Firm Linkages from the Wisdom of Crowds

by *

Huaixin Wang Weichen Zhang

January 15, 2023

Abstract

Social media activity reflects the collective perception of firms' interactions, thereby providing an implicit representation of the financial network. Using data from the Reddit platform, we identify linkages between firms implied by social media. We construct firm connections by either shared Reddit posts or shared Reddit authors, and show that Reddit-linked firms have correlated fundamentals for a battery of firm characteristics. We also show that the returns of Reddit peer firms have a strong predictive ability for focal firms' future returns. The lead-lag effect of Reddit-linked firms is distinct from other cross-firm return predictability and consistent with the limited attention hypothesis. The result suggests that Reddit linkage helps to capture comprehensive and dynamic information about firm connections and provides a new dimension for analyzing the role of social media in financial markets.

JEL classification: G11, G12, G14

Keywords: social media, connected firms, comovements, return predictability

^{*}Preliminary, comments welcomed. We thank Jianfeng Yu and Xiaoyan Zhang for valuable comments and suggestions. All errors are our own. Author affiliation/contact information: Huaixin Wang and Weichen Zhang are from PBC School of Finance, Tsinghua University. E-mail address: wanghx.19@pbcsf.tsinghua.edu.cn (Huaixin Wang), zhangwch.18@pbcsf.tsinghua.edu.cn (Weichen Zhang)

Firm Linkages from the Wisdom of Crowds

Abstract

Social media activity reflects the collective perception of firms' interactions, thereby providing an implicit representation of the financial network. Using data from the Reddit platform, we identify linkages between firms implied by social media. We construct firm connections by either shared Reddit posts or shared Reddit authors, and show that Reddit-linked firms have correlated fundamentals for a battery of firm characteristics. We also show that the returns of Reddit peer firms have a strong predictive ability for focal firms' future returns. The lead-lag effect of Reddit-linked firms is distinct from other cross-firm return predictability and consistent with the limited attention hypothesis. The result suggests that Reddit linkage helps to capture comprehensive and dynamic information about firm connections and provides a new dimension for analyzing the role of social media in financial markets.

JEL classification: G11, G12, G14

Keywords: social media, connected firms, comovements, return predictability

1 Introduction

The rise of social media has attracted growing attention in financial economics studies. Starting with early work by Turmakin and Whitelaw (2001) and Antweiler and Frank (2004), researchers have examined the information content of online messages (Bartov, Faurel, and Mohanram, 2018) and investigated topics such as information transmission (Chen, De, Hu, and Hwang, 2014; Rakowski, Shirley, and Stark, 2021), dispersion of opinion (Cookson and Niessner, 2020), and investor sentiment (Das and Chen, 2007; McGurk, Nowak, and Hall, 2020). Most recently, the extreme price movement of *GameStop* induces a new wave of research studying how social media affects investors' trading behavior and price efficiency (Diangson and Jung, 2021; Hu, Jones, Zhang, and Zhang, 2021; Long, Lucey, and Yarovaya, 2021; Vasileiou, Bartzou, and Tzanakis, 2021; Lyócsa, Baumöhl, and Výrost, 2022).

This paper attempts to provide a different perspective on the role of social media in financial markets. What motivates our study is that the network structure, which is one of the most salient features, of social media, could contain useful information about underlying connections among individual firms. Consider the scenario that social media users subscribe to online communities and share information, opinions, and ideas on some platform. If two firms appear in the same "post", a collection of comments from users about a particular theme, then such co-mention relation would indicate a potential linkage between the stocks featured. Similarly, if two firms are followed by the same social media user, then such overlapping preference also suggests that these firms should not be mutually independent in the financial network.

We explore financial implications of linkage relations underlying social media, using *Red*dit as the reseach subject. Reddit, the so called "front page of the Internet," is a social news and discussion website.¹ Figure 1 shows the brief structure of this platform. Typically, Reddit is organized into "subreddits", which are communities with specific topics.² Our focus is the r/wallstreetbets subreddit, in which users discuss about equity and option trading. It also becomes well-known after the "GameStop Short Squeeze" occurred in January 2021.³ In the r/wallstreetbets subreddit, users can post any content that is related to the trading or that they believe will be of interest to other members of the community, such as texts, images, videos, or links to external websites. The content that users post in a subreddit

¹As of November 2022, Reddit is the 9th most-visited website in the world and the 6th most-visited website in the United States. Details and the most recent statistics can be found at https://www.semrush.com/website/reddit.com/overview/

²For example, a topic can be financial news, politics, sports, or personal hobbies.

 $^{^{3}}$ The short squeeze caused GameStop's share price to rise by 1,500% over two weeks, reaching an all-time intraday high of \$483.00 on the NYSE on January 29.

is intended to spark discussion and engagement among members of the community. Users can exchange their opinions about specific stocks or the market trend in posts through comments, and comments are formed by "reply-to" relations between them, as shown in Figure 1. The Reddit community r/wallstreetbets is devoted to equities with significant market risk and is credited with a concerted effort for the majority of this effect. The GameStop event was a notable example of the power of online communities in influencing market trends.

[Place Figure 1 here]

We choose Reddit as our research subject for three reasons. First, the "GameStop Short Squeeze" event demonstrates how social media comments and peer opinions are associated with stock price movement. A growing body of literature shows the predictive ability of social media activities for future stock returns.⁴ Our construction of the social media linkage is well supported by this branch of studies. Second, Reddit has millions of users from all over the world, which enables us to gain insights into how people use social media and the types of discussions and interactions that take place on the platform. Third, unlike other peerreviewed websites, r/wallstreetbets administrators do not actively engage in the discussion of each post. Instead, registered Reddit users can upvote or downvote posts and comments to affect their ranking position. The posts must receive votes and comments to stay at the top of the page; otherwise, they might quickly disappear in the flood of fresher posts. This setting helps to prevent some social media platform regulators from having the power to censor speech and steer users in certain directions.

Using detailed text data from the Reddit platform, we start our analysis by defining two types of "Reddit peers" for individual firms. First, two firms are identified as *Reddit postlinked* each month if there was at least one post on Reddit mentioning both firms in the past six months. This construction captures the collective perspective of linkage relationships between firms. Second, we defined two firms as *Reddit author-linked* if at least one Reddit user published comments on both firms in the past six months. This definition reflects the aggregated preference of Reddit users for potentially connected firms. Our final Reddit linkage sample period spans from December 2018 to December 2021.

The first part of our result shows that Reddit-linked firms are economically connected. Specifically, we document significant fundamental comovements between focal firms and their Reddit peers, such as quarterly stock return, return on assets, gross profit, earnings-to-price ratio, asset growth, and asset turnover. We also find that the comovement becomes stronger

⁴See, for example, Chen et al. (2014) and Hu et al. (2021).

on average when refining peer firms to have higher numbers of shared Reddit posts or shared Reddit authors. In addition, the fundamental comovements of asset growth and leverage ratio become less evident after the refinement, suggesting that both strongly linked peers and weakly linked peers correlate with the focal firm in fundamentals. These results suggest that firm linkages identified by Reddit help to capture real economic connections.

Next, we examine the cross-firm return predictability in the context of Reddit linkage. Prior research such as Cohen and Frazzini (2008) shows that economically connected firms exhibit a "lead-lag" effect in stock returns. The intuition is that investors are likely to overlook value-relevant information implied by peer firms' price movement, due to limited attention; as a result, the focal firm's stock price would delay in reacting to peer stocks' returns. To test this asset pricing implication, we calculate the weighted average daily return of Reddit peers, *REDDIT RET*, for each focal stock. The weights in calculations are determined by the number of shared Reddit posts (authors) with the focal firm. The portfolio analysis reveals a significant lead-lag returns relation among Reddit-linked firms. Under the Reddit post linkage setting, the strategy that buying stocks in the top *REDDIT RET* decile and selling stocks in the bottom *REDDIT RET* decile earns a value-weighted return of 5.86 basis points per day and an equal-weighted return of 6.58 basis points per day, on average. We find weaker but positive returns for strategies under the Reddit author linkage.

The predictive ability of Reddit peer return for focal stocks' daily returns remains significant after controlling for other firm characteristics in regressions, and is also robust to controlling for the industry lead-lag effect and shared analyst coverage (Ali and Hirshleifer, 2020). We further confirm the limited attention channel in driving the Reddit lead-lag effect by showing that the predictive ability of *REDDIT RET* is weaker for focal stocks with more attention or higher institutional ownership. We also find that the momentum spillover is stronger for stocks with more Reddit peers, which is consistent with the hypothesis that investors underreact to peer stocks' price movement due to limited capability in processing value-relevant information.

In robustness tests, we first examine the information content of Reddit peers' stock returns. We find that Reddit-linked stocks' quarterly returns positively predict the focal firm's future earnings surprise, suggesting the price movement of Reddit peers does contain useful information about future fundamentals of the focal stock. Second, the momentum spillover returns are higher on earnings announcement or news release days, which is consistent with the initial mispricing story. Third, we explore the time-variation of the lead-lag effect of Reddit-linked firms. Consistent with the limited attention hypothesis, the strategy return of Reddit peer momentum is higher following the spike of uncertainty, measured by the Chicago Board Options Exchange Volatility Index (VIX). Next, we test the predictive ability of *REDDIT RET* for long-horizon returns. It turns out that the Reddit post peer return persistently predicts the focal stock's future returns for up to 40 days; by contrast, the predictive power of the Reddit author peer return is more transient. Finally, we address the concern that the cross-firm return predictability from Reddit linkages is driven by investors' continued overreaction. Using various constructions of order imbalance, we find no predictive ability of *REDDIT RET* for future excess demand. This result suggests that the lead-lag effect documented in this paper is more likely to be driven by limited attention rather than overreaction.

This paper contributes to the emerging literature that studies the impact of social media on financial markets. Pedersen (2022) provides a theoretical framework of how investment ideas transmitted through a social network can affect investor behavior and market prices. The belief spillover in social network interactions implied in Pedersen (2022) is closely related to our study. Using data from *Seeking Alpha*, Chen et al. (2014) is one of the earliest papers that discuss how investor opinions are propagated through social media. Hu et al. (2021) show that Reddit activities have a predictive power for future stock returns. Dim (2022) discusses how Social Media Investment Analysts, the influencers on social media platforms, affect the form of others' beliefs. Hosseini, Jostova, Philipov, and Savickas (2020) use Twitter data to investigate the social media risk premium in stocks and bonds. These theoretical and empirical findings reveal the existence of information contained in social media activities and enable us to construct inter-firm social media linkages.

Our work is also closely related to the literature on social networks. Kuchler and Stroebel (2021) provide a comprehensive review of recent contributions to this field. Studies in this area typically discuss the impact of social networks on financial decisions; for example, Bailey, Cao, Kuchler, and Stroebel (2018a) show how Facebook connections influence house price expectations. Recent studies also use social media networks to indicate connections in real society. Bailey, Cao, Kuchler, Stroebel, and Wong (2018b) use Facebook administrative data and create the Social Connectedness Index to measure social connectedness between two locations; our method is conceptually similar to theirs.

To be specific, we treat social media as a venue for information dissemination and extract economic connections between firms using social media's built-in network structure. From this angle, our approach to the joint study of social media and financial markets differs greatly from previous literature. Instead of treating individual firms as independent, we take advantage of the natural network structure of social media and integrate the collective perspective about economic connections. Therefore, this paper would suggest a new dimension over which social media takes effect in the transmission, aggregation, and incorporation of information in the financial market.

This paper also adds to the growing literature on discovering economic connections and studying their asset pricing implications. These connections manifest in multiple ways, including, for example, customer-supplier relation (Cohen and Frazzini, 2008), upstreamdownstream industries (Menzly and Ozbas, 2010), conglomerate firms (Cohen and Lou, 2012), co-searches (Lee, Ma, and Wang, 2015), alliance partners (Cao, Chordia, and Lin, 2016), technological closeness (Lee, Sun, Wang, and Zhang, 2019), geographic links (Parsons, Sabbatucci, and Titman, 2020), shared analyst coverage (Ali and Hirshleifer, 2020), and competition mentions (Eisdorfer, Froot, Ozik, and Sadka, 2022). This stream of research documents strong cross-firm return predictability among connected firms, which is usually referred to as momentum spillovers or the lead-lag returns puzzle. We contribute to this branch of literature by showing that the lead-lag returns relation also exists in the context of social media and the Reddit linkage provides incremental information about economic connections.

We are certainly not the first paper that uses social media information to identify firm linkages. For example, Al Guindy and Riordan (2019) collect data about firm-related tweets from Twitter and construct stock connections they called "The Social Internetwork". Our work is considerably different from Al Guindy and Riordan (2019) and other papers discussing "wisdom of the crowd" in social media along several important dimensions.

First, Reddit is different from other social media sites in nature, such as Twitter or Stocktwits. Twitter is organized around individual profiles and tweets, whereas Reddit is structured as communities called "subreddits" of particular topics. Additionally, Twitter and other Twitter-like platforms are more like "follower social media", where users follow another user rather than content. These follower social media platforms have a built-in social network, which is different from Reddit's recommendation algorithm based on comments and votes and is more like a reflection of the collective wisdom of the group. Accordingly, Twitter users could mainly get information and opinions from influencers,⁵ or those referred to as social media investment analysts. Wang (2022) demonstrates that a change in Reddit comment display influence stocks' absolute abnormal returns and retail trading volumes in the five minutes after comment publication. These studies suggest that the heterogeneity in how users allocate their attention may impact the network structure and how information is disseminated.

⁵Bakshy, Hofman, Mason, and Watts (2011) analyzes 74 million diffusion events in Twitter and indicates that the biggest cascades frequently come from influential users.

Second, our research questions are different from related studies. For example, Al Guindy and Riordan (2019) focus on *comovements* in stock returns and trading volume, and study the propagation of information and noise from the network perspective. While our paper also examines comovements, we document a *cross-period* return predictability of Reddit-linked firms. Especially, our result provides new evidence of market inefficiency in the context of social media.

Finally, the richness of Reddit data allows us to explore linkages in various forms and at diverse granular levels. It is important to note that Twitter is constrained by the volume and bias of its API. For example, the current version of the Twitter API, Twitter API v2, has a monthly cap of 10 million Tweets and a rate cap of 100 requests per 15 minutes. ⁶ Financial tweets on Twitter quickly (within a week) become irretrievable, according to Al Guindy and Riordan (2019). Furthermore, the structure of Reddit enables us to identify Reddit peers at both the shared posts and shared authors levels. Comparing different types of social media linkage helps deepen the understanding of firms' interaction effect in the economy.

The rest of the paper is organized as follows. Section 2 describes the Reddit data and the construction of Reddit linkages. Section 3 examines the fundamental comovement of Reddit-linked firms. Section 4 presents the results on the lead-lag returns relation among Reddit-linked firms. Section 5 provides additional robustness tests. Section 6 concludes.

2 Data and variable construction

Reddit is made up of over 3 million subreddits, each of which is named "r/subreddit_name" and usually focused on a certain field. The subreddit itself is constituted of different posts, each of which focuses on a more specific topic launched by a Reddit member, and the comments in a post are structured by the "reply-to" relation between them. This paper focuses on the r/wallstreetbets subreddit.

2.1 Data source

We collect text-form comments data from the r/wallstreetbets subreddit via the Pushshift.⁷ To find comments that are related to stocks, we first tokenize the body of all comments in r/wallstreetbets and then cross-reference these tokenized texts with the list of tickers to

 $^{^6{\}rm For}$ more details of Twitter API, please check https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api

⁷Pushshift is a social media data collection and archiving platform, and it has collected Reddit data since 2015. Baumgartner, Zannettou, Keegan, Squire, and Blackburn (2020) introduce this dataset in detail.

infer which might be connected to stock tickers. The list of tickers consists of common stocks traded on NYSE/AMEX/NASDAQ. We obtain stock-related data from CRSP and accounting information from COMPUSTAT. We describe our approach in more detail in Appendix A.

Following prior literature, we impose several filters to the Reddit dataset: (1) Drop comments that are deleted and removed by the author or Reddit; (2) Drop comments missing information about the author, the content, or the timestamp when comments are created; (3) Drop comments that do not mention any ticker symbol;⁸ (4) Drop comments with the author being "AutoModerator" or have been deleted;⁹ (5) Keep only one comment for duplicated author-time-content observations. In addition, we exclude GameStop (stock ticker: "GME") from our sample due to its abnormal frequency of mentions on Reddit.¹⁰ Our final sample consists of 4,751,796 text-form observations,¹¹ written by 479,733 Reddit users under 217,775 posts, and ranges from July 1, 2018 to December 31, 2021.

2.2 Definition of Reddit peers

Our primary work is to identify stock linkages in the context of social media. As described previously, a post on Reddit is a collection of comments with a specific topic, where comments are made by Reddit authors (i.e., users). Therefore, we consider two forms of Reddit connection: the Reddit *post* link and the Reddit *author* link. Specifically, we define two stocks as Reddit post (author)-connected if there was at least one post (author) on Reddit mentioning both firms in the past six months. The co-mention of firms in the same post reflects the collective interest of Reddit crowds about the underlying stocks; the co-track of firms by the same author reflects Reddit users' aggregated attention and knowledge about the stocks featured.

We use a screenshot of a real Reddit post to help better understand the construction of Reddit peers, as shown in Figure $2.^{12}$ The original poster expressed its viewpoint on INTC

¹²The complete conversation thread can be seen at https://www.reddit.com/r/wallstreetbets/

⁸The definition of mentioning stock tickers is based on whether the comment contains any string of a ticker and we keep tickers with at least 3 letters to avoid ambiguities in the matching process. For example, if a comment describes "AAPL is soaring", we match it to *Apple Inc.* By contrast, we do not associate a comment like "BA raises its airline capacity" with *Boeing Co.*

⁹ "AutoModerator" refers to the Reddit bot with the objective of maintaining the community order.

 $^{^{10}\}mathrm{Our}$ analysis is not affected by this filter and we get quantitatively similar results if including GameStop in the sample.

¹¹In each post on Reddit, the original submission is initiated by a user about a particular topic (a Level 1 comment). Then users could comment on this submission about this topic (Level 2 comments). They could also comment on existing Level 2 comments (Level 3 comments). The depth of the comments in the post could extend indefinitely. In our study, we do not differentiate between initial submissions and various levels of comments and treat them all as comments.

(Intel Corp.) in the original submission. Another user "u/sockalicious," as captured in Figure 2, mentioned AAPL (Apple Inc.), AMD (Advanced Micro Devices, Inc.), and MSFT (Microsoft Corp.). Therefore, in our construction, we add one to the number of shared Reddit posts to each of the six stock pairs for the corresponding window.¹³

[Place Figure 2 here]

By construction, the Reddit post (author) peer for each focal firm is the set of firms that are also mentioned in the same post (by the same author) in the preceding six months. We identify and update Reddit connections each month. Therefore, our Reddit peer data starts from December 2018 to December 2021.

2.3 Summary statistics

Table 1 presents summary information on the composition of our Reddit comments data. On average, each post in our sample mentions 4.9 stocks, collectively by 21.8 replies from 13.2 Reddit users. Each post (author/comment) in our sample on average mentions 4.9 (4.2/1.2) stocks. Panel B shows that 13.98% of comments, 53.02% of posts, and 50.99% of authors in our sample mention at least two stocks, which helps to illustrate the feasibility of the construction of Reddit linkages. Furthermore, The distribution of users, comments, and mentioned stocks in each post is not uniform and is more like following a power-law trend. The phenomenon of "superstars", discussed in Rosen (1981), seems to be at play similarly in Reddit, where relatively small numbers of authors and opinions earn enormous amounts of attention and dominate the topics in which they engage. Combined with some views that Reddit user Keith Patrick Gill (often referred to as "u/Deep****ingValue" in Reddit) trigger the well-known "GameStop Short Squeeze", we can gain further insight into the structure of Reddit.

[Place Table 1 here]

Table 2 reports summary statistics of Reddit post connections in Panel A, and the Reddit author connection in Panel B. The first three rows of each panel report distributions of the number and size of Reddit-linked stocks across the sample period. The remaining four rows of each panel report the time-series averages of cross-sectional statistics of the corresponding

comments/i45ao7/intc_is_actually_a_pretty_good_deal.

¹³Specifically, the stock connections identified through this post are INTC-AMD, INTC-AAPL, INTC-MSFT, AMD-AAPL, AMD-MSFT, and AAPL-MSFT.

variables. On average, our Reddit post (author) connection sample covers 52.6% (50.8%) of the stock universe in terms of number and 76.9% (76.2%) in terms of market capitalization, indicating that the discussion on the forum frequently focuses more on large firms. We also rank stocks based on size and assign them to deciles using NYSE breakpoints. The average size rank of firms with Reddit connections is 4.6, and the 10, 50, and 90 percentiles are respectively in the first, fourth, and ninth groups. The distribution of size rank is nearly uniform, demonstrating that our sample is not dominated by particularly large or small companies.

Each focal firm connected by a post on average links to 516 firms, and the most central focal firm has 1179 peer firms. Each focal firm connected by an author on average links to 229 firms, while the most central focal firm has 587 peer firms. This demonstrates that Reddit post-connected firms often have more connections than those connected by authors. Each post-connected stock pair has an average of 7 common posts, and each author-connected stock pair has an average of 4 common authors. We also calculate the proportion of Reddit peers in the same industry to set our linkage apart from previous industry links. This calculation reveals that only 9.2% of Reddit post peers and 13.2% of Reddit author peers are in the same industry.

[Place Table 2 here]

Figure 3 shows the time-series average of cross-sectional percentiles of the number of Reddit connections for stock pairs based on shared posts and shared authors, respectively. Following the pattern that we mentioned above, the distribution of shared Reddit posts and shared Reddit authors in the stock-pair level resembles a power-law distribution. While most of the observations are 1 or 2 below the 80th percentile, those that are above the 80th percentile start to show a significant increase. Given the high skewness of the data, we use the logarithm transformation of the number of shared Reddit posts/authors, instead of the raw number, as the weight in calculating Reddit peer firms' variables in later analyses.

[Place Figure 3 here]

3 Fundamental comovement of Reddit-linked firms

We begin by testing whether Reddit-linked firms have correlated fundamentals. We begin by testing whether Reddit-linked firms have correlated fundamentals. Reddit users are often associated with retail investors, gamification, and "YOLO trading" in some views.¹⁴ However, an amount of studies have already discuss how collective judgment, or the "Wisdom of Crowds", are different from personal opinions, and how the coordination among small participants impact the whole market.¹⁵ Studies based on Reddit, like Hu et al. (2021), also shows supporting empirical evidence that variables derived from Reddit activities are informative. And Pedersen (2022) uses a model to predict how social networks influence prices before the information is fully known. These studies support our use of Reddit information to find economic connections between firms.

We consider a broad set of characteristics that capture the profitability, valuation, and growth of a firm: quarterly stock return (QRET), return on assets (ROA), return on equity (ROE), gross profit (GP), earnings-to-price ratio (EP), cash flow-to-price ratio (CP), sales-toprice ratio (SP), sales growth (SG), profit growth (PG), revenue growth (RG), asset growth (AG), leverage ratio (LEV), asset turnover (AT), and R&D expense-to-sales ratio (RDS). Detailed description of the fundamental variables is provided in the Appendix.

3.1 Baseline result

Following Parsons et al. (2020) and Peng, Titman, Yönaç, and Zhou (2022), we conduct the following panel regression:

$$F_{i,j,t} = \beta_1 F_{i,t}^R + \beta_2 F_{j,t}^I + \gamma_t + \varepsilon_{i,j,t}$$

$$\tag{1}$$

in which $F_{i,j,t}$ is a fundamental variable of firm *i* in industry *j* at quarter *t*. The independent variable of interest is the weighted average of Reddit peers' corresponding fundamentals, $F_{i,t}^R$, and the weights are determined by the number of common Reddit posts (authors) shared with focal firm *i*:

$$F_{i,t}^{R} = \frac{1}{\sum_{p=1}^{N} w_{p,t}} \sum_{p=1}^{N} w_{p,t} F_{p,t}, \quad w_{p,t} = \log(1 + \# \text{Reddit connection}).$$
(2)

In regressions, we also control for the average fundamental of industry peers $F_{j,t}^{I}$ and include time fixed effects (γ_t) .

Panel A of Table 3 shows that focal firms comove positively and significantly with their

 $^{^{14}}$ YOLO, stands for "you only live once", is always related to high-risk or lottery-like trading in r/wallstreet bets subreddit. Lyócsa et al. (2022) discuss "YOLO trading" in detail.

¹⁵For example, Chen et al. (2014) discuss how investors' opinions transmission impact the stock market, and Allen, Haas, Nowak, Pirovano, and Tengulov (2022) discuss how investor coordination on social media platforms potential may have fueled a series of short squeezes.

Reddit post peers along seven dimensions, including QRET, ROA, GP, EP, PG, AG, and AT. The economic magnitudes are also significant. For example, a one standard deviation increase in Reddit post peer's quarterly return and ROA is associated with an increase of the focal stock's contemporaneous quarterly return and ROA of about 2.56% and 0.6%, respectively. For fundamental variables such as ROE, CP, SP, SG, RG, LEV, and RDP, we find less evident similarities among Reddit post peers.

The fundamental comovement of Reddit author peers is slightly stronger. Panel B of Table 3 shows that focal firms have significantly similar fundamentals with their Reddit author peers along ten dimensions: QRET, ROA, ROE, GP, EP, CP, SP, AG, LEV, and AT. For fundamental variables such as sales growth (SG), revenue growth (RG), and R&D expense-to-sales ratio (RDS), we find weak comovement among peer firms for both the Reddit post link and the Reddit author link. On average, the economic magnitude of the Reddit author peers' fundamental comovement is larger relative to that of the Reddit post peer.

Overall, the result suggests that the Reddit linkage helps to capture fundamental connectedness between firms, especially in terms of profitability, valuation, and investment.

[Place Table 3 here]

3.2 Linkage refinement and fundamental comovement

Firms that are Reddit-connected do not necessarily imply that they have uniformly correlated fundamentals. For example, a pair of customer-supplier firms might correlate in profitability and sales, while technology-linked firms have similar valuation ratios and investments. Nevertheless, given the fact that the average number of Reddit peer firms is high,¹⁶ it is reasonable to conjecture that firms with more Reddit co-mentions have stronger fundamental comovement. Therefore, we examine a refined version of the Reddit linkage when calculating peer firms' average fundamentals. Specifically, we rank Reddit peers in descending order based on the number of Reddit connections with the focal firm. Then, we keep the closest 50 firms when applying equation (2).

Table 4 presents regression results using the refined Reddit linkage. We find strengthened fundamental comovement of Reddit peer firms. For example, Panel A of Table 4 shows that the estimated coefficient on Reddit post peer's fundamental is significantly positive for cash flow-to-price ratio (CP), sales-to-price ratio (SP), sales growth (SG), and revenue growth

 $^{^{16}\}mathrm{As}$ shown in Table 2, the average number of peer firms is 516 for the Reddit post link and 229 for the Reddit author link.

(RG), whereas those in Table x are insignificant. For Reddit author peers, we find similar improvements. We also note that the comovement in asset growth (AG) and leverage ratio (LEV) is weaker when restricted to the closest 50 Reddit peers. Due to the complexity of the formation of Reddit linkage, it is difficult to use a unified theory to explain the heterogeneity in the variation of fundamental comovement. However, our result strongly confirms that Reddit-linked firms are fundamentally connected, and both close and distant peer firms contain useful information that is related to the focal firm.

[Place Table 4 here]

4 Lead-lag returns relation of Reddit-linked firms

In this section, we examine the asset pricing implication of Reddit linkage. Specifically, we test the cross-period relation of Reddit-linked stocks' returns. Prior studies document significant lead-lag effects among economically connected firms, starting with early work by Cohen and Frazzini (2008). Following this literature, if investors comprehend the Reddit connection and the pricing of Reddit-linked stocks is efficient, then the Reddit peer's price movement should be unrelated to the focal stock's future return. Instead, if investors fail to fully incorporate the information implied by Reddit peers' returns, then we expect to see a momentum spillover effect among Reddit-linked stocks.

4.1 Portfolio analysis

To examine whether returns spillover across Reddit peer firms, we construct a return signal for each type of Reddit link:

$$REDDIT \ RET_{i,d} = \frac{1}{\sum_{p=1}^{N} w_{p,t}} \sum_{p=1}^{N} w_{p,t} Ret_{p,d}, \quad w_{p,t} = \log(1 + \# \text{Reddit connection}), \quad (3)$$

where $Ret_{p,t}$ is stock p's daily return on day t, month t. We use the most recent linkage information to calculate $REDDIT RET_{i,d}$, hence the weight for each Reddit peer $(w_{p,t})$ remains constant throughout month t. Each day, stocks are sorted into decile portfolios based on REDDIT RET. We then calculate value-weighted and equal-weighted returns for each portfolio on the next day. We require stocks to have a share price of at least \$1 and use the average market value in the past five days when calculating value-weighted returns.

Panel A of Table 5 reports the result of the Reddit post link. We find that Reddit post peer's return positively predicts the focal stock's future return. A long-short strategy based on *REDDIT RET* generates a return of 5.8 basis points (t=2.89) per day. The return remains significant after adjusting for the five-factor model (Fama and French, 2015) or the momentum-augmented model (FF6). Panel B presents the result of the Reddit author link, which suggests a relatively weaker lead-lag effect. For example, the equal-weighted strategy based on *REDDIT RET* earns a return of 6.9 pbs on average, whereas the value-weighted return is only 2.4 pbs and insignificant (t=1.25).

[Place Table 5 here]

The portfolio analysis suggests an evident daily momentum spillover effect among Redditlinked stocks. The result that the Reddit post link generates a stronger lead-lag relation than the Reddit author link is potentially ascribed to the fact that the former type of link is associated with more peer firms. As a consequence, investors have to pay more attention in order to incorporate the news underlying Reddit peers' price movement, which leads to a stronger daily momentum spillover effect. We will test this limited attention hypothesis in later analyses.

4.2 Regression analysis

Next, we control for other firm characteristics in regressions and examine the predictive ability of *REDDIT RET*. Since our sample is restricted by data availability and only has a three-year length, we follow Hu et al. (2021) and estimate the following panel regression:

$$Ret_{i,d+1} = \beta_0 + \beta_1 REDDIT RET_{i,d} + \beta_2 Controls_{i,d} + \gamma_d + \varepsilon_{i,d+1}.$$
(4)

The dependent variable $Ret_{i,d+1}$ is focal stock *i*'s return on day d + 1. The main dependent variable of interest is $REDDIT RET_{i,d}$, the one-day lagged Reddit peer return. We include day fixed effects (γ_t) and control for focal stock's past performance, including the return on day *d*, the cumulative return during [t - 20, t - 2], and the cumulative return during [t - 120, t - 21]. We also control for characteristics such as illiquidity, idiosyncratic volatility, log of market capitalization, and log of book-to-market ratio measured at the end of last month. ¹⁷

Table 6 shows the estimated coefficients. Consistent with the portfolio results, we find significant daily momentum spillover effects under the Reddit linkage. Columns (1) and (5) of Table 6 indicate that 4.09% (2.43%) of the prior day's Reddit post (author) peer return

 $^{^{17}{\}rm We}$ cluster standard errors at the firm level. However, our results remain robust under the two-way clustering by both firm and date.

carries over into the following day's return of the focal stock, after controlling for other firm characteristics. We have shown that, on average, the number of Reddit peers is sufficiently large and only about 10% of them are within the same industry as the focal firm. However, one could still worry that the Reddit lead-lag effect is driven by industry momentum. In columns (2) and (6), we control for industry peer return (*IND RET*) and the result shows that the estimated Reddit momentum remains significant.¹⁸

We further include co-analyst peer return ($CF \ RET$) in columns (3)-(4) and columns (7)-(8) to control for the effect from shared analyst coverage (Ali and Hirshleifer, 2020).¹⁹ Notably, while the industry peer return is dominated by $CF \ RET$, the predictive ability of $REDDT \ RET$ remains significant. For example, column (4) of Table 6 shows that the estimated coefficient on Reddit post peer return is 2.660 (t=3.04), whereas the coefficient on industry peer return is only -0.238 and insignificant (t=-0.45). This result suggests that the Reddit linkage helps to capture incremental information beyond industry classification and analyst coverage.

[Place Table 6 here]

4.3 Tests of the limited attention hypothesis

The prevailing explanation for the cross-firm return predictability is the underreaction hypothesis. Specifically, since investors' attention is generally limited, they tend to underreact to peer firms' news. Therefore, the information contained in Reddit peers' price movement would be sluggishly disseminated into the market, which gives rise to the lead-lag returns relation. This section tests the limited attention hypothesis in explaining the lead-lag effect of Reddit-linked firms.

We use analyst coverage and media coverage as attention proxies.²⁰ If the underreaction to the information contained in Reddit peer's return generates the cross-firm predictabil-

 $^{^{18}}$ For each focal stock, the industry peer return is calculated as the value-weighted one-day return of other stocks in the same Fama-French 49 industry classification.

¹⁹We follow the same method of Ali and Hirshleifer (2020) in calculating CF RET. Each month, we define two stocks as connected if at least one analyst issues FY1 or FY2 earnings forecast for both stocks in the past 12 months. For each focal stock, the co-analyst peer return is computed as the average of linked stocks' one-day return, weighted by the number of analysts shared with the focal stock. We use the most recent month's analyst linkage information for calculations.

²⁰For analyst coverage, a stock is covered by an analyst if the analyst issues at least one FY1 or FY earnings forecast for the underlying firm over the past year. The analyst forecast data is obtained through the International Brokers Estimate System (IBES). For media coverage, we use news data from the RavenPack database. Following previous literature, we require an Event Novelty Score (ENS) and a relevance score of 100. Each month, the media coverage of a stock is the number of news stories mentioning the stock.

ity, then the predictive power of *REDDIT RET* should be weaker when investor attention is high. We also consider the effect of institutional ownership.²¹ The prediction is that underreaction-induced mispricing should be mitigated when there are more rational investors. In addition, the predictive ability of *REDDIT RET* would be stronger when the number of Reddit peer firms is high since in that case investors are required more attention to comprehend the linkage information. Throughout all specifications, we control for the interaction effect of size since larger firms tend to attract more attention, have more institutional investor participation, and are connected to more Reddit peers.

Table 7 reports the estimated coefficients of regressions. Panel A presents the result of the Reddit post link and panel B shows the Reddit author link. Consistent with the hypothesis, we find a stronger predictive ability of Reddit peer return when the number of Reddit linkages is high. The interaction of *REDDIT RET* and the dummy indicating a high linkage number is positively significant. The stock return of Reddit peers carries over into the focal stock's future return by 1.83% (2.21%) more if the number of Reddit post (author) linkages is above the medium. When the attention to the focal stock is high, measured by a high level of analyst coverage or media coverage, the estimated effect of *REDDIT RET* is 31%-35% weaker for the Reddit post link and 33%-43% weaker for the Reddit author link. Finally, the estimated coefficient on the interaction term of *REDDIT RET* and *Inst High* is negative, suggesting that the Reddit peer momentum is less evident for firms with more institutional holding.

[Place Table 7 here]

5 Robustness tests

5.1 The information content of Reddit peer returns

We have shown that Reddit-linked firms are fundamentally connected and the Reddit linkage generates cross-firm return predictability. The underlying premise of the momentum spillover effect is that Reddit peers' price movement contains useful information about the focal firm's future performance. In this section, we examine the information content of Reddit peer returns using focal firms' standardized unexpected earnings (SUE), which captures unanticipated changes in cash flow. We calculate SUE as unexpected earnings scaled by the standard deviation of unexpected earnings over the preceding eight quarters. The unexpected

²¹We obtain quarterly data on institutional investor (13F) holding from Thomson-Reuters.

earnings are measured by year-over-year changes in quarterly earnings before extraordinary items.

Table 8 presents the result of regressing the focal firm's SUE on lagged Reddit peer quarterly return (*REDDIT QRET*). Panel A shows that Reddit post peer return has a significant predictive ability for the focal firm's future SUE. In the univariate regression, column (1) shows that a one standard deviation increase in *REDDIT QRET* implies an increase of SUE in the next quarter of 0.023. The estimated coefficient remains significant after controlling for the focal firm's lagged SUE. Panel B reports the result of the Reddit author link, and we again find a strong predictive power of *REDDIT QRET*. This result suggests that the stock return of Reddit peer firms does contain fundamental information about the focal firm.

[Place Table 8 here]

5.2 Returns around earnings and news days

To further inspect the mechanism underlying the cross-firm return predictability of Reddit linkage, we examine stock returns around information release events such as earnings announcements or news stories. For each stock, we first identify the day with an earnings announcement or news release. ²² Following the method of Engelberg, McLean, and Pontiff (2018), we examine the stock's trading volume (scaled by the market trading volume) for a three-day window centered on the event date and define an earnings day (Eday) or news day (Nday) as the day with the highest trading volume. Then, we perform the following regression:

$$Ret_{i,d+1} = \beta_0 + \beta_1 REDDIT RET_{i,d} + \beta_2 REDDIT RET_{i,d} \times Eday_{i,d+1} + \beta_3 REDDIT RET_{i,d} \times Nday_{i,d+1} + \beta_4 Eday_{i,d+1} + \beta_5 Nday_{i,d+1} + \delta Controls + \gamma_d + \varepsilon_{i,d+1}.$$
(5)

As specified, the variable Eday (Nday) is a dummy variable equal to one on earnings (news) days and zero otherwise. The set of control variables includes lagged values for each of the past ten days of stock returns, return squared, and trading volume. If the predictive ability of *REDDIT RET* is driven by the initial underpricing, then we should observe higher anomaly returns around earnings or news days, since the release of information helps for correcting the biased expectation about the focal stock's cash flows.

 $^{^{22}{\}rm The}$ earnings announcement dates are obtained from the COMPUSTAT quarterly database. We obtain news story dates from the RavenPack database.

Table 9 presents our estimation results. We find that subsequent returns on earnings or news days are significantly larger for stocks with higher *REDDIT RET*. Column (1) shows that the effect of Reddit peer returns is about 4.12 times higher on earnings days than on non-earnings days. In column (3), the estimated coefficient on *REDDIT RET* × *Nday* is 3.969, suggesting the effect is about 3.39 times stronger on news days than on non-news days. Overall, we find higher momentum spillover returns on both earnings days and news days, although the statistical significance varies depending on specifications. This result is consistent with the finding of Engelberg et al. (2018) and supports the underreaction story of the lead-lag effect among Reddit-linked firms.

[Place Table 9 here]

5.3 Uncertainty and Reddit peer momentum

As examined in previous sections, the lead-lag effect of Reddit peer firms is likely to be driven by investors' underreaction to peer firms' news, which generates a delayed price reaction of the focal stock. In this section, we examine the time variation of Reddit-linked firms' momentum spillover effect. Specifically, we use the Chicago Board Options Exchange Volatility Index (VIX) to predict the Reddit peer momentum strategy. Since investors are likely to be distracted in periods of enhanced uncertainty, the Reddit momentum effect should be stronger following a high level of VIX.

Table 10 reports the time-series regression results. Column (1) shows that a one standard deviation increase in the last day's VIX implies that the Reddit peer momentum return increases by 6.4 pbs. This effect is robust to controlling for the six-factor model. For the Reddit author link, column (3) shows that the following day's strategy return increases by 5.89 bps when VIX increases by one standard deviation. Although the estimate is only marginal for value-weighted strategies, we still find a positive relationship between uncertainty and the Reddit lead-lag effect.

[Place Table 10 here]

5.4 Long-horizon return prediction

We have thus far focused on predicting one-day ahead returns, and the result suggests a significant daily momentum spillover effect of Reddit-linked firms. In this section, we examine the predictive ability of *REDDIT RET* for long-horizon returns. Specifically, we use the regression specification of equation (4) and replace the dependent variable $Ret_{i,d+1}$ with $Ret_{i,d+n}$.

Table 11 reports the estimated coefficients. We find that peer returns from the Reddit post link can predict the focal stock's long-term returns, up to 40 days. As a comparison, the estimated coefficient on *REDDIT RET* is positive but insignificant for the Reddit author link. This result suggests that Reddit post linkage might be more difficult for investors to analyze and the information conveyed by peer stocks' returns is gradually incorporated into the focal stock's price over time. By contrast, the peer price movement from Reddit author linkage reflects transient variation in mispricing and hence is less persistent in predicting the focal stock's future returns.

[Place Table 11 here]

5.5 Order imbalance and Reddit peer returns

Finally, we test an alternative explanation for the Reddit momentum spillover effect. In addition to the underreaction hypothesis, researchers also document significant lead-lag relations that is driven by *continued overreaction*. For example, Chen, An, Yu, and Wang (2022) present the "attention spillover" effect of stocks with adjacent listing codes in China's stock market; He, Wang, and Yu (2021) show that stocks with similar characteristics exhibit cross-firm return predictability, which is consistent with experience effect. In this context, the predictive ability of Reddit peer return is the result of excess demand for the focal stock. That is, the momentum spillover of Reddit-linked firms reflects the build-up of mispricing.

We examine this channel by testing whether Reddit peer return can predict the order imbalance of the focal stock. If the lead-lag effect of Reddit-linked firms is driven by investors' continued overreaction, then we should observe a significantly positive relationship between Reddit peer return and subsequent excess demand. We calculate daily order imbalances using data from TAQ. The trade imbalance is calculated by the difference in the number of buys and number of sells divided by the total number of buys and sells. Similarly, we also calculate the volume imbalance and the dollar imbalance.²³

²³Specifically, the volume imbalance is shares of buy trades minus shares of sell trades divided by the total volume of buys and sells. The dollar imbalance is the difference in the dollar value of buys and the dollar value of sells divided by the total dollar value of buys and sells. A buyer-initiated or seller-initiated trade is determined based on the Lee and Ready (1991) test. The data on total buys, total sells, retail buys, and retail sells are available through the Wharton Research Data Services (WRDS) Millisecond Intraday Indicators.

Table 12 reports the regression result that uses *REDDIT RET* to predict one-day ahead order imbalance. Across all specifications, we do not detect a significant predictive ability of Reddit peer return for future order imbalances. Although we cannot completely rule out the overreaction hypothesis, this result suggests that the momentum spillover effect from Reddit linkage should not be simply ascribed to the excess demand for the focal stock. Instead, we find supporting evidence in previous analyses that limited attention is more likely to be the source of the cross-firm predictability of Reddit-linked firms.

[Place Table 12 here]

6 Conclusion

Social media platforms are no doubt becoming an increasingly influential venue for the dissemination of opinions and information in today's world. And just like a company's location in geography, or in the industry chain, a firm's location in social media is an implicit representation of the financial network. If two firms are considered peers in the context of social media, then it implies that they would be economically connected. This paper shows that economic linkages can indeed be inferred from the social media platform.

Using data from the Reddit platform, we construct firm-by-firm connections through shared Reddit posts and shared Reddit authors. Our results show that focal firms and their Reddit peers have significant fundamental comovements, suggesting that Reddit-linked firms are economically connected. Then, we document a significant lead-lag relation of Redditconnected firms, which is not likely to be driven by industry linkages or shared analyst coverage. Finally, we provide evidence showing that the momentum spillover effect under Reddit linkages is consistent with the limited attention hypothesis.

Overall, we show that the collective wisdom of Reddit crowds helps to capture comprehensive and dynamic information about financial connections. It turns out that market participants do not fully comprehend such implicit linkage relations, leading to cross-firm return predictability. Given the ongoing attention to social media's role in information propagation in modern society, further research may need to consider using various layers of social media data, such as the author level or the topic level, to provide a heuristic overview of the nested networks of social media.

References

- Al Guindy, M., and R. Riordan. 2019. The Social Internetwork and Stock Returns. Working paper .
- Ali, U., and D. Hirshleifer. 2020. Shared analyst coverage: Unifying momentum spillover effects. Journal of Financial Economics 136:649–675.
- Allen, F., M. Haas, E. Nowak, M. Pirovano, and A. Tengulov. 2022. Squeezing Shorts Through Social Media Platforms. Working paper .
- Antweiler, W., and M. Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. The Journal of finance 59:1259–1294.
- Bailey, M., R. Cao, T. Kuchler, and J. Stroebel. 2018a. The Economic Effects of Social Networks: Evidence from the Housing Market. *Journal of Political Economy* 126:2224–2276.
- Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong. 2018b. Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives* 32:259–280.
- Bakshy, E., J. M. Hofman, W. A. Mason, and D. J. Watts. 2011. Everyone's an influencer: quantifying influence on twitter p. 65.
- Bartov, E., L. Faurel, and P. S. Mohanram. 2018. Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review* 93:25–57.
- Baumgartner, J., S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. 2020. The Pushshift Reddit Dataset. Available at arXiv:2001.08435.
- Cao, J., T. Chordia, and C. Lin. 2016. Alliances and return predictability. Journal of Financial and Quantitative Analysis 51:1689–1717.
- Chen, H., P. De, Y. J. Hu, and B.-H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27:1367–1403.
- Chen, X., L. An, J. Yu, and Z. Wang. 2022. Attention spillover in asset pricing. *Journal of Finance, forthcoming*.
- Cohen, L., and A. Frazzini. 2008. Economic links and predictable returns. *The Journal of Finance* 63:1977–2011.
- Cohen, L., and D. Lou. 2012. Complicated firms. Journal of financial economics 104:383-400.
- Cookson, J. A., and M. Niessner. 2020. Why don't we agree? Evidence from a social network of investors. The Journal of Finance 75:173–228.
- Das, S. R., and M. Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management science 53:1375–1388.
- Diangson, B., and N. Jung. 2021. Bet it on Reddit: The Effects of Reddit Chatter on Highly Shorted Stocks. Working paper .
- Dim, C. 2022. Should Retail Investors Listen to Social Media Analysts? Evidence from Text-Implied Beliefs. Available at SSRN 3813252.
- Eisdorfer, A., K. Froot, G. Ozik, and R. Sadka. 2022. Competition links and stock returns. The Review of Financial Studies 35:4300–4340.
- Engelberg, J., R. D. McLean, and J. Pontiff. 2018. Anomalies and news. The Journal of Finance 73:1971– 2001.
- Fama, E. F., and K. R. French. 2015. A five-factor asset pricing model. Journal of financial economics 116:1–22.

He, W., Y. Wang, and J. Yu. 2021. Similar stocks. Available at SSRN 3815595.

- Hosseini, A., G. Jostova, A. Philipov, and R. Savickas. 2020. The Social Media Risk Premium. Available at SSRN 3514826.
- Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. The Review of Financial Studies 33:2019–2133.
- Hu, D., C. M. Jones, V. Zhang, and X. Zhang. 2021. The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery. *Available at SSRN 3807655*.
- Kuchler, T., and J. Stroebel. 2021. Social Finance. Annual Review of Financial Economics 13:37–55.
- Lee, C. M., P. Ma, and C. C. Wang. 2015. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics* 116:410–431.
- Lee, C. M., and M. J. Ready. 1991. Inferring trade direction from intraday data. The Journal of Finance 46:733-746.
- Lee, C. M., S. T. Sun, R. Wang, and R. Zhang. 2019. Technological links and predictable returns. Journal of Financial Economics 132:76–96.
- Li, F. 2022. Retail Trading and Asset Prices: The Role of Changing Social Dynamics. Available at SSRN 4236966.
- Long, C., B. M. Lucey, and L. Yarovaya. 2021. 'I Just Like the Stock'versus' Fear and Loathing on Main Street': The Role of Reddit Sentiment in the GameStop Short Squeeze. Available at SSRN 3822315.
- Lyócsa, S., E. Baumöhl, and T. Výrost. 2022. YOLO trading: Riding with the herd during the GameStop episode. *Finance Research Letters* 46:102359.
- McGurk, Z., A. Nowak, and J. C. Hall. 2020. Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics and Finance* 44:458–485.
- Menzly, L., and O. Ozbas. 2010. Market segmentation and cross-predictability of returns. The Journal of Finance 65:1555–1580.
- Parsons, C. A., R. Sabbatucci, and S. Titman. 2020. Geographic lead-lag effects. The Review of Financial Studies 33:4721–4770.
- Pedersen, L. H. 2022. Game On: Social Networks and Markets. Forthcoming in Journal of Financial Economics p. 54.
- Peng, L., S. Titman, M. Yönaç, and D. Zhou. 2022. Social Ties, Comovements, and Predictable Returns. Working paper.
- Rakowski, D., S. E. Shirley, and J. R. Stark. 2021. Twitter activity, investor attention, and the diffusion of information. *Financial Management* 50:3–46.
- Turmakin, R., and R. Whitelaw. 2001. News or Noise? Internet Message Board Activity and Stock Prices. Financial Analyst Journal 573:41–51.
- Vasileiou, E., E. Bartzou, and P. Tzanakis. 2021. Explaining Gamestop Short Squeeze using Intraday Data and Google Searches. Available at SSRN 3805630.
- Wang, M. 2022. Comment Display Matters: Evidence from Reddit's WallStreetBets. Working paper .



Figure 1. Structure of Reddit.

This figure depicts the brief structure of Reddit. Reddit is a social media aggregation containing over three million communities discussing a variety of topics. Each community is made up by diverse posts made by different users. And in each post, the comments are structured by the reply-to relation between them.

pt	Posted by u/iBifteki 2 years ago
76	INTC is actually a pretty good deal
P.	Discussion
	х У
	💭 77 Comments 🧪 Share 🔲 Save 🕲 Hide 🏳 Report
	Sort By: Best 👻
	View all comments View discussions in 0 other communities
Ô	sockalicious 🤯 +2 · 2 yr. ago · edited 2 yr. ago 🎭
Ī	INTC still has dominant market share in both consumer and server markets
	This isn't actually a forward looking bull pitch. Once you have dominant market
	share, you can lose market share - say, half your market share - but you cannot
	aspire to gain much market share. If you have over 50% of the market you cannot double your percent market share, for instance - it's mathematically impossible.
	This pitch: "Oh, the competitor beats us on price and quality, but they're too small

 \triangle

This pitch: "Oh, the competitor beats us on price and quality, but they're too small to threaten us" has sunk many ships. Anyone who thinks AMD and ARM don't want the business, or can't ramp production timely enough, is wearing blinders.

There's a difference between being 🌈 🐯 and being able to see a broken company.

EDIT: Let's go into detail. I have 40 or so friends who worked at AAPL, made a 20+ year career there. Interestingly in the last 5 years most have left. Some retired and now tool around Atherton in their McLarens with wife #3. But let's look at INTC from an AAPL perspective. INTC once owned processor. They put Motorola, a great company, out of business about 5 years after Apple switched from PowerPC to x86 for their laptops. Intel has always been thick as thieves with MSFT and Apple had to spend years at the negotiating table getting Intel to unlock the full power of their chips for an EFI based architecture instead of legacy Win 3.1 compatible DOS-baed bios. I swear INTC was optimizing their chips for CP/M well into the 90s.

I digress. Apple came to x86 but about 5 years ago hired Chris Lattner to finish building llvm which is now the nasty nuts and bolts basis of Xcode, the free compiler for AAPL developers. Only once that project was done and Lattner set adrift with his \$20m in options was AAPL free to announce that, with their newly acquired platform-independence, they'd be moving off x86 to ARM. ARM is a successful new power-friendly architecture that AAPL had a large hand in developing over the past 15 years as it has always powered the iPhone, and they have assessed (correctly imo) that the ARM platform is more flexible and friendly for the computing goals of AAPL's next 10 years of product pipeline. Remember what happened the last time AAPL switched platforms for its big iron? You still holding shares in Motorola Mobility?

Figure 2. Example of Reddit r/wall street bets comments.

This figure is a screenshot of selected comments from the thread https://www.reddit.com/ r/wallstreetbets/comments/i45ao7/intc_is_actually_a_pretty_good_deal on Reddit. The original poster talked about Intel Corp. in the original submission, the captured part of comments shown in Figure 2 also mentioned Advanced Micro Devices, Inc., Apple Inc., and Microsoft Corp. We marked these ticker symbols with red boxes in this screenshot.



Figure 3. Distribution of shared Reddit posts and shared Reddit authors.

These two graphs plot the time-series averages of cross-sectional percentiles of the number of shared Reddit posts and the number of shared Reddit authors, measured at the stock-pair level. We do not show the maximum value in graphs because of existence of outliers. For example, the last point in the left-hand-side graph shows the monthly time series average of the cross-section 99th percentile of the number of common posts for each stock pair. The sample period is from December 2018 to December 2021.

Table 1. Summary statistics of the Reddit data.

This table reports summary statistics of the pooled Reddit observation. Our list of stock tickers include common shares in NYSE/AMEX/NASDAQ. Comments in our sample are related to at least one stock ticker. Panel A reports the statistics of the number of authors and comments in each post and the stocks mentioned in each post, author, and comment. Panel B reports the percentage of posts, authors, comments of the whole sample mentioning at least two stocks. The sample is from July 1, 2018 to December 31, 2021.

Panel A. Reddit and stock mentions								
	Ν	Mean	Std .Dev	Min	P10	Median	P90	Max
# of authors each post	217775	13.2	130.9	0	1	2	10	9369
# of comments each post	217775	21.8	254.1	1	1	2	12	19096
# of stocks mentioned each post	217775	4.9	21.8	1	1	2	7	843
# of stocks mentioned each author	479733	4.2	13.5	1	1	2	8	1968
# of stocks mentioned each comment	4642481	1.2	3.2	1	1	1	2	437
Panel B. Co-mention on Reddit								
	Percentage							
Posts mentioning at least two stocks	53.0							
Authors mentioning at least two stocks	51.0							
Comments mentioning at least two stocks	14.0							

Table 2. Summary statistics of the Reddit linkage.

This table presents summary statistics of the Reddit linkage. Our list of stock tickers include common shares in NYSE/AMEX/NASDAQ. Comments in our sample are related to at least one stock ticker. Each month, two stocks are defined as Reddit-linked if there was at least one post (Panel A) or author (Panel B) on Reddit mentioning both stocks in the past six months. In each Panel, the first three rows report monthly time-series statistics of the respective variables across the sample period. The remaining four rows in each Panel report the time-series averages of cross-sectional statistics of the respective variables. Stocks are ranked into ten groups based on firm size monthly, and size rank of firms with Reddit linkages represents the ranked group number of corresponding firms. The sample period is December 2018 to December 2021.

Panel A. Reddit post linkage								
	Ν	Mean	Std.Dev	Min	P10	Median	P90	Max
# of stocks with Reddit linkage	37	1999	517	1329	1377	2194	2659	2736
% of total number of stocks covered	37	52.6	11.8	36.9	38.2	54	66.4	67.9
% of total market capitalization covered	37	76.9	4.2	70.1	71.5	79	81.1	81.4
Size rank of firms with Reddit linkage	73953	4.6	3	1	1	4.4	9.1	10
# of Reddit linkage for each focal firm	73953	516	429	1	102	383	1179	1873
# of common posts for each stock pair	21552741	7	19.4	1	1	2	14	1442
% of Reddit peers in the same industry	73953	9.2	11.7	0	0.2	3.5	24.1	99.2
Panel B. Reddit author linkage								
Panel B. Reddit author linkage	N	Mean	Std.Dev	Min	P10	Median	P90	Max
Panel B. Reddit author linkage # of stocks with Reddit linkage	N 37	Mean 1933	Std.Dev 528.9	Min 1271	P10 1302	Median 2129	P90 2581	Max 2656
Panel B. Reddit author linkage # of stocks with Reddit linkage % of total number of stocks covered	N 37 37	Mean 1933 50.8	Std.Dev 528.9 12	Min 1271 35.4	P10 1302 36	Median 2129 55.2	P90 2581 63.7	Max 2656 65.4
Panel B. Reddit author linkage # of stocks with Reddit linkage % of total number of stocks covered % of total market capitalization covered	N 37 37 37 37	Mean 1933 50.8 76.2	Std.Dev 528.9 12 4.4	Min 1271 35.4 69.6	P10 1302 36 70.2	Median 2129 55.2 78.6	P90 2581 63.7 80.7	Max 2656 65.4 80.9
Panel B. Reddit author linkage # of stocks with Reddit linkage % of total number of stocks covered % of total market capitalization covered Size rank of firms with Reddit linkage	N 37 37 37 71528	Mean 1933 50.8 76.2 4.7	Std.Dev 528.9 12 4.4 3	Min 1271 35.4 69.6 1	P10 1302 36 70.2 1	Median 2129 55.2 78.6 4.4	P90 2581 63.7 80.7 9.3	Max 2656 65.4 80.9 10
Panel B. Reddit author linkage # of stocks with Reddit linkage % of total number of stocks covered % of total market capitalization covered Size rank of firms with Reddit linkage # of Reddit linkage for each focal firm	N 37 37 37 71528 71528	Mean 1933 50.8 76.2 4.7 229	Std.Dev 528.9 12 4.4 3 264.4	Min 1271 35.4 69.6 1 1	P10 1302 36 70.2 1 6	Median 2129 55.2 78.6 4.4 106	P90 2581 63.7 80.7 9.3 587	Max 2656 65.4 80.9 10 1537
Panel B. Reddit author linkage # of stocks with Reddit linkage % of total number of stocks covered % of total market capitalization covered Size rank of firms with Reddit linkage # of Reddit linkage for each focal firm # of common posts for each stock pair	N 37 37 37 71528 71528 71528 9032945	Mean 1933 50.8 76.2 4.7 229 4	Std.Dev 528.9 12 4.4 3 264.4 23.4	Min 1271 35.4 69.6 1 1 1	P10 1302 36 70.2 1 6 1	Median 2129 55.2 78.6 4.4 106 1	P90 2581 63.7 80.7 9.3 587 7	Max 2656 65.4 80.9 10 1537 2823

Table 3. Fundamental comovement.

This table reports fundamental comovements between the focal firm and its Reddit peers. Each month, two stocks are defined as Reddit-linked if there was at least one post (Panel A) or author (Panel B) on Reddit mentioning both stocks in the past six months. The independent variables are focal firms' quarterly fundamentals, including quarterly stock return (QRET), return on assets (ROA), return on equity (ROE), gross profit (GP), earnings-to-price ratio (EP), cash flow-to-price ratio (CP), sales-to-price ratio (SP), sales growth (SG), profit growth (PG), revenue growth (RG), asset growth (AG), leverage ratio (LEV), asset turnover (AT), and R&D expense-to-sales ratio (RDS). The independent variable of interest is the corresponding fundamentals of Reddit-connected firms. We control for average fundamentals of industry peers and include time fixed effects. All independent variables are winsorized at 1% and 99% each cross-section and standardized to have zero mean and unit variance. Standard errors are clustered by both time and firm. *t*-statistics are reported in parentheses. The sample period is from Q4 2018 to Q4 2021.

	Panel A	. Reddit pe	er: post			Panel B	. Reddit pe	eer: author		
	Reddit	Industry	Time FE	Ν	$\mathrm{Adj.}R^2$	Reddit	Industry	Time FE	Ν	$\mathrm{Adj.}R^2$
QRET	2.563	5.548	Yes	18091	0.28	1.882	5.706	Yes	17140	0.28
	(3.47)	(6.66)				(3.36)	(6.59)			
ROA	0.597	1.166	Yes	15530	0.03	0.612	1.140	Yes	14621	0.03
	(3.80)	(5.06)				(4.81)	(5.11)			
ROE	0.531	1.078	Yes	14432	0.01	0.809	0.947	Yes	13556	0.01
	(1.65)	(2.04)				(2.56)	(1.98)			
GP	0.246	2.409	Yes	14756	0.13	0.393	2.422	Yes	13917	0.13
	(3.09)	(16.15)				(4.86)	(16.18)			
\mathbf{EP}	0.507	1.848	Yes	15579	0.06	0.408	1.809	Yes	14672	0.06
	(5.95)	(4.44)				(4.98)	(4.21)			
CP	0.063	1.171	Yes	11186	0.07	0.119	1.174	Yes	10474	0.07
	(1.05)	(4.98)				(2.19)	(4.92)			
SP	0.616	14.895	Yes	15062	0.09	1.178	14.859	Yes	14187	0.09
	(1.00)	(7.16)				(2.43)	(7.21)			
SG	2.425	16.311	Yes	13868	0.04	0.239	16.363	Yes	13029	0.04
	(1.10)	(2.50)				(0.23)	(2.42)			
\mathbf{PG}	0.055	0.773	Yes	13924	0.06	0.039	0.778	Yes	13079	0.06
	(2.21)	(5.52)				(1.03)	(5.74)			
RG	2.804	17.705	Yes	13146	0.04	0.100	17.792	Yes	12373	0.04
	(1.16)	(2.39)				(0.08)	(2.32)			
AG	2.696	8.338	Yes	14379	0.01	4.305	8.810	Yes	13507	0.01
	(2.09)	(2.74)				(2.86)	(2.80)			
LEV	4.493	18.624	Yes	14454	0.00	4.773	19.192	Yes	13583	0.00
	(1.25)	(3.68)				(2.53)	(3.64)			
AT	0.423	6.941	Yes	15530	0.19	0.465	6.949	Yes	14621	0.18
	(3.06)	(17.28)				(2.73)	(16.84)			
RDP	9.720	72.557	Yes	15060	0.02	4.336	72.614	Yes	14184	0.02
	(1.51)	(3.79)				(1.19)	(3.71)			

Table 4. Fundamental comovement: refined linkage

This table reports fundamental comovements between the focal firm and its Reddit peers. Each month, two stocks are defined as Reddit-linked if there was at least one post (Panel A) or author (Panel B) on Reddit mentioning both stocks in the past six months. The independent variables are focal firms' quarterly fundamentals, defined identically as in Table 3. The independent variable of interest is the corresponding fundamentals of the closest 50 Reddit-connected firms. We control for average fundamentals of industry peers and include time fixed effects. All independent variables are winsorized at 1% and 99% each cross-section and standardized to have zero mean and unit variance. Standard errors are clustered by both time and firm. t-statistics are reported in parentheses. The sample period is from Q4 2018 to Q4 2021.

	Panel A	. Reddit pe	eer: post			Panel B	. Reddit pe	eer: author		
	Reddit	Industry	Time FE	Ν	$\mathrm{Adj.}R^2$	Reddit	Industry	Time FE	Ν	$\mathrm{Adj.}R^2$
QRET	3.005	5.542	Yes	18091	0.28	2.793	5.680	Yes	17140	0.28
	(3.69)	(6.60)				(3.32)	(6.62)			
ROA	0.730	1.172	Yes	15530	0.04	0.716	1.130	Yes	14621	0.04
	(4.34)	(5.14)				(5.38)	(5.06)			
ROE	0.678	1.089	Yes	14432	0.01	0.988	0.934	Yes	13556	0.01
	(2.05)	(2.07)				(2.98)	(1.96)			
GP	0.257	2.410	Yes	14756	0.13	0.442	2.418	Yes	13917	0.13
	(3.59)	(16.15)				(4.91)	(16.15)			
\mathbf{EP}	0.558	1.850	Yes	15579	0.06	0.428	1.805	Yes	14672	0.06
	(4.16)	(4.49)				(5.92)	(4.21)			
CP	0.177	1.165	Yes	11186	0.07	0.197	1.170	Yes	10474	0.07
	(3.15)	(4.96)				(3.49)	(4.90)			
SP	1.877	14.863	Yes	15062	0.09	1.875	14.810	Yes	14187	0.09
	(3.72)	(7.13)				(3.90)	(7.18)			
\mathbf{SG}	5.309	16.275	Yes	13868	0.04	2.335	16.323	Yes	13029	0.04
	(3.65)	(2.52)				(2.16)	(2.42)			
\mathbf{PG}	0.073	0.771	Yes	13924	0.06	0.054	0.777	Yes	13079	0.06
	(3.42)	(5.52)				(1.32)	(5.74)			
RG	5.833	17.673	Yes	13146	0.04	2.423	17.750	Yes	12373	0.04
	(3.67)	(2.41)				(1.88)	(2.31)			
AG	1.051	8.321	Yes	14379	0.01	2.663	8.759	Yes	13507	0.01
	(1.27)	(2.75)				(1.82)	(2.79)			
LEV	-2.779	18.584	Yes	14454	0.00	3.066	19.263	Yes	13583	0.00
	(-0.86)	(3.68)				(1.30)	(3.64)			
AT	0.505	6.941	Yes	15530	0.19	0.577	6.945	Yes	14621	0.19
	(3.71)	(17.29)				(3.43)	(16.85)			
RDP	15.792	72.513	Yes	15060	0.02	1.132	72.723	Yes	14184	0.02
	(1.76)	(3.78)				(0.21)	(3.72)			

Table 5. One-sort portfolios performance by Reddit-peer returns.

This table presents the performance of portfolios based on Reddit peer firms' returns. Reddit peers are identified by shared Reddit post (Panel A) or shared Reddit author (Panel B). Each day, stocks are sorted into decile portfolios based on *REDDIT RET*, the weighted average one-day return of stocks that are linked through Reddit. Value-weighted and equal-weighted returns with a one-day holding period are calculated for each portfolio. The sample includes common stocks listed on NYSE, AMEX, and NASDAQ with a share price of at least \$1 at of portfolio formations. The table reports average excess returns (Return) and alphas using the Fama-French five-factor model (FF5) and the momentum-augmented factor model (FF6). *t*-statistics with Newey-West adjustment are reported in parentheses. The sample period is from January 2019 to December 2021.

	1	2	3	4	5	6	7	8	9	10	10-1
Panel A	. Reddit j	post peers	8								
					Va	lue-weigh	ted				
Return	7.02	9.14	7.98	7.57	7.86	7.60	10.42	11.74	11.95	12.88	5.86
	(1.34)	(1.85)	(1.65)	(1.76)	(1.81)	(1.68)	(2.17)	(2.55)	(2.32)	(2.53)	(2.89)
FF5	-1.89	-0.08	-1.38	-1.84	-1.27	-1.97	0.18	2.39	2.41	3.73	5.61
	(-1.11)	(-0.05)	(-0.91)	(-1.43)	(-0.95)	(-1.49)	(0.12)	(2.02)	(1.54)	(2.31)	(2.73)
FF6	-1.75	-0.18	-1.43	-1.91	-1.32	-1.96	0.24	2.33	2.39	3.80	5.55
	(-1.07)	(-0.11)	(-0.95)	(-1.50)	(-1.00)	(-1.46)	(0.16)	(1.98)	(1.54)	(2.38)	(2.74)
					Eq	ual-weigh	ted				
Return	7.33	9.77	8.83	9.95	8.76	9.70	11.89	13.60	12.48	13.91	6.58
	(1.07)	(1.44)	(1.26)	(1.50)	(1.25)	(1.45)	(1.71)	(2.04)	(1.80)	(2.07)	(2.45)
FF5	-0.99	1.19	0.25	1.00	-0.15	0.71	3.11	4.89	4.07	5.87	6.87
	(-0.46)	(0.59)	(0.12)	(0.58)	(-0.08)	(0.38)	(1.48)	(2.57)	(1.91)	(2.53)	(2.53)
FF6	-1.30	0.78	-0.11	0.72	-0.39	0.41	2.85	4.59	3.78	5.54	6.85
	(-0.60)	(0.40)	(-0.05)	(0.43)	(-0.21)	(0.23)	(1.38)	(2.49)	(1.82)	(2.44)	(2.52)
Panel B	. Reddit a	author pe	ers								
					Va	lue-weigh	ted				
Return	7.42	6.23	9.21	8.88	9.09	8.48	9.74	12.11	11.00	9.86	2.44
	(1.34)	(1.19)	(1.96)	(1.93)	(2.05)	(1.85)	(2.05)	(2.52)	(2.22)	(1.94)	(1.25)
FF5	-1.81	-2.96	-0.28	-0.41	-0.43	-1.19	-0.45	2.28	1.75	1.04	2.85
	(-1.05)	(-1.75)	(-0.16)	(-0.30)	(-0.29)	(-0.87)	(-0.34)	(1.42)	(1.00)	(0.72)	(1.43)
FF6	-1.85	-2.87	-0.13	-0.42	-0.46	-1.21	-0.42	2.30	1.80	1.00	2.84
	(-1.10)	(-1.72)	(-0.08)	(-0.31)	(-0.31)	(-0.90)	(-0.31)	(1.43)	(1.03)	(0.69)	(1.46)
					Eq	ual-weigh	ted				
Return	7.82	6.54	7.78	10.54	11.27	12.57	13.17	12.16	11.19	14.75	6.93
	(1.11)	(0.96)	(1.13)	(1.53)	(1.69)	(1.86)	(1.96)	(1.78)	(1.64)	(2.18)	(3.11)
FF5	-0.41	-2.03	-0.88	2.18	2.10	3.71	4.43	3.39	2.70	6.39	6.81
	(-0.20)	(-1.03)	(-0.46)	(1.10)	(1.17)	(2.09)	(2.23)	(1.75)	(1.24)	(2.86)	(3.04)
FF6	-0.71	-2.30	-1.04	2.00	1.84	3.31	4.09	3.11	2.34	5.89	6.60
	(-0.35)	(-1.16)	(-0.55)	(1.01)	(1.05)	(1.92)	(2.13)	(1.65)	(1.12)	(2.81)	(3.01)

Table 6. Momentum spillover among Reddit-linked firms.

This table presents the estimated regression coefficients predicting one-day ahead returns using Reddit peer returns (REDDIT RET). For each focal stock. the Reddit peer return is the weighted average one-day return of stocks that are linked through Reddit post (Panel A) or Reddit author (Panel B). Control variables include the focal stock's one-day return, one-month return (skipping the most recent day), cumulative 100-day return (skipping the most recent month), as well as illiquidity, idiosyncratic volatility, log of market capitalization, and log of book-to-market ratio measured at the end of last month. We further control for peer returns from other linkage relationships such as industry (INDFF49 RET) and shared-analyst coverage (CF RET) of Ali and Hirshleifer (2020). Day fixed effects are included in regressions and t-statistics with standard errors clustered at the firm level are reported in parentheses. The sample period is from January 2019 to December 2021.

	Par	el A. Red	dit post pe	ers		Pane	el B. Redd	it author p	eers
	(1)	(2)	(3)	(4)	_	(5)	(6)	(7)	(8)
REDDIT RET	4.090	3.893	2.607	2.660		2.431	2.355	1.396	1.460
	(4.04)	(3.84)	(2.99)	(3.04)		(4.35)	(4.22)	(2.65)	(2.77)
INDFF49 RET		3.404	-0.552	-0.238			3.345	-0.405	-0.108
		(6.96)	(-1.06)	(-0.45)			(6.73)	(-0.77)	(-0.20)
CF RET			8.414	10.618				8.138	10.227
			(12.17)	(13.56)				(11.62)	(12.87)
Controls	Yes	Yes	No	Yes		Yes	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Intercept	0.219	0.210	0.112	0.101		0.231	0.221	0.114	0.112
	(5.42)	(5.19)	(32.48)	(2.66)		(5.69)	(5.42)	(33.68)	(2.91)
Observations	1296870	1296865	1184149	1184149		1257943	1257937	1148186	1148186
Adjusted R-squared	0.14	0.14	0.18	0.18		0.14	0.14	0.18	0.18

hypothesis.
attention
limited
of the
7. Tests
Table 7

(Controls) are defined identically as in Table 6, except that we replace the focal firm's log market firm value with the dummy variable This table reports the results of regressions that explore the limited attention hypothesis of Reddit peers' momentum spillover effect. The RET) and its interaction terms. We consider interactions using the number of Reddit peers (Link), analyst coverage (Analyst), media coverage (Media), and institutional ownership (Inst). In regressions, we include dummy variables that are equal to 1 if the underlying interaction variable is above the cross-sectional median and 0 otherwise. Reddit peers are identified by shared Reddit posts in Panel A and shared Reddit authors in Panel B. In all specifications, we control for the interaction effect from firm size. Other control variables dependent variable is the focal stock's one-day ahead return. The main dependent variables of interest are Reddit peer return (REDDIT Size High. We include time-fixed effects and t-statistics with standard errors clustered at the firm level are reported in parentheses. The sample period is from January 2019 to December 2021.

	(1)	(2)	(3)	(4)	(2)	(9)	(2)	(8)
	Paı	nel A. Red	dit post pe	ers	Pane	el B. Redd	it author p	eers
REDDIT RET	7.004	7.160	7.487	8.412	4.676	4.803	5.063	5.576
	(6.61)	(6.77)	(6.99)	(7.68)	(7.42)	(7.58)	(7.84)	(8.33)
REDDIT RET×Link High	1.834 (3.00)				2.212 (3.75)			
REDDIT RET×Analyst High	r	-2.236			r.	-1.619		
		(-3.09)				(-2.49)		
REDDIT RET×Media High			-2.639				-2.193	
			(-4.37)				(-4.02)	
REDDIT RET×Inst High				-4.440				-3.303
				(-7.53)				(-6.14)
REDDIT RET×Size High	-8.546	-6.810	-7.171	-6.654	-6.940	-5.758	-5.879	-5.489
	(-13.90)	(-9.10)	(-11.67)	(-11.22)	(-13.09)	(-8.74)	(-10.64)	(-10.21)
Contorls	Yes	Yes	Yes	$\mathbf{Y}_{\mathbf{es}}$	Yes	Yes	Yes	Yes
Time FE	\mathbf{Yes}	\mathbf{Yes}	\mathbf{Yes}	$\mathbf{Y}_{\mathbf{es}}$	\mathbf{Yes}	\mathbf{Yes}	\mathbf{Yes}	Yes
Intercept	0.116	0.113	0.106	0.107	0.124	0.117	0.111	0.112
	(10.99)	(10.71)	(10.12)	(9.53)	(11.28)	(10.87)	(10.31)	(9.83)
Observations	1296870	1296870	1296870	1296870	1257943	1257943	1257943	1257943
Adjusted R-squared	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

Table 8. Reddit peer return and future earnings surprise.

This table presents regressions of the next quarter's standardized unexpected earnings (SUE) of Reddit peer's quarterly return (REDDIT QRET). Reddit peers are identified by shared Reddit post in Panel A and shared Reddit author in Panel B. We include firm fixed effect and time fixed effect and control for the focal firm's own lagged SUEs. The sample period is from December 2018 to December 2021.

	(1)	(2)	(3)	(4)
	Panel A. Re	ddit post peers	Panel B. Redd	it author peers
REDDIT QRET	0.023	0.025	0.036	0.049
	(2.36)	(2.67)	(2.92)	(5.03)
SUE(t-1)		0.143		0.143
		(3.40)		(3.39)
SUE(t-2)		0.062		0.062
		(1.66)		(1.65)
SUE(t-3)		-0.017		-0.014
		(-0.44)		(-0.36)
SUE(t-4)		-0.605		-0.604
		(-15.19)		(-15.10)
Firm FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Observations	19137	18037	18407	17341
Adjusted R-squared	0.16	0.28	0.16	0.28

Table 9. Returns around information days.

This table presents results from regressions of one-day ahead returns on lagged Reddit peer returns (REDDIT RET), the earnings day and news day dummy variables, and interaction terms. The dependent variable is multiplied by 100. The control variables include lagged values for each of the past ten days of stock returns, return squared, and trading volume. The Reddit peer return is the weighted average one-day return of stocks that are linked through Reddit post (Panel A) or Reddit author (Panel B). Eday (Nday) is a dummy variable equal to one on earnings (news) days and zero otherwise. Following Engelberg et al. (2018), the earnings (news) day is defined as the day with the highest trading volume around the three-day window centered on an earnings announcement (news release). Regressions include day fixed effects. Standard errors are clustered on time and t-statistics are in parentheses. The sample period is from January 2019 to December 2021.

	(1)	(2)	(3)	(4)
	Panel A. Red	ldit post peers	Panel B. Reddi	t author peers
REDDIT RET	3.588	4.015	1.658	1.791
	(2.61)	(2.83)	(2.62)	(2.86)
REDDIT RET \times Eday	11.190	11.422	6.663	6.813
	(2.20)	(2.24)	(1.70)	(1.73)
REDDIT RET×Nday	3.490	3.588	3.969	4.100
	(1.67)	(1.72)	(2.20)	(2.28)
Eday	0.199	0.198	0.242	0.241
	(2.45)	(2.43)	(2.97)	(2.96)
Nday	0.530	0.535	0.522	0.528
	(19.61)	(19.51)	(19.47)	(19.41)
Controls	No	Yes	No	Yes
Day FE	Yes	Yes	Yes	Yes
Observations	1453439	1441060	1407436	1395411
Adjusted R-squared	0.13	0.14	0.14	0.14

Table 10. Uncertainty and Reddit peer momentum.

This table presents predictive regressions on Reddit peer momentum on lagged VIX. The Reddit peer momentum is measured by the equal-weighted strategy return (Panel A) and value-weighted strategy return (Panel B) as implemented in Table 5. Reddit peers are identified using shared Reddit post or shared Reddit author. VIX is standardized to have zero mean and unit variance. Control variables include factor returns of the market excess return, SMB, HML, CMA, RMW, and Momentum (Fama and French, 2015). *t*-statistics with Newey-West adjusted standard errors are reported in parentheses. The sample period is January 2019 to December 2021.

	Pane	l A. Equal-v	veighted por	rtfolio	Pane	l B. Value-v	veighted por	tfolio
	(1) Reddi	(2) t post	(3) Reddit	(4) author	(5) Reddi	(6) t post	(7) Reddit	(8) author
Lagged VIX	6.403 (3.03)	6.627 (3.19)	5.887 (2.27)	5.708 (2.26)	3.250 (1.50)	3.058 (1.36)	4.648 (1.48)	5.124 (1.75)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Intercept	6.654	6.993	6.924	6.648	5.828	5.513	2.400	2.839
	(2.55)	(2.65)	(3.19)	(3.06)	(2.92)	(2.75)	(1.26)	(1.49)
Observations	755	755	755	755	755	755	755	755

Table 11. Long-term return prediction.

This table presents the estimated regression coefficients predicting future cumulative returns during t+n using Reddit peer returns. For each focal stock, the Reddit peer return is the weighted average one-day return of stocks that are linked through Reddit post (Panel A) or Reddit author (Panel B). Day fixed effects are included in regressions and t-statistics with standard errors clustered at the firm level are reported in parentheses. The sample period is from January 2019 to December 2021.

Panel A. Reddit post peers										
	t+5	t+10	t+20	t+30	t+40	t+60				
	(1)	(2)	(3)	(4)	(5)	(6)				
REDDIT RET	5.913	8.937	11.280	12.071	11.305	9.160				
	(2.81)	(2.89)	(3.16)	(2.75)	(2.12)	(1.33)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes				
Time FE	Yes	Yes	Yes	Yes	Yes	Yes				
Intercept	0.978	1.777	3.147	4.305	5.450	8.369				
	(4.40)	(4.05)	(3.86)	(3.78)	(3.82)	(4.21)				
Observations	1295876	1294624	1292111	1289604	1287215	1282543				
Adjusted R-squared	0.14	0.14	0.16	0.17	0.17	0.18				
Panel B. Reddit author peers										
	t+5	t+10	t+20	t+30	t+40	t+60				
	(1)	(2)	(3)	(4)	(5)	(6)				
REDDIT RET	1.753	1.951	2.804	0.980	2.062	2.827				
	(1.22)	(1.22)	(1.52)	(0.41)	(0.73)	(0.83)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes				
Time FE	Yes	Yes	Yes	Yes	Yes	Yes				
Intercept	1.049	1.984	3.556	4.693	5.797	8.767				
	(4.63)	(4.39)	(4.28)	(4.09)	(4.05)	(4.39)				
Observations	1256968	1255743	1253298	1250849	1248483	1243886				
Adjusted R-squared	0.14	0.14	0.16	0.16	0.17	0.18				

Table 12. Order imbalance and Reddit peer returns.

This table presents the estimated regression coefficients predicting one-day ahead order imbalance using Reddit peer returns. For each focal stock. the Reddit peer return is the weighted average one-day return of stocks that are linked through Reddit post (Panel A) or Reddit author (Panel B). We consider three types of order imbalance: trade imbalance, volume imbalance, and dollar imbalance. The trade imbalance is calculated by the difference in the number of buys and number of sells divided by the total number of buys and sells; the volume imbalance is shares of buy trades minus shares of sell trades divided by the total volume of buys and sells; the dollar imbalance is the difference in the dollar value of buys and the dollar value of sells divided by the total dollar value of buys and sells. Control variables are defined identically as in Table 6. Day fixed effects are included in regressions and t-statistics with standard errors clustered at the firm level are reported in parentheses. The sample period is from January 2019 to December 2021.

Panel A. Reddit post peers										
	Total order imbalance			Retail order imbalance						
	(1) Trade	(2) Volume	(3) Dollar	(4) Trade	(5) Volume	(6) Dollar				
REDDIT RET	-0.040 (-1.63)	-0.030 (-1.01)	-0.031 (-1.04)	$0.053 \\ (1.24)$	$0.096 \\ (1.60)$	$0.095 \\ (1.58)$				
Controls	Yes	Yes	Yes	Yes	Yes	Yes				
Time FE	Yes	Yes	Yes	Yes	Yes	Yes				
Observations	1268652	1268652	1268652	1265776	1265776	1265776				
Adjusted R-squared	0.03	0.02	0.02	0.01	0.00	0.00				
Panel B. Reddit author peers										
	Total order imbalance			Retail order imbalance						
-	(1) Trade	(2) Volume	(3) Dollar	(4) Trade	(5) Volume	(6) Dollar				
REDDIT RET	-0.021 (-1.69)	-0.012 (-0.82)	-0.012 (-0.82)	$0.007 \\ (0.36)$	$0.007 \\ (0.24)$	$0.007 \\ (0.24)$				
Controls	Yes	Yes	Yes	Yes	Yes	Yes				
Time FE	Yes	Yes	Yes	Yes	Yes	Yes				
Observations	1229737	1229737	1229737	1226858	1226858	1226858				
Adjusted R-squared	0.03	0.02	0.02	0.01	0.00	0.00				

Appendix to "Firm Linkages from the Wisdom of Crowds"

A Matching between stocks and texts

Here, we discuss in detail the method to match a comment in r/wallstreetbets to a stock. In this paper, if the tokenized text¹ matched our list of stock tickers², we would label this comment as a stock-related comment and identify any related stocks. For example, the sentence "INTC still has dominant market share in both consumer and server markets", shown in Figure 2 is tokenized in our method to a list of 12 tokens ['INTC', 'still', 'has', 'dominant', 'market', 'share', 'in', 'both', 'consumer', 'and', 'server', 'markets'], then we cross-reference all these tokens of this list with the ticker list and find "INTC" meeting our requirments. So, we label the comment containing this sentence as stock-related and tag this comment as "Intel Corp." related.

Naturally, this matching method may raise several concerns. First, removing tickers with fewer than 3 letters may result in the omission of many stock-related comments, referred to as false negative matches. Second, some matches may include comments not discussing stocks—the false positive matches. We discuss our opinions about these two concerns respectively. The deletion of tickers makes the sample not contain stocks with less than three letters, and these ticker symbols account for about 3% of the number of stocks in the list. But the inclusion of these tickers with one or two letters will cause much confusion because of the acronym overload and add excessive noise to the construction of our linkage. However, the criterion of at least three letters cannot solve semantic confusion problems completely. These ticker symbols could be used by certain Reddit users to represent something other than stock. For example, they can use "CAKE" to refer to the cake they are now eating today as without talking about The Cheesecake Factory Inc., which uses the ticker symbol list as a robustness check, deleting all the 5000 most frequent words according to the Google Web Trillion Word Corpus³, and we get similar empirical results.

An alternative matching method is to use "\$" before the ticker symbol, which means we tag the comment as "Intel Corp." related when "\$INTC" emerges in it instead of when

¹Tokenization is the process of substituting text into words, phrases, symbols, or other non-sensitive equivalents, known as tokens.

 $^{^{2}}$ The list contains stocks in NYSE/AMEX/NASDAQ and with share code 10 or 11. And we keep tickers with at least 3 letters to avoid ambiguities in the matching process.

³The word counts data file is obtained from http://norvig.com/ngrams/.

"INTC" does. The majority of the semantic misunderstanding might be avoided by using this tighter approach, which is also widely used in literature, like in Hu et al. (2021) and in Li (2022). The main reason we refuse this method is because it may result in a significant lack of comments about stocks and may affect the representativeness of our Reddit linkages. Reddit users are not forced to use "\$TICKER", with a dollar sign before the ticker, to refer to the stock, and they also usually use the ticker symbol without dollar signs while discussing the fundamental information of the firm.

B Description of fundamental variables

We use CRSP monthly stock returns and COMPUSTAT quarterly data for the sample period December 2018-December 2021. We drop firm-quarter observations with missing or negative total assets (item atq). We follow Lee et al. (2015) and Hou, Xue, and Zhang (2020) in constructing the fundamental variables.

QRET: quarterly stock return. We use CRSP monthly stock return data to calculate quarterly stock returns.

ROA: return on assets. Quarterly net income before extraordinary items (item ibq) scaled by 1-quarter-lagged total assets (item atq).

ROE: return on equity. Quarterly net income before extraordinary items (item ibq) scaled by 1-quarter-lagged book equity. Book equity is shareholders' equity, plus balance sheet deferred taxes and investment tax credit (item txditcq) if available, minus the book value of preferred stock (item pstkq). Depending on availability, we use stockholders' equity (item seqq), or common equity (item ceqq) plus the book value of preferred stock, or total assets (item atq) minus total liabilities (item ltq) in that order as shareholders' equity.

GP: gross profit. Total revenue (item revtq) minus cost of goods sold (item cogsq) divided by 1-quarter-lagged total assets (item atq).

EP: earnings-to-price ratio. Quarterly net income before extraordinary items (item ibq) divided by market equity.

CP: cash flow-to-price ratio. Quarterly cash flows divided by the market equity. Quarterly cash flows are income before extraordinary items (item ibq) plus depreciation (item dpq).

SP: sales-to-price ratio. Quarterly sales (item saleq) divided by the market equity.

SG: sales growth. Quarterly sales (item saleq) divided by the four-quarter-lagged value,

and then minus one. We require both the current sales and the lagged sales to be positive in order to calculate the growth rate.

PG: profit growth. Quarterly operating income before depreciation (item oibdpq) minus its value four quarters ago divided by 1-quarter-lagged total assets (item atq).

RG: revenue growth. Quarterly total revenue (item revtq) divided by its value four quarters ago, then minus one.

AG: asset growth. Total assets (item atq) divided by its value four quarters ago and then minus one.

LEV: leverage ratio. Long-term debt (item dlttq) scaled by total shareholder' equity (item seqq).

AT: asset turnover. Quarterly sales (item saleq) divided by total assets (item atq).

RDS: R&D expense-to-sales ratio. Quarterly R&D expense (item xrdq) divided by quarterly sales (item saleq). Firms with nonpositive sales are excluded. If a firm's R&D expense is missing, we assume the value to be zero.