(2.324)

Table 9: Predicting Future Earnings Surprises, Analyst Forecast Errors, and Earnings 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 composite social-peer return obtained by optimally combining *SCIRET*, *PROXRET*, and *J2JRET* to predict next month's focal stock return. *SPFMOM* is the composite social-peer return over months $t-11$ through $t-1$ obtained by combining the compound *SCIMOM*, *PROXMOM*, and *J2JMOM* using the same weights in *SPFRET*. In Panel B, we also control for *INDRET* and *INDMOM*, which are the short- and long-term industry momentum, respectively. 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 Announcement 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	33.804***	25.476***	17.835***	10.811***	0.147***	0.175***	0.189***	0.192***	0.043	0.028	−0.038	−0.092
	(22.209)	(17.197)	(12.113)	(7.408)	(5.910)	(4.721)	(4.459)	(4.486)	(0.756)	(0.479)	(−0.662)	(−1.616)
SPFRET	11.024***	12.276***	10.651***	8.812***	0.053***	0.042*	0.055**	0.093***	0.111**	−0.001	−0.025	0.019
	(10.348)	(11.913)	(9.778)	(8.266)	(3.330)	(1.880)	(2.086)	(3.139)	(2.330)	(−0.011)	(−0.511)	(0.376)
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	2,021,995	2,020,772	2,012,550	1,995,798	1,052,668	1,014,547	979,339	945,153	1,776,588	1,725,261	1,696,776	1,671,727
R^2	0.170	0.171	0.171	0.168	0.177	0.164	0.195	0.200	0.050	0.054	0.055	0.057

Panel B: Controlling for Industry Effects

	Earnings Growth				Forecast Error				Earnings Announcement 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	12.162***	10.588***	8.713***	7.257***	0.068**	0.074*	0.118**	0.133***	0.094	0.074	0.083	0.011
	(7.630)	(6.623)	(5.318)	(4.521)	(2.297)	(1.903)	(2.553)	(2.762)	(1.567)	(1.174)	(1.371)	(0.176)
SPFRET	0.273	1.626*	2.812***	2.649***	0.005	−0.005	0.020	0.028	0.065	0.013	−0.002	0.069
	(0.293)	(1.806)	(2.928)	(2.769)	(0.262)	(−0.233)	(0.739)	(1.028)	(1.593)	(0.301)	(−0.052)	(1.641)
INDMOM	28.466***	19.265***	11.675***	4.232**	0.102***	0.131***	0.091**	0.073	−0.074	−0.061	−0.161**	−0.134*
	(17.237)	(11.566)	(6.947)	(2.512)	(3.178)	(3.074)	(2.181)	(1.532)	(−1.169)	(−0.909)	(−2.238)	(−1.935)
INDRET	15.050***	15.005***	11.082***	8.770***	0.066***	0.066**	0.048	0.090***	0.065	−0.018	−0.031	−0.071
	(12.933)	(13.137)	(9.132)	(7.496)	(2.751)	(2.252)	(1.324)	(2.679)	(1.242)	(−0.361)	(−0.599)	(−1.345)
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	2,021,995	2,020,772	2,012,550	1,995,798	1,052,668	1,014,547	979,339	945,153	1,776,588	1,725,261	1,696,776	1,671,727
R^2	0.172	0.172	0.171	0.168	0.177	0.164	0.195	0.200	0.050	0.054	0.055	0.057

Appendix

- Figure A.1: Illustration of SCI
- Table A.1: Variable Descriptions
- Table A.2: Summary Statistics and Correlations for Return Variables
- Table A.3: Other Socially-Weighted Predictors

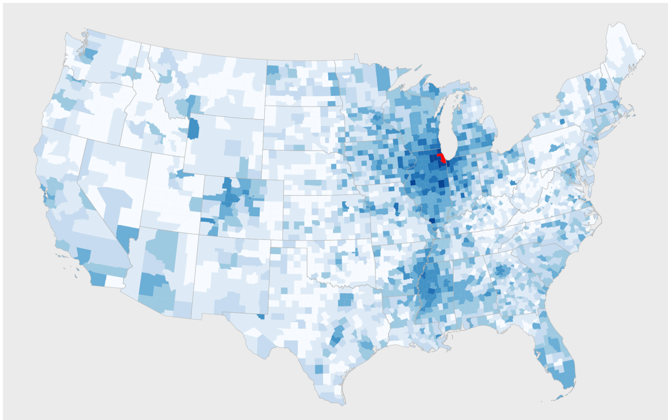
Illustration of SCI

Figure A.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.

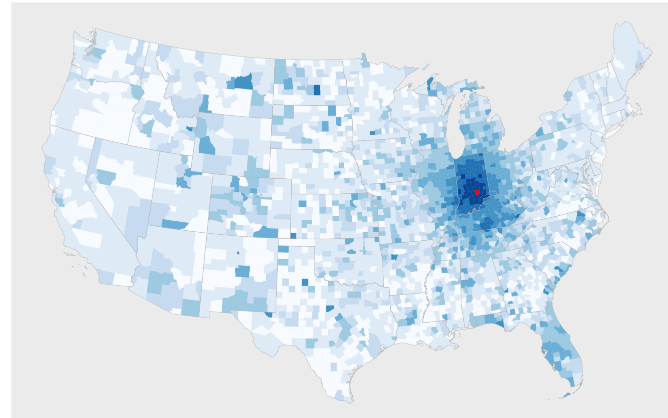
Figure A.1: Examples of Social Connectedness

This figure shows examples of social connectedness—as measured by the Social Connectedness Index (SCI)—and scaled job-to-job employee flows (J2J) using county-level heat maps of continental U.S. Panels A and B present social connectedness to Cook County, IL, and Bartholomew County, IN respectively. 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. CMI belongs to the "machinery" category of FF48 industry classification, and examples of its 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. Panel D shows the scaled job-to-job employee flow (J2J) for the machinery firms in Bartholomew County as measured in April 2016—time of measurement for SCI. Darker blue indicates higher J2J between two locations. In panels C and D, counties without any peer firms' presence are presented in dark grey.

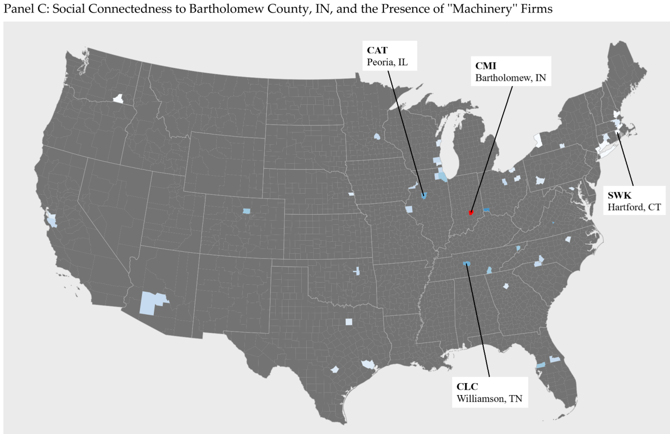
Panel A: Social Connectedness to Cook County, IL



Panel B: Social Connectedness to Bartholomew County, IN



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



Panel D: Job-to-Job Employee Flow for Machinery Firms in Bartholomew County, IN

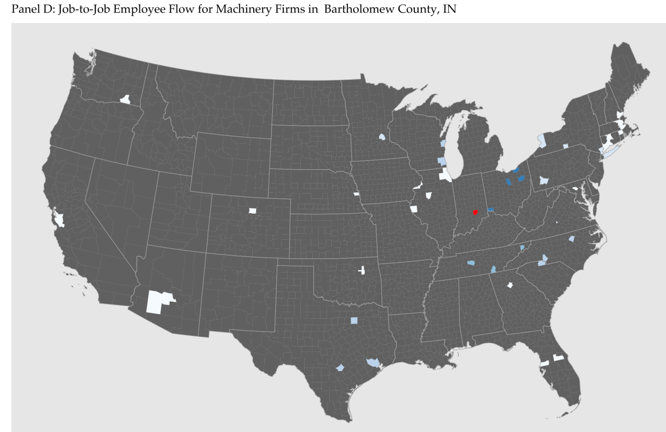


Table A.1: Variable Descriptions

Variable	Definition
Social Connectedness Index (SCI)	Number of Facebook friends links between firms' headquarters counties, scaled by the product of the populations of the two counties.
Geographic Proximity (PROX)	Inverse of the geodesic distance (in miles) plus 1 between two counties.
Scaled Job-to-Job Employee Flow (J2J)	Total number of employee movements between two counties divided by the total (unique) employee flows into or out of the two counties. Job-to-job employee flows data is obtained from the U.S. Census Bureau's Longitudinal Household-Employer Dynamics (LEHD) database.
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.
Weighted Average of Fundamentals	Fundamentals (ΔEPS, $\Delta Sales$, $\Delta Employees$, $NewCapital$) of an SCI-, PROX-, or J2J-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.
SCI-Weighted Peer Firm Return (SCIRET)	SCI-weighted returns based on all firms in the focal firm's industry, excluding same-state firms. *SCIMOM* is *SCIRET* compounded between months $t-11$ and $t-1$.
PROX-Weighted Peer Firm Return (PROXRET)	PROX-weighted returns based on all firms in the focal firm's industry, excluding same-state firms. *PROXMOM* is *PROXRET* compounded between months $t-11$ and $t-1$.
J2J-Weighted Peer Firm Return (J2JRET)	J2J-weighted returns based on all firms in the focal firm's industry, excluding same-state firms. *J2JMOM* is *J2JRET* compounded between months $t-11$ and $t-1$.
Social Peer Firm Return (SPFRET)	Composite social-connectedness-related peer firm return constructed by optimally combining *SCIRET*, *PROXRET*, and *J2JRET* to explain the contemporaneous focal firm return. $SPFRET_{No\ SCI}$ is an alternative social peer firm return which is constructed without using *SCIRET*. $SPFRET_{\perp}$ is *SPFRET* orthogonalized against all return variables and characteristics used in the Fama-MacBeth regressions. $SPFRET_{XIND}$ is the version of *SPFRET* constructed by using out-of-state peers that belong to industries that are dominant in the focal firm's state, except the focal industry itself.
Social Peer Firm Momentum (*SPFMOM*)	The long-run version of *SPFRET*, which is obtained by interacting the weights used in the construction of *SPFRET* with the rank-normalized *SCIMOM*, *PROXMOM*, and *J2JMOM*.
Industry Return (INDRET)	The equal-weighted average return of stocks with the same Fama-French 48 (FF48) 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).
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).
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).

Variable	Definition
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).
Text-Based Industry Momentum (TNICRET)	Equal-weighted stock return of peer firms identified through 10-K product text, constructed following Hoberg and Phillips (2016). $TNICRET_{SPF}$ is constructed the same way as *SPFRET* except FF48 peers are replaced by the peer firms that share the same TNIC category as the focal firm.
10K Strategy-Based Momentum (10KRET)	The weighted average stock return of strategy-linked peer firms, where the weights are the cosine similarities between the peer firm and the focal firm identified through 10-K text related to business strategy.
Composite Strategy-Based Momentum (STRATRET)	Composite strategy-related peer firm return constructed by optimally combining *TECHRET*, *TNICRET*, and *10KRET* to explain the contemporaneous focal firm return.
Analyst Momentum (CFRET)	The weighted average return of stocks that have shared 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). $CFRET_{SPF}$ is derived the same way as *SPFRET* except FF48 peers are replaced by peers that have shared at least one analyst with the focal firm over the previous 12 months.
Input-Ouput Weighted Peer Return (IORET)	The weighted average peer return where the weights are obtained using the industry links from the Historical Input-Output Accounts database of the Bureau of Economic Analysis (BEA). The link between two industries is taken to be the average of the (normalized) input and output links. $IORET_{SPF}$ is the composite of SCI-, PROX-, and J2J-weighted IORET, where the weights of individual components are taken from *SPFRET*.
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.*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. *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. All returns are reported in percentages. All of the reported statistics and correlations are averages over monthly cross sections. *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. *10KRET* is the 10-K based strategy similarity weighted peer return.

Panel A: Summary statistics of Return Variables

	N	Mean	Std. Dev.	Min.	25%	75%	Max.	Skewness	Kurtosis
RET	3,423	1.685	12.736	-55.007	-4.775	6.628	203.049	3.364	74.192
INDRET	3,345	0.980	3.129	-8.147	-1.032	2.863	12.945	0.360	3.118
INDMOM	3,345	14.521	15.522	-23.087	4.222	23.539	73.734	0.499	2.473
GEORET	1,988	1.065	3.837	-24.808	-0.532	2.511	40.753	1.518	33.182
CRET	538	0.946	7.839	-31.502	-3.257	4.878	48.272	0.717	10.182
CONGRET	830	0.937	4.037	-15.432	-1.351	3.141	20.697	0.287	8.401
CFRET	2,659	0.960	4.179	-20.956	-1.428	3.233	30.417	0.446	10.626
TNICRET	2,186	1.239	5.852	-31.126	-1.685	3.769	66.715	1.774	36.348
TECHRET	926	1.301	2.664	-11.821	-0.087	2.569	21.297	0.930	19.955
10KRET	1,088	1.512	0.665	-1.552	1.139	1.872	5.045	0.216	9.165

Panel B: Correlations

	RET	INDRET	INDMOM	GEORET	CRET	CONGRET	CFRET	TNICRET	TECHRET
INDRET	0.222								
INDMOM	0.023	0.145							
GEORET	0.006	-0.024	-0.001						
CRET	0.108	0.153	0.010	0.026					
CONGRET	0.139	0.389	0.038	-0.004	0.137				
CFRET	0.271	0.503	0.046	0.021	0.223	0.376			
TNICRET	0.175	0.368	0.044	0.002	0.161	0.227	0.437		
TECHRET	0.112	0.269	0.030	0.007	0.113	0.155	0.332	0.220	
10KRET	0.063	0.186	0.009	0.013	0.064	0.100	0.249	0.158	0.136

Table A.3: Other Socially-Weighted Predictors

The table shows the predictability results for the composite versions of three alternative peer returns, $TNICRET_{SPF}$, $CFRET_{SPF}$, and $IORET_{SPF}$. $TNICRET_{SPF}$ is found analogously with $SPFRET$, where instead of peers in the same FF48 classification, we use peers in the same TNIC class. For $CFRET_{SPF}$, we use peers that share an analyst with the focal firm. $IORET_{SPF}$, we use the industry links obtained from Bureau of Economic Analysis' Historical Input-Output Accounts as peer weights. For all three returns, to optimally combine SCI-, $PROX$-, and $J2J$-weighted versions, we use the same OLS weights used in the construction of $SPFRET$. In Panel A, we report the results of the dependent bivariate portfolio sorting based on these three sets of variables, where stocks are first sorted into quintiles on the original (non-socially-weighted) variable and then on the socially-weighted version within each of these quintiles. We then calculate the value-weighted returns for each of the resultant 25 portfolios (per variable set) over the month following the portfolio formation, and report the average FF5 alphas for the lowest and highest SPF-weighted quintiles, as well as the 5-1 LS portfolios. The final row reports the FF5 alphas of the average portfolio returns across the columns. In Panel B, we use Fama-MacBeth regressions to regress next month's focal stock return on $TNICRET_{SPF}$, $CFRET_{SPF}$, $IORET_{SPF}$, $TNICRET$, $CFRET$, and $IORET$. We also include the following control variables in all regressions: RET, $SIZE$, $BMKT$, BM, MOM, $IVOL$, $ILLIQ$, MAX, $SKEW$, and $COSKEW$. The sample period for both panels is July 1989 to November 2023 due to the availability of $TNICRET_{SPF}$ data. All returns are reported in percentages. In regressions, missing values of independent variables are imputed with the monthly medians. In order to standardize the independent variables in regressions, 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.

Panel A: Bivariate Sort

	$TNICRET_{SPF}$			$CFRET_{SPF}$			$IORET_{SPF}$		
	(1 Low)	(5 High)	(5-1)	(1 Low)	(5 High)	(5-1)	(1 Low)	(5 High)	(5-1)
1 (Low Row)	-0.442***	-0.081	0.361*	-0.908***	-0.288	0.620**	-0.473**	0.089	0.562**
	(-2.619)	(-0.543)	(1.909)	(-3.528)	(-1.460)	(2.496)	(-2.327)	(0.460)	(2.583)
2	0.029	0.087	0.059	-0.163	-0.077	0.085	0.089	0.252*	0.162
	(0.200)	(0.611)	(0.319)	(-0.897)	(-0.560)	(0.442)	(0.516)	(1.721)	(0.828)
3	-0.229**	0.233*	0.462***	-0.013	0.137	0.150	-0.026	0.128	0.154
	(-2.094)	(1.850)	(2.948)	(-0.110)	(1.173)	(0.971)	(-0.216)	(0.731)	(0.845)
4	0.033	-0.045	-0.079	-0.054	0.382***	0.436**	0.019	-0.107	-0.126
	(0.210)	(-0.356)	(-0.431)	(-0.423)	(2.994)	(2.276)	(0.135)	(-0.753)	(-0.721)
5	0.265*	0.417*	0.152	0.043	0.663*	0.620*	-0.155	0.150	0.305*
	(1.802)	(1.960)	(0.609)	(0.295)	(1.807)	(1.836)	(-0.932)	(0.790)	(1.755)
Average	-0.069	0.122	0.191*	-0.219***	0.163**	0.382***	-0.109	0.102	0.212**
	(-0.927)	(1.555)	(1.811)	(-3.137)	(2.201)	(3.722)	(-1.301)	(1.521)	(2.302)

Panel B: Fama-MacBeth Regressions

	RET$_{t+1}$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TNICRET	1.416***						1.369***		
	(5.953)						(5.745)		
CFRET		1.764***						1.486***	
		(7.041)						(6.128)	
IORET			1.188***						0.605**
			(5.508)						(2.362)
TNICRET$_{SPF}$				0.899***			0.067		
				(5.817)			(0.906)		
CFRET$_{SPF}$					1.495***			0.329***	
					(6.820)			(2.767)	
IORET$_{SPF}$						1.147***			0.660***
						(6.254)			(4.909)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	413	413	413	413	413	413	413	413	413
# Stocks	3,717	3,717	3,717	3,717	3,717	3,717	3,717	3,717	3,717
R^2	0.060	0.061	0.061	0.059	0.060	0.060	0.061	0.062	0.061

Internet Appendix

Internet Appendix I

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.*

Internet Appendix II

- Table IA1: Controlling for Product Similarity
- Table IA2: Comovements in Firm Fundamentals and Stock Returns with Industry Portfolio
- Table IA3: Return Predictability of Social Peer Firm Returns: Within Industry Ranking
- Table IA4: Understanding *SPFRET* Portfolio Alphas: Spanning Regressions
- Table IA5: Long-Run Predictability, Earnings Surprises, Analyst Forecast Errors, and Earnings Returns using Orthogonalized Social Peer Firm Return Measures
- Table IA6: Return Predictability of Peer Firms' Returns: Robustness Checks
- Table IA7: Return Predictability of In-State Peer Firms and Non-Peer Firms
- Figure IA1: Strategy Alphas for Post-2000

Figure IA1: Strategy Alphas for Post-2000

The figure presents the Long-Short FF5 alphas for the lead-lag predictors (*SPFRET*, *INDRET*, *GEORET*, *CFRET*, *CRET*, *TECHRET*, *TNICRET*, and *CONGRET*, and *STRATRET*) for the 2000/01 - 2023/12 sample period. Portfolios are value-weighted. Panel A uses all stocks (after the usual share code, exchange code, and price filters), while Panel B excludes the largest 100 stocks. The orange bars show the 90% confidence intervals.

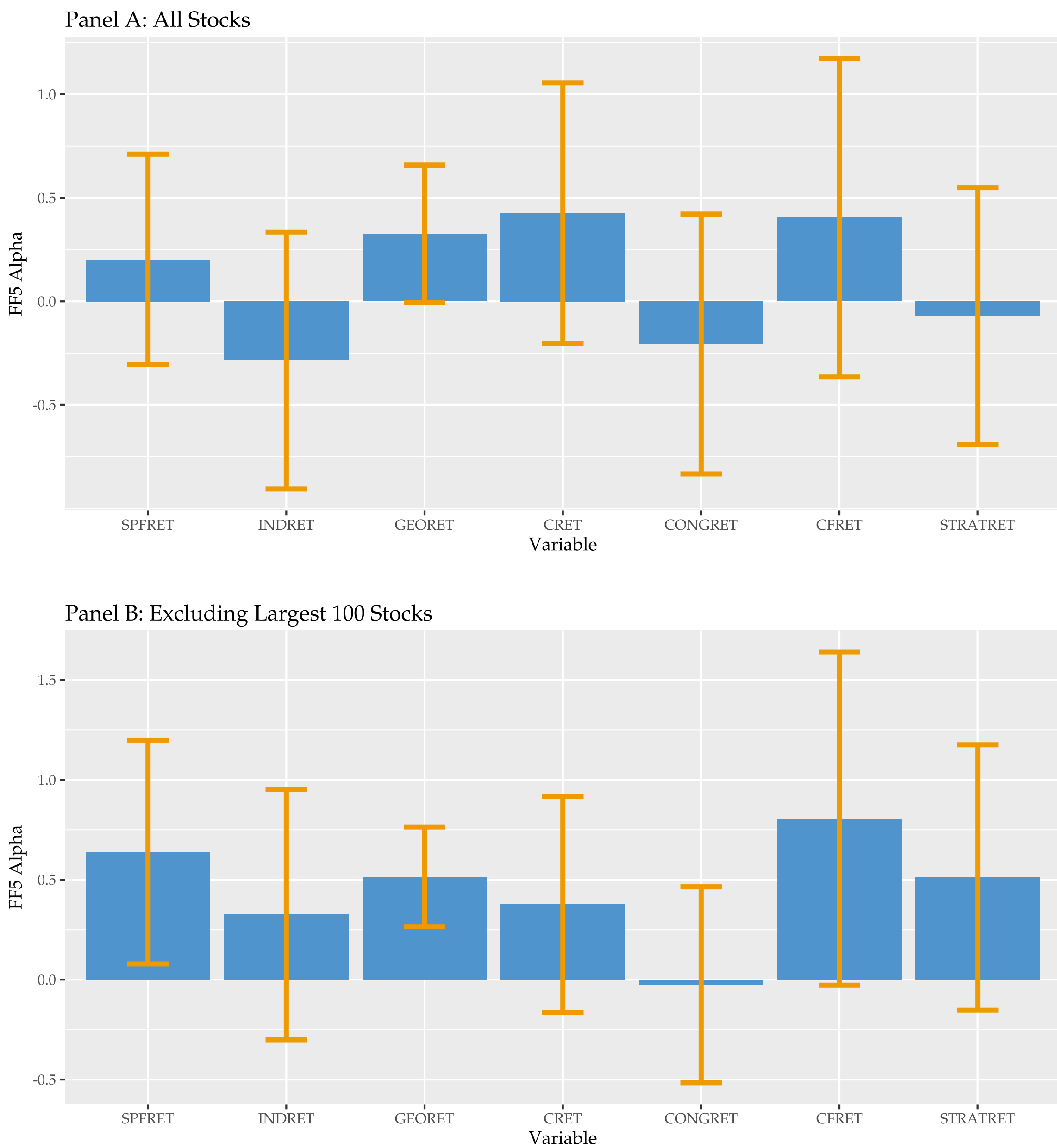


Table IA1: Controlling for Product Similarity

The table reports the results of the panel regressions of strategy similarities between firm pairs on SCI and $STRATSIM^{PROD}$, the product-related similarity. *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-Ks (Hoberg and Phillips, 2016, 2018). Similarities are calculated only for firms that are in the same Fama-French 48 industry. The sample period is 1993-2023 for $STRATSIM^{10K}$ and $STRATSIM^{PROD}$ regressions, while it is 1993-2010 for the $STRATSIM^{TECH}$ regression. *SCI* is the social connectedness index between the firms' headquarters counties. *PROX* is geographic proximity between the firms' headquarters counties, measured as the inverse of the geodesic distance. *J2J* is is the scaled employee flow between the state-industry pairs that the firms belong to. All three measures are rank normalized within each industry each year. *Avg. Conn.* is the average of SCI, $PROX$, and $J2J$, whichever is available. *Similar Size* is a dummy variable that equals one if the market capitalizations of the two firms are within 10 times of each other. 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.

	$STRATSIM^{10K}$	$STRATSIM^{TECH}$	$STRATSIM^{10K}$	$STRATSIM^{TECH}$	$STRATSIM^{10K}$	$STRATSIM^{TECH}$	$STRATSIM^{10K}$	$STRATSIM^{TECH}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SCI	0.007***	0.021***						
	(5.930)	(2.920)						
PROX			0.003***	0.012**				
			(3.362)	(2.111)				
J2J					0.004**	0.015**		
					(2.019)	(2.463)		
Avg. Conn.							0.007***	0.019**
							(3.625)	(2.387)
$STRATSIM^{PROD}$	0.053***	0.252***	0.053***	0.252***	0.053***	0.253***	0.053***	0.252***
	(5.456)	(8.344)	(5.487)	(8.378)	(5.430)	(8.336)	(5.387)	(8.391)
Similar Size	0.002***	0.017***	0.002***	0.017***	0.002***	0.017***	0.002***	0.017***
	(3.407)	(6.723)	(3.488)	(6.738)	(3.538)	(6.737)	(3.459)	(6.737)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,956,219	522,009	4,956,219	522,009	4,954,677	521,817	4,956,219	522,009
R^2	0.313	0.255	0.313	0.255	0.313	0.255	0.313	0.255

Table IA2: Comovements in Firm Fundamentals and Stock Returns: Robustness Tests

The table presents robustness tests for the results given in Table 3. In Panel A (B), we regress a focal firm's fundamentals and return on the contemporaneous equal-weighted (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. In Panel C, the independent variables are the corresponding fundamental of the contemporaneous SCI-weighted, J2J-weighted, and PROX-weighted industry portfolio (excluding firms from the same state); equal- and value-weighted industry portfolios are excluded from the controls. 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: 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

Panel C: Fundamental and Return Comovements without Equal- or Value-Weighted Controls

Peer Type	ΔEPS (1)	ΔSales (2)	ΔEmployees (3)	NewCapital (4)	Capx (5)	Returns (6)
SCI-FF48	0.640	4.586**	2.393**	2.188***	3.615***	3.169***
	(1.536)	(2.455)	(2.279)	(4.352)	(4.215)	(14.910)
PROX-FF48	0.763*	7.705***	2.933***	0.351	4.192***	1.465***
	(1.936)	(4.261)	(2.931)	(0.719)	(5.116)	(8.420)
J2J-FF48	1.324***	5.028***	3.331***	2.070***	7.375***	1.866***
	(6.048)	(5.599)	(5.443)	(6.264)	(8.732)	(13.717)
TNIC	4.285***	33.387***	16.259***	8.271***	18.566***	7.589***
	(10.244)	(13.522)	(11.465)	(13.387)	(19.854)	(26.842)
TECH	−0.180	−2.275*	−2.700***	−0.481	−2.533***	2.388***
	(−0.682)	(−1.681)	(−2.945)	(−0.950)	(−4.966)	(8.936)
10K	−1.528***	3.073*	2.114*	1.758**	5.002***	1.330***
	(−3.319)	(1.756)	(1.647)	(2.177)	(5.724)	(10.841)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76,086	75,089	90,507	76,074	76,044	1,305,855
R^2	0.042	0.097	0.064	0.064	0.197	0.142

Table IA3: Return Predictability of Social Peer Firm Returns: Within Industry Ranking

The table reports the results of the univariate portfolio sorts based on *SPFRET*, rank-normalized over the FF48 industries. Each month, for each FF48 industry, we rank all common stocks based on *SPFRET* and normalize the rankes into [0,1]. Then, we sort all stocks into decile portfolios based on their normalized within-industry *SPFRET* ranks and calculate both the equal-weighted and value-weighted one-month-ahead returns for these portfolios, as well as the return of the portfolio that long the decile 10 portfolio and short the decile 1 portfolio. We present both the average (raw) returns and the FF5 alphas for each portfolio. 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	0.927***	0.941***	0.989***	1.032***	1.052***	1.099***	1.216***	1.185***	1.235***	1.309***	0.382***
	(4.528)	(4.347)	(4.682)	(4.884)	(4.894)	(5.155)	(5.793)	(5.886)	(6.004)	(6.104)	(7.436)
FF5 Alpha EW	−0.171***	−0.126***	−0.079*	−0.063	−0.027	0.026	0.146***	0.099**	0.137***	0.194***	0.364***
	(−3.277)	(−2.934)	(−1.830)	(−1.319)	(−0.564)	(0.526)	(3.377)	(2.146)	(2.680)	(4.043)	(6.702)
Raw Return VW	0.689***	0.961***	0.813***	0.842***	0.916***	0.887***	1.023***	1.121***	0.960***	1.073***	0.384***
	(3.930)	(5.207)	(4.564)	(4.700)	(5.024)	(5.053)	(5.820)	(6.016)	(5.143)	(5.824)	(4.385)
FF5 Alpha VW	−0.254***	0.077	−0.135*	−0.083	0.006	−0.026	0.119	0.204***	0.073	0.150**	0.404***
	(−4.046)	(1.207)	(−1.914)	(−1.235)	(0.066)	(−0.345)	(1.415)	(2.813)	(0.951)	(2.237)	(3.956)

Table IA4: Understanding *SPFRET* Portfolio Alphas: Spanning Regressions

The table presents the results of spanning regressions where the long-short (LS) portfolio return of *SPFRET* is regressed against the LS portfolio returns obtained using other variables that capture economic linkages between firms. *SPFRET* is the composite social-peer return, obtained by combining *SCIRET*, *PROXRET*, and *J2JRET* in an optimal way to predict next month's focal stock return. The other variables that are used to form LS portfolios 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*), and the analyst-linked firm return (*CFRET*). We control for the Fama-French five-factor model but do not report those coefficients for brevity. The sample period is August 1989 to December 2023 due to the availability of *TNICRET*. 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.

	RET							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	1.025***	0.510***	0.504***	0.470***	0.467***	0.306***	0.318***	0.297***
	(3.939)	(4.147)	(3.973)	(3.771)	(3.825)	(2.705)	(2.788)	(2.665)
INDRET		0.769***	0.767***	0.729***	0.681***	0.379***	0.365***	0.309***
		(20.184)	(19.180)	(16.071)	(12.260)	(6.835)	(6.674)	(5.105)
GEORET			0.024	0.030	0.044	−0.054	−0.050	−0.064
			(0.307)	(0.379)	(0.562)	(−0.875)	(−0.841)	(−1.087)
CRET				0.085**	0.066**	0.025	0.027	0.024
				(2.301)	(2.020)	(0.914)	(1.001)	(0.891)
CONGRET					0.114**	0.047	0.042	0.029
					(2.031)	(1.010)	(0.963)	(0.692)
CFRET						0.324***	0.296***	0.248***
						(7.933)	(7.044)	(5.528)
TECHRET							0.069	0.047
							(1.635)	(1.196)
TNICRET								0.159**
								(2.547)
Observations	413	413	413	413	413	413	413	413
R^2	0.066	0.763	0.763	0.768	0.775	0.823	0.825	0.829

Table IA5: 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)

	$\text{SPFRET}_{\perp}$			$\text{SPFMOM}_{\perp}$		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Months 1–3	-0.180^{**}	-0.095	0.085	-0.131	-0.096	0.035
	(-2.057)	(-1.046)	(1.095)	(-1.441)	(-0.755)	(0.346)
Months 4–6	-0.122	-0.106	0.016	-0.096	0.056	0.152
	(-1.606)	(-1.351)	(0.220)	(-1.235)	(0.507)	(1.368)
Months 7–12	-0.138^{*}	-0.132	0.005	-0.227^{**}	0.092	0.320^{***}
	(-1.870)	(-1.623)	(0.110)	(-2.242)	(1.152)	(2.999)
Months 13–24	-0.098	0.015	0.113^{**}	-0.058	0.128	0.186^{*}
	(-1.286)	(0.276)	(2.404)	(-0.650)	(1.645)	(1.724)

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)
$\text{SPFMOM}_{\perp}$	3.154**	2.585*	1.484	2.651*	1.927	4.548**	3.804*	0.535	0.103*	0.191***	0.124**	0.144**
	(2.125)	(1.842)	(0.991)	(1.802)	(1.222)	(2.356)	(1.934)	(0.273)	(1.803)	(3.233)	(2.164)	(2.398)
SPFRET	−1.278	0.499	0.951	1.484*	−0.535	−0.492	1.573	1.011	0.079**	0.046	0.079**	0.006
	(−1.460)	(0.565)	(1.071)	(1.690)	(−0.573)	(−0.547)	(1.303)	(0.812)	(2.202)	(1.372)	(2.112)	(0.155)
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	1,038,809	1,036,019	1,033,289	1,030,450	706,751	636,432	579,685	533,141	1,109,748	1,083,827	1,058,781	1,034,051
R^2	0.173	0.176	0.179	0.177	0.142	0.164	0.174	0.179	0.059	0.064	0.066	0.065

Table IA6: Return Predictability of Peer Firms' Returns: Robustness Checks

The table presents robustness checks for Table 6, column 10 with alternative specifications. Section 2 provides detailed variable descriptions. Column 1 uses the same specification as Table 6, column 10. Column 2 replaces the equal-weighted industry return with its value-weighted version. Column 3 considers the social peer firm return constructed without using *SCIRET*, $SPFRET_{No\ SCI}$. In column 4, *SPFRET* is calculated excluding peer firms headquartered in states where the focal firm has economic presence. In column 5, we add *SIMRET* to the controls, which is the socioeconomicsimilarity-weighted peer firm return. In column 6, we construct *SPFRET* by excluding peer firms within 100 miles of the focal firm's headquarters. Column 7 replaces county-based *SPFRET* with its ZIP code-based version. In column 8, we use an orthogonalized *SPFRET*, defined as a regression residual of *SPFRET* on all the other independent variables. Column 9 uses an alternative standardization method and transforms all the explanatory variables by subtracting the sample mean and dividing by its standard deviation. In column 10, 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 10, 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)	$(2)_{VW}$	$(3)_{No\ SCI}$	$(4)_{No\ Overlap}$	$(5)_{SIM}$	$(6)_{100}$	$(7)_{ZIP}$	$(8)_{\perp}$	$(9)_{SD}$	$(10)_{Panel}$
SPFRET	0.317***	0.618***	0.279**	0.202**	0.345***	0.302***	0.262**	0.328***	0.080**	0.471***
	(2.957)	(5.398)	(2.499)	(2.013)	(2.646)	(2.664)	(2.117)	(3.900)	(2.411)	(3.308)
STRATRET	0.605***	0.650***	0.609***	0.614***	0.610***	0.606***	0.618***	0.639***	0.151***	0.814***
	(4.628)	(4.574)	(4.621)	(4.670)	(4.658)	(4.632)	(4.694)	(4.663)	(4.375)	(4.550)
INDRET	0.636***	0.153	0.660***	0.706***	0.659***	0.647***	0.674***	0.701***	0.174***	0.738***
	(4.276)	(1.034)	(4.242)	(4.791)	(4.005)	(4.311)	(4.992)	(5.196)	(3.460)	(2.886)
GEORET	0.243***	0.225***	0.242***	0.242***	0.239***	0.242***	0.240***	0.224***	0.048***	0.327***
	(3.777)	(3.560)	(3.754)	(3.773)	(3.719)	(3.773)	(3.721)	(3.468)	(3.943)	(2.964)
CRET	−0.068	−0.034	−0.063	−0.059	−0.070	−0.067	−0.053	−0.090	−0.010	−0.014
	(−0.635)	(−0.351)	(−0.595)	(−0.558)	(−0.659)	(−0.630)	(−0.496)	(−0.847)	(−0.733)	(−0.102)
CONGRET	0.467***	0.461***	0.470***	0.468***	0.468***	0.468***	0.473***	0.457***	0.045***	0.438***
	(3.923)	(3.911)	(3.923)	(3.933)	(3.929)	(3.892)	(3.939)	(3.911)	(3.443)	(3.120)
CFRET	1.066***	1.090***	1.070***	1.093***	1.069***	1.066***	1.087***	1.132***	0.288***	1.243***
	(8.192)	(8.032)	(8.231)	(8.432)	(8.201)	(8.174)	(8.588)	(8.147)	(7.751)	(7.696)
SIMRET					−0.048					
					(−0.309)					
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	503	503	503	503	503	503	503	477	500	503
# Stocks	3,761	3,761	3,761	3,761	3,761	3,761	3,761	3,759	3,762	3,761
Time FE	–	–	–	–	–	–	–	–	–	Yes
Observations	–	–	–	–	–	–	–	–	–	1,891,663
R^2	0.065	0.064	0.064	0.064	0.065	0.065	0.064	0.065	0.065	0.196

Table IA7: Return Predictability of In-State Peer Firms

The table presents robustness checks for Table 6, column 9 with an additional control. $SPFRET_{STATE}$ is obtained by reconstructing *SPFRET* using only the peers from the same state as the focal firm (excluding the focal firm itself). Section 2 provides detailed descriptions of the other variables. The sample period is from January 1982 to December 2023 due to the availability of *CFRET*. 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)	(3)	(4)
SPFRET	1.278^{***}		1.225^{***}	0.317^{***}
	(7.255)		(7.326)	(3.013)
$SPFRET_{STATE}$		0.731^{***}	0.363^{***}	0.127^{*}
		(5.487)	(4.010)	(1.692)
STRATRET				0.596^{***}
				(4.583)
INDRET				0.614^{***}
				(4.096)
GEORET				0.239^{***}
				(3.741)
CRET				-0.067
				(-0.634)
CONGRET				0.467^{***}
				(3.928)
CFRET				1.064^{***}
				(8.368)
Controls	Yes	Yes	Yes	Yes
# Periods	503	503	503	503
# Stocks	3,761	3,761	3,761	3,761
R^2	0.058	0.056	0.058	0.065