CFR Working Paper No. 24-03

Traditional Investment Research and Social Networks: Evidence from Facebook Connections

T. Dyer • G. Köchling • P. Limbach

centre for financial research
cologne

Traditional Investment Research and Social Networks: Evidence from Facebook Connections*

Travis Dyer
Brigham Young University

Gerrit Köchling
TU Dortmund University

Peter Limbach
University of Bielefeld and Centre for Financial Research

March 2024

Abstract: We show that investors acquire more public information about firms to which they are more socially proximate. On average, a standard deviation increase in the Social Connectedness Index (Bailey et al., 2018) between a firm's headquarter county and a searcher county is associated with 30% more EDGAR filing downloads from the searcher county. The effect of social proximity on traditional investment research is distinct from the effect of geographic proximity. We find similar results studying headquarter relocations, investor-level data, and EDGAR downloads from European regions, for which physical distance should be irrelevant. Social proximity matters more during times of high market-wide uncertainty and for firms with weaker information environments. Finally, information gathered by socially proximate investors predicts short-term earnings and stock returns, but also heightened volatility. Collectively, the evidence indicates that social networks mitigate informational frictions and foster information acquisition in financial markets.

Keywords: Corporate disclosures, EDGAR, Geography, Information acquisition, Social networks, Social connections

JEL codes: D80, D83, G10, G41, M40

---

* Travis Dyer (corresponding author, contact: travis.dyer@byu.edu) is with Brigham Young University. Gerrit Köchling (contact: gerrit.koechling@udo.edu) is with the Technical University of Dortmund. Peter Limbach (contact: peter.limbach@uni-bielefeld.de) is with the University of Bielefeld and the Centre for Financial Research, Cologne. We are grateful for comments from Nicholas Guest, Bret Johnson, Jed Neilson, Johannes Stroebel, and workshop participants at the BYU Accounting Symposium and Radboud University. We also thank John Christensen for excellent research assistance.

1. Introduction

Traditional investment research, such as reading company press releases, financial reports, and proxy statements, has been a cornerstone of functioning equity markets for decades. However, with the advent of the internet and social media, investors increasingly encounter a wide variety of additional information, which could either replace or enhance traditional research due to their substitutive or complementary nature (Blankespoor et al. 2020). In this regard, recent evidence suggests that information obtained from social media connections is particularly influential, materially affecting how investors allocate their capital (Ammann et al., 2022; Kuchler et al., 2022). While these networks are recognized as important to capital flows, the influence that social ties have on the quality and depth of traditional investment research, and investors' resulting informational gains, remains largely unknown. In fact, social ties may provide information that enhances or substitutes traditional information research. Furthermore, conditional on investing in a firm, it is unclear whether social connection to an investment influences the ongoing level of investor information acquisition. Accordingly, this paper aims to address a crucial question: To what extent do social ties affect the level of traditional investment research by investors?

Social connections have the potential to play a significant role in shaping traditional investment research. On one hand, social networks can facilitate the discovery of new information about investment opportunities by simply increasing awareness.[1] For instance, if a friend from Boston comments on a work-related LinkedIn post, it can heighten awareness of the mentioned company, and thereby lead to information acquisition, even if the observer is geographically distant. This concept aligns with Merton (1987), who highlights the importance of investor awareness in reducing informational frictions and making investment decisions.

[1] Social networks in this paper represent a construct of human connection (e.g., social interactions), rather than a specific mode of connection (e.g., social platforms, emails, group texts, personal visits, etc.).

Beyond awareness and attention, social connections may also offer soft information cues about a company that encourage traditional investment research (i.e., complementary signals). For example, if someone sees a post from a connection showcasing a new Porsche purchase, it might hint at the individual's recent professional successes, leading the observer to research the company the friend is associated with, or to research Porsche and its competitors. Similarly, direct messages between connections discussing a challenging work week or problems at a corporate facility might lead to information gathering and potential adjustments in investment strategies for that particular company. Notably, online social connections do not have to reside with one another or personally meet to be informative. Thus, while social ties are influential in local economies, they are not limited by geographic boundaries.

On the other hand, if social ties provide substitutive signals, they may actually dampen investors' incentives to incur the costs of acquiring traditional information. Specifically, if the insights gained from social connections overlap with insights that can be obtained from traditional research (e.g., information on planned product innovations), the benefits of such research would diminish. Additionally, social connections may spur in-group bias and instill a sense of trust, leading to over-reliance on friends' information and even information "free riding" (e.g., Han and Yang, 2013). These effects can decrease the weight investors place on traditional information sources and, by extension, reduce investors' incentives to obtain further information. Ultimately, how social ties influence investors' decisions to conduct research and gather information is an open empirical question that we hope to shed light on in this paper.

We employ the Social Connectedness Index (SCI), as provided by Bailey et al. (2018) and used in Kuchler et al. (2022), to examine how social connections relate to investors' decision to perform traditional investment research. The SCI assesses the strength of social connections between two counties by tallying the number of Facebook friends linked between them. To measure traditional investment research by investors, we use the internet log files that track the download activity of company filings on the Securities and Exchange Commission's EDGAR (Electronic Data Gathering, Analysis, and Retrieval) server. These downloads represent requests to view major filing types, such as forms 3/4 (i.e., insider transactions) (Chen et al., 2020), form 8-K (i.e., material corporate events) (Ben-Rephael et al., 2022), forms 10-K/Q (i.e., annual and quarterly reports) (Drake et al., 2020), and forms DEF 14A/C (i.e., proxy statements) (Iliev, Kalodimos, and Lowry, 2021). To identify the location of the information requestor, we leverage the ownership record data for the requestor's IP address. We then aggregate the download activity for a specific security by county. The combination of these steps results in a panel of 14,188,331 observations at the county-firm pair year-quarter level that constitute our primary sample. The panel comprises 2,458 unique firms and 2,075 unique counties between Q1 2007 to Q2 2017.

Our analyses reveal that the download activity for a firm's information tends to concentrate in counties with greater social proximity (i.e., search concentrates in counties with heightened Facebook connections to the firm's headquarter county). On average, we find that a standard deviation increase in the SCI for a firm county-searcher county dyad is associated with 30.1% more EDGAR downloads of company filings from the searcher county. Importantly, this associational effect is distinct from local bias in information acquisition (Dyer, 2021) and endowment effects.[2] Specifically, our coefficient estimates for SCI account for whether firms and investors reside in the same county, and hence share the same neighborhood or experience similar

[2] As with recent studies by Ammann et al. (2022) and Kuchler et al. (2022), we view social proximity as a related yet distinct construct from the effects documented about geographic proximity. In our case, while it is evident that distance influences decisions related to information gathering, it is unclear if this is due to social ties, as highlighted in Dyer (2021). Accordingly, we follow prior studies and control for distance effects to underscore the incremental and important role of social proximity (Kuchler et al., 2022).

local shocks, and do not hinge on whether we control for physical distance – either via a continuous variable or using 500-tile distance dummies. They are also robust to the inclusion of a control for the dollar amount invested in each firm by each searcher county as well as various controls for firm county×searcher county similarities. We further add varying combinations of fixed effects, with our most saturated regression model including county×year-quarter and firm×year-quarter fixed effects, which should capture most time-varying unobserved heterogeneity. We find robust evidence that social proximity is related to heightened traditional information acquisition.

The above results suggest that social connections may play an important role in shaping investor behavior, particularly when it comes to traditional investment research. However, despite the large number of observations yielding considerable variation in the data and the use of various fixed effects, omitted variable bias as well as the high correlation between SCI and physical distance may still distort statistical inference. To aid in refining our inferences, we perform several analyses. First, we exploit changes to firm county-searcher county dyads resulting from corporate headquarter relocations that cause plausible exogenous variation in SCI and physical distance. Since SCI and physical distance do not correlate perfectly, at least some headquarter relocations will cause the two to change disproportionately, mitigating both multicollinearity and endogeneity concerns.[3] The results from the difference-in-differences analysis suggest that when a firm relocates, changes in social connectedness between the firm's county and the searcher's county influence information acquisition. This is consistent with our initial regression findings, even when accounting for the effects on physical distance.

[3] We acknowledge that this approach improves parameter identification under the assumption that firms' headquarter relocation decisions are unaffected by expected changes to social connections. In this regard, prior research suggests that firms typically relocate their headquarters based on strategic business decisions, such as lowering employee wages or corporate taxes (e.g., Strauss-Kahn and Vives, 2009).

Second, we exploit granular investor-level data on information acquisition and portfolio holdings of institutional investors using data from Dyer (2021). Consistent with our county-firm-level results, we show that investors conduct significantly more traditional investment research on portfolio firms to which they are socially connected, independent of physical distance. This test allows us to control for investors' specific investment in a company as well as their specialization in certain industries. Importantly, we include firm×year-quarter and investor IP×year-quarter fixed effects, mitigating unobserved heterogeneity concerns (e.g., investors' location, skill, and wealth). We conclude that social connections foster the *ongoing* level of investor information acquisition, even after accounting for the level of their investment.

Our third test exploits the availability of SCI data for dyads of U.S. counties and European (NUTS 3) regions, which allows us to further disentangle social from physical proximity. Specifically, physical distance to specific U.S. counties where firms are headquartered should not matter to investors across Europe because all Europeans have to fly over the Atlantic for site visits or personal meetings with management. Our results support this expectation. While physical distance does not seem to influence the demand for information by Europeans, we observe that social ties across European regions do. This evidence suggests that our results for SCI extend beyond the effect of physical distance. Additionally, these tests highlight that the effects of social ties not only apply within the United States, but also generalize to regions outside the United States.

To further augment our inference that these effects stem from social ties, we conduct "intensity of treatment" tests based on institutional features in Germany. In so doing, we can exploit within-country variation in the *strength* of social ties to the U.S. that resulted from Germany's role in World War II, specifically its separation into East and West Germany and the presence of U.S. military bases across the west. We anticipate that conditional on having a social connection with the U.S., the strength of such ties will be enhanced for Germanic regions in the west and nearer to U.S. military bases. We find empirical support for these predictions. Importantly, the cross-sectional results within Germany help rule out a variety of alternative explanations, since any omitted variable would have to cause similar empirical patterns.[4]

Collectively, these results indicate that social connections foster traditional investment research and that the effects of social ties are not confined to physical proximity. To provide tighter inference on the prevalence and role of social connections in financial markets, we conduct complementary analyses. First, we probe the generalizability of these inferences. We show that the positive association between SCI and EDGAR downloads is prevalent across major filing types, as well as across a variety of information-user types (i.e., financial firms, financial centers, and all other IP addresses) and over time.

Second, to better understand whether social connections mitigate informational frictions, we exploit cross-sectional and time-series variation in market-wide uncertainty and asymmetric information. We find the positive relationship between social connections and information acquisition is amplified during times of greater uncertainty (i.e., during the global financial crisis and during peaks in the CBOE Volatility Index (VIX)). On the contrary, social connections matter less for firms with strong information environments (e.g., larger and older firms, firms with greater analyst following, and firms headquartered in large metropolitan areas). The collection of evidence is consistent with social ties playing an enhanced role when informational frictions are larger.

Finally, we examine whether socially connected counties appear to have an informational edge over less connected counties. Adopting the approach by Drake et al. (2020), we assess if

[4] Please note that the resulting patterns cannot simply reflect differences between East and West Germany or the fact that until today several thousand Americans (e.g., military and their families) live in or close to the U.S. military bases because our regressions include German county×year-quarter fixed effects.

EDGAR download activity from socially connected counties is incrementally predictive of firms' future performance, compared to the counties with weaker social ties. Our findings indicate that information acquisition by socially proximate investors is predictive of unexpected earnings and stock returns over the next month, next quarter, and next two quarters. Such predictive ability is generally stronger than and incremental to counties with lower social connections, suggesting that the average investor with strong social ties has an informational edge. However, socially connected search also concentrates in riskier stocks (i.e., those with higher return volatility). Accordingly, we find little predictive ability for risk-adjusted returns, consistent with Kuchler et al. (2022) who document little portfolio alpha to socially allocated capital.

The results of our study contribute to the emerging literature on social connections and investor behavior (e.g., Amman et al., 2022; Hirshleifer, Peng, and Wang, 2023; Kuchler et al., 2022), highlighting the importance of social networks in shaping traditional investment research. Our findings suggest that social connection enhances traditional information gathering, leading to more informed investment decisions, even if those decisions may not improve portfolio alpha. We thereby contribute more generally to the literature on investors' information choice and market outcomes (see Blankespoor, deHaan, and Marinovic, 2020, for an overview).[5] We believe understanding the impact of modern social networks and connections on aspects of investment decisions is critical for policymakers and market participants alike, as it sheds light on the current factors shaping information markets and can inform efforts to improve market efficiency.[6]

[5] We also add to the growing literature on information acquisition via EDGAR (e.g., Drake, Roulstone, and Thornock. 2015; Chen et al., 2020; Drake et al., 2020; Gibbons, Iliev, and Kalodimos, 2021).

[6] Our study is related to but significantly different from extant work on social interactions in neighborhoods (e.g., Hong, Kubik, and Stein, 2004 and 2005; Pool, Stoffman, and Yonker, 2015). In contrast to this work, we study social connections in general – not geographically limited, direct personal interactions with neighbors – and examine their role with information acquisition. Extant studies examine households' decisions to participate in the stock market or investment decisions by mutual fund managers.

The results of our study complement recent evidence on information gathering decisions of local vs. non-local investors (Chi and Shanthikumar, 2017; Dyer, 2021) and the declining informational advantage of physical proximity since the early 2000s (Bernile et al., 2019). In aggregate, our results suggest that social connections appear to play a contributing role in the documented effects of physical proximity, particularly instances in which investors and firms do not reside in the same county. We thereby provide a potential response to extant work wondering why local bias in investor portfolios has not decreased in the modern information age (e.g., Van Nieuwerburgh and Veldkamp, 2009).

2. Hypotheses Development

Social networks, as highlighted by Granovetter (2005), play a pivotal role in shaping economic outcomes by enhancing the flow of information. Demonstrating the importance of social connection, individuals spend, on average, about two and a half hours *each day* using social media,[7] with the main motivation being a desire to stay in touch with family, current and old friends, and those that share similar interests.[8] Given the importance of social networks and how such networks tend to augment the flow of information, our study aims to understand how such networks impact traditional investment research. In particular, we investigate how social networks shape investors' demand for information that firms publicly disclose.

Ex ante, it is unclear whether and how social networks shape investors' information-gathering decisions. These ties may affect investors' screening process by enhancing the discovery of new investment opportunities. For instance, an acquaintance mentioning a lesser-known firm from a rural area, might stimulate an investor's curiosity to study the firm further. This aligns with Merton's (1987) theory that heightened investor awareness can diminish informational frictions

[7] See https://datareportal.com/reports/digital-2023-july-global-statshot.
[8] See https://www.pewresearch.org/internet/2011/11/15/why-americans-use-social-media/.

and promote investment. Such increased awareness may describe why investors allocate more capital to socially connected firms (Ammann et al., 2022; Kuchler et al., 2022).

Social ties may help serve as a useful monitoring mechanism. An investor can learn about ongoing structural challenges or internal dynamics of a portfolio company through a connection, especially if the connection is associated with that firm. Such soft information may prompt an investor to seek corroborative hard data from traditional information sources, such as annual reports or other corporate disclosures. Additionally, these connections may also elevate significant corporate events to investors' attention, such as mergers or financial misstatements. To the extent that information signals from social ties are complements to traditional investment research, social connections may encourage additional due diligence research by market participants.

While unlawful, social connections are also known to periodically have access to illegal insider information, as suggested in Ahern (2017). Such information certainly gives connections a competitive advantage, but it is unclear how that informational advantage would change investor's demand for traditional market information. It's possible that such inside information may either discourage or encourage further research by connected investors. On one hand, possessing insider information (or perceived insider information) might make investors complacent, believing they already have a decisive edge. Some investors might even avoid seeking more information to prevent the appearance of impropriety. On the other hand, it could pique their curiosity, prompting them to validate or contextualize this information with further research.

There are also several reasons why social proximity may discourage traditional investment research. For example, if social connections tend to offer insights that are commonly obtained from traditional investment research (i.e., substitutive signals), the incentives for traditional information research would decrease. Additionally, social ties are known to incentivize information "free riding" (e.g., Han and Yang, 2013) and over-reliance on their ties' information. This may be because social connections tend to foster trust and in-group bias, which is associated with higher levels of perceived competence (Baskett, 1973; Hewstone, Rubin, and Willis, 2002). Such free riding and over-reliance on social ties may reinforce confirmation bias and cause investors to place less weight on alternative information, and by extension, decrease investors' incentive to gather alternative information via traditional investment research.

The aforementioned evidence and incentives suggest that, ultimately, it is an empirical question whether and how social connections affect investors' demand for corporate information. Accordingly, we formulate our main hypothesis (H1) in its null form:

H1: *Social connections do not affect investors' traditional information acquisition.*

To the extent that social connections help reduce informational frictions, we anticipate that the impact of these connections will be more pronounced under conditions of heightened uncertainty and information asymmetry. In other words, we expect the value of social connection to be higher in market conditions where there are heightened informational frictions. In the absence of such conditions, or if uncertainty increases the demand for verified or verifiable information, the significance of these connections may diminish. We hence pose our second hypothesis (H2) in its null form:

H2: *The effect of social connections does not depend on uncertainty and asymmetric information.*

3. Data, Methodology, and Summary Statistics

3.1 Data

To construct our sample, we merge data from various sources. We start by obtaining data on investors' information acquisition activity from the SEC EDGAR system (i.e., downloads of company filings as recorded on the EDGAR log files database). To keep our sample manageable, we constrain our data collection in two ways: (1) we focus on firms in the S&P 1500, plus firms that were once part of the index, and (2) we only obtain EDGAR download activity for the following major filing types – forms 3 and 4 (i.e., insider transactions), form 8-K (i.e., current reports of material corporate events), forms 10-K and 10-Q (i.e., annual and quarterly reports), and forms DEF 14A and 14C (i.e., proxy statements for annual and special shareholder meetings).[9] For perspective, Chen et al. (2020) show that these filing types account for the majority of downloads from the EDGAR server, both in general and by institutional investors.

The EDGAR log file database generally covers downloads between January 2003 and June 2017. However, due to errors in the data from mid-2005 to mid-2006 (Bauguess, Cooney, and Weiss Hanley, 2018), we start our sample in January 2007. We follow prior literature and filter robot searches following the procedure described in Ryans (2017): for each day, we keep requests with successful delivery by the EDGAR server, remove index page observations, remove requests from self-identified web crawlers, and remove requests from IP addresses with requests per minute greater than 25, or with a number of CIKs downloaded per minute greater than 3, or with more than 500 downloads during the day.

A unique feature of the EDGAR log file data is that it retains sufficient information about the requesting IP address, such that geolocation data can be attributed. Accordingly, we merge our sample with geolocation information on IP addresses from IP2Location. This data allows us to assign IP addresses to US county FIPS codes (or NUTS3 European region codes), which we label the “searcher county.” For each firm year-quarter, we aggregate the number of filings downloaded

[9] Specifically, we use the following form types: "3", "3/A", "4", "4/A", "8-K", "8-K/A", "10-K", "10-K/A", "10-K405", "10-K405/A", "10-Q", "10-Q/A", and "DEF 14A", "DEFA14A", "DEF 14C", "DEFR14A", "DEFM14A", "DEFS14A", "DEFC14A", "DEFN14A", "DEFM14C", "DEFR14C", "DEFA14C".

from EDGAR from IP addresses located in the searcher county. We then assign corporate headquarter location details based on disclosed annual reports.[10]

We obtain data on social connectivity between counties and regions across the globe (the Social Connectedness Index) from Humanitarian Data Exchange.[11] We use annual macroeconomic county-level data from the United States Census Bureau, quarterly firm-level accounting data from Compustat, and investor holdings from the Thomson Reuters 13F Institutional Holdings database. Investor location data is obtained from Nelson's Directory of Investment Managers and by searching SEC documents and websites of institutional managers.[12] Our final sample comprises 14,188,331 observations at the county-firm pair year-quarter level, covering all quarters from Q1 2007 to Q2 2017. The sample constitutes 2,458 unique firms and 2,075 unique counties. On average, the sample includes 308,442 observations per quarter. There is no county in our sample without any downloads of company filings from the EDGAR server.

3.2 Methodology

We use the county-firm-quarter panel described above to conduct regressions using the following regression model:

$$\begin{aligned}\mathrm{Search\ Volume}_{i,j,t} &= \beta_0 + \beta_1 \ln(\mathrm{Social\ Connectedness})_{j,t} + \beta_2 \ln(\mathrm{Distance})_{j,t} + \beta_3 \ln(\mathrm{AUM})_{j,t} \\ &+ \beta_4 \mathrm{Firm\ controls}_{i,j,t-1} + \beta_5 \mathrm{County\ controls}_{j,t-1} + \mathrm{FEs} + \varepsilon_{i,j,t}\end{aligned} \quad (1)$$

The dependent variable *Search Volume* is the sum of daily EDGAR downloads of firm *j* disclosures by IP addresses located in county *i* during the year-quarter *t*. The use of a count variable of EDGAR

[10] We use the 10-K header information files from Loughran-McDonald Software Repository for Accounting and Finance. See https://sraf.nd.edu/.
[11] We retrieve the data from https://data.humdata.org/dataset/social-connectedness-index.
[12] We thank Gennaro Bernile, Alok Kumar, and Johan Sulaeman for sharing this data with us.

downloads as a measure of investors' information acquisition follows several prior studies (e.g., Drake, Roulstone, and Thornock, 2015; Drake et al., 2020; Iliev and Lowry, 2021).

Our variable of interest is *Social Connectedness*, which is the Social Connectedness Index proposed by Bailey et al. (2018). It measures the social proximity (or strength of social connection) between two counties based on anonymized cross-sectional data on Facebook friendship networks, i.e., friendship links between U.S.-based Facebook users.[13] Formally, the social connectedness between county *i*, the searcher county from where an IP address downloads a company filing, and county *k*, in which the company is headquartered, is defined as:

$$Social\ Connectedness_{i,k} = \frac{FB\_Connections_{i,k}}{FB\_Users_i * FB\_Users_k}.$$

Here, FB_Users_i and FB_Users_k are the number of Facebook users in counties *i* and *k*, and $FB\_Connections_{i,k}$ is the total number of Facebook friendship connections between individuals in the two counties. The measure is scaled to have a minimum value of 1 and a maximum value of 1,000,000,000. As a result, *Social Connectedness* measures the relative probability of a Facebook friendship link between a given Facebook user in county *i* and a given user in county *k*. Put differently, if this measure doubles in size, a given Facebook user in county *i* is about twice more likely to be connected with a given Facebook user in county *k*. We use the

[13] There is also data on friendship links between U.S.-based and international Facebook users. For a detailed description of the data, we refer the reader to Bailey et al. (2018). Facebook is the world's largest online social network with about 3 billion active global users and ca. 180 million active U.S. users as of early 2023. In January 2021, Facebook was the platform on which users in the U.S. spent the most time per day, 33 minutes on average (https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/). According to a survey of Facebook users, 79% of online adults in the U.S. across income groups, education levels, race, and location use Facebook (Duggan et al., 2016). In the U.S., Facebook serves as a social network for real-world friends and acquaintances, and users usually only add connections on Facebook to individuals whom they know in the real world (Kuchler et al., 2022). Bailey et al. (2018), among others, provide evidence that friendships observed on Facebook are a good proxy for real-world U.S. social connections. Thus, Facebook networks resemble real-world social networks but are not subject to the same frictions, such as geographical distance, that hamper social interaction.

natural logarithm of the Social Connectedness Index, i.e., *ln(Social Connectedness)*, in most regressions because the index data is strongly right-skewed and takes very large values.

Our regression model includes numerous control variables, which we define in Appendix Table A1. Importantly, since Dyer (2021) shows that physical distance between firms and investors explains information acquisition, we also control for the natural logarithm of physical distance (*ln*(*Distance)*) between the county *k*, where a firm is headquartered, and the searcher county *i*. We measure distance using great-circle distances, calculated using the Haversine formula and based on internal points in the geographic area (NBER County Distance Database). Following Kuchler et al. (2022), we also include an indicator that equals one if firm *j* is headquartered in searcher county *i* in each regression (*Same county FE*) to capture county-level heterogeneity, such as firms and nearby searchers locating in the same neighborhood, which may affect information and investment (Coval and Moskowitz, 1999; Hong, Kubik, and Stein, 2005), or both being affected by the same local economic shocks or regulations.

As endowment effects might affect search activity, we also control for the natural logarithm of the U.S. dollar amount invested in each firm *j* by searcher county *i* (*ln(AUM)*). We use data on investor locations from Bernile, Kumar, and Sulaeman (2015) and Bernile et al. (2019). We aggregate the amount invested in each firm across all investors located in the same U.S. county in each year-quarter.

Furthermore, we control for several firm and county characteristics. Quarterly firm-level controls include a firm's age (*ln(Age)*) and size (*ln(Assets)*), defined as the natural logarithm of a firm's years since IPO and its total assets, respectively; a firm's total liabilities divided by assets (*Leverage*); return on assets (*ROA*), defined as the operating income before depreciation divided by assets; and market-to-book ratio (*MarketBook*), defined as the market value of equity divided by its book equity. Annual county-level controls include the unemployment rate (*Unemployment*), the logarithm of a county's population (*ln(Population)*), and its per capita income (*Income*). In additional analyses, we include a large set of controls for county-pair similarities.

We saturate the regression model with various fixed effects (FE) to mitigate concerns of omitted variable bias resulting from unobserved heterogeneity across counties, firms, and time. In particular, we use varying combinations of fixed effects, namely (1) county, firm, and year-quarter FE, (2) firm and county×year-quarter FE, (3) county and firm×year-quarter FE, and (4) county×year-quarter and firm×year-quarter FE. Combination (4) should capture most unobserved time-varying heterogeneity at the county and firm level.

To estimate the regression model, we follow Kuchler et al. (2022) and use Poisson Pseudo Maximum Likelihood (PPML) regressions since *Search Volume* is a left-censored count variable with many zeros (i.e., no EDGAR downloads for firm *j* by searcher county *i* in quarter *t*).[14] In this regard, Cohn, Liu, and Wardlaw (2022) demonstrate that regressions of the log of 1 plus the outcome variable yield coefficient estimates that have no natural interpretation and can have the wrong sign in expectation. They recommend the use of Poisson models with fixed effects to get consistent and reasonably efficient estimates. We winsorize all continuous variables at the 1st and 99th percentiles to mitigate the impact of outliers in our regressions. Again following Kuchler et al. (2022), we double cluster standard errors at the dyad levels (i.e., county and firm).

In supplemental analyses, we use several variables as alternative outcomes or moderating variables. This includes variation in information acquisition by filing type (i.e., forms 3 and 4, 8-K, 10K/Q, and DEF 14) and investor type (i.e., financial firms and financial centers). *Financial*

[14] PPML regressions are widely used in the trade literature because of the left-censoring of trade flows between countries (see Santos Silva and Tenreyro, 2006, for a discussion). Our results are qualitatively similar using OLS and log of (1+ the count variable).

Crisis is an indicator that equals one if an observation is in the year-quarters between Q3 2008 and Q1 2009. *VIX* is the average value of the CBOE Volatility Index (VIX) each year-quarter. *Analysts* is the number of analysts covering a firm as reported in I/B/E/S. *Metropolitan Area* is an indicator that measures if a firm is headquartered in a metropolitan area with a population of more than a million people.

3.3 Summary Statistics

Panel A of Table 1 provides summary statistics for the previously described variables based on the 14,188,331 county-firm-quarter observations in our sample. The left-censoring, skewness, and many zeros of our dependent variable *Search Volume* can directly be inferred from its distribution. The variable has a mean of 5.29, a median of 0, and a standard deviation of 15.33. The search volume by IPs of financial firms or financial centers shows an equal pattern.

Social Connectedness has a mean and median of 6,032 and 2,128, respectively, and a standard deviation of 17,795, indicating that the variable is right-skewed and takes large values, which is why we use its natural logarithm in all of our regressions. *ln(Social Connectedness)* has a mean and median of 7.83 and 7.66, respectively. *Distance* has a mean (median) of 1,686 (1,433), indicating that the average firm is about 1,686 kilometers away from each searcher county on average. *ln(Social Connectedness)* and *ln(Distance)* exhibit a correlation of -72%, consistent with Kuchler et al. (2022) who report -69%. Figure 1 illustrates this correlation by displaying social connectedness and distances between Montgomery County in Ohio, as a representative county, and all other counties in the U.S. Despite the significant correlation, the figure shows that there are many distant counties across the U.S. with a high connectedness to Montgomery County (e.g., parts of Arizona and Colorado as well as southern Florida), consistent with social connections capturing something distinct from geographic distance.

4. Social Proximity and Investors' Information Acquisition: Empirical Evidence

4.1 Baseline Estimates

Panel B of Table 1 presents univariate tests comparing mean and median values for *Search Volume* when we partition the sample based on the sample median of *Social Connectedness*. Observations with above-median values for social connectedness show significantly larger mean and median values for the variable *Search Volume* as well as for the natural logarithm of 1 plus this variable. The same holds true when we account for the high correlation between social and physical proximity by partitioning the sample based on the residual of social connectedness that stems from a regression of *Social Connectedness* on *Distance*. Consistently, Figure 2 shows a strict positive relation between *Social Connectedness*, measured in deciles on the x-axis, and average *Search Volume* per decile. Thus, descriptively, investors located in counties with more social connections to a firm's headquarter county download significantly more of that firm's filings from EDGAR.

The above univariate evidence implies a positive relation between investors' information acquisition and social proximity. We next test whether this relation is robust to multivariate regressions. To this end, we estimate the regression model in equation (1), varying fixed effects across columns. Panel A of Table 2 shows the regression results. In columns (1) and (2), we regress *Search Volume* on *ln(Distance)* and *ln(Social Connectedness)*, respectively, omitting all control variables and fixed effects. In columns (3) and (4), we repeat the regressions analogously, including firm and county controls as well as county, same county, firm, and year-quarter fixed effects. In the four columns, the coefficients on both variables are statistically significant at the 1% level. Both also have the expected signs, negative for *ln(Distance)* in columns (1) and (3) and positive for *ln(Social Connectedness)* in columns (2) and (4), consistent with the aforementioned univariate evidence. As shown in columns (5) to (8), we observe a strong positive association between *Search Volume* and *ln(Social Connectedness)*, with a virtually identical economic magnitude, even after controlling for geographic distance effects via *ln(Distance)* and the *Same County* fixed effect. While the coefficient on the former is statistically insignificant, the coefficient on the latter (not tabulated for brevity) is positive and significant at the 1% level in all columns. This result is consistent with the evidence on local investors' information demand in Dyer (2021).

Panel B of Table 2 shows the results from re-estimating the regressions in columns (3) to (8) of Panel A, adding a large set of control variables that account for similarities between firm and searcher counties. Specifically, because social connections are more likely to form between individuals who have something in common (e.g., a similar age or ethnicity), and because similar individuals may share more information, we control for various aspects of socio-demographic similarity between firm and searcher counties.[15] To account for economic connectedness, we also control for industry similarity between searcher counties and firms using the variable *% Firms Same SIC3*. It is defined as the number of firms headquartered in a searcher county that have the same three-digit SIC code as the focal firm, divided by the total number of firms in the three-digit SIC industry.[16] Controlling for industry and socio-demographic similarity, the coefficient on *ln(Social Connectedness)* remains virtually unchanged, both statistically (t-value $\geq$ 12.8) and economically ($\beta \geq 0.255$).

In additional regressions, shown in Appendix Table A3, we follow Kuchler et al. (2022) and re-estimate the regressions in Panel A of Table 2 controlling for physical distance between

[15] We control for similarity in terms of age, education, employment, family, income, mobility, and race. We use data from the American Consumer Survey, which is available from 2009 onwards. Thus, the regressions include fewer observations in Panel B compared to Panel A. We measure similarity between county pairs via the absolute value of the difference between county mean values (e.g., the different share of people between 45 and 64 years in the firm and the searcher county, i.e., *Abs. Diff. % Age 45-64*). Appendix Table A2 shows the coefficient estimates for the various socio-demographic variables.

[16] In untabulated tests, we use industry classifications based on pairwise similarity scores from text analysis of firms' 10-K product descriptions (TNIC) from the Hoberg-Phillips Data Library (see Hoberg and Phillips, 2016) and find similar results.

county pairs using 500-tile dummies instead of the continuous variable *ln(Distance)*. The coefficient on *ln(Social Connectedness)* remains positive and significant at the 1% level. Furthermore, in untabulated regressions, we find qualitatively similar results when we re-estimate our regressions (i) omitting the ten counties with the greatest search volume (e.g., Dallas County, New York county, and Los Angeles County) or (ii) clustering standard errors at the county level only. Collectively, the results suggest that the estimates for *ln(Social Connectedness)* are not driven by certain counties or county-pair similarities and do not hinge on the choice of fixed effects, standard errors, or the inclusion of *ln(Distance)* as a control variable, speaking to the robustness of our evidence. In terms of the economic magnitude of social connectedness, the multivariate estimates indicate that a change between the 25th and 75th percentile of social connectedness is equivalent to approximately a 33.2% increase in information acquisition.[17] Similarly, a standard deviation increase in *Social Connectedness* from the mean is associated with an increase in *Search Volume* by about 30.3%.[18]

The observed positive association between social proximity and traditional information gathering could be driven by two underlying forces, namely (i) increased investor awareness (i.e., more IP addresses searching) and (ii) increased search, conditional on awareness (i.e., more filings searched per IP address). To provide insight into these underlying forces, we construct two additional measures of information gathering that map to these two constructs. Appendix Table A4 Panel A (Panel B) presents the association between social connectedness and the number of unique IP addresses searching a given firms' filings (the average number of filings searched per IP address). We find evidence consistent with social proximity raising investor awareness (i.e., more IPs searching) and amplifying interest once that awareness exists (i.e., more search per IP).

[17] $(e^{0.2625} - 1) * (\ln(3{,}934) - \ln(1{,}300)) = 33.2\%$.
[18] $(e^{0.2625} - 1) * 1.01 = 30.3\%$.

4.2 Assessment of Robustness

The above results consistently suggest that social proximity affects investors' traditional investment research. While it is unlikely, given our stringent fixed effects design, we acknowledge that an omitted variable may still be descriptive of our findings. Given the substantial correlation between physical and social proximity, it is also possible that multicollinearity may inadvertently influence our results. Accordingly, we provide several tests that address these issues.

The first test exploits the relocation of corporate headquarters as changes to firm county-searcher county dyads that cause plausibly exogenous variation in investors' physical and social proximity. Specifically, when firm *j* moves its headquarters from county *k* to county *h*, both the firm's physical distance and the social connectedness to searcher county *i* have the potential to change. Importantly, because physical distance and social connectedness do not correlate perfectly, at least some headquarter relocations will cause the two to change disproportionately, so that physical proximity changes more than social proximity, or vice versa. This mitigates multicollinearity concerns. Also, to the extent that firms' headquarter relocation decisions are unaffected by expected changes to social proximity, such tests also enhance identification and mitigate endogeneity concerns (e.g., endowment effects). In this regard, prior research suggests that firms tend to relocate their headquarters to medium-sized service-oriented metropolitan areas, based on strategic business decisions, such as lowering employee wages or corporate taxes (e.g., Strauss-Kahn and Vives, 2009).[19]

We identify corporate headquarter relocations from our panel dataset and use them in a difference-in-differences (DID) setting. Following Hasan et al. (2017), we focus on the sample of

[19] We do not anticipate that executives select headquarter locations based on investors' preference for enhanced social proximity. This belief is strengthened by the fact that corporate investors are scattered both nationwide and internationally. For example, a geographic headquarter move from county *a* to county *b* affects the firm's social connection with both county *i* (an increase in social ties) and county *k* (a decrease in social ties).

firms that relocate and exclude the year during which firms relocated. We then study the four years that surround the relocation year (i.e., -/+16 quarters). We re-estimate the regressions shown in columns (3) to (8) of Table 2 with the variables Δ *Social Connectedness* and Δ *Distance*. These variables represent our treatment variables and capture the extent to which relocation-induced changes affected social and geographic proximity. The indicator variable *Post* equals one for all year-quarters after a firm relocated its headquarters. Table 3 presents the results, which echo those from Table 2. In particular, throughout columns (2) to (6), the coefficient on *Post*×Δ *Social Connectedness* is negative and statistically significant at the 5% level, except for column (5) where it is significant at only the 10% level. This holds true even when controlling explicitly for simultaneous changes in geographic distance. Additionally, in column (7) we provide evidence that the parallel trends assumption of our DID approach is not violated. Collectively, the findings support the inference that social connectivity affects traditional information acquisition, even after accounting for distance, endowment, and fixed attributes between firm county-searcher county dyads.

Our second test provides investor-level evidence based on granular information acquisition and portfolio holdings data from Dyer (2021). The sample comprises up to 4,108,417 observations at the investor-firm pair year-quarter level, covering all quarters from Q1 2007 to Q1 2015. It includes 3,092 unique firms and 9,810 unique investor IPs. Table 4 presents the regression results. We successively include additional controls between column (1) and (5). Importantly, this test allows us to control for investors' specific investment in a company as well as investors' specialization in certain industries. Consistent with our county-firm-level results, we find that institutional investors conduct significantly more traditional investment research on portfolio firms to which they are socially connected, independent of physical distance. The coefficient on our variable of interest, *ln(Social Connectedness)*, is statistically significant at the 5% level or better in all columns. Regarding the economic magnitude of our coefficient estimates, a one standard deviation increase in *ln(Social Connectedness*) is associated with a 13.1% increase in *Search Volume*. Further, since we include firm×year-quarter and investor IP×year-quarter fixed effects, we can mitigate several unobserved heterogeneity concerns (e.g., investors' location, skill, and wealth). The results indicate that social connections influence the *ongoing* level of investor information acquisition, even after accounting for the level of their investment.

Our third test exploits the fact that Facebook is a global social network and that the EDGAR server can be accessed from around the globe, independent of an investor's location. Bailey et al. (2018) also provide data on the social connectedness between U.S. counties and European regions, where the latter are based on the NUTS 3 classification.[20] This data allows us to examine whether investors from European region *i* download more of firm *j*'s corporate filings if firm *j* is headquartered in a U.S. county that is socially more proximate to European region *i*. The primary advantage of this test is its ability to distinguish between physical and social proximity more effectively. For European investors, the actual physical distance to specific U.S. counties should not be a significant factor, given that they all need to cross the Atlantic. However, social proximity is likely to be influential, as European investors can gain valuable insights from their social networks in the U.S. regardless of geographic constraints. This test ensures that the construct of social connectedness is not conflated with physical distance, and vice versa. Supporting this idea,

[20] According to the Eurostat website (https://ec.europa.eu/eurostat/web/nuts/background), the NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU and the UK. NUTS 3 is the most granular classification. As of January 2021, it lists 1,166 different European regions. The following link provides a map of all Dutch NUTS 3 regions as an example: https://ec.europa.eu/eurostat/documents/345175/7451602/2021-NUTS-3-map-NL.pdf.

the correlation between the variables *ln(Distance)* and *ln(Social Connectedness)* attenuates by 78%, relative to the U.S. setting.

Table 5 presents the results from the above test. Panel A shows the results from regressions that are identical to those in columns (3) to (8) of Table 2, except that the variable *AUM* is omitted (due to lack of data for Europe) and that the variables *ln(Distance)* and *ln(Social Connectedness)* now capture proximity between U.S. counties and European NUTS 3 regions. Supporting our reasoning that physical distance to U.S. counties should not matter to European investors, we find a statistically insignificant coefficient on *ln(Distance)* in column (1), where we omit the variable *ln(Social Connectedness)*, as well as in columns (3) to (6). In contrast, we find a positive coefficient on *ln(Social Connectedness)*, which is significant at the 1% level throughout columns (2) to (6). As in the U.S. setting, the coefficient estimate for *ln(Social Connectedness)* does not hinge on the inclusion of *ln(Distance)* as a control variable, nor on our choice of fixed effects. In economic terms, the coefficient on *ln(Social Connectedness)* amounts to 0.1554 in column (6), which is approximately 60% of the corresponding coefficient estimate for the U.S. setting.[21]

The results for Europe support our U.S. evidence. To further augment our inference that these effects stem from social ties, we conduct "intensity of treatment" tests. We begin by recognizing that, conditional on having a social tie, not all social ties are equally strong. We capitalize on this intuition using a German setting. Germany provides a unique laboratory for examining social ties with the U.S. given its role in World War II. Following the war, the country was divided into East and West Germany until the reunification in 1990. East Germany's identity was significantly molded by the Soviet Union and its Cold War interactions with the U.S. Conversely, West Germany was influenced predominantly by the Allied powers, especially the

[21] The magnitude of the coefficient is smaller than in the U.S. setting, which appears reasonable since European investors invest relatively less in U.S. firms due to home bias.

United States. Even today, the U.S. maintains military bases in West Germany, including bases like Ramstein and the Stuttgart Army Air Field.

We exploit the above exogenous features of the German setting by conducting two "intensity of treatment" tests. First, we expect that conditional on social ties, the *strength* of such social connections with the U.S. will attenuate for regions in East Germany.[22] As a result, we use the interaction term *ln(Social Connectedness)×East Germany* and expect a weakening of the effects from social connectedness on information gathering. The indicator *East Germany* is equal to one for all NUTS 3 regions in East Germany, except for Berlin (which has in part belonged to the West). Second, we expect that conditional on social ties, the *strength* of such social connections with the U.S. will amplify for regions near U.S. military bases in Germany. This rationale is based on the idea that Germans in these areas have had significant interactions with Americans, fostering a heightened trust with U.S. connections. We use *ln(Social Connectedness)×Distance US* $Base_{<25km}$ to examine this second intensity of treatment test. *Distance US* $Base_{<25km}$ is an indicator that equals one if a German NUTS 3 region is no more than 25 kilometers away from a U.S. military basis.

Panel B of Table 5 presents the results based on the German regions. We conduct regressions similar to that in column (6) of Panel A. As before, we find the coefficient on *ln(Distance)* to be statistically insignificant, while that on *ln(Social Connectedness)* is positive and significant at the 1% level. The size of the coefficients compare well to those for the U.S. setting. Most important, in columns (5) and (6) we estimate our intensity of treatment tests. We find that the coefficient on *ln(Social Connectedness)×East Germany* is negative and marginally significant (p-value = 0.0524), while the coefficient on *ln(Social Connectedness)×Distance US* $Base_{<25km}$ is

[22] We note that according to Brodbeck and Frese (2007) East and West German cultures are generally very much alike and characterized by similar values and attitudes.

positive and significant at the 1% level. As our regressions include German county×quarter fixed effects, these results cannot simply reflect differences between East and West Germany or the fact that several thousand Americans (e.g., military and their families) live and work close to or in the military bases. Rather, the results are consistent with societal imprints that affect the *strength* of social connections between Germans and the U.S.

The results presented in Tables 3 to 5 support our baseline estimates from Table 2. In additional untabulated tests, we exclude all firms reporting foreign income from Table 5 and find qualitatively similar results. Thus, the results are unlikely to reflect (i) international firms whose employees download their employers' disclosures while working abroad or (ii) social connections from business operations abroad. Collectively, the evidence indicates that social connections do appear to affect investors' information acquisition activity, whereby we reject H1. This result holds when accounting for endowment effects, county selection effects, and distance effects. It not only applies to U.S. counties and investors, but also European regions. When we increase the strength of given social connections, we observe the effect enhances. As such, any remaining omitted variable would have to cause empirical patterns similar to the ones reported in each of our tests.

Despite this robust evidence, we present the results of two supplemental tests. First, while we can and do exploit intertemporal variation in *firm x searcher* county pairs by using headquarter relocations (Table 3), the Facebook data on social connectedness does not provide information on time-series variation in social connectivity for county pairs. However, in the documentation of the SCI data, the group of authors state: *"the underlying object captured by the SCI […] is highly stable over time. As a result, social connectedness as measured today is likely to predict interactions over other time horizons."* [23] Consistently, Kuchler et al. (2022) show that social

[23] The SCI data documentation is available at https://data.humdata.org/dataset/social-connectedness-index.

connectedness as measured today predicts mutual fund investments in the 2000s equally well as it predicts them today. Nonetheless, to address the concern that the SCI data might not accurately reflect the dynamic historical patterns of social connectedness, we use county-to-county migration information from 1990 to 2011 based on Statistics of Income data from the IRS to identify county pairs with stable migration patterns. Arguably, county pairs with very little time-series change in migration flows would be the county pairs in which our fixed social connectedness measure is most representative. We calculate the rates of change in migration for each county-pair and repeat our baseline test (Table 2) using the 80% (or alternatively 50%) of the county-pairs with the lowest absolute change in migration flows. We find qualitatively similar results in this alternative setting.

Second, we present the results from supplemental 2SLS regressions in Appendix Table A5. We employ a *Same Highway* indicator, based on Baum-Snow (2007), as an instrumental variable to illicit plausibly exogenous variation in *ln(Social Connectedness)*. This instrument indicates if a searcher county and a firm's headquarter county share a highway. Historically, U.S. highways, planned during WWII and built in the Cold War era, inadvertently fostered social connections by facilitating commuting, travel, and suburbanization. We posit that some of these historical social ties persist, making *Same Highway* a significant predictor for *ln(Social Connectedness)*. Note that, by design, our instrumental variable cannot explain physical distance. The first-stage regression in column (1) confirms the instrument's relevance, with a positive coefficient significant at the 1% level. The presence of a connecting highway boosts social connectedness between counties by approximately 22%. Second-stage regressions in column (2) align with our earlier findings, with the instrumented social connectedness significant at the 1% level. The magnitude of the IV estimate for *ln(Social Connectedness)* is only slightly larger than our baseline estimates in Table 2 or untabulated OLS estimates, supporting the quality of our instrument (see Jiang, 2017). While we cannot test the exclusion restriction, it seems improbable that highway connectivity, beyond the controls we use, influences investors' information acquisition decisions except via improvements in social connectivity.

4.3 Social vs. Geographic Proximity

Geographic proximity is a common empirical measure in academic work, despite its lack of a clear underlying construct. Scholars have argued that it captures a variety of constructs like transaction cost differences, informational frictions, behavioral biases, media coverage differences, and social connection differences (see Cooper, Sercu, and Vanpée, 2013). Our results suggest that social proximity – distinct from geographic proximity – appears to drive investors' information acquisition. However, one obvious question is to what extent earlier studies on information demand and geographic effects may have captured the construct of social connectedness.

To explore this line of questioning, we provide an additional test. We first conduct an OLS regression of *ln(Distance)* on *ln(Social Connectedness)*, which yields fitted values and residuals denoted as $ln(Distance)_{Fitted\ Values}$ and $ln(Distance)_{Residuals}$, respectively. The coefficient on *ln(Social Connectedness)* is significantly negative and the regression's R-squared is 51.9%, indicating that social connectedness explains more than half of the variation in geographic proximity. We then re-estimate our baseline regressions shown in Table 2 substituting the variables *ln(Social Connectedness)* and *ln(Distance)* with the fitted values and residuals for the latter. Table 6 presents the regression results. Throughout all columns, and independent of whether we control for the *Same County* fixed effect, the coefficient on $ln(Distance)_{Fitted\ Values}$ is negative and significant at the 1% level, while the coefficient on $ln(Distance)_{Residuals}$ is positive and statistically insignificant. Hence, only variation in geographic proximity that relates to variation in social connections has explanatory power for investors' information acquisition. Geographic proximity orthogonalized for social connections does not explain search volume. While relatively simple, this test suggests that earlier academic work connecting geographic effects to information acquisition (Dyer, 2021) may have captured the underlying construct of social connectedness.[24]

4.4 Assessment of Generalizability

Having established a significant and robust link between investors' information acquisition and social proximity to firms, we next attempt to address the question of how pervasive, and hence generalizable, our results are. To this end, we study the cross-section of corporate filings and investor identities. Table 7 presents the results. They are based on estimations of regressions that are identical to that in column (8) of Table 2 (i.e., the most saturated regression in terms of fixed effects), yet for different dependent variables.

Columns (1) to (4) of Panel A, respectively, show the results from regressions that separately explain the variable *Search Volume* for forms 3/4, 8-K, 10-K/Q, DEF 14. In all four columns, the coefficient on *ln(Social Connectedness)* is positive and statistically significant at the 1% level, indicating that social proximity fosters the acquisition of different types of information (i.e., insider transactions, material news, quarterly performance, and governance information). Relative to the average baseline estimates in Table 2, the coefficient on *ln(Social Connectedness)* does not considerably differ for 8-K, 10-K/Q, and DEF 14 filings but is about 40% larger for form 3/4-related search volume. Thus, while the link between information acquisition and social proximity is pervasive across all filing types, it is more pronounced for insider transactions.

In Panel B, we present results from regressions that explain the dependent variable *Search Volume* for different investor identities, as approximated by their IP addresses. The regressions in

[24] Note that this test *does not* suggest that constructs like behavioral biases or media coverage differences are not important to information demand. Rather, it suggests that variation in geographic distance that does not correlate with social proximity is less important to information demand.

columns (1) and (2) examine the search volume by financial firms (*Search Volume*$_{Financial\ Firm\ IPs}$) and non-financial IP addresses (*Search Volume*$_{Other\ IPs}$), identified using IP ownership records from ARIN WhoIs data. The dependent variables in columns (3) and (4) measure search volume from financial centers (*Search Volume*$_{Financial\ Center\ IPs}$) and non-financial centers (*Search Volume*$_{Non\text{-}Financial\ Center\ IPs}$), where financial center IP addresses are those from Boston, Chicago, Los Angeles, New York, Philadelphia, or San Francisco (as per Christoffersen and Sarkissian, 2009). Throughout columns (1) to (4), we find a positive coefficient on *ln(Social Connectedness)*, which is statistically significant at the 1% level in columns (1), (2) and (4), and at the 5% level in column (3). Hence, the link between information acquisition and social proximity is pervasive across many IP addresses, including those related to finance professionals. The magnitude of the coefficient on *ln(Social Connectedness)* is smaller for finance-related IP addresses, consistent with the notion that finance professionals are more aware of diverse investment opportunities, and also rely more on other information, such as conversations with firms' investor relations or top managers.

In Appendix Table A6, we examine how pervasive the social proximity effects are across time. Therein, we re-estimate the main empirical design for each year of our sample period, using the most stringent fixed effects design from Table 2. In all annual regressions, we find the coefficient on *ln(Social Connectedness)* to be positive and significant at the 1% level, with almost constant magnitude in the 2010s. Additionally, the coefficient estimates are almost identical to the estimates based on the full sample, which further supports the robustness of our results.

The above evidence supports the generalizability of our results beyond the international evidence presented in Table 5. Overall, this section provides robust evidence on the breadth of generalizability for the link between traditional investment research and social proximity.

5. Complementing Analyses

5.1 Cross-sectional and Time-series Variation in Uncertainty and Asymmetric Information

In the following section, we attempt to understand how the effect of social connection on information acquisition varies with overall market uncertainty and firm-level asymmetric information (*H2*). To the extent that social connections reduce informational frictions, we anticipate that the impact of social connections on information gathering will be especially pronounced under conditions of heightened market uncertainty and information asymmetry.

To test this hypothesis, we re-estimate the regression in column (8) of Table 2, interacting the variable *ln(Social Connectedness)* with different measures of market-wide uncertainty or the information environment of firms. Panel A of Table 8 presents the results from regressions in which we exploit time-series variation in market-wide uncertainty, namely the global financial crisis and peaks in expected market volatility measured by the VIX index.[25] In columns (1) and (2), respectively, we study the interaction of *ln(Social Connectedness)* with the dummy variables *Financial Crisis* and $VIX_{>Median}$. The coefficients on both interaction terms as well as that on *ln(Social Connectedness)* are positive and significant at the 1% level, indicating that social proximity is associated with more information acquisition by investors during times of greater market-wide uncertainty. Economically, such periods of market uncertainty tend to increase the effect of social proximity on information demand by about 10% (column 1) and 6.6% (column 2).

Panel B of Table 8 presents the results from regressions that exploit cross-sectional variation in the information environment of firms. In columns (1) to (3) of Panel B, we interact *ln(Social Connectedness)* with a dummy that equals one if a firm's analyst coverage, age, and size

[25] According to the CBOE website (https://www.cboe.com/tradable_products/vix/), the "VIX Index is a calculation designed to produce a measure of constant, 30-day expected volatility of the U.S. stock market, derived from real-time, mid-quote prices of S&P 500 Index (SPX) call and put options."

are above the sample median, respectively. In line with the literature (e.g., Frankel and Li, 2004; Harford et al., 2019), we assume that firms with greater analyst coverage as well as older and larger firms have a stronger information environment. Consistentwith this intuition, we find that the coefficients on all three interaction terms (i.e., *ln(Social Connectedness)* ×$Analysts_{>Median}$, etc.) are negative and significant at the 1% level, indicating that the link between investors' information acquisition and social proximity is weaker for firms with fewer informational frictions. In untabulated regressions, we find similar evidence for firms belonging to the S&P 500. We also find corroborating results using firm locations to measure a firm's information environment. Specifically, in column (4), we interact *ln(Social Connectedness)* with an indicator variable that equals one for firms headquartered in metropolitan areas (with a population of more than a million). Firms in rural areas tend to face greater challenges related to information flow. This is because such firms have less analyst coverage, are owned by fewer institutional investors, and trade less (Loughran and Schultz, 2005). Additionally, while site visits lower information asymmetry between firms and market participants (e.g., Cheng et al., 2016; Cheng et al., 2019), the cost of rural site visits is higher, and as such is less common. Consequently, firms in large metropolitan areas should have a stronger information environment. The results support our expectations. The coefficient on *ln(Social Connectedness)* ×*HQ in Metropolitan Area* is negative and significant at the 1% level.

Together, the results in Table 8 suggest that social connections are particularly important when there is greater asymmetric information, consistent with the notion that social proximity mitigates informational frictions.

5.2 Is Information Acquisition by Socially Proximate Investors Informed?

While the previous results indicate that social connections help overcome information frictions and foster information acquisition, they ignore the question of whether information acquisition by socially proximate investors contains information. To address this question, we create a panel at the firm-year quarter level to directly mimic the approach in Drake et al. (2020). We regress measures of future firm performance on the two variables *Socially Connected Search Volume* and *Socially Unconnected Search Volume*, along with firm controls as well as firm and year-quarter fixed effects.[26] *Socially Connected Search Volume* follows Drake et al. (2020) and is the decile ranking (1-10) of the number of clicks from socially connected U.S. counties. We define socially connected counties as those with above sample median *ln(Social Connectedness)*. To measure future firm performance, we use *Unexpected Earnings*, which is defined as the difference in diluted EPS before extraordinary items in year-quarter *t* and year-quarter *t-1*, scaled by stock price at the end of fiscal year-quarter *t*, as per Drake et al. (2020). We calculate this variable for quarter t+1 and t+2. We also use raw buy-and-hold returns for the next month (i.e., $Return_{Next\ Month}$), next quarter, and next two quarters. Since information demand effects are enhanced for firms with informational frictions (Section 5.1), socially connected search may also concentrate in higher-risk stocks. Therefore, we additionally examine how socially connected search relates to stock volatility and Sharpe ratios, i.e., stock returns divided by volatility.

Table 9 presents summary statistics for the firm-level data in Panel A and the results from regressions predicting unexpected earnings and stock returns in Panels B and C, respectively. Panel

[26] We use the same year-quarter-level firm controls as in our previous regressions but additionally include controls for annual analyst coverage and institutional ownership. We also control for *Social Proximity to Capital*, which measures the investment-weighted social connections between a firm's headquarter county k and a searcher county j, as per Kuchler et al. (2022). Analogously, we also control for *Physical Proximity to Capital*. These controls account for the effects that socially invested and geographically proximate capital may have on future firm outcomes.

D shows the results for stock volatility and Sharpe ratios. The results in Panel B show that the coefficient on *Socially Connected Search Volume* is positive and statistically significant when we predict unsigned as well as positive unexpected earnings for the next quarter and the next two quarters in columns (1) and (2) and columns (3) and (4), respectively. In all four columns, the coefficient on *Socially Connected Search Volume* is significant at the 5% level or better, and it is significantly different from the coefficient on *Socially Unconnected Search Volume*. In columns (5) and (6), in which we explain negative unexpected earnings, the coefficient on *Socially Connected Search Volume* is negative and significant, hence predicting negative earnings, but statistically indistinguishable from the coefficient on *Socially Unconnected Search Volume*.

Regarding future stock returns, in Panel C *Socially Connected Search Volume* has a positive (negative) sign and is significant at the 1% when used to explain positive (negative) stock returns for the next month, the next quarter, and the next two quarters in columns (1) to (3) (in columns (4) to (6)). Also, *Socially Connected Search Volume* is significantly different from *Socially Unconnected Search Volume* in all six columns. Taken together, the results in Panels B and C indicate that EDGAR search volume by socially proximate investors appears to contain information that predicts short-term firm performance. In contrast, information acquisition by socially less proximate investors does not predict firm performance. Lastly, the results in Panel D show that socially connected search is associated with greater stock volatility, and significantly more so than socially unconnected search, while it does not relate to Sharpe ratios.

The results indicate that, in aggregate, information acquisition by more socially proximate investors contains valuable information. Hence, in a classical Grossman and Stiglitz (1980) setting, while the average investor is compensated for incurring the costs of investment research, our results indicate that such compensation appears enhanced for socially connected individuals.

However, the average socially connected investor also faces greater stock volatility, i.e., enhanced risk, and does not outperform the average investor with fewer social connections when this risk is accounted for. This result is consistent with the investor-level evidence in Kuchler et al. (2022).

6. Conclusion

This study provides evidence that investors acquire more public information about firms to which they are more socially proximate. In particular, the Social Connectedness Index (Bailey et al., 2018) between a firm's headquarter county and another U.S. county is significantly positively associated with EDGAR downloads of company filings from that county. Studying firms' headquarter relocations as a source of plausibly exogenous variation in social proximity yields similar results. Such effects are distinct from physical distance. When we examine EDGAR downloads from investors in European counties, for which physical distance should be irrelevant, we find consistent social proximity effects on information acquisition.

We provide complementing evidence that indicate social connections particularly matter during times of high market-wide uncertainty and for firms with weaker information environments. Also, enhanced EDGAR search volume by more socially proximate investors predicts future short-term earnings and stock returns, consistent with the notion that those investors, on average, are compensated incrementally for their costly investment research. Collectively, the results presented in this study suggest that social networks mitigate informational frictions and foster information acquisition in financial markets. Our results also suggest that prior evidence on the role of physical distance may, in part, have captured social proximity. Thereby, our study contributes to the literature on how social networks shape capital market outcomes as well as the literature on the role that physical distance plays for capital market participants.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly and ChatGPT in order to ensure grammatical accuracy, improve sentence structure, refine sentence clarity, and enhance the overall coherence of the text. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

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

Amman, M., A. Cochardt, L. Cohen, and S. Heller, 2022, Hidden alpha, Working paper, Harvard Business School.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong, 2018, Social connectedness: Measurements, determinants, and effects, *Journal of Economic Perspectives* 32, 259-80.

Baskett, G.D., 1973, Interview decisions as determined by competency and attitude similarity, *Journal of Applied Psychology* 57, 343-345.

Bauguess, S.W., J.W. Cooney, Jr., K. Weiss Hanley, 2018, Investor demand for information in newly issued securities, Working paper, Lehigh University.

Baum-Snow, N., 2007, Did highways cause suburbanization?, *Quarterly Journal of Economics* 122, 775-805.

Ben-Rephael, A., Z. Da, P. Easton, and R.D. Israelsen, 2022, Who pays attention to SEC Form 8-K?, *The Accounting Review* 97, 59-88.

Bernile, G., A. Kumar, and J. Sulaeman, 2015, Home away from home: Geography of information and local investors, *Review of Financial Studies* 28, 2009-2049.

Bernile, G., A. Kumar, J. Sulaeman, and Q. Wang, 2019, Has local informational advantage disappeared?, *Review of Financial Economics* 37, 38-60.

Blankespoor, E., E. deHaan, I. Marinovic, 2020, Disclosure processing costs, investors' information choice, and equity market outcomes: A review, *Journal of Accounting and Economics* 70, 101344.

Brodbeck, F.C., and M. Frese, 2007, Societal culture and leadership in Germany, in: Chhokar, J.S., F.C. Brodbeck, and R.J. House (eds.), Culture and Leadership Across the World: The GLOBE Book of In-Depth Studies (GLOBE Book 2). Lawrence Erlbaum, Mahwah, New Jersey, 147-214.

Chen, H., L. Cohen, U. Gurun, D. Lou, and C. Malloy, 2020, IQ from IP: Simplifying search in portfolio choice, *Journal of Financial Economics* 138, 118-137.

Cheng, Q., F. Du, X. Wang, and B.Y. Wang, 2016, Seeing is believing: Analysts' corporate site visits, *Review of Accounting Studies* 21, 1245-1286.

Cheng, Q., F. Du, B.Y. Wang, and X. Wang, 2019, Do corporate site visits impact stock prices?, *Contemporary Accounting Research* 36, 359-388.

Chi, S.S., and D.M. Shanthikumar, 2017, Local bias in Google search and the market response around earnings announcements, *The Accounting Review* 92, 115-143.

Christoffersen, S.E.K., and S. Sarkissian, 2009, City size and fund performance, *Journal of Financial Economics* 92, 252-275.

Cohn, J.B., Z. Liu, and M.I. Wardlaw, 2022, Count (and count-like) data in finance, *Journal of Financial Economics* 146, 529-551.

Cooper, I., P. Sercu, and R Vanpée, 2013, The equity home bias puzzle: A survey, *Foundations and Trends® in Finance* 7, 289-416.

Coval, J.D., and T.J. Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 2045-2073.

Drake, M.S., D.T. Roulstone, and J.R. Thornock, 2015, The determinants and consequences of information acquisition via EDGAR, *Contemporary Accounting Research* 32, 1128-1161.

Drake, M.S., B.A. Johnson, D.T. Roulstone, and J.R. Thornock, 2020. Is there information content in information acquisition?, *The Accounting Review* 95, 113-139.

Duggan, M., S. Greenwood, and A. Perrin, 2016, Social media update 2016, Pew Research Center.

Dyer, T., 2021, The demand for public information by local and nonlocal investors: Evidence from investor-level data, *Journal of Accounting and Economics* 72, 101417.

Frankel, E., and X. Li, 2004, Characteristics of a firm's information environment and the information asymmetry between insiders and outsiders, *Journal of Accounting and Economics* 37, 229-259.

Gibbons, B., P. Iliev, and J. Kalodimos, 2021, Analyst information acquisition via EDGAR, *Management Science* 67, 769-793.

Granovetter, M., 2005, The impact of social structure on economic outcomes, *Journal of Economic Perspectives* 19, 33-50.

Grossman, S., and J. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Han, B., and L. Yang, 2013, Social networks, information acquisition, and asset prices, *Management Science* 59, 1444-1457.

Harford, J., F. Jiang, R. Wang, and F. Xie, 2019, Analyst career concerns, effort allocation, and firms' information environment, *Review of Financial Studies* 32, 2179-2224.

Hasan, I., C.-K. Hoi, Q. Wu, and H. Zhang, 2017, Does social capital matter in corporate decisions? Evidence from corporate tax avoidance, *Journal of Accounting Research* 55, 629-668.

Hewstone, M., M. Rubin, and H. Willis, 2002, Intergroup bias, *Annual Review of Psychology* 53, 575-604.

Hirshleifer, D., L. Peng, and Q. Wang, 2023, News diffusion in social networks and stock market reactions, Working paper, University of Southern California.

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

Hong, H., J.D. Kubik, and J.C. Stein, 2004, Social interaction and stock-market participation, *Journal of Finance* 59, 137-163.

Hong, H., J.D. Kubik, and J.C. Stein, 2005, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *Journal of Finance* 60, 2801-2824.

Iliev, P., J. Kalomidis, and M. Lowry, 2021, Investors' attention to corporate governance, *Review of Financial Studies* 34, 5581-5628.

Jiang, W., 2017, Have instrumental variables brought us closer to the truth, *Review of Corporate Finance Studies* 6, 127-140.

Kuchler, T., Y. Li, L. Peng, J. Stroebel, and D. Zhou, 2022, Social proximity to capital: Implications for investors and firms, *Review of Financial Studies* 35, 2743-2789.

Loughran, T., and P. Schultz, 2005, Liquidity: Urban versus rural firms, *Journal of Financial Economics* 78, 341-374.

Merton, R.C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483-510.

Pool, V.K., N. Stoffman, and S.E. Yonker, 2015, The people in your neighborhood: Social interactions and mutual fund portfolios, *Journal of Finance* 70, 2679-2732.

Ryans, J., 2017, Using the EDGAR log file data set, White paper, London Business School.

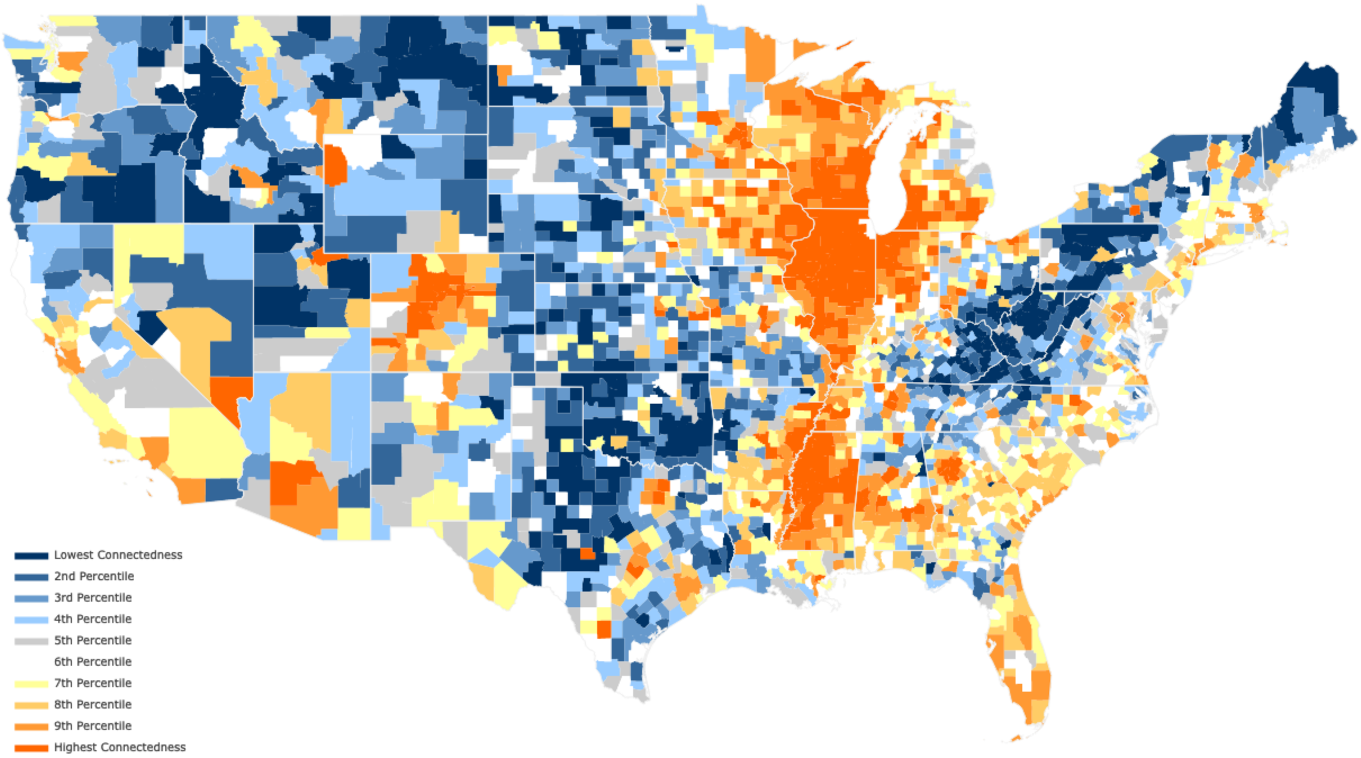
Santos Silva, J.M.C., and S. Tenreyro, 2006, The log of gravity, *Review of Economics and Statistics* 88, 641-658.

Strauss-Kahn, V., and X. Vives, 2009, Why and where do headquarters move?, *Regional Science and Urban Economics* 39, 168-186.

Van Nieuwerburgh, S., and L. Veldkamp, 2009, Information immobility and the home bias puzzle. *Journal of Finance* 64, 1187-1215.

Figures

a) Social Connectedness



b) Distance

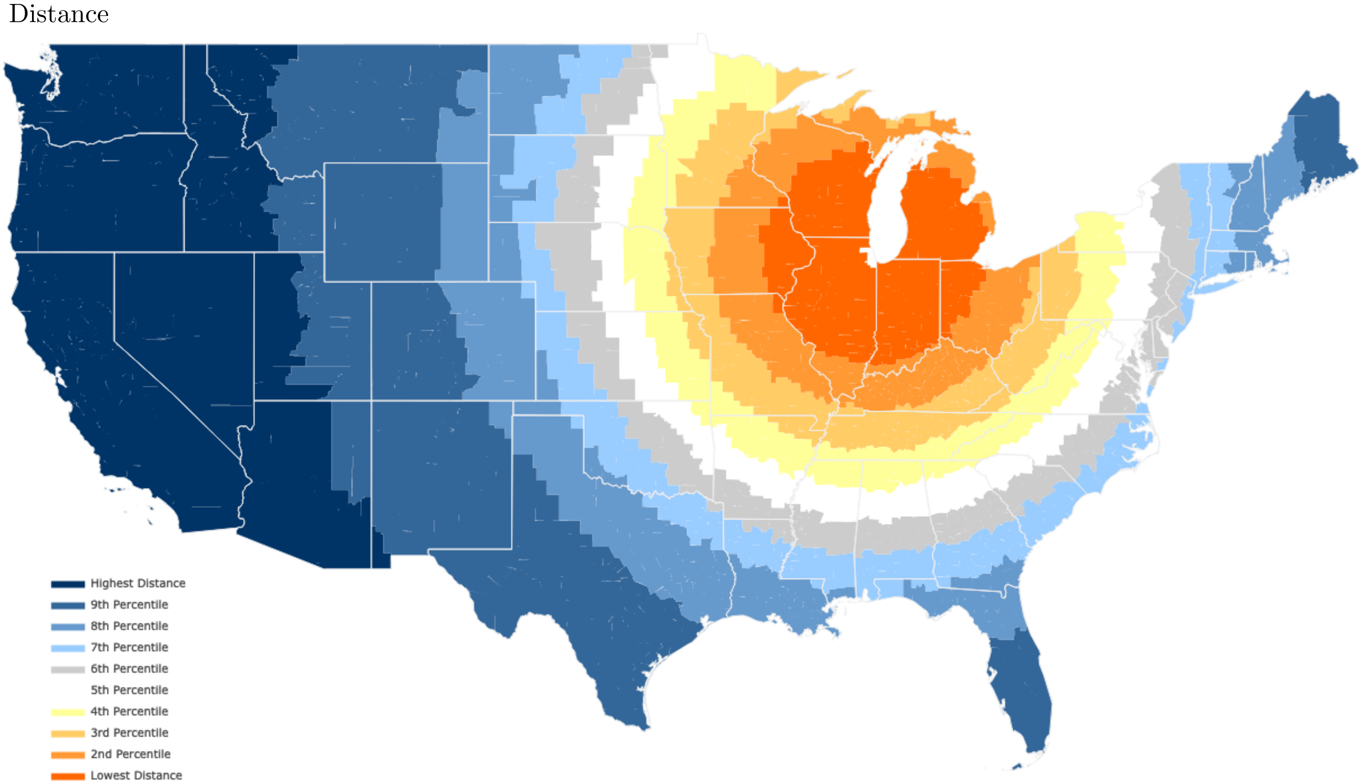


Figure 1: The relation between social and physical proximity. This figure illustrates the correlations between the two variables *Social Connectedness* and *Distance* by displaying the connectedness and distances between Cook County in Illinois and all other counties in the United States.

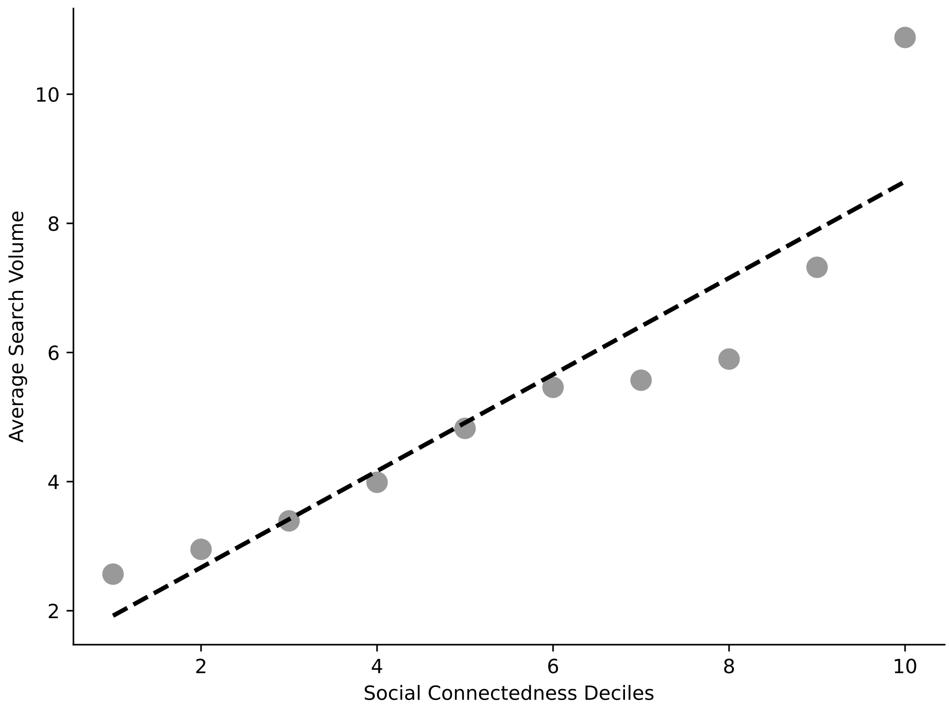


Figure 2: Univariate Sorts. The x-axis shows deciles of *Social Connectedness* and the y-axis shows the average value of *Search Volume* for each decile. The dashed line shows a linear regression model fit.

Tables

Table 1: Summary Statistics

Panel A: Descriptive Statistics

	Mean	P25	P50	P75	Std.
Dependent Variables:					
Search Volume	5.29	0.00	0.00	3.00	15.33
Search Volume$_{\text{Financial Firm IPs}}$	0.06	0.00	0.00	0.00	0.45
Search Volume$_{\text{Financial Center IPs}}$	0.63	0.00	0.00	0.00	4.30
Main Independent Variable:					
Social Connectedness	6032.14	1300.00	2128.00	3934.00	17795.20
ln(Social Connectedness)	7.83	7.17	7.66	8.28	1.01
Controls:					
Distance	1685.68	758.46	1433.03	2484.79	1169.59
ln(Distance)	7.08	6.63	7.27	7.82	1.01
AUM (in millions)	14.40	0.00	0.00	0.00	112.25
ln(AUM)	3.18	0.00	0.00	0.00	6.34
Assets (in thousands)	29.53	0.955	3.443	13.10	105.07
ln(Assets)	8.26	6.86	8.14	9.48	1.93
Leverage	0.25	0.08	0.22	0.36	0.20
ROA	0.01	0.00	0.01	0.02	0.03
MarketBook	3.13	1.29	2.13	3.66	4.78
Age	29.94	15.00	24.00	46.00	18.69
ln(Age)	3.21	2.77	3.22	3.85	0.71
Unemployment	6.68	4.90	6.30	8.20	2.35
Population (in thousands)	694.90	217.29	468.68	844.32	802.86
ln(Population)	12.94	12.29	13.06	13.65	1.09
Income (in thousands)	47.99	38.95	44.57	53.55	13.78
Moderators:					
Analysts	14.07	6.00	12.00	21.00	10.03
HQ in Metropolitan Area	0.81	1.00	1.00	1.00	0.39
Financial Crisis	0.06	0.00	0.00	0.00	0.23
VIX	19.86	14.28	17.03	22.03	8.31

Panel B: Univariate Tests

	Social Connectedness					
	≤ Median		> Median			
Variable:	Mean	Median	Mean	Median	Diff(Mean)	Diff(Median)
ln(Search Volume+1)	0.6919	0.0000	0.9934	0.6931	0.3015***	0.6931***
Search Volume	3.5300	0.0000	7.0440	1.0000	3.5140***	1.0000***

	Social Connectedness$_{\text{Residuals}}$					
	≤ Median		> Median			
Variable:	Mean	Median	Mean	Median	Diff(Mean)	Diff(Median)
ln(Search Volume+1)	0.6889	0.0000	0.9963	0.6931	0.3073***	0.6931***
Search Volume	3.5645	0.0000	7.0077	1.0000	3.4432***	1.0000***

Notes: This table shows summary statistics based on 14,188,331 county-firm-quarter observations. Appendix Table A1 provides variable definitions. *Social Connectedness*$_{\text{Residuals}}$ are the residuals from an OLS regression of *ln(Social Connectedness)* on *ln(Distance)*.

Table 2: Social Connectedness and Information Acquisition: Baseline Results

Panel A: Baseline Results
Dependent Variable: Search Volume

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(Social Connectedness)		**0.3636*****		**0.2493*****	**0.2589*****	**0.2604*****	**0.2610*****	**0.2625*****
		(24.60)		**(17.35)**	**(12.66)**	**(12.64)**	**(12.67)**	**(12.71)**
ln(Distance)	-0.2208***		-0.1815***		0.0100	0.0101	0.0115	0.0114
	(-9.150)		(-12.49)		(0.8340)	(0.8338)	(0.9428)	(0.9281)
ln(AUM)			0.0061***	0.0055***	0.0055***	0.0051***	0.0070***	0.0065***
			(4.428)	(4.040)	(4.040)	(3.861)	(4.721)	(4.544)
ln(Assets)			0.2035***	0.2033***	0.2033***	0.2038***		
			(17.13)	(17.16)	(17.17)	(17.29)		
Leverage			0.3230***	0.3247***	0.3248***	0.3179***		
			(9.330)	(9.436)	(9.441)	(9.120)		
ROA			-1.157***	-1.146***	-1.146***	-1.153***		
			(-11.87)	(-11.71)	(-11.70)	(-11.75)		
MarketBook			0.0009*	0.0009*	0.0009*	0.0008*		
			(1.850)	(1.773)	(1.768)	(1.662)		
ln(Age)			-0.2266***	-0.2284***	-0.2285***	-0.2328***		
			(-6.024)	(-6.055)	(-6.055)	(-6.302)		
Unemployment			-0.0008	-0.0011	-0.0011		-0.0099	
			(-0.0435)	(-0.0566)	(-0.0572)		(-0.4285)	
ln(Population)			0.1846	0.1771	0.1771		0.1844	
			(0.1360)	(0.1301)	(0.1301)		(0.1351)	
Income			0.0027	0.0027	0.0027		0.0027	
			(0.4476)	(0.4470)	(0.4473)		(0.4293)	
Firm FE			✓	✓	✓	✓		
County FE			✓	✓	✓		✓	
Year-Qtr FE			✓	✓	✓			
Same County FE			✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE							✓	✓
County×Year-Qtr FE						✓		✓
Observations	14,188,331	14,188,331	14,188,331	14,188,331	14,188,331	14,172,693	14,161,448	14,145,833
Pseudo R^2	0.01265	0.03099	0.42183	0.42836	0.42852	0.44675	0.44767	0.46606

Panel B: Additional Controls for County-Pair Similarities
Dependent Variable: Search Volume

	(3)	(4)	(5)	(6)	(7)	(8)
ln(Social Connectedness)		**0.2549*****	**0.2686*****	**0.2737*****	**0.2698*****	**0.2752*****
		(15.87)	**(12.81)**	**(13.07)**	**(12.81)**	**(13.09)**
ln(Distance)	-0.1786***		0.0153	0.0180	0.0162	0.0190
	(-12.60)		(1.326)	(1.541)	(1.389)	(1.606)
ln(AUM)	0.0058***	0.0052***	0.0052***	0.0051***	0.0065***	0.0063***
	(4.416)	(4.144)	(4.147)	(4.081)	(4.875)	(4.818)
% Firms Same SIC3	2.634***	2.413***	2.406***	2.422***	2.407***	2.426***
	(4.526)	(4.261)	(4.244)	(4.248)	(4.250)	(4.247)
Controls as in Panel A	✓	✓	✓	✓	✓	✓
County-Pair Age Similarity	✓	✓	✓	✓	✓	✓
County-Pair Education Similarity	✓	✓	✓	✓	✓	✓
County-Pair Employment Similarity	✓	✓	✓	✓	✓	✓
County-Pair Family Similarity	✓	✓	✓	✓	✓	✓
County-Pair Income Similarity	✓	✓	✓	✓	✓	✓
County-Pair Mobility Similarity	✓	✓	✓	✓	✓	✓
County-Pair Racial Similarity	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓		
County FE	✓	✓	✓		✓	
Year-Qtr FE	✓	✓	✓			
Same County FE	✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE					✓	✓
County×Year-Qtr FE				✓		✓
Observations	12,644,761	12,644,761	12,644,761	12,630,090	12,622,988	12,608,353
Pseudo R^2	0.45089	0.45662	0.45681	0.47229	0.47487	0.49107

Notes: Panels A and B of this table show results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *ln(Social Connectedness)* along with county- and firm-level controls and varying sets of fixed effects. The sample covers U.S. firms and counties. The sample period for the regressions in Panel A is Q1 2007 to Q2 2017. The regressions shown in Panel B include additional control variables for county-pair similarities, which are not shown for brevity (see Appendix Table A2 for detailed estimates). The sample period for the regressions in Panel B is Q1 2009 to Q2 2017 due to lack of data availability. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 3: Difference-in-Differences Regressions Based on Headquarter Relocations

Dependent Variable: Search Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post × Δ Social Connectedness		**0.0047****	**0.0043****	**0.0052****	**0.0037***	**0.0048****	**0.0064****
		(2.329)	**(2.119)**	**(2.628)**	**(1.793)**	**(2.370)**	**(2.618)**
Post × Δ Distance	-0.0299**		-0.0117	-0.0115	-0.0202*	-0.0199	-0.0297*
	(-2.158)		(-0.8810)	(-0.8392)	(-1.774)	(-1.645)	(-1.890)
Post	-0.0724*	-0.0751*	-0.0755*	-0.0816*			
	(-1.664)	(-1.718)	(-1.730)	(-1.869)			
Δ Social Connectedness		0.0004	0.0011	0.0002	0.0014	0.0005	-0.0011
		(0.2015)	(0.5064)	(0.1190)	(0.6766)	(0.2324)	(-0.4456)
Δ Distance	0.0196		0.0242**	0.0227*	0.0285**	0.0270**	0.0370**
	(1.595)		(2.051)	(1.893)	(2.598)	(2.346)	(2.593)
ln(AUM)	0.0010	0.0012	0.0013	0.0014	0.0019	0.0020	0.0020
	(0.6019)	(0.7265)	(0.7832)	(0.6566)	(1.115)	(0.8986)	(0.8959)
ln(Assets)	0.2841***	0.2799***	0.2811***	0.2854***			
	(9.437)	(9.373)	(9.371)	(9.530)			
Leverage	-0.1550	-0.1357	-0.1394	-0.1138			
	(-0.8154)	(-0.7177)	(-0.7365)	(-0.6058)			
ROA	-1.643***	-1.614***	-1.616***	-1.597***			
	(-3.710)	(-3.572)	(-3.580)	(-3.479)			
MarketBook	0.0024	0.0023	0.0023	0.0020			
	(1.299)	(1.294)	(1.291)	(1.085)			
ln(Age)	-0.4367***	-0.4493***	-0.4484***	-0.4442***			
	(-2.919)	(-3.020)	(-3.012)	(-3.017)			
Unemployment	-0.0098	-0.0117	-0.0114		-0.0069		
	(-0.4098)	(-0.4883)	(-0.4742)		(-0.2702)		
ln(Population)	0.1868	0.1028	0.0994		0.1280		
	(0.1250)	(0.0682)	(0.0660)		(0.0847)		
Income	0.3314	-0.0892	-0.0709		0.3304		
	(0.0508)	(-0.0136)	(-0.0108)		(0.0501)		
Pre_1 × Δ Social Connectedness							0.0028
							(1.461)
Pre_2 × Δ Social Connectedness							0.0017
							(1.034)
Pre_1 × Δ Distance							-0.0162
							(-1.372)
Pre_2 × Δ Distance							-0.0151
							(-1.233)
Firm FE	✓	✓	✓	✓			
County FE	✓	✓	✓		✓		
Year-Qtr FE	✓	✓	✓				
Same County FE	✓	✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE					✓	✓	✓
County×Year-Qtr FE				✓		✓	✓
Observations	405,516	405,516	405,516	389,058	404,751	388,329	388,329
Pseudo R^2	0.46733	0.46859	0.46878	0.50524	0.49612	0.53570	0.53563

Notes: This table shows results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *Post* × Δ *Social Connectedness* along with its base terms as well as county- and firm-level controls and varying sets of fixed effects for the sample period Q1 2007 to Q2 2017. The sample contains U.S. firms that relocate their corporate headquarters for the four years surrounding the relocation year (i.e., -/+16 quarters), excluding the year of relocation. Δ *Social Connectedness* and Δ *Distance* are the treatment variables capturing the extent to which relocation-induced changes affected social and geographic proximity. The indicator variable *Post* equals one for all year-quarters after a firm relocated its headquarters. To test for parallel trends, column (7) includes interactions with the indicator variables Pre_1 and Pre_2, which equal one for the quarters -4 to -1 and -8 to -5 before the relocation year, respectively. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 4: Social Connectedness and Information Acquisition: Evidence from Investor IP-Level Data

Dependent Variable: Search Volume					
	(1)	(2)	(3)	(4)	(5)
ln(Social Connectedness)		**0.1384***** **(6.968)**	**0.0794***** **(4.727)**	**0.0771**** **(2.436)**	**0.0730**** **(2.440)**
ln(Distance)	-0.1328*** (-7.266)		-0.0794*** (-3.803)	-0.0282 (-1.309)	-0.0328 (-1.586)
ln(AUM)				0.0659*** (3.701)	0.0641*** (3.772)
Specialization					0.3538*** (2.803)
Investor IP×Year-Qtr FE	✓	✓	✓	✓	✓
Firm×Year-Qtr FE	✓	✓	✓	✓	✓
Same ZIP FE	✓	✓	✓	✓	✓
Observations	4,108,417	4,108,417	4,108,417	1,278,803	1,278,803
Pseudo R^2.	0.37581	0.37512	0.38042	0.36519	0.37008

Notes: This table shows results from PPML (Poisson Pseudo Maximum Likelihood) regressions for a sample of institutional investors from 2007Q1 to 2015Q1 based on Dyer (2021). The indicator variable *Specialization* equals one if a firm is in the industry most researched by the investor over the sample period. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by investor IP and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 5: European County (NUTS3 classification) Setting

Panel A: Europe
Dependent Variable: Search Volume from European County i for Disclosures by Firm j in Year-Quarter t

	(1)	(2)	(3)	(4)	(5)	(6)
ln(Social Connectedness)		**0.1665***** **(4.963)**	**0.1672***** **(4.940)**	**0.1443***** **(4.447)**	**0.1741***** **(4.955)**	**0.1554***** **(5.105)**
ln(Distance)	-0.0934 (-0.4535)		0.0870 (0.3897)	0.0072 (0.0333)	0.1764 (0.4345)	0.0961 (0.2423)
Controls as before	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓		
European County FE	✓	✓	✓		✓	
Year-Qtr FE	✓	✓	✓			
Firm×Year-Qtr FE					✓	✓
European County×Year-Qtr FE				✓		✓
Observations	2,992,027	2,992,027	2,992,027	2,979,344	2,980,912	2,968,261
Pseudo R^2	0.34062	0.34089	0.34087	0.41831	0.38393	0.47930

Panel B: Germany
Dependent Variable: Search Volume from German County i for Disclosures by Firm j in Year-Quarter t

	(2)	(3)	(4)	(5)	(6)
ln(Social Connectedness)		**0.2455***** **(3.702)**	**0.2437***** **(3.681)**	**0.2460***** **(3.697)**	**0.2023***** **(2.804)**
ln(Distance)	-2.438 (-0.8479)		-1.996 (-0.6864)	-2.587 (-0.9319)	-2.284 (-0.6826)
ln(Social Connectedness)×East Germany				**-0.3122*** **(-1.946)**	
ln(Social Connectedness)×Distance US Base$_{<25km}$					**0.1219***** **(3.395)**
Firm×Year-Qtr FE	✓	✓	✓	✓	✓
German County×Year-Qtr FE	✓	✓	✓	✓	✓
Observations	470,790	470,790	470,790	470,790	470,790
Pseudo R^2	0.48686	0.48828	0.48838	0.48832	0.48844

Notes: Panels A and B of this table show results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *ln(Social Connectedness)* along with region- and firm-level controls (as in Panel A of Table 2) and varying sets of fixed effects for the sample period Q1 2007 to Q2 2017. In Panel A (Panel B), the sample covers U.S. firms and European (German) NUTS-3 regions. *East Germany* is an indicator variable that equals one if a German NUTS-3 region is located in East Germany (i.e., the former Soviet controlled part of Germany). *Distance US Base*$_{<25km}$ is an indicator variable that equals one if a NUTS-3 region is located within a 25km radius from a U.S. military base (e.g., Ramstein). Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 6: Social Connectedness, Geographic Proximity, and Information Acquisition

Dependent Variable: Search Volume					
	(1)	(2)	(3)	(4)	(5)
ln(Distance)$_{\textbf{Fitted Values}}$	**-0.4006*****	**-0.4026*****	**-0.4019*****	**-0.4041*****	**-0.3535*****
	(-14.26)	**(-14.19)**	**(-14.29)**	**(-14.24)**	**(-16.53)**
ln(Distance)$_{\textbf{Residuals}}$	**0.0010**	**0.0013**	**0.0025**	**0.0026**	**0.0114**
	(0.0827)	**(0.1065)**	**(0.2039)**	**(0.2138)**	**(0.9281)**
Controls as before	✓	✓	✓	✓	✓
Firm FE	✓	✓			
County FE	✓		✓		
Year-Qtr FE	✓				
Same County FE					✓
Firm×Year-Qtr FE			✓	✓	✓
County×Year-Qtr FE		✓		✓	✓
Observations	14,188,331	14,172,693	14,161,448	14,145,833	14,145,833
Pseudo R^2	0.44653	0.46519	0.46646	0.48490	0.46606

Notes: This table shows results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *ln(Distance)$_{Fitted\ Values}$* and *ln(Distance)$_{Residuals}$* along with county- and firm-level controls and varying sets of fixed effects (as in Panel A of Table 2). The sample covers U.S. firms and counties between Q1 2007 to Q2 2017. *ln(Distance)$_{Fitted\ Values}$* and *ln(Distance)$_{Residuals}$* are the fitted values and residuals, respectively, from an untabulated OLS regression of *ln(Distance)* on *ln(Social Connectedness)*. The adjusted R^2 of this regression is 51.9%. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 7: Social Connectedness and Information Acquisition: Form Types and Searcher Identity

Panel A: Form Types
Dependent Variable: Search Volume for Form Type

	3/4	8K	10K/Q	DEF 14
ln(Social Connectedness)	**0.3708*****	**0.2931*****	**0.2226*****	**0.2598*****
	(12.92)	**(12.20)**	**(12.42)**	**(11.08)**
ln(Distance)	0.0474**	0.0213	-0.0031	-0.0013
	(2.482)	(1.514)	(-0.2873)	(-0.0914)
ln(AUM)	0.0055**	0.0067***	0.0066***	0.0060***
	(2.464)	(5.375)	(4.515)	(4.181)
Same County FE	✓	✓	✓	✓
Firm×Year-Qtr FE	✓	✓	✓	✓
County×Year-Qtr FE	✓	✓	✓	✓
Observations	13,286,340	14,011,000	14,011,428	13,648,133
Pseudo R^2	0.16944	0.37959	0.43570	0.35538

Panel B: Searcher Identity
Dependent Variable: Search Volume from Different IP Clusters for Disclosures by Firm j in Year-Quarter t

	Search $\text{Volume}_{\text{Financial Firm IPs}}$	Search $\text{Volume}_{\text{Other IPs}}$	Search $\text{Volume}_{\text{Financial Center IPs}}$	Search $\text{Volume}_{\text{Non-Financial Center IPs}}$
ln(Social Connectedness)	**0.1082*****	**0.2641*****	**0.0828****	**0.2712*****
	(2.857)	**(12.68)**	**(2.516)**	**(12.97)**
ln(Distance)	-0.0424	0.0122	-0.0288**	0.0075
	(-0.9527)	(0.9860)	(-2.277)	(0.5288)
ln(AUM)	0.0066**	0.0065***	0.0145***	0.0052***
	(2.041)	(4.551)	(4.132)	(3.752)
Same County FE	✓	✓	✓	✓
Firm×Year-Qtr FE	✓	✓	✓	✓
County×Year-Qtr FE	✓	✓	✓	✓
Observations	3,028,620	14,145,830	736,276	14,145,793
Pseudo R^2	0.42942	0.46308	0.78323	0.39439

Notes: Panels A and B of this table show results from PPML (Poisson Pseudo Maximum Likelihood) regressions of variations of the dependent variable *Search Volume*, i.e., different form types and different IP clusters, on *ln(Social Connectedness)* along with county- and firm-level controls and a set of fixed effects (as in Panel A of Table 2). The sample covers U.S. firms and counties between Q1 2007 to Q2 2017. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 8: Social Connectedness and Information Acquisition: Market-wide Uncertainty and Firms' Information Environment

Panel A: Market-wide Uncertainty
Dependent Variable: Search Volume

	(1)	(2)
ln(Social Connectedness)	**0.2611***** **(12.72)**	**0.2545***** **(12.85)**
ln(Distance)	0.0114 (0.9290)	0.0113 (0.9234)
ln(AUM)	0.0065*** (4.545)	0.0065*** (4.546)
ln(Social Connectedness) × Financial Crisis	**0.0267***** **(2.982)**	
ln(Social Connectedness) × $\mathbf{VIX}_{>\mathbf{Median}}$		**0.0168***** **(2.620)**
Same County FE	✓	✓
Firm×Year-Qtr FE	✓	✓
County×Year-Qtr FE	✓	✓
Observations	14,145,833	14,145,833
Pseudo R^2	0.46626	0.46640

Panel B: Firms' Information Environment
Dependent Variable: Search Volume

	(1)	(2)	(3)	(4)
ln(Social Connectedness)	**0.3486***** **(14.44)**	**0.3112***** **(13.00)**	**0.3441***** **(14.17)**	**0.3516***** **(13.13)**
ln(Distance)	0.0064 (0.5330)	0.0051 (0.4095)	0.0052 (0.4112)	(-0.7865)
ln(AUM)	0.0076*** (5.029)	0.0065*** (4.815)	0.0066*** (5.100)	0.0065*** (4.846)
ln(Social Connectedness) × $\mathbf{Analysts}_{>\mathbf{Median}}$	**-0.1002***** **(-7.956)**			
ln(Social Connectedness) × Firm $\mathbf{Age}_{>\mathbf{Median}}$		**-0.0311***** **(-3.090)**		
ln(Social Connectedness) × Firm $\mathbf{Size}_{>\mathbf{Median}}$			**-0.0845***** **(-7.174)**	
ln(Social Connectedness) × HQ in Metropolitan Area				**-0.0846***** **(-4.296)**
Same County FE	✓	✓	✓	✓
Firm×Year-Qtr FE	✓	✓	✓	✓
County×Year-Qtr FE	✓	✓	✓	✓
Observations	11,210,177	14,145,833	14,145,833	14,145,833
Pseudo R^2	0.49951	0.48705	0.49547	0.48680

Notes: Panels A and B of this table show results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *ln(Social Connectedness)* along with county- and firm-level controls and a set of fixed effects. The sample covers U.S. firms and counties between Q1 2007 to Q2 2017. The regressions shown in Panel A includes variables to exploit market-wide time-series variation, namely *Financial Crisis* and $VIX_{>Median}$. Panel B includes variables to exploit firm-level cross-sectional variation, namely $Analyst_{>Median}$, $Firm\ Age_{>Median}$, $Firm\ Size_{>Median}$, and *HQ in Metropolitan Area.* Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table 9: Social Connectedness and Information Acquisition: Future Company Performance

Panel A: Summary Statistics

	Mean	P25	P50	P75	Std.
Dependent Variables:					
Unexpected Earnings	0.00	-0.004	0.00	0.01	0.06
Pos. Unexpected Earnings	0.02	0.00	0.01	0.01	0.06
Neg. Unexpected Earnings	-0.02	-0.01	-0.01	-0.002	0.04
Pos. Return $_{\text{Next Month}}$	0.08	0.02	0.05	0.11	0.09
Pos. Return $_{\text{Next Quarter}}$	0.15	0.04	0.10	0.19	0.17
Pos. Return $_{\text{Next Two Quarters}}$	0.24	0.06	0.14	0.28	0.31
Neg. Return $_{\text{Next Month}}$	-0.06	-0.09	-0.05	-0.02	0.05
Neg. Return $_{\text{Next Quarter}}$	-0.10	-0.15	-0.08	-0.04	0.08
Neg. Return $_{\text{Next Two Quarters}}$	-0.14	-0.20	-0.12	-0.06	0.10
Vola $_{\text{Next Month}}$	0.09	0.05	0.07	0.11	0.06
Vola $_{\text{Next Quarter}}$	0.16	0.09	0.13	0.19	0.10
Vola $_{\text{Next Two Quarters}}$	0.22	0.13	0.19	0.27	0.14
Sharpe Ratio $_{\text{Next Month}}$	0.01	-0.71	-0.03	0.71	1.04
Sharpe Ratio $_{\text{Next Quarter}}$	0.06	-0.67	-0.01	0.73	1.01
Sharpe Ratio $_{\text{Next Two Quarters}}$	0.09	-0.64	-0.02	0.73	1.03
Main Independent Variable:					
Socially Connected Search Volume	5.65	3.00	6.00	8.00	2.89
Socially Unconnected Search Volume	5.73	3.00	6.00	8.00	2.91
Controls:					
Analysts	11.37	5.00	9.00	17.00	8.58
ln(Analysts)	2.21	1.79	2.30	2.89	0.88
Institutional Ownership	0.76	0.66	0.81	0.92	0.20
Assets (in thousands)	14.567	0.771	2.569	8.818	38.947
ln(Social Proximity to Capital)	23.11	22.36	22.95	23.66	1.12
ln(Physical Proximity to Capital)	10.96	9.85	10.70	11.68	1.49
ln(Assets)	7.94	6.65	7.85	9.08	1.79
Leverage	0.23	0.06	0.20	0.34	0.20
ROA	0.01	0.00	0.01	0.02	0.03
MarketBook	2.98	1.33	2.12	3.56	4.17
Age	28.94	15.00	24.00	43.00	17.96
ln(Age)	3.19	2.77	3.22	3.78	0.68

Panel B: Unexpected Earnings

Dependent Variable:	Unexpected Earnings		Pos. Unexpected Earnings		Neg. Unexpected Earnings	
	Quarter $t+1$	Quarter $t+2$	Quarter $t+1$	Quarter $t+2$	Quarter $t+1$	Quarter $t+2$
	(1)	(2)	(3)	(4)	(5)	(6)
Socially Connected Search Volume	**0.0006*****	**0.0007*****	**0.0022*****	**0.0015*****	**-0.0016*****	**-0.0008***
	(2.696)	**(3.108)**	**(6.332)**	**(3.795)**	**(-3.838)**	**(-1.912)**
Socially Unconnected Search Volume	**-0.0001**	**0.0002**	**0.0004***	**0.0004**	**-0.0010*****	**-0.0004**
	(-0.5893)	**(0.8504)**	**(1.743)**	**(1.159)**	**(-3.127)**	**(-1.144)**
ln(Analysts)	-0.0004	0.0002	0.0000	0.0011	0.0003	0.0006
	(-0.3705)	(0.3665)	(0.0492)	(0.4100)	(0.1110)	(0.2275)
Inst. Ownership	0.0043	0.0030	-0.0089	-0.0048	0.0155*	0.0134*
	(1.199)	(1.075)	(-1.307)	(-0.6231)	(1.954)	(1.721)
Social Proximity to Capital	-0.0010	0.0017	0.0019	0.0055	-0.0049	-0.0043
	(-0.3483)	(0.7267)	(0.2240)	(0.5911)	(-0.5763)	(-0.4684)
Physical Proximity to Capital	0.0038	0.0001	0.0063	0.0019	-0.0006	-0.0032
	(1.525)	(0.0828)	(1.460)	(0.4053)	(-0.1226)	(-0.6355)
ln(Assets)	-0.0091***	-0.0034***	-0.0004	0.0117***	-0.0106***	-0.0134***
	(-6.564)	(-4.281)	(-0.2167)	(4.641)	(-4.306)	(-5.506)
Leverage	-0.0017	0.0044	0.0209***	0.0214***	-0.0233***	-0.0236***
	(-0.3025)	(1.136)	(2.989)	(2.637)	(-3.208)	(-3.454)
ROA	-1.106***	-0.1329***	-1.200***	-0.3046***	-0.0255	0.2532***
	(-26.08)	(-5.462)	(-27.99)	(-9.632)	(-0.6067)	(7.565)
MarketBook	0.0008***	0.0000	0.0002	-0.0006***	0.0010***	0.0006***
	(6.766)	(-1.103)	(1.340)	(-4.008)	(6.254)	(4.629)
ln(Age)	0.0093**	0.0000	0.0105**	-0.0030	0.0026	0.0004
	(2.555)	(-0.0210)	(2.046)	(-0.4311)	(0.4045)	(0.0567)
p-value (Con.=Uncon.)	0.020	0.086	0.000	0.039	0.211	0.307
Firm FE	✓	✓	✓	✓	✓	✓
Year-Qtr FE	✓	✓	✓	✓	✓	✓
Observations	48,815	48,787	25,105	25,236	22,156	22,015
Adjusted R^2	0.18842	0.00336	0.36105	0.03604	0.01719	0.03032

Table is continued on the next page...

Table is continued from the last page...

Panel C: Returns

Dependent Variable:	Pos. Returns			Neg. Returns		
	Next Month	Next Quarter	Next Two Quarters	Next Month	Next Quarter	Next Two Quarters
	(1)	(2)	(3)	(4)	(5)	(6)
Socially Connected Search Volume	**0.0021******	**0.0035******	**0.0063******	**-0.0013******	**-0.0028******	**-0.0041******
	(4.543)	**(4.291)**	**(4.820)**	**(-3.641)**	**(-4.897)**	**(-4.697)**
Socially Unconnected Search Volume	**0.0007****	**0.0008**	**0.0012**	**-0.0003**	**0.0000**	**-0.0012****
	(1.779)	**(1.253)**	**(1.231)**	**(-1.110)**	**(0.0814)**	**(-1.787)**
ln(Analysts)	-0.0062***	-0.0131***	-0.0169***	-0.0030**	-0.0007	-0.0034
	(-3.609)	(-4.051)	(-3.148)	(-2.321)	(-0.2983)	(-0.9797)
Inst. Ownership	-0.0230***	-0.0533***	-0.1075***	0.0019	-0.0096	-0.0126
	(-3.060)	(-4.099)	(-5.011)	(0.4047)	(-1.106)	(-0.9095)
Social Proximity to Capital	0.0106*	0.0048	0.0252	-0.0076	-0.0053	-0.0063
	(1.818)	(0.5620)	(1.641)	(-1.580)	(-0.6782)	(-0.5105)
Physical Proximity to Capital	-0.0052	-0.0027	-0.0123	0.0042	0.0029	0.0018
	(-1.148)	(-0.4413)	(-1.053)	(1.253)	(0.4689)	(0.1864)
ln(Assets)	-0.0063***	-0.0201***	-0.0392***	-0.0021	-0.0083***	-0.0173***
	(-2.791)	(-4.959)	(-5.416)	(-1.380)	(-3.139)	(-4.126)
Leverage	0.0488***	0.0905***	0.1619***	-0.0262***	-0.0498***	-0.0466***
	(7.303)	(6.772)	(7.272)	(-4.696)	(-5.402)	(-3.449)
ROA	-0.0850***	0.0182	-0.0523	0.1040***	0.2104***	0.2233***
	(-2.893)	(0.3423)	(-0.6128)	(4.903)	(6.284)	(5.037)
MarketBook	0.0000	-0.0004	-0.0011**	0.0000	0.0001	0.0000
	(-0.1566)	(-1.450)	(-2.546)	(-0.5046)	(0.5639)	(0.0080)
ln(Age)	0.0010	0.0037	0.0093	-0.0011	0.0162*	0.0165
	(0.1554)	(0.3177)	(0.4349)	(-0.2181)	(1.913)	(1.236)
p-value (Con.=Uncon.)	0.031	0.013	0.003	0.040	0.000	0.011
Firm FE	✓	✓	✓	✓	✓	✓
Year-Qtr FE	✓	✓	✓	✓	✓	✓
Observations	23,808	24,186	24,090	24,904	24,527	24,630
Adjusted R^2	0.01032	0.01253	0.01717	0.00636	0.01028	0.01057

Table is continued on the next page...

Table is continued from the last page...

Panel D: Volatility and Sharpe Ratio

Dependent Variable:	Volatility			Sharpe Ratio		
	Next Month	Next Quarter	Next Two Quarters	Next Month	Next Quarter	Next Two Quarters
	(1)	(2)	(3)	(4)	(5)	(6)
Socially Connected Search Volume	**0.0026*****	**0.0049*****	**0.0060*****	**-0.0031**	**-0.0014**	**-0.0052**
	(9.054)	**(10.56)**	**(9.159)**	**(-0.7055)**	**(-0.3384)**	**(-1.101)**
Socially Unconnected Search Volume	**0.0003**	**0.0003**	**0.0006**	**-0.0009**	**-0.0044**	**-0.0019**
	(1.162)	**(0.8423)**	**(1.203)**	**(-0.2477)**	**(-1.204)**	**(-0.4693)**
ln(Analysts)	-0.0022*	-0.0037*	-0.0044	-0.0992***	-0.1212***	-0.1418***
	(-1.834)	(-1.817)	(-1.442)	(-5.970)	(-7.080)	(-6.456)
Inst. Ownership	-0.0137***	-0.0255***	-0.0256**	-0.0808	-0.0837	-0.1521**
	(-2.694)	(-2.962)	(-2.044)	(-1.509)	(-1.612)	(-2.151)
Social Proximity to Capital	0.0097**	0.0102	0.0145	0.0632	0.0479	0.0475
	(2.235)	(1.491)	(1.421)	(1.134)	(0.6687)	(0.5189)
Physical Proximity to Capital	-0.0035	-0.0024	-0.0028	-0.0526	-0.0192	-0.0031
	(-1.149)	(-0.5251)	(-0.4057)	(-1.295)	(-0.4660)	(-0.0553)
ln(Assets)	-0.0047***	-0.0077***	-0.0045	-0.0822***	-0.1773***	-0.2604***
	(-3.022)	(-2.985)	(-1.173)	(-4.622)	(-9.454)	(-10.54)
Leverage	0.0483***	0.0891***	0.1244***	0.0084	-0.0309	0.0776
	(10.12)	(11.25)	(10.86)	(0.1398)	(-0.5044)	(1.019)
ROA	-0.1217***	-0.2399***	-0.3472***	1.033***	2.019***	1.346***
	(-7.811)	(-9.839)	(-9.953)	(4.823)	(9.446)	(5.971)
MarketBook	0.0000	0.0000	0.0000	-0.0011	-0.0018	-0.0033*
	(0.0484)	(-0.1522)	(-0.0580)	(-0.6832)	(-1.116)	(-1.737)
ln(Age)	-0.0066	-0.0240***	-0.0402***	0.0517	0.1455***	0.2291***
	(-1.490)	(-3.188)	(-3.473)	(0.9348)	(2.628)	(3.031)
p-value (Con.=Uncon.)	0.000	0.000	0.000	0.371	0.346	0.348
Firm FE	✓	✓	✓	✓	✓	✓
Year-Qtr FE	✓	✓	✓	✓	✓	✓
Observations	48,715	48,728	48,742	48,712	48,713	48,720
Adjusted R^2	0.02228	0.03675	0.03919	0.00270	0.00756	0.01070

Notes: Panel A of this table presents summary statistics for the firm-level panel dataset used to conduct the regressions tabulated in Panels B to D. Panels B and C show the results from OLS regressions of future unexpected earnings and stock returns, respectively, on *Socially Connected Search Volume* and *Socially Unconnected Search Volume* along with firm-level controls as well as firm and year fixed effects. In Panel D, the dependent variables are future stock volatility and the Sharpe ratio, i.e., stock return divided by stock volatility. The sample covers U.S. firms between Q1 2007 to Q2 2017. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Appendices

Table A1: Variable Definitions

Variable	Definition
Dependent Variables:	
Search Volume	Number of clicks from US county i for disclosures by firm j in year-quarter t
Search Volume$_{\text{Financial Firm IPs}}$	Number of clicks from financial firms (SIC1 = 6) from US county i for disclosures by firm j in year-quarter t
Search Volume$_{\text{Financial Center IPs}}$	Number of clicks from New York City, Boston, Philadelphia, Chicago, Los Angeles or San Francisco for disclosures by firm j in year-quarter t
Search Volume$_{\text{\# Unique IP Addresses}}$	Number of daily unique IP addresses from US county i accessing disclosures by firm j in year-quarter t
Search Volume$_{\text{per IP Address}}$	*Search Volume* divided by *Search Volume*$_{\textit{\# Unique IP Addresses}}$
(Pos./Neg.) Return$_{\text{Next } T}$	(Positive/Negative) buy and hold return of company j's stock over the period T after year-quarter t
(Pos./Neg.) Unexpected Earnings$_{\text{Quarter t}}$	(Positive/Negative) diluted earnings before extraordinary items per share in year-quarter t minus diluted earnings before extraordinary items per share in year-quarter $t-1$, scaled by stock price at the end of fiscal year-quarter t as per Drake et al. (2020)
Vola$_{\text{Next } T}$	Return volatility of company j's stock over the period T after year-quarter t multiplied by the square root of T.
Sharpe Ratio$_{\text{Next } T}$	Return$_{\text{Next } T}$ divided by Vola$_{\text{Next } T}$.
Main Independent Variable:	
ln(Social Connectedness)	Natural logarithm of the number of Facebook friends connected between county i and the county in which the headquarter of company j is located divided by the product of the population of the two counties, scaled up by a factor of 10^9
Socially Connected Search Volume	Decile ranking (1-10) of the number of clicks from socially connected US counties for disclosures by firm j in year-quarter t. Socially connected US counties are defined as counties with above median *Social Connectedness*
Socially Unconnected Search Volume	Decile ranking (1-10) of the number of clicks from socially unconnected US counties for disclosures by firm j in year-quarter t. Socially unconnected US counties are defined as counties with below median *Social Connectedness*
Controls:	
ln(Distance)	Natural logarithm of the physical distance between the midpoint of county i and company j's headquarter location
ln(AUM)	Natural logarithm of the aggregate amount of USD holdings of company j by institutional investors located in county i
ln(Assets)	Natural logarithm of total assets of company j
Leverage	Total liabilities divided by total assets of company j
ROA	Return on assets of company j calculated as operating income before depreciation divided by total assets
MarketBook	Market value of equity divided by book equity of company j
ln(Age)	Natural logarithm of company j's age
Unemployment	Unemployment rate of county i
ln(Population)	Natural logarithm of the population of county i
Income (1000 USD)	Income per capita of county i

Table is continued on the next page...

Table is continued from the last page...

ln(Social Proximity to Capital)	Natural logarithm of the social proximity to capital of the county in which company j is headquartered. County k's social proximity to capital in year-quarter t is defined as $\sum_l(County\ AUM_{l,t} \times Social\ Connectedness_{l,k})$, where county AUM is the sum of AUM of all the institutions located in a given county in year-quarter t. Based on Kuchler et al. (2022)
Institutional Ownership	Percentage of shares outstanding of company j held by institutional investors
Moderators:	
Analysts	Number of analysts following company j
HQ in Metropolitan Area	Binary variable indicating whether the headquarter of company j is located in a county in metro areas of 1 million population or more (2013 Rural-urban Continuum Code "1")
Financial Crisis	Binary variable for the year-quarters 2008Q3, 2008Q4, and 2009Q1
VIX	The average value of the VIX index per year-quarter
Instruments:	
Same Highway	Binary variable equal one if county i and the county in which company j is headquartered are connected by the same highway based on Baum-Snow (2007)

Note: This table provides variable definitions.

Table A2: Social Connectedness and Information Acquisition: County-Pair Similarity Controls

Dependent Variable: Search Volume						
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Social Connectedness)		**0.2549*****	**0.2686*****	**0.2737*****	**0.2698*****	**0.2752*****
		(15.87)	**(12.81)**	**(13.07)**	**(12.81)**	**(13.09)**
ln(Distance)	-0.1786***		0.0153	0.0180	0.0162	0.0190
	(-12.60)		(1.326)	(1.541)	(1.389)	(1.606)
ln(AUM)	0.0058***	0.0052***	0.0052***	0.0051***	0.0065***	0.0063***
	(4.416)	(4.144)	(4.147)	(4.081)	(4.875)	(4.818)
ln(Assets)	0.1929***	0.1933***	0.1933***	0.1928***		
	(15.32)	(15.46)	(15.46)	(15.46)		
Leverage.	0.3291***	0.3309***	0.3311***	0.3248***		
	(9.215)	(9.307)	(9.310)	(8.998)		
ROA	-1.058***	-1.048***	-1.048***	-1.055***		
	(-9.698)	(-9.626)	(-9.625)	(-9.655)		
MarketBook	0.0006	0.0006	0.0006	0.0005		
	(1.206)	(1.171)	(1.168)	(1.072)		
ln(Age)	-0.2273***	-0.2298***	-0.2299***	-0.2141***		
	(-5.002)	(-5.026)	(-5.025)	(-4.766)		
Unemployment	-0.0218	-0.0214	-0.0213		-0.0259	
	(-0.9388)	(-0.9192)	(-0.9155)		(-1.059)	
ln(Population)	0.9292	0.8907	0.8882		0.9204	
	(0.7178)	(0.6882)	(0.6866)		(0.7085)	
Income	-0.0005	-0.0006	-0.0006		-0.0006	
	(-0.0761)	(-0.0967)	(-0.1012)		(-0.0921)	
County-Firm Industry Similarity:						
% Firms Same SIC3	2.634***	2.413***	2.406***	2.422***	2.407***	2.426***
	(4.526)	(4.261)	(4.244)	(4.248)	(4.250)	(4.247)
County-Pair Age Similarity:						
Abs. Diff. % Age 0-17	0.0058*	0.0101***	0.0104***	0.0104***	0.0108***	0.0107***
	(1.824)	(2.884)	(2.913)	(2.945)	(2.989)	(3.017)
Abs. Diff. % Age 18-29	0.0035	0.0039	0.0041	0.0041	0.0039	0.0041
	(0.9797)	(1.063)	(1.124)	(1.599)	(1.059)	(1.598)
Abs. Diff. % Age 30-44	0.0009	-0.0019	-0.0017	-0.0020	-0.0022	-0.0025
	(0.1723)	(-0.3747)	(-0.3344)	(-0.4879)	(-0.4141)	(-0.5700)
Abs. Diff. % Age 45-64	-0.0124**	-0.0109*	-0.0107*	-0.0109**	-0.0106*	-0.0110**
	(-1.980)	(-1.955)	(-1.952)	(-2.397)	(-1.877)	(-2.359)
Abs. Diff. % Age >65	-0.0044	-0.0009	-0.0013	-0.0000	-0.0014	-0.0003
	(-1.031)	(-0.2239)	(-0.3120)	(-0.0256)	(-0.3386)	(-0.0911)
County-Pair Education Similarity:						
Abs. Diff. % College Degree	0.0025	0.0030	0.0026	0.0022	0.0029	0.0024
	(0.9397)	(1.296)	(1.145)	(0.9627)	(1.230)	(1.030)
Abs. Diff. % No High School Degree	0.0063***	0.0039*	0.0037*	0.0041**	0.0037*	0.0041**
	(2.948)	(1.887)	(1.804)	(2.086)	(1.738)	(2.035)
Abs. Diff. % Limit English	-0.0031	0.0039	0.0044	0.0040	0.0040	0.0037
	(-0.9630)	(1.314)	(1.487)	(1.338)	(1.340)	(1.206)
County-Pair Employment Similarity:						
Abs. Diff. % Administration	-0.0025	-0.0010	-0.0013	-0.0022	-0.0008	-0.0015
	(-0.3365)	(-0.1465)	(-0.1868)	(-0.3029)	(-0.1056)	(-0.2055)
Abs. Diff. % Art	-0.0028	-0.0015	-0.0015	-0.0038	-0.0028	-0.0056
	(-0.3570)	(-0.2140)	(-0.2164)	(-0.6348)	(-0.3999)	(-0.8893)
Abs. Diff. % Construction	-0.0014	0.0035	0.0037	0.0069	0.0038	0.0069
	(-0.2240)	(0.6337)	(0.6822)	(1.413)	(0.6782)	(1.388)
Abs. Diff. % Education	0.0028	-0.0019	-0.0026	-0.0030	-0.0024	-0.0027
	(1.059)	(-0.7889)	(-0.9926)	(-1.279)	(-0.8808)	(-1.140)
Abs. Diff. % Finance	0.0203***	0.0157***	0.0152***	0.0140***	0.0158***	0.0146***
	(3.257)	(2.936)	(2.876)	(2.621)	(3.030)	(2.772)
Abs. Diff. % Government	-0.0055*	-0.0043	-0.0041	-0.0023	-0.0042	-0.0023
	(-1.812)	(-1.445)	(-1.398)	(-0.8268)	(-1.373)	(-0.7952)
Abs. Diff. % Information	0.0340***	0.0315***	0.0315***	0.0251**	0.0356***	0.0298***
	(3.153)	(2.997)	(2.976)	(2.382)	(3.331)	(2.821)
Abs. Diff. % Manufacturing	-0.0048	-0.0018	-0.0017	-0.0018	-0.0022	-0.0024
	(-1.472)	(-0.5779)	(-0.5613)	(-0.5707)	(-0.6863)	(-0.7185)
Abs. Diff. % Nature	-0.0814***	-0.0584***	-0.0580***	-0.0583***	-0.0653***	-0.0654***
	(-7.718)	(-5.806)	(-5.787)	(-5.795)	(-6.295)	(-6.248)
Abs. Diff. % Other	-0.0051	-0.0135	-0.0145	-0.0238**	-0.0156	-0.0251**
	(-0.3742)	(-1.052)	(-1.113)	(-2.265)	(-1.149)	(-2.308)

Table is continued on the next page...

Table is continued from the last page...

Abs. Diff. % Professional	-0.0148***	-0.0080***	-0.0074**	-0.0053*	-0.0079***	-0.0058*
	(-4.225)	(-2.755)	(-2.533)	(-1.786)	(-2.622)	(-1.902)
Abs. Diff. % Retail	0.0049	0.0003	0.0003	-0.0033	0.0002	-0.0035
	(0.5647)	(0.0343)	(0.0372)	(-0.4652)	(0.0273)	(-0.4816)
Abs. Diff. % Transport	0.0055	0.0114*	0.0120*	0.0134**	0.0124**	0.0137**
	(0.8190)	(1.845)	(1.923)	(2.284)	(1.977)	(2.313)
Abs. Diff. % Wholesale	-0.0084	-0.0019	-0.0019	-0.0054	-0.0033	-0.0065
	(-0.8016)	(-0.1962)	(-0.1866)	(-0.6854)	(-0.3255)	(-0.7961)
Abs. Diff. % Unemployment	-0.0127***	-0.0083*	-0.0081*	-0.0065	-0.0072	-0.0053
	(-2.930)	(-1.952)	(-1.895)	(-1.608)	(-1.637)	(-1.283)
County-Pair Family Similarity:						
Abs. Diff. % Single Parent	0.0003	0.0004	0.0005	0.0004	0.0005	0.0004
	(0.2086)	(0.2952)	(0.3496)	(0.3036)	(0.3911)	(0.3111)
Abs. Diff. % Home Work	0.0013***	0.0014***	0.0014***	0.0014***	0.0015***	0.0014***
	(2.652)	(3.158)	(3.156)	(3.230)	(3.241)	(3.346)
County-Pair Income Similarity:						
Abs. Diff. % Gini Index	1.133**	1.081***	1.088***	1.149***	1.101***	1.135***
	(2.338)	(2.852)	(2.898)	(3.083)	(2.826)	(2.936)
Abs. Diff. % Income <10,000	0.0064	0.0036	0.0032	0.0032	0.0042	0.0044
	(1.250)	(0.7495)	(0.6709)	(0.6926)	(0.8826)	(0.9548)
Abs. Diff. % Income >100,000	0.0021	0.0023	0.0025	0.0017	0.0025	0.0016
	(0.6898)	(0.8002)	(0.8750)	(0.6578)	(0.8234)	(0.6163)
Abs. Diff. % Income 10,000-14,999	-0.0147	-0.0125	-0.0121	-0.0107	-0.0141	-0.0127
	(-1.543)	(-1.363)	(-1.332)	(-1.363)	(-1.486)	(-1.604)
Abs. Diff. % Income 15,000-24,999	-0.0312***	-0.0266***	-0.0265***	-0.0271***	-0.0261***	-0.0268***
	(-4.199)	(-3.718)	(-3.685)	(-5.533)	(-3.469)	(-5.307)
Abs. Diff. % Income 25,000-49,999	0.0033	0.0095	0.0094	0.0115**	0.0091	0.0113*
	(0.4965)	(1.462)	(1.464)	(1.970)	(1.350)	(1.856)
Abs. Diff. % Income 50,000-99,999	-0.0057	-0.0039	-0.0039	-0.0009	-0.0031	0.0002
	(-1.370)	(-0.8525)	(-0.8334)	(-0.2549)	(-0.6358)	(0.0603)
County-Pair Mobility Similarity:						
Abs. Diff. % Moved State	-0.0146*	-0.0185**	-0.0188**	-0.0214***	-0.0200***	-0.0225***
	(-1.828)	(-2.459)	(-2.539)	(-3.136)	(-2.612)	(-3.207)
Abs. Diff. % Moved Within County	0.0059*	0.0018	0.0011	0.0009	0.0012	0.0012
	(1.792)	(0.5100)	(0.3323)	(0.3122)	(0.3388)	(0.3923)
Abs. Diff. % Moved Within State	-0.0009	-0.0013	-0.0017	-0.0038	-0.0007	-0.0027
	(-0.1379)	(-0.2564)	(-0.3349)	(-0.7441)	(-0.1276)	(-0.5129)
County-Pair Racial Similarity:						
Abs. Diff. % Asian	0.0095**	0.0069*	0.0063	0.0064	0.0065	0.0066*
	(2.123)	(1.728)	(1.597)	(1.631)	(1.610)	(1.649)
Abs. Diff. % Black	0.0007	0.0019*	0.0018*	0.0018*	0.0018*	0.0017
	(0.5577)	(1.733)	(1.708)	(1.704)	(1.670)	(1.636)
Abs. Diff. % Hispanic	-0.0023**	-0.0024***	-0.0025***	-0.0024***	-0.0025***	-0.0024***
	(-2.397)	(-2.616)	(-2.726)	(-2.748)	(-2.754)	(-2.783)
Abs. Diff. % Native	-0.1478***	-0.1320***	-0.1348***	-0.1343***	-0.1394***	-0.1405***
	(-4.506)	(-4.695)	(-4.841)	(-4.808)	(-4.864)	(-4.917)
Abs. Diff. % White	-0.0016**	0.0000	0.0002	0.0000	0.0001	0.0000
	(-2.400)	(0.0734)	(0.3001)	(0.1754)	(0.2170)	(0.1483)
Abs. Diff. % Population	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(3.469)	(4.282)	(4.237)	(4.011)	(4.598)	(4.314)
Controls as in Table 2 Panel A	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓		
County FE	✓	✓	✓		✓	
Year-Qtr FE	✓	✓	✓			
Same County FE	✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE					✓	✓
County×Year-Qtr FE				✓		✓
Observations	12,644,761	12,644,761	12,644,761	12,630,090	12,622,988	12,608,353
Pseudo R^2	0.45089	0.45662	0.45681	0.47229	0.47487	0.49107

Notes: This table shows the detailed estimates of the PPML regressions incompletely tabulated in Panel B of Table 2. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table A3: Social Connectedness and Information Acquisition: 500 Distance Percentiles

Dependent Variable: Search Volume				
	(1)	(2)	(3)	(4)
ln(Social Connectedness)	**0.2328*****	**0.2343*****	**0.2352*****	**0.2370*****
	(11.20)	**(11.21)**	**(11.18)**	**(11.27)**
ln(AUM)	0.0054***	0.0050***	0.0069***	0.0064***
	(4.148)	(4.049)	(4.895)	(4.805)
ln(Assets)	0.2035***	0.2040***		
	(17.22)	(17.36)		
Leverage	0.3259***	0.3200***		
	(9.475)	(9.194)		
ROA	-1.143***	-1.148***		
	(-11.68)	(-11.75)		
MarketBook	0.0009*	0.0008*		
	(1.811)	(1.706)		
ln(Age)	-0.2293***	-0.2323***		
	(-6.085)	(-6.302)		
Unemployment	-0.0010		-0.0098	
	(-0.0534)		(-0.4244)	
ln(Population)	0.1747		0.1833	
	(0.1284)		(0.1345)	
Income	0.0027		0.0027	
	(0.4434)		(0.4262)	
Distance 500-tile FE	✓	✓	✓	✓
Firm FE	✓	✓		
County FE	✓		✓	
Year-Qtr FE	✓			
Same County FE	✓	✓	✓	✓
Firm×Year-Qtr FE			✓	✓
County×Year-Qtr FE		✓		✓
Observations	14,188,331	14,172,693	14,161,448	14,145,833
Pseudo R^2	0.43847	0.45681	0.45774	0.47615

Notes: This table shows results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *ln(Social Connectedness)* along with county- and firm-level controls and varying sets of fixed effects. The regressions are identical to those shown in Panel A of Table 2, except that the control variable *ln(Distance)* is replaced by 500-tile fixed effects for physical distance (as per Kuchler et al., 2022). The sample covers U.S. firms and counties. The sample period for the regressions is Q1 2007 to Q2 2017. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table A4: Social Connectedness and Information Acquisition: Alternative Dependent Variables

Panel A: Dependent Variable: Search Volume$_{\text{\# Unique IP Addresses}}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(Social Connectedness)		**0.2751*****		**0.2164*****	**0.2225*****	**0.2240*****	**0.2245*****	**0.2257*****
		(13.96)		**(15.93)**	**(11.95)**	**(11.91)**	**(11.94)**	**(11.93)**
ln(Distance)	-0.0839**		-0.1567***		0.0064	0.0062	0.0077	0.0071
	(-2.507)		(-11.74)		(0.5859)	(0.5641)	(0.6991)	(0.6354)
ln(AUM)			0.0064***	0.0059***	0.0059***	0.0053***	0.0072***	0.0066***
			(4.428)	(4.089)	(4.087)	(3.925)	(4.566)	(4.420)
ln(Assets)			0.1796***	0.1795***	0.1795***	0.1803***		
			(16.72)	(16.80)	(16.81)	(16.90)		
Leverage			0.2431***	0.2452***	0.2453***	0.2366***		
			(7.899)	(8.003)	(8.008)	(7.541)		
ROA			-0.9833***	-0.9737***	-0.9736***	-0.9813***		
			(-10.80)	(-10.66)	(-10.66)	(-10.77)		
MarketBook			0.0013***	0.0012***	0.0012***	0.0012***		
			(2.883)	(2.784)	(2.780)	(2.650)		
ln(Age)			-0.2587***	-0.2603***	-0.2603***	-0.2652***		
			(-6.727)	(-6.752)	(-6.753)	(-7.005)		
Unemployment			-0.0007	-0.0009	-0.0010		-0.0101	
			(-0.0382)	(-0.0482)	(-0.0485)		(-0.4347)	
ln(Population)			0.3042	0.2981	0.2981		0.3104	
			(0.2229)	(0.2180)	(0.2180)		(0.2265)	
Income			0.0044	0.0043	0.0043		0.0042	
			(0.6589)	(0.6560)	(0.6561)		(0.6201)	
Firm FE			✓	✓	✓	✓		
County FE			✓	✓	✓		✓	
Year-Qtr FE			✓	✓	✓			
Same County FE			✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE							✓	✓
County×Year-Qtr FE						✓		✓
Observations	14,188,331	14,188,331	14,188,331	14,188,331	14,188,331	14,172,693	14,161,448	14,145,833
Pseudo R^2	0.04266	0.05286	0.51965	0.52435	0.52446	0.55061	0.54268	0.56844

Panel B: Dependent Variable: Search Volume$_{\text{per IP Address}}$

		(1)	(2)	(3)	(4)			
ln(Social Connectedness)		**0.0178*****		**0.0215*****	**0.0195*****	**0.0186*****	**0.0193*****	**0.0185*****
		(6.397)		**(16.17)**	**(10.68)**	**(10.67)**	**(10.63)**	**(10.65)**
ln(Distance)	-0.0163***		-0.0184***		-0.0023	-0.0024	-0.0024	-0.0025
	(-7.974)		(-14.68)		(-1.324)	(-1.472)	(-1.397)	(-1.514)
ln(AUM)			-0.0002	-0.0003	-0.0003	-0.0002	-0.0002	-0.0000
			(-1.096)	(-1.353)	(-1.344)	(-1.033)	(-0.7878)	(-0.2636)
ln(Assets)			0.0275***	0.0275***	0.0275***	0.0268***		
			(7.861)	(7.871)	(7.872)	(7.813)		
Leverage			0.0584***	0.0585***	0.0585***	0.0587***		
			(5.383)	(5.399)	(5.399)	(5.558)		
ROA			-0.1401***	-0.1391***	-0.1392***	-0.1443***		
			(-4.586)	(-4.556)	(-4.558)	(-4.824)		
MarketBook			-0.0003*	-0.0003*	-0.0003*	-0.0003*		
			(-1.807)	(-1.815)	(-1.814)	(-1.706)		
ln(Age)			0.0248*	0.0248*	0.0248*	0.0246**		
			(1.955)	(1.959)	(1.959)	(2.013)		
Unemployment			-0.0020	-0.0020	-0.0020		-0.0014	
			(-0.5890)	(-0.5885)	(-0.5890)		(-0.3656)	
ln(Population)			0.0737	0.0727	0.0727		0.0651	
			(0.7271)	(0.7170)	(0.7176)		(0.6486)	
Income			-0.0021*	-0.0021*	-0.0021*		-0.0024*	
			(-1.731)	(-1.735)	(-1.735)		(-1.888)	
Firm FE			✓	✓	✓	✓		
County FE			✓	✓	✓		✓	
Year-Qtr FE			✓	✓	✓			
Same County FE			✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE							✓	✓
County×Year-Qtr FE						✓		✓
Observations	6,958,209	6,958,209	6,958,209	6,958,209	6,958,209	6,958,209	6,958,209	6,958,209
Pseudo R^2	0.00060	0.00062	0.02186	0.02192	0.02192	0.04884	0.03604	0.06536

Notes: This table shows results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variables *Search Volume*$_{\text{\# Unique IP Addresses}}$ (Panel A) and *Search Volume*$_{\text{per IP Address}}$ (Panel B), respectively, on *ln(Social Connectedness)* along with county- and firm-level controls and varying sets of fixed effects. The sample covers U.S. firms and counties. The sample period for the regressions is Q1 2007 to Q2 2017. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table A5: 2SLS-IV Regression Results

Dependent Variable: (1) ln(Social Connectedness) (2) ln(Search Volume)	First Stage	Second Stage
	(1)	(2)
Same Highway	**0.222***** **(10.004)**	
Social Connectedness (IV)		**0.231***** **(7.158)**
ln(Distance)	-0.828*** (-54.388)	0.090*** (3.296)
ln(AUM)	0.003*** (7.136)	0.010*** (9.185)
Same County FE	✓	✓
Firm×Year-Qtr FE	✓	✓
County×Year-Qtr FE	✓	✓
Kleibergen-Paap F-statistic	100.08	
Ratio Social Connectedness (IV) / Social Connectedness (OLS)		2.42
Observations	14,188,331	14,188,331
Adjusted R^2	0.789	0.488

Notes: This table shows results from a two-stage least squares regression. Column (1) shows the results from the first-stage regression of *ln(Social Connectedness)* on the instrument *Same Highway* and column (2) shows the results from the second-stage regression of *Search Volume* on the fitted values from the first-stage regression *Social Connectedness (IV)* along with a set of controls and fixed effects. *Same Highway* indicates if a searcher county and a firm's headquarter county share a highway based on Baum-Snow (2007). *Ratio Social Connectedness (IV) / Social Connectedness (OLS)* is the ratio of the coefficient on *Social Connectedness (IV)* from the second-stage (OLS) regression as shown in column (2) divided by the coefficient on *ln(Social Connectedness)* from an OLS regression of *Search Volume* on *ln(Social Connectedness)* with the same set of controls and fixed effects as in column (2). The sample covers U.S. firms and counties. The full sample period for the regressions is Q1 2007 to Q2 2017. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

Table A6: Information Acquisition and Social Connectedness: Annual Estimates

Dependent Variable: Search Volume											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Subsample Based on Observations from	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
ln(Social Connectedness)	**0.2384*****	**0.2814*****	**0.2635*****	**0.2723*****	**0.2530*****	**0.2690*****	**0.2585*****	**0.2619*****	**0.2652*****	**0.2639*****	**0.2569*****
	(8.029)	**(8.981)**	**(8.388)**	**(9.373)**	**(11.83)**	**(12.43)**	**(12.42)**	**(13.00)**	**(13.35)**	**(13.86)**	**(14.00)**
ln(Distance)	-0.0180	0.0044	0.0005	0.0141	0.0061	0.0086	0.0114	0.0145	0.0193	0.0228*	0.0220*
	(-0.9744)	(0.2329)	(0.0292)	(0.8083)	(0.4540)	(0.5679)	(0.8233)	(1.097)	(1.535)	(1.836)	(1.837)
ln(AUM)	0.0033***	0.0052***	0.0052***	0.0060***	0.0066***	0.0079***	0.0073***	0.0068***	0.0068***	0.0071***	0.0066***
	(2.596)	(4.445)	(3.539)	(3.983)	(4.102)	(5.532)	(4.634)	(3.860)	(3.647)	(3.340)	(3.115)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Same County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm×Year-Qtr FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County×Year-Qtr FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	571,950	909,873	1,189,906	1,429,970	1,583,238	1,659,215	1,714,655	1,712,368	1,580,600	1,316,767	471,061
Pseudo R^2	0.41575	0.43376	0.47338	0.47381	0.45462	0.43302	0.47267	0.47688	0.48302	0.48371	0.47227

Notes: This table shows results from PPML (Poisson Pseudo Maximum Likelihood) regressions of the dependent variable *Search Volume* on *ln(Social Connectedness)* along with county- and firm-level controls and a set of fixed effects. Each regression is based on a subsample with observations from a single year. The sample covers U.S. firms and counties. The full sample period for the regressions is Q1 2007 to Q2 2017. Appendix Table A1 provides variable definitions. The t-statistics (shown below the coefficient estimates) are on the basis of standard errors clustered by searcher county and firm. The following significance levels are indicated by asterisks: *** (1%), ** (5%), * (10%).

cfr working paper series

centre for financial research cologne

cfr working papers are available for download from www.cfr-cologne.de.

2024

No.	Author(s)	Title
24-03	T. Dyer, G. Köchling, P. Limbach	Traditional Investment Research and Social Networks: Evidence from Facebook Connections
24-02	A. Y. Chen, A. Lopez-Lira, T. Zimmermann	Does Peer-Reviewed Research Help Predict Stock Returns?
24-01	G. Cici, P. Schuster, F. Weishaupt	Once a Trader, Always a Trader: The Role of Traders in Fund Management

2023

No.	Author(s)	Title
23-08	A. Braun, J. Braun, F. Weigert	Extreme Weather Risk and the Cost of Equity
23-07	A. G. Huang, R. Wermers, J. Xue	"Buy the Rumor, Sell the News": Liquidity Provision by Bond Funds Following Corporate News Events
23-06	J. Dörries, O. Korn, G. J. Power	How Should the Long-term Investor Harvest Variance Risk Premiums?
23-05	V. Agarwal, W. Jiang, Y. Luo, H. Zou	The Real Effect of Sociopolitical Racial Animus: Mutual Fund Manager Performance During the AAPI Hate
23-04	V. Agarwal, B. Barber, S. Cheng, A. Hameed, H. Shanker, A. Yasuda	Do Investors Overvalue Startups? Evidence from the Junior Stakes of Mutual Funds
23-03	A. Höck, T. Bauckloh, M. Dumrose, C. Klein	ESG Criteria and the Credit Risk of Corporate Bond Portfolios
23-02	T. Bauckloh, J. Dobrick, A. Höck, S. Utz, M. Wagner	In partnership for the goals? The level of agreement between SDG ratings
23-01	F. Simon, S. Weibels, T. Zimmermann	Deep Parametric Portfolio Policies

2022

No.	Author(s)	Title
22-12	V. Agarwal, A. Cochardt, V. Orlov	Birth Order and Fund Manager's Trading Behavior: Role of Sibling Rivalry

22-11	G. Cici, S. Gibson, N. Qin, A. Zhang	The Performance of Corporate Bond Mutual Funds and the Allocation of Underpriced New Issues
22-10	E. Theissen, C. Westheide	One for the Money, Two for the Show? The Number of Designated Market Makers and Liquidity
22-09	R. Campbell, P. Limbach, J. Reusche	Once Bitten, Twice Shy: Failed Deals and Subsequent M&A Cautiousness
22-08	M. Gehde-Trapp, L. Klingler	The Effect of Sentiment on Institutional Investors: A Gender Analysis
22-07	T. Bauckloh, V. Beyer, C. Klein	Does it Pay to Invest in Dirty Industries? – New Insights on the Shunned-Stock Hypothesis
22-06	J. Balthrop and G. Cici	Conflicting Incentives in the Management of 529 Plans
22-05	I. T. Ivanov, T. Zimmermann, N. W. Heinrich	Limits of Disclosure Regulation in the Municipal Bond Market
22-04	M. Ammann, A. Cochardt, S. Straumann, F. Weigert	Back to the Roots: Ancestral Origin and Mutual Fund Manager Portfolio Choice
22-03	A. Betzer, J. Gider, P. Limbach	Do Financial Advisors Matter for M&A Pre-Announcement Returns?
22-02	S. Lesmeister, P. Limbach, P.R. Rau, F. Sonnenburg	Indexing and the Performance-Flow Relation of Actively Managed Mutual Funds
22-01	T. Bauckloh, C. Klein, T. Pioch, F. Schiemann	Under Pressure: The Link between Mandatory Climate Reporting and Firms' Carbon Performance

This document only covers the most recent cfr working papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
fon +49(0)221-470-6995
fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de