International Review of Financial Analysis 91 (2024) 103036

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

Social media information diffusion and excess stock returns co-movement

Zhang-HangJian Chen [a,g], Wang-Long Wu [a], Sai-Ping Li [b,c], Kun Bao [d,*], Kees G. Koedijk [e,f]

[a] School of Economics, Anhui University, Hefei 230601, China
[b] Institute of Physics, Academia Sinica, Taipei 115, Taiwan
[c] Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
[d] School of Economics and Management, Beijing Jiaotong University, Beijing 10009, China
[e] School of Economics, Utrecht University, Utrecht, the Netherlands
[f] CEPR, London EC1V 0DX, UK
[g] Research Center for Financial and Statistical, Anhui University, Hefei 230601, China

ARTICLE INFO

Keywords:
Information diffusion
Social media
Excess co-movement
Individual investors
Information interaction

ABSTRACT

This research investigates the dynamic interplay between information diffusion on social media platforms and the co-movement of excess stock returns through a comprehensive methodology encompassing the multilayer complex network analysis, panel vector autoregression (PVAR) modeling, and the thermal optimal path (TOP) approach. Utilizing weekly data spanning from January 1, 2016, to December 31, 2021, our research finds a significant interrelationship between information diffusion and excess co-movement, notably shaped by exogenous shocks, such as the COVID-19 outbreak. We investigate the microcosmic mechanism, revealing that variations in excess co-movement significantly impact the information interaction behaviors of individual investors within sub-forums, subsequently influencing their trading activities across related stocks. Moreover, stocks characterized by a heightened strength of information diffusion exhibit swifter responsiveness to new information and demonstrate superior performance in hedging strategies involving the IC500 stock index futures. These findings hold potential to aid regulators and investors in comprehending risk transmission within the stock market and refining portfolio management. A heightened understanding of the role played by information interaction among individual investors via social media in the co-movement of excess stock returns empowers informed decision-making and risk mitigation.

1. Introduction

A profound comprehension of the microcosmic mechanisms underpinning enigmatic asset pricing phenomena, such as the pronounced co-movement of stock returns, holds paramount significance for the effective management of risk in the stock market and the optimization of investment portfolios. In the contemporary landscape, the advent of social media has instigated a transformative paradigm shift in the global diffusion of information (Agarwal, Kumar, & Goel, 2019), thereby amplifying its consequential impact on financial markets. The perspectives of individual investors pertaining to fluctuations in stock prices can rapidly disseminate across diverse social media platforms. As the Chinese stock market continues its expansion, its prominence within the global capital markets escalates, with individual investors predominantly shaping the trajectory of this market while institutional investors adopt a relatively subdued role[1] (Wang, Xie, Zhao, & Jiang, 2018). Furthermore, spurred by the rapid proliferation of the internet in China, individual investors have become increasingly active in articulating and disseminating their viewpoints concerning stocks through various social media platforms.[2] This phenomenon not only expedites and enriches the dissemination of information related to stocks (Huang, Sun, & Chu, 2022; Liu, Wu, Li, & Li, 2015), but also makes a substantial contribution to the overarching enhancement of the information milieu accessible to individual investors.

* Corresponding author at: No.3 Shangyuancun, Haidian District, Beijing 10009, China.
E-mail address: kunbao@bjtu.edu.cn (K. Bao).

[1] Based on the 2020 Shanghai Stock Exchange Statistics Annual (http://www.sse.com.cn/), it is noteworthy that retail investors constitute a significant majority, contributing to 85% of the daily trading volume on the Shanghai Stock Exchange, with institutional investors representing a mere 15%.

[2] As reported in the 51st Statistical Report on Internet Development in China, a publication by the China Internet Network Information Center (CNNIC), the year 2022 witnessed a remarkable achievement with the total count of internet users in China surging to an impressive figure of 1.067 billion. This surge in internet users corresponds to an equally noteworthy internet penetration rate of 75.6%.

https://doi.org/10.1016/j.irfa.2023.103036
Received 11 July 2023; Received in revised form 20 October 2023; Accepted 10 November 2023
Available online 15 November 2023

In this study, we employ a dataset obtained from a stock message board, which encompasses dedicated sub-forums for individual stocks. Our primary objective is to devise a novel metric that provides a more direct and precise measure of the potency of information dissemination stemming from individual investors. Owing to the constraint of limited attention, as expounded in extant literature (Hirshleifer, Lim, & Teoh, 2011; Peng, Xiong, & Bollerslev, 2007), individual investors confront the impossibility of actively participating in all sub-forums. They predominantly engage in interactions with fellow participants by means of posting or responding within the sub-forums that align with their particular areas of interest.

In instances where investors exhibit activity across two discrete sub-forums within a defined temporal framework—hereinafter termed as "co-investors"—they create a conduit for deliberating upon a specific stock within each other's respective forums through the act of posting or responding. This mode of interaction serves as a catalyst for the dissemination of information between these two sub-forums. Furthermore, within the context of social networks, it has been empirically established that users boasting a substantial following possess enhanced capabilities for information diffusion, as substantiated by prior research (Li, Cai, Li, & Du, 2019; Li, Zhang, Feng, & An, 2019). In light of the discerned information exchange behaviors among co-investors, we have consequently opted to employ the information entropy framework to devise a metric that quantifies the magnitude of information diffusion.

Most previous studies have primarily concentrated their efforts on examining the impact of information diffusion on excess co-movement, predominantly through the lenses of the word-of-mouth effect and the social embeddedness theory. The process of information diffusion facilitated by "word-of-mouth" communication is known to enhance investors' informational milieu and bolster informed trading (Christos & Wang, 2017; Ivkovich & Weisbenner, 2007). This, in turn, can expedite the adjustment of stock prices in response to common information (Badrinath, Kale, & Noe, 1995; Holden & Subrahmanyam, 1992). Informed investors, driven by the mutual information they acquire, may concurrently engage in trading multiple stocks, thereby inducing co-movement in stock returns (Koch, Ruenzi, & Starks, 2016; Li, Cai, et al., 2019; Li, Zhang, et al., 2019). In contrast, social media platforms offer individual investors a more expeditious and convenient means of information exchange, thus significantly enhancing the informational environment for individual investors when compared to the word-of-mouth effect.

The social embeddedness theory posits that firms are influenced by their entrenched social relationships (Granovetter, 1985), which inherently possess resource value and facilitate information diffusion. Investors may categorize risky assets into distinct styles based on these attributes, ultimately culminating in the co-movement of stock returns (Barberis & Shleifer, 2003; Wahal & Yavuz, 2013). Empirical research has underscored the profound impact of firms' embedded social relationships on stock return co-movement. Such relationships encompass equity cross-holdings, shared individual ownership, and director interlocks (Khanna & Thomas, 2009), common ownership among institutional investors (Anton & Polk, 2014; Fricke & Savoie, 2017; Gong & Du, 2020; Koch et al., 2016), as well as shared analyst coverage (Hameed, Morck, Shen, & Yeung, 2015; Muslu, Rebello, & Xu, 2014). Moreover, social media platforms have the capacity to facilitate information exchange among market participants, thereby enhancing the formation of embedded social relationships among firms in the market ecosystem.

There are also a limited number of studies that have delved into the ramifications of investor attention (Drake, Jennings, Roulstone, & Thornock, 2017) and investor communication on social media platforms (Ding, Zhou, & Li, 2020; Jiang, Liu, & Yang, 2019; Liu et al., 2015) with respect to the phenomenon of excess co-movement in stock returns. However, there remains a paucity of comprehensive empirical evidence pertaining to the intricate interplay between information diffusion via social media platforms by individual investors and the occurrence of excess co-movement in stock returns, as well as the underlying microcosmic mechanisms governing this interrelationship. The primary objective of this study is to bridge these existing gaps.

Two facets of the existing literature provide a compelling rationale for investigating the intricate interplay between the diffusion of information via social media and the co-movement of excess stock returns. Firstly, it is well-documented that individual investors' anticipated returns on stocks can be significantly influenced by the diffusion of information through social media platforms, thereby potentially contributing to the observed excess co-movement in stock returns. Individual investors often collectively allocate their attention to various stocks, as evidenced by their concurrent engagement with information derived from corresponding online stock forums. This co-attention behavior, expediting the adjustment of prices in response to common information (Badrinath et al., 1995; Christos & Wang, 2017), may induce a convergence of individual investors' expectations across multiple stocks, predicated on the shared information they receive, consequently resulting in excess co-movement (Koch et al., 2016; Li, Cai, et al., 2019; Li, Zhang, et al., 2019).

Moreover, investors have the capacity to acquire information regarding specific stocks by perusing or responding to posts authored by others within the corresponding forum. The opinions expressed by their peers in these forums can exert a discernible influence on investors' expected returns for a particular stock (Challet, Chicheportiche, Lallouache, & Kassibrakis, 2018). The frequent exchange of information within online stock forums offers a convenient means for investors to monitor the behaviors of their peers, a circumstance that can readily precipitate information cascades and amplify the herding effect (Bikhchandani, Hirshleifer, & Welch, 1998; Kimberly & Michael, 2013). When investors collectively exhibit pronounced herding behavior in their trading activities across multiple stocks, it inevitably engenders excess co-movement in stock returns (Li, Cai, et al., 2019; Li, Zhang, et al., 2019). Furthermore, heterogeneity in investors' trading decisions may arise from variations in the information accessible to them, which can either augment or diminish the correlation between two stocks stemming from the herding effect.

On the other hand, stock price fluctuations can attract online attention, especially in the aftermath of a stock market crash, often accompanied by heightened stock price co-movements. Consequently, the dynamics of the stock market can induce fluctuations in internet attention on specific trading days (Gao, Zhao, Wang, & Liu, 2020). The attention garnered from the online sphere has the potential to significantly influence investors' sentiments and decision-making processes, with the ability to rapidly propagate and amplify through social media channels (Rizkiana et al., 2018). These fluctuations in excess co-movement of stock returns may, in turn, have the potential to impact information diffusion on social media platforms, thereby giving rise to a multifaceted interrelationship between the two phenomena.

Furthermore, the information interaction behaviors exhibited by individual investors on social media platforms serve as a micro-level determinant that propels and facilitates information diffusion. Conversely, the trading behaviors of individual investors across different stocks represent the foundational rationale behind the occurrence of excess co-movement in stock returns. Consequently, it becomes imperative to delve deeper into the microcosmic mechanisms underpinning the interaction between information diffusion and excess co-movement, underpinned by a comprehensive understanding of their interrelationship.

First, we embark on an exploration of the dynamic interplay between information diffusion and excess co-movement through the utilization of a Panel Vector Autoregression (PVAR) model. The empirical findings illuminate a significant causal relationship between variables associated with information diffusion and the presence of excess co-movement. This noteworthy relationship persists even after a lag of three periods has been considered. Furthermore, we delve into the non-linear dynamics governing the relationship between information diffusion and excess co-movement. Our investigations reveal that information diffusion does not invariably dominate the excess co-movement of stock returns, and conversely, the latter does not always exert significant influence.

To augment our understanding, we construct a multilayer network encompassing layers representing information diffusion within corresponding sub-forums of stocks and the excess co-movement of stock returns. This approach allows us to explore the dynamic evolution patterns characterizing these layers. Notably, we observe certain structural similarities between the two layers and substantial fluctuations in network topology indicators over time. Remarkably, the information diffusion layer exhibits higher global efficiency compared to the excess co-movement layer. Moreover, we find that the correlation coefficient of node degree series and the ratio of edge overlap between the two layers are highly sensitive to shifts in the external stock market environment. Financial stress intensifies the interconnectedness between the layers of information diffusion and excess co-movement.

Subsequently, we delve into the microcosmic mechanisms underpinning the aforementioned interrelationship, scrutinizing co-investors' information interaction behaviors within sub-forums. We also expound upon how these information interaction behaviors shape the trading decisions of individual investors. We segment pairs of stocks into high and low co-movement groups, predicated on the level of excess co-movement, and unearth noteworthy findings. In instances characterized by high excess co-movement among pairs, co-investors within corresponding sub-forums tend to exhibit greater connectivity within the information interaction network. Moreover, they display elevated tweeting quantity and frequency. Additionally, in scenarios marked by a high positive correlation among pairs, co-investors manifest a more consistent sentiment towards both stocks. Furthermore, when co-investors possess a substantial following within sub-forums, exhibit heightened tweeting activity, maintain increased tweeting frequency, and evince minimal disparities in sentiment between the two stocks, individual investors are more inclined to display heightened consistency in their trading behavior across the two stocks. These results unveil the microcosmic behavioral characteristics governing the information interaction and investment decision-making of individual investors, thereby contributing to a more profound comprehension of the dynamic interrelationship between information diffusion and excess co-movement.

Finally, we embark on an assessment of the hedging performance of a comprehensive selection of sample stocks with various stock index futures, including the CSI 500 Stock Index Futures (IC500), the CSI 300 Stock Index Futures (IF300), and the Shanghai 50 Stock Index Futures (IH50). Our investigations reveal that stocks characterized by a heightened strength of information diffusion exhibit swifter responsiveness to information and fare more favorably in hedging strategies involving IC500 stock index futures. We conduct robustness tests, exploring the microcosmic mechanisms underpinning the interplay between information diffusion and excess co-movement from the vantage point of the information interaction behaviors of all sub-forum users. Consistent conclusions are obtained, affirming the robustness of our findings.

Our research makes several noteworthy contributions to existing literature. To the best of our knowledge, this study represents the inaugural endeavor to scrutinize the dynamic interplay between the diffusion of information through social media and the occurrence of excess co-movement in stock returns. In doing so, it supplements existing research focused on elucidating the impact of information diffusion engendered by the word-of-mouth effect on stock price dynamics (Christos & Wang, 2017; Colla & Mele, 2010; Hong, Jeffrey, & Stein, 2005). Given the inherent unobservability of word-of-mouth information diffusion (Rantala, 2019), our contribution lies in the development of a novel indicator that offers a more direct and accurate means of gauging the magnitude of information diffusion.

Furthermore, our study augments the body of literature devoted to understanding the co-movement of stock returns and its nexus with the embedded social ties among firms, as forged by market participants (Anton & Polk, 2014; Fricke & Savoie, 2017; Hameed et al., 2015; Khanna & Thomas, 2009; Koch et al., 2016; Muslu et al., 2014), particularly from the perspective of individual investors.

Our research also aligns with prior work exploring the role of investor behaviors in shaping the dynamics of stock return co-movement (Barber, Odean, & Zhu, 2009; Jiang et al., 2019; Kumar & Lee, 2006). Kumar and Lee (2006) have demonstrated that retail investors' trades exhibit systematic correlations, which can elucidate the co-movement observed in stock returns. We extend this line of inquiry by revealing that alterations in excess co-movement can exert a substantial impact on the information interaction behaviors of individual investors within sub-forums. These behaviors, in turn, influence their trading activities across two stocks.

The remainder of this paper is structured as follows: Section 2 introduces the sample data and delineates the construction of key variables. Section 3 presents the empirical findings of this study and outlines its application in portfolio construction. Finally, Section 4 provides the concluding remarks.

2. Data and methodology

2.1. Data

The sample stocks in this study are sourced from the Shanghai Stock Exchange, and our sampling period spans from January 1, 2016, to December 31, 2021. Recognizing that the influence of investors' trading behavior on stock return co-movement may vary across firms with differing market capitalizations (Kumar & Lee, 2006), we have excluded stocks with intermittent trading activity during the sampling period. Ultimately, we have selected a total of 429 sample stocks to ensure sufficient liquidity for the analysis.

For the measure of information diffusion emanating from individual investors, we have collected post data generated by individual investors by monitoring all messages posted on the Eastmoney stock sub-forum.[3] The Eastmoney stock forum stands as China's preeminent and most influential financial website, boasting a daily active user base exceeding 19,000 and witnessing over four million daily clicks on posts. The platform provides a dedicated sub-forum for investors to exchange opinions on any given stock. Our dataset encompasses post data for each stock from its respective sub-forum, encompassing user IDs, post content, posting times, as well as replier IDs, reply content, and reply times for each post. Throughout the sampling period, we have accumulated a staggering 64,103,724 posts contributed by 4,677,488 users across 429 sub-forums on Eastmoney.

To measure the excess co-movement of stock returns, we have sourced daily closing prices and data on the five-factor model from the RESSET databases. Additionally, pertinent data concerning stock attributes have been extracted from the Wind and CSMAR databases, encompassing annual market capitalization, industry affiliations, quarterly P/E ratios, quarterly long-term debt, total assets, institutional investor shareholdings, and analyst coverage.

2.2. Methodology

2.2.1. Measures of co-movement

Given the phenomenon of limited attention (Egeth & Kahneman, 1973), wherein investors tend to process a greater volume of market and sector-wide information, it leads to stock prices adjusting to common information, thereby increasing the positive correlation among stock returns. Conversely, when investors focus more on firm-specific information, it results in stock prices deviating and increases the negative correlation of stock prices (Peng et al., 2007). Consequently, the

[3] https://guba.eastmoney.com/.

diffusion of various information types among sub-forums may yield diverse price correlation effects, all of which possess the potential to elevate stock co-movement levels. To measure this co-movement, we employ the absolute value of the Pearson correlation coefficient among stocks' excess returns.

Pasquariello (2007) defines excess co-movement as the covariation between two assets beyond what can be accounted for by fundamental factors. Kallberg and Pasquariello (2008) further utilize the Pearson correlation coefficient of excess stock returns to quantify excess co-movement. Anton and Polk (2014), Li, Zhang, et al. (2019) and Li, Cai, et al. (2019) also adopt this methodology to investigate the impact of common ownership among institutional investors on excess co-movement. Following their approach, we initially calculate excess stock returns based on a five-factor model (Fama & French, 2015) as follows:

$$r_{lt} - r_{fm} = \alpha_0 + \alpha_1 mkt_t + \alpha_2 smb_t + \alpha_3 hml_t + \alpha_4 rmw_t + \alpha_5 cma_t + e_{lt}, \tag{1}$$

where r_{lt} represents stock l's return in week t, $r_{fm}$is the risk-free rate for that week, mkt_t denotes the value-weight market portfolio return in excess of the risk-free rate, and smb_t, hml_t, rmw_t, cma_t are respectively represent the size, value, profitability and investment factors. e_{lt} represents the excess return of stock l.

We next calculate the weekly Pearson correlation coefficient ($Cor_{lk,t}$) of stock excess returns using the rolling-window method, employing a window size of 60 trading days and a step size of 5 trading days.[4] The level of excess co-movement between stocks l and k is quantified as $Dis_{lk,t} = |pearson(e_{lt}, e_{kt})|$.

2.2.2. Measures of information diffusion

According to the theory of investor attention, investors face limitations in their ability to pay attention to all stocks due to attention being a finite resource (Hirshleifer et al., 2011). Consequently, investors tend to be selective in the stocks they follow and actively participate in the sub-forums that align with their interests. In these sub-forums, investors can express their views by posting, and others can engage in communication through replies, thereby facilitating the diffusion of information from posters to repliers. When co-investors are active on two different sub-forums during a specific period, discussions about one stock can occur on the forum of another stock through posting or replying, further promoting information diffusion between the sub-forums.

Inspired by Li, Zhang, et al. (2019) and Li, Cai, et al. (2019) and, who identified key nodes in complex networks, we consider users with many secondary neighbors in an information interaction network as central to the information flow. These users can be likened to the root node of a full binary tree, enabling information to reach a wider audience. When users with many secondary neighbors are active on two sub-forums, information is more likely to flow between them. Following Li, Zhang, et al. (2019) and Li, Cai, et al. (2019), we employ the information entropy approach (Maes, Collignon, Vandermeulen, Marchal, & Suetens, 1997) to measure the strength of information diffusion among sub-forums based on the information interaction behaviors of co-investors, who play a crucial role in facilitating information diffusion among sub-forums.

We begin by constructing an information interaction network, denoted as g_l (g_k), using data from the corresponding sub-forum associated with stock l (stock k). In this network, users of the sub-forum represent nodes, and if user v replies to a post by user u, an edge is established from v to u. We define $N_{lk}=\{v_1^{lk}, v_2^{lk}, \dots, v_i^{lk}, i = 1, 2, \dots, n_{lk}\}$ as the set of co-investors participating in sub-forums l and k. Moreover, we denote $N_{l,i}=\{v_1^{l}, v_2^{l}, \dots, v_i^{l}, i = 1, 2, \dots, n_{l}\}$ and $N_{k,i}=\{v_1^{k}, v_2^{k}, \dots, v_i^{k}, i = 1, 2, \dots, n_{k}\}$ as the sets of neighbors of co-investor v_i^{lk} in the information interaction networks G_l and G_k, respectively.

We then calculate both the total in-degree of neighbors of co-investor v_i^{lk} in the information interaction networks g_l (denoted as $Deg_{l,i}^{in}$) and the total out-degree of neighbors of co-investor v_i^{lk} in the information interaction networks g_k (denoted as $Deg_{k,i}^{out}$) as follows:

$$Deg_{l,i}^{in} = \sum_{m=1}^{n_l} d_{l,m}^{in}, \tag{2}$$

$$Deg_{k,i}^{out} = \sum_{m=1}^{n_k} d_{k,m}^{out}, \tag{3}$$

where $d_{l,m}^{in}$ represents the in-degree of neighbor v_m^l of co-investors v_i^{lk} in G_l, while $d_{k,m}^{out}$ represents the out-degree of neighbor v_m^k of co-investor v_i^{lk} in G_k. We further define the in-degree entropy of co-investors in G_l (s_l^{in}) and the out-degree entropy of co-investors in G_k (s_k^{out}) as follows:

$$s_l^{in} = -\sum_{i=1}^{n_{lk}} \frac{Deg_{l,i}^{in}}{\sum_{i=1}^{n_{lk}} Deg_{l,i}^{in}} log \frac{Deg_{l,i}^{in}}{\sum_{i=1}^{n_{lk}} Deg_{l,i}^{in}}, \tag{4}$$

$$s_k^{out} = -\sum_{i=1}^{n_{lk}} \frac{Deg_{k,i}^{out}}{\sum_{i=1}^{n_{lk}} Deg_{k,i}^{out}} log \frac{Deg_{k,i}^{out}}{\sum_{i=1}^{n_{lk}} Deg_{k,i}^{out}}. \tag{5}$$

To quantify the strength of information diffusion between sub-forums l and k, we calculate the entropy of information flow from sub-forum l to sub-forum k is $s_{lk} = s_l^{in} + s_k^{out}$ and define the strength of information diffusion as $Mi_{lk} = (s_{lk} + s_{kl})/2$.

For the purpose of aligning with the co-movement calculation window, we employ a rolling-window approach with a window size of 60 trading days and step size of 5 trading days. This enables us to compute the weekly $Mi_{lk,t}$ for each pair of sub-forums. During each window, we utilize post and reply data within the defined period to construct the corresponding information interaction networks, identify co-investors for each pair of sub-forums, and subsequently compute $Mi_{lk,t}$.

2.2.3. Control variables

To ensure the robustness of our results, we incorporate control variables that have previously been identified as influential factors in co-movement, as substantiated by prior research. Specifically, we include the following control variables: size difference Si (Kumar & Lee, 2006), price difference Pr (Green & Hwang, 2009), P/E ratio similarity Pe (Campbell & Shiller, 1998), financial leverage difference Fin (Anton & Polk, 2014), common ownership of institutional investors Ins (Pareek, 2012), and analyst coverage Ana (Francisco, 2017). Si represents the extent of difference in stock sizes and is defined as:

$$Si_{lk,t} = |ln(Mv_{lt}/Mv_{kt})|, \tag{6}$$

where Mv_{lt} signifies the market capitalization of stock l during week t.

Pr quantifies the level of difference in the stock prices and is expressed as:

$$Pr_{lk,t} = |ln(P_{lt}/P_{kt})|, \tag{7}$$

where P_{lt} represents the closing price of stock l during week t.

Pe gauges the degree of similarity in the P/E ratio of stocks through:

$$Pe_{lk,t} = \begin{cases} min\left(Per_{lk,t}, 1/Per_{lk,t}\right), Per_{lk,t} \geq 0 \\ max\left(Per_{lk,t}, 1/Per_{lk,t}\right), Per_{lk,t} < 0 \end{cases}, \tag{8}$$

where $Per_{lk,t}$ signifies the ratio of stock l's P/E ratio to stock k's for the quarter in which week t falls. Financial leverage difference $Fin_{lk,t}$ on week t is measured by the absolute difference of the financial leverage ratio, defined as long-term debt divided by total assets, between the two stocks. Common ownership of institutional investors $Ins_{lk,t}$ is denotes as

[4] We have computed the degree of excess co-movement, denoted as $Cor_{lk,t}$, for window sizes of both 90 and 120 trading days. Importantly, the findings and conclusions remain consistent across these two window sizes. However, for the sake of brevity, we have chosen not to present the results for both window sizes.

$$Ins_{lk,t} = \frac{1}{M}\sum\nolimits_{M}\left(ro_{lt}^{m} + ro_{kt}^{m}\right) \Big/ 2, \tag{9}$$

where M represents the set of institutional investors who hold both stocks l and k during the quarter that encompasses week t, and ro_{lt}^{m} signifies the portion of stock l held by institution investors m. Lastly, analyst coverage $Ana_{lk,t}$ represents the number of common analysts covering each pair of stocks l and k during the quarter that includes week t.

3. Empirical results

3.1. Relationship between information diffusion and excess co-movement through the PVAR model

In this section, we employ the panel vector autoregression (PVAR) model to rigorously examine the quantitative relationship between information diffusion within corresponding stock sub-forums and the occurrence of excess co-movement in stock returns. The PVAR model is a tool capable of considering all economic variables as endogenous factors. It not only accounts for the interdependencies among these variables but also seamlessly integrates the strengths of both panel data and vector autoregression (VAR) models. This enables us to systematically analyze how economic variables dynamically respond to shocks while controlling for individual and time-specific effects (Abrigo & Love, 2016; Love & Zicchino, 2006; Zhang, Zhang, Xu, & Chen, 2023).

Following the framework of Abrigo and Love (2016), we formulate a homogenous PVAR of order p with panel-specific effects for k variables as follows:

$$Y_{lt} = \sum A_p Y_{lt-p} + \sum B_p X_{lt-p} + \mu_l + \varepsilon_{lt}, \quad l \in \{1, 2, \dots, L\}, \quad t \in \{1, 2, \dots, T\}, \tag{10}$$

where Y_{lt} represents a vector of endogenous variables, which encompasses the strength of information diffusion (*Mi*) and excess co-movement (*Dis*), while X_{lt} comprises a vector of exogenous variables including the level of difference in stocks size (*Si*), the level of difference in stock prices (*Pr*), the level of similarity in stocks *P/E* ratio (*Pe*), the level of difference in stocks financial leverage (*Fin*), the common ownership of institutional investors of two stocks (*Ins*), and the number of analysts in common for two stocks (*Ana*). Parameters A and B are matrices. μ_l denotes a vector of dependent variable-specific panel fixed effects, and ε_{lt} is a vector of idiosyncratic errors with the properties $E(\varepsilon_{lt}) = 0$, $\sum = E\left(\varepsilon_{lt}^{'}\varepsilon_{lt}\right)$ and $E\left(\varepsilon_{lt}^{'}\varepsilon_{lt}\right) = 0$ for all $t > s$. The optimal lag length p for the PVAR model is determined using criteria such as AIC (Akaike information criterion), BIC (Bayesian information criterion), and HQIC (Hannan–Quinn information criterion). In line with Abrigo and Love (2016), we employ the Generalized Method of Moments (GMM) estimation approach to estimate the model parameters, effectively mitigating the over-identification issue.

The dataset used in this study is weekly, encompassing a total of 91,806 pairs of stocks for each variable, with each stock pair comprising 277 observations. Observations where no co-investors were active on the two sub-forums in the t-th week were excluded. Therefore, the data for the PVAR model constitutes an unbalanced panel dataset with a total of 19,303,769 observations for each variable. The summary statistics of all standardized variables are presented in Table 1. Notably, the statistics reveal that the average, minimum, and maximum values of *Dis* are 0.1272, 0.0000, and 0.1190, respectively, indicating a substantial variation in the level of excess co-movement across different pairs of stocks. Additionally, there are noticeable disparities in the strength of information diffusion among various pairs of sub-forums, with *Mi* ranging from 0.0000 to 0.1190.

We commence our empirical analysis by examining the stationarity of variables. To accomplish this, we employ panel unit root tests, specifically the Levin-Lin-Chu (LLC) and the cross-sectional augmented Im-Pesaran-Shin (IPS) tests (Levin, Lin, & Chu, 2002; Pesaran, 2007). The results of these tests are presented in Table 2. All variables exhibit stationary behavior, as evidenced by the rejection of the null hypothesis of a unit root at the 1% significance level.

Subsequently, we determine the optimal lag length (p) for the Panel Vector Auto Regression (PVAR) model using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC). Table 3 displays the results, indicating that a lag order of 3 is the optimal choice according to all three criteria.

Furthermore, we conduct a panel Granger causality test to explore the causal relationship between information diffusion and excess co-movement. The results of this test are summarized in Table 4. Significantly, the null hypothesis of no causality between Mi and Dis is firmly rejected at the 1% significance level, providing compelling evidence of a substantial bidirectional causal relationship between information diffusion and excess co-movement.

Table 5 displays the results of the PVAR model estimated using the GMM approach. When considering *Dis* as the dependent variable, the coefficients of lagged *Mi* for lags 1, 2, and 3 are all highly significant at the 1% level, with respective values of 0.8424, 0.0119, and 0.0067. This indicates that an increase in the strength of information diffusion among the corresponding sub-forums of stocks significantly contributes to the excess co-movement among stock returns. These findings align with those of Jiang et al. (2019), who also observed that more frequent discussions of related stocks on social media lead to greater co-movement. Notably, the estimated coefficients of *Mi* decline as the lag period increases, suggesting that the impact of historical information diffusion on current excess co-movement diminishes over time, which aligns with intuitive expectations.

Conversely, when *Mi* is treated as the dependent variable, the coefficients of lagged *Dis* for lags 1, 2, and 3 are also highly significant at the 1% level. This implies that the excess co-movement of stock returns significantly influences the strength of information diffusion among the

Table 1
Summary statistics of all variables.

	Mean	Std.Dev	Min	25%	Median	75%	Max
Dis	0.1272	0.1028	0.0000	0.0487	0.1040	0.1805	0.9412
Mi	0.0170	0.0106	0.0000	0.0101	0.0155	0.0217	0.1190
Si	1.2248	1.0732	0.0000	0.4206	0.9393	1.7219	7.9428
Pr	0.8025	0.6884	0.0000	0.2891	0.6329	1.1314	6.9316
Pe	0.2874	0.4156	−1.0000	0.0717	0.3062	0.5908	1.0000
Fin	0.1032	0.1034	0.0000	0.0236	0.0700	0.1508	0.5841
Ins	0.3531	1.3412	0.0000	0.0000	0.0000	0.0495	47.8850
Ana	0.0615	0.7782	0.0000	0.0000	0.0000	0.0000	52.0000

Note: *Dis* represents the absolute value of the correlation coefficient measuring excess co-movement in stock returns. *Mi* measures the strength of information diffusion using the mutual information approach. *Si* measures the level of difference in stock sizes. *Pr* measures the level of disparity in stock prices. *Pe* measures the level of similarity in stocks' P/E ratios. *Fin* measures the level of variation in stocks' financial leverage. *Ins* measures the common ownership by institutional investors in two stocks. *Ana* counts the number of analysts shared between two stocks.

Table 2
Results of the panel unit root tests.

	Dis	*Mi*	*Si*	*Pr*	*Pe*	*Fin*	*Ins*	*Ana*
LLC	−890.00*** (0.00)	−1080.00*** (0.00)	−560.00*** (0.00)	−650.00*** (0.00)	−850.00*** (0.00)	−320.00*** (0.00)	910.00*** (0.00)	1200.00*** (0.00)
IPS	−920.00*** (0.00)	−990.00*** (0.00)	−1500.00*** (0.00)	−38.00*** (0.00)	−490.00*** (0.00)	150.00*** (0.00)	−370.00*** (0.00)	−1600.00*** (0.00)

Note: *Dis* represents the absolute value of the correlation coefficient measuring excess co-movement in stock returns. *Mi* measures the strength of information diffusion using the mutual information approach. *Si* measures the level of difference in stock sizes. *Pr* measures the level of disparity in stock prices. *Pe* measures the level of similarity in stocks' P/E ratios. *Fin* measures the level of variation in stocks' financial leverage. *Ins* measures the common ownership by institutional investors in two stocks. *Ana* counts the number of analysts shared between two stocks.

Table 3
Results of selection of the optimal lag length *p* of the PVAR model.

	L.1	L.2	L.3	L.4	L.5	L.6
AIC	−11.2008	−11.2017	−11.2081*	−11.2075	−11.2058	−11.2042
BIC	−11.0801	−11.0804	−11.0864*	−11.0853	−11.083	−11.0809
HQIC	−11.1701	−11.1708	−11.1771*	−11.1764	−11.1745	−11.1728

Table 4
Results of the panel granger causality test.

Null Hypothesis	Test stat	*p*-value
Excess co-movement (*Dis*) does not cause information diffusion (*Mi*)	81.709	0.00
Information diffusion (*Mi*) does not cause excess co-movement (*Dis*)	564.54	0.00

Note: *Dis* represents the absolute value of the correlation coefficient measuring excess co-movement in stock returns. *Mi* measures the strength of information diffusion using the mutual information approach.

corresponding sub-forums. This outcome is consistent with the findings of Rizkiana et al. (2018) and Gao et al. (2020), who identified that stock price movements can capture investor attention and subsequently affect social media sentiment. The convenience of information interaction on platforms like stock forums may accelerate the propagation of such sentiment.

In conclusion, these results underscore the presence of a significant dynamic interrelationship between information diffusion among the corresponding sub-forums of stocks and the excess co-movement of stock returns.

3.2. Non-linear relationship between information diffusion and excess co-movement

Due to the inherent complexity and frictions present in real financial markets, there typically exists a complex non-linear relationship between stock prices and other financial variables (Ren, Cai, Li, Xiong, & Chen, 2023). In this section, we delve further into the non-linear association between information diffusion and excess co-movement, employing the Thermal Optimal Path (TOP) method. Initially introduced by Zhou and Sornette (2006), the TOP method was devised to explore lead-lag relationships between the S&P 500 index and U.S. Treasury bonds. TOP is a nonparametric estimation technique that transforms conventional economic inquiries into classical probabilistic

Table 5
Results of the PVAR model based on the GMM estimation.

Dependent variable: *Dis*

	Dis	*Mi*	*Si*	*Pr*	*Pe*	*Fin*	*Ins*	*Ana*
L.1	0.8424*** (1861.9472)	0.4398*** (61.0654)	0.0275*** (12.5715)	0.1108*** (27.9698)	0.0002 (1.0252)	0.0114*** (5.0468)	0.0001** (2.0509)	−0.0002*** (−2.0559)
L.2	0.0119*** (23.1295)	0.0455*** (8.4324)	−0.0338*** (−9.3761)	−0.1435*** (−23.7075)	0.0001 (0.1392)	−0.0011 (−0.4369)	0.0001 (0.0581)	0.0003** (2.3552)
L.3	0.0067*** (16.6185)	0.0280*** (5.3384)	0.0099*** (5.8294)	0.0469*** (17.8790)	0.0009*** (5.2374)	0.0080*** (4.0270)	0.0002*** (5.0223)	−0.0005*** (−5.1987)

Dependent variable: *Mi*

	Dis	*Mi*	*Si*	*Pr*	*Pe*	*Fin*	*Ins*	*Ana*
L.1	0.0015*** (45.2361)	0.9720*** (479.3063)	0.0023*** (14.0075)	0.0046*** (14.6989)	0.0000 (0.7554)	0.0016*** (11.9844)	0.0001*** (29.4382)	−0.0001*** (−2.9363)
L.2	0.0001*** (2.3074)	0.0800*** (39.9200)	−0.0026*** (−9.9967)	−0.0060*** (−12.4088)	0.0001*** (7.3148)	0.0002 (1.2098)	0.0001*** (0.3620)	−0.0001 (−1.2543)
L.3	0.0001*** (4.2172)	0.0406*** (21.6885)	0.0007*** (5.6738)	0.0019*** (8.8883)	0.0001*** (13.3920)	0.0003*** (2.8224)	−0.0001*** (−20.8897)	0.0001*** (3.7015)

Note: The PVAR model is represented as $Y_{it} = \sum A_p Y_{it-p} + \sum B_p X_{it-p} + \mu_i + \varepsilon_{it}$, where Y_{it} is a vector of endogenous variables, and X_{it} is a vector of exogenous variables. *Dis* represents the absolute value of the correlation coefficient measuring excess co-movement in stock returns. *Mi* measures the strength of information diffusion using the mutual information approach. *Si* measures the level of difference in stock sizes. *Pr* measures the level of disparity in stock prices. *Pe* measures the level of similarity in stocks' P/E ratios. *Fin* measures the level of variation in stocks' financial leverage. *Ins* measures the common ownership by institutional investors in two stocks. *Ana* counts the number of analysts shared between two stocks.

transfer models within the realm of statistical physics. This transformation is achieved by introducing a distance matrix between time series X and Y through recursive computations of the partition function. The TOP method's fundamental purpose is to reveal the dynamic lead-lag structure between time series X and Y, making it a widely adopted tool for investigating non-linear relationships among financial variables (Wang, Tu, Chang, & Li, 2017; Chen et al., 2020; Gao et al., 2020; Yang & Shao, 2020; Yao & Li, 2020; Ren et al., 2023; Chen, Ren, Yang, Lu, & Li, 2023).

Following the methodology outlined by Chen et al. (2023), we construct a matrix of 19 × 19 thermal paths, each characterized by distinct starting and ending points. Subsequently, we select the path with the minimum energy as the final heat path. To account for the influence of the temperature parameter (*Temp*) on the results, we compute lead-lag orders (*Ord*) between information diffusion and excess co-movement for Temp values of 5, 10, and 15. Specifically, *Ord* measures whether the sequence of information diffusion (*Mi*) precedes or lags behind the sequence of excess co-movement (*Dis*). A positive *Ord* indicates that *Mi* leads *Dis*, while a negative *Ord* suggests the opposite.

Fig. 1 illustrates the dynamic evolution of lead-lag orders between information diffusion and excess co-movement for various temperature values. Notably, these lead-lag orders exhibit variations with changes in *Temp*. When *Temp* decreases, the absolute magnitude of *Ord* increases, but the overall trend remains consistent. Therefore, for the sake of simplicity and to maintain generality, we fix *Temp* at a constant value of 10 for subsequent analysis. Additionally, from Fig. 1, it becomes apparent that the lead-lag order $Ord_{Temp_{10}}$ is not consistently positive or negative. This observation implies that information diffusion does not consistently dominate the excess co-movement of stock returns, nor does excess co-movement always exert dominance. Specifically, $Ord_{Temp_{10}}$ tends to be >0 before February 2020, signifying that social media information diffusion predominantly leads the excess co-movement of stock returns on most trading days during this period. However, post-February 2020, $Ord_{Temp_{10}}$ experiences a decline, reaching levels below 0, with a minimum value approaching −2. During episodes of external shocks in the stock market that lead to crashes, excess co-movement in stock returns takes on a leading role, causing social media information diffusion to lag behind changes in asset price behavior.

3.3. Multilayer network analysis of information diffusion and excess co-movement

In this section, we employ a multilayer network approach to delve into the characteristics of information diffusion and excess co-movement of stock returns. The objective is to gain an intuitive understanding of the dynamic evolution patterns of information diffusion among

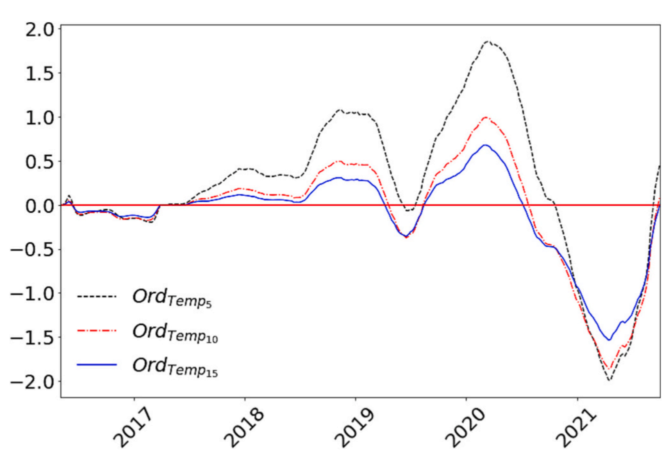

Fig. 1. Dynamic lead–lag orders between information diffusion and excess co-movement under various temperature values.

corresponding sub-forums of stocks and the excess co-movement among stock returns. The multilayer network approach serves as a powerful tool for comprehensively analyzing complex systems by simultaneously encompassing multiple distinct relationships. It provides an accurate representation of interactions among variables within real-world systems (Aldasoro & Alves, 2018; Musmeci, Nicosia, Aste, Matteo, & Latora, 2017).

Each week, we derive information diffusion or excess co-movement matrices denoted as C_t^α based on the measures of *Mi* or *Dis*, where $\alpha = 1$ and 2 represent the information diffusion matrix and excess co-movement matrix, respectively. Consequently, we map these matrices into the multilayer network $G_t = \{G_t^1, G_t^2\}$ comprising two layers, each with 429 nodes. $G_t^1 = G(V, E^1, W^1)$ represents the information diffusion layer, where $V = \{l, l = 1, 2, \ldots, 429\}$ is the set of nodes, $E^1 = \{(l,k), l = 1, 2, \ldots, 429, k = 1, 2, \ldots, 429, l \neq k\}$ represents the set of edges in the information diffusion layer, and $W^1 = (Mi_{lk}, l = 1, 2, \ldots, 429, k = 1, 2, \ldots, 429, l \neq k)$ denotes the set of weights for the edges in the information diffusion layer. Similarly, $G_t^2 = G(V, E^2, W^2)$ represents the excess co-movement layer, where $E^2 = \{(l,k), l = 1, 2, \ldots, 429, k = 1, 2, \ldots, 429, l \neq k\}$ is the set of edges in the excess co-movement layer, and $W^2 = (Dis_{lk}, l = 1, 2, \ldots, 429, k = 1, 2, \ldots, 429, l \neq k)$ represents the set of weights for the edges in the excess co-movement layer.

To eliminate redundant and noisy information in each layer, we employ the Planar Maximally Filtered Graph approach (PMFG) to filter the information diffusion and excess co-movement networks (Tumminello, Di Matteo, Aste, & Mantegna, 2007). This approach retains the most important and significant edges, resulting in the filtered multilayer network denoted as $G_{t,PMFG} = \left\{G^1_{t,PMFG}, G^2_{t,PMFG}\right\}$. Here, $G^\alpha_{t,PMFG} = G(V, E^\alpha_{PMFG}, W^\alpha_{PMFG})(\alpha = 1, 2)$ represents the filtered multilayer network for each week.

To analyze the topological properties of a filtered single-layer network for information diffusion or excess co-movement, we introduce three standard network measures: the average shortest path length (*Aspl*) among nodes, the average clustering coefficient (*Ac*) of nodes, and the global efficiency (*Ge*) of the network (Latora & Marchiori, 2001). Additionally, we apply two categories of multiplex measures to study the connectedness properties in the multilayer network. Following Szell, Lambiotte, and Thurner (2010), we employ the Pearson correlation coefficient $\rho\left(k^1, k^2\right)$ to assess the structural similarity of node degree series between the two filtered single-layer networks. As defined in Bianconi (2013), we also calculate the global overlap of edges denoted as $M^{voerlap}$, representing the total number of edges that appear simultaneously in both the information diffusion layer and the excess co-movement layer. We compute the ratio of edge overlap $O^{1,2} = M^{voerlap}/\left(\left(M^1 + M^2\right)/2\right)$, where M^α $(\alpha = 1, 2)$ is the total number of edges in layer α.

To gain a more intuitive understanding of the intricate relationship between information diffusion and the excess co-movement of stock returns, we initially offer a snapshot of the multilayer network for the 46th week. This multilayer network comprises two distinct layers: one representing information diffusion among corresponding sub-forums of stocks and the other depicting the excess co-movement among stocks, as shown in Fig. 2.

Upon closer examination of Fig. 2, it becomes apparent that both the information diffusion layer and the risk contagion layer exhibit certain similarities in terms of their network topology. In both layers, a select few nodes emerge as focal points with a notably higher number of connections. This phenomenon is particularly pronounced in the case of the information diffusion layer.

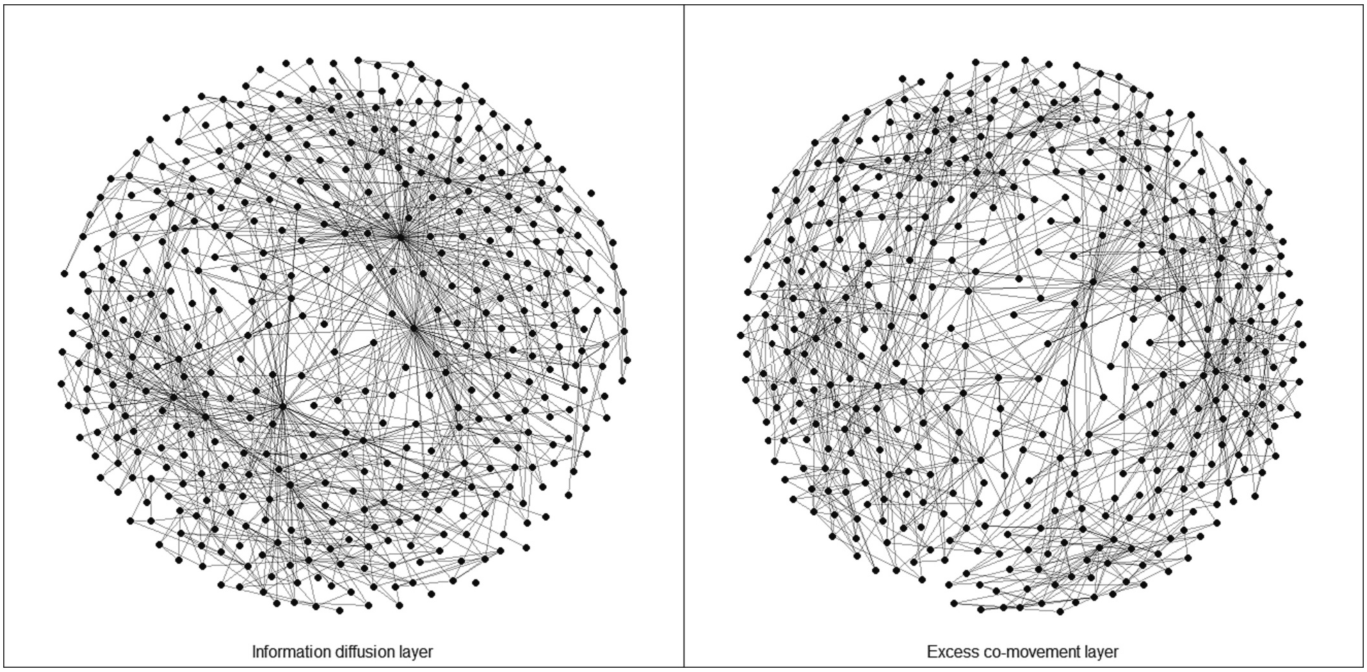

Fig. 2. Multilayer network includes an information diffusion layer (left) and an excess co-movement layer (right), utilizing data from the 46th week as an example. The multilayer network is constructed using data from the initial week, with nodes representing stocks in both layers. The information diffusion layer is established based on the strength of information diffusion within the corresponding sub-forums of stocks, while the excess co-movement layer is founded on the excess co-movement among stocks. Both the information diffusion and excess co-movement networks are subjected to filtering using the PMFG approach.

The node degree distributions of the information diffusion and excess co-movement networks, as illustrated in Fig. 2, follow power-law exponents of 2.78 and 3.5, respectively.[5] These findings suggest that both networks adhere to the characteristics commonly associated with scale-free networks. Notably, these results align with the observations made by Kim, Lee, Kahng, and Kim (2002), who also identified scale-free characteristics within financial correlation networks.

In comparison to the excess co-movement network, the information diffusion network exhibits higher efficiency. It showcases a more discernible network topology and boasts a smaller power-law exponent value. This distinction can be attributed to the fact that social media platforms offer individual investors a swifter and more convenient means of exchanging information. Consequently, this enhances the overall information environment for individual investors when contrasted with traditional word-of-mouth channels (Agarwal et al., 2019).

Furthermore, it's worth noting that for information diffusion among the corresponding sub-forums of stocks to exert an influence on the excess co-movement among stocks, an intermediate step comes into play. This intermediate step involves individual investors digesting and processing the disseminated information. It is likely that this intermediary phase is the primary factor contributing to the observed outcomes.

We proceed to illustrate the dynamic evolution of topological measures within both the information diffusion layer and the excess co-movement layer across the entire sampling period, as depicted in Fig. 3. In Fig. 3(a), we observe the average shortest path length ($Aspl^1$) of the information diffusion layer, which ranges between 1.5 and 3.5. In contrast, Fig. 3(b) presents the average shortest path length ($Aspl^2$) of the excess co-movement layer, spanning from 5 to 8. A lower average shortest path length implies a more efficient transmission of information or risk between nodes.

Moving to Fig. 3(c) and (d), these figures showcase the average clustering coefficients of the information diffusion layer (Ac^1) and the excess co-movement layer (Ac^2), respectively. Notably, Ac^1 fluctuates between 0.7 and 0.82, surpassing the range observed for Ac^2, which falls between 0.62 and 0.72. During crisis periods, especially within the excess co-movement layer, we observe an increase in clustering coefficients. Fig. 3(d) further highlights distinct spikes that correspond to highly significant events in the stock market, including the initial outbreak of the COVID-19 pandemic in early 2020. These findings align with those of Bartesaghi, Clemente, and Grassi (2022), who identified a similar pattern in the multiplex correlation network of financial assets during the latter half of 2008.

In Fig. 3(e) and (f), we explore the global efficiency (Ge^1) of the information diffusion network compared to that of the excess co-movement network. It becomes evident that the global efficiency of the information diffusion network surpasses that of the excess co-movement network throughout the sampling period. This indicates that the efficiency of information diffusion within the corresponding sub-forums of stocks exceeds that of risk contagion within the excess co-movement network, corroborating our earlier findings from Fig. 2.

Our investigation extends to examining the interconnected properties of the multilayer network throughout the entire sampling period. Fig. 4(a) and (b) present the dynamic evolution of two critical metrics: the correlation coefficient ($\rho\left(k^1, k^2\right)$) of node degree series and the ratio of edge overlap ($O^{1,2}$) within the information diffusion and excess co-movement layers, respectively.

Notably, the values of both $\rho\left(k^1, k^2\right)$ and $O^{1,2}$ exhibit pronounced fluctuations across the entire sampling period. Specifically, $\rho\left(k^1, k^2\right)$ fluctuates within the range of −0.04 to 0.3, while $O^{1,2}$ spans from 0.12 to 0.25. An intriguing trend emerges as $O^{1,2}$ gradually ascends from July 2017 to January 2019. Furthermore, both $\rho\left(k^1, k^2\right)$ and $O^{1,2}$ display a rapid increase during the first half of 2019 and the first half of 2020. It is worth noting that these fluctuations in the interconnectedness of the information diffusion and excess co-movement networks are likely

[5] The average power-law exponents of the degree distributions for the information diffusion and excess co-movement networks throughout the entire sampling period are 2.83 and 3.35, respectively.

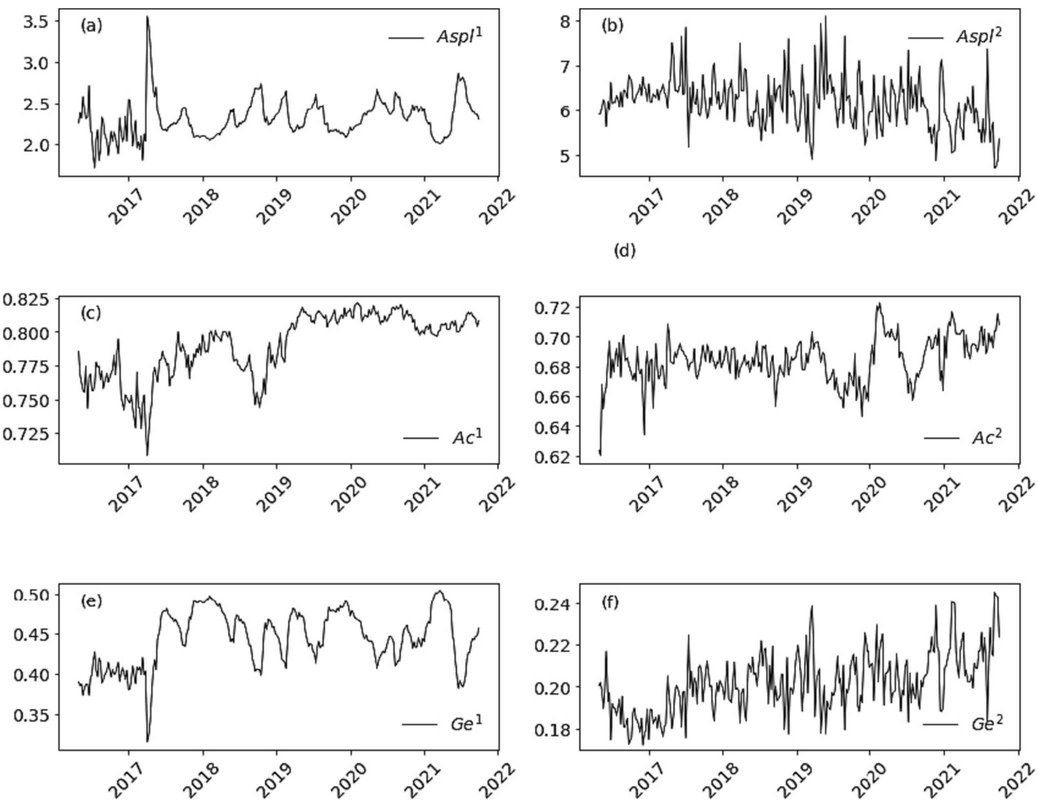

Fig. 3. Dynamic evolution of topological measures within both the information diffusion layer and the excess co-movement layer. (a) and (b) depict the changes in the average shortest path length of the information diffusion layer (*Aspl1*) and the excess co-movement layer (*Aspl2*) over time. (c) and (d) display the progression of average clustering coefficients for both layers, while (e) and (f) showcase the evolution of global efficiencies within the two layers.

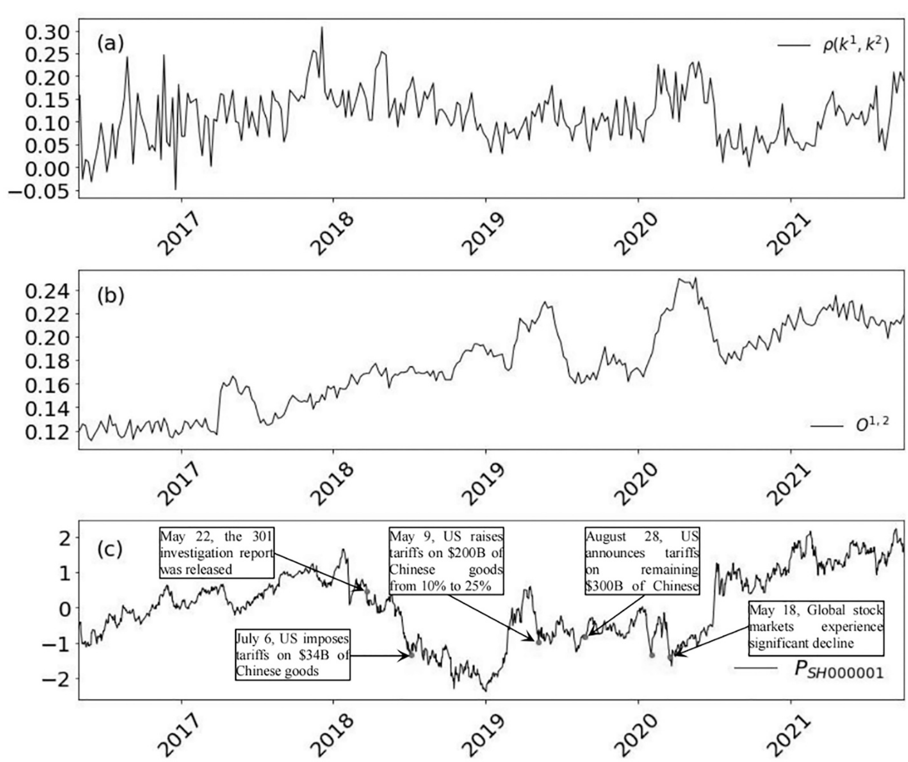

Fig. 4. Dynamic evolution of connectedness measures within the multilayer network alongside the SSE Composite Index SH000001. In particular, (a) provides insight into the evolving correlation coefficient of node degree series between the information diffusion layer and the excess co-movement layers. (b) presents the changing ratio of edge overlap between these two layers. Additionally, (c) displays the progressive evolution of the standardized closing prices of the SSE Composite Index SH000001.

attributable to the changing external environment of the stock market.

In Fig. 4(c), we provide insight into the evolution of the standardized closing prices of the Shanghai Securities (SSE) Composite Index SH000001. Multiple episodes of China-US trade frictions between 2018 and 2019 have played a pivotal role in inducing significant fluctuations within the Chinese stock market, as documented by Li, Zhuang, Wang, and Zhang (2020). Additionally, the onset of the COVID-19 pandemic in 2020 triggered multiple market crashes, impacting both the global and Chinese stock markets, as reported by Chen et al. (2023). These exogenous shocks have consistently captured the attention of investors, resulting in the widespread dissemination of information across social media platforms and imposing considerable stress on the financial system. It is noteworthy that Wang et al. (2018) have previously explored the relationship between stock market uncertainty or stress and volatility connectedness, arriving at a similar conclusion that such connectedness is sensitive to distress-induced factors.

3.4. Microcosmic mechanism of the dynamic interrelationship

In this section, we delve into the microcosmic mechanisms underlying the dynamic interplay between information diffusion and excess co-movement. Our focus shifts to the perspective of individual investors' information interaction behaviors within sub-forums. To comprehensively assess these behaviors, we construct four key indicators.

Firstly, we establish an information interaction network denoted as g_{lk}, utilizing data from the corresponding sub-forums of stocks l and k. In this network, all users of both sub-forums become nodes, and if user v replies to a post by user u, an edge is created from v to u. Within this network, we compute the average node degree (*Deg*) of co-investors actively participating in both stock forums. A higher average node degree indicates a quicker and wider dissemination of information across sub-forums.

Additionally, we calculate the number of tweets (*Pst*) generated by co-investors, shedding light on their overall activity levels. Furthermore, we assess the tweeting frequency of co-investors by measuring the time interval between two consecutive tweeting activities for each user, providing us with the average time intervals of co-investor activities (*Frq*) on a weekly basis. Increased tweet volume and higher tweeting frequency among individual investors facilitate the flow of information between the two forums, potentially attracting greater attention from investors and, consequently, strengthening the level of excess co-movement between the corresponding stocks.

Moreover, it's important to consider that the nature of the information diffused among sub-forums may lead to varying price correlation effects. If the information consistently conveys a positive or negative sentiment, it may cause the prices of the two corresponding stocks to move in the same direction, resulting in increased stock price correlation. Conversely, when individual investors express contrasting sentiments in the two sub-forums, it may lead to divergence between the stock prices of the two corresponding stocks. To explore the impact of information content on excess co-movement, we calculate the sentiment difference of co-investors between stocks i and j using data from the relevant sub-forums. We employ text mining techniques on co-investors' tweets in the sub-forums to gauge their sentiment towards the corresponding stocks.

We use the Cnsenti[6] Python library, which supports seven emotion statistics, including “good,” “happy,” “sad,” “angry,” “fear,” “disgust,” and “shock” (Fang, Jia, Li, & Liu, 2022). According to Werner and Murray (2004), the sentiment of individual investors (Sti_{it}) for stock l is defined as follows:

$$Sti_{lt} = (Pos_{lt} - Neg_{lt})/(Pos_{lt} + Neg_{lt}), \tag{11}$$

where Pos_{lt} represents the number of positive words in the text messages of the corresponding sub-forum of stock l in week t, which is equal to the sum of the counts of “good” and “happy” words. Neg_{it} denotes the number of negative words, which is equal to the sum of the counts of “sad,” “angry,” “fear,” “disgust,” and “shock” words. We then calculate the sentiment difference of co-investors towards stocks l and k as $Emo_{lk,t} = |Sti_{lt} - Sti_{kt}|$. A smaller value of *Emo* indicates greater consistency in investor sentiment towards the two stocks.

The frequent information interaction within online stock forums offers investors the convenience of observing the actions of others, a phenomenon that can easily trigger information cascades and amplify the herding effect (Bikhchandani et al., 1998; Musciotto, Marotta, Piilo, & Mantegna, 2018). When investors exhibit a strong herding effect in trading multiple stocks, it leads to excess co-movement in stock returns (Li, Cai, et al., 2019; Li, Zhang, et al., 2019). Additionally, there may be variations in investors' trading decisions (Musciotto et al., 2018) due to the diversity of information available to them, which can either increase or decrease the correlation between two stocks resulting from the herding effect.

To investigate the impact of individual investors' information interaction behaviors on the herding effect, we construct an indicator to measure the consistency of investors' trading behavior on stocks l and k. Following the methodology of Kumar and Lee (2006), we first calculate a daily buy-sell imbalance indicator to assess the movements of individual investors entering and exiting stock l:

$$Ibm_{lt'} = (Vb_{lt'} - Vs_{lt'})/(Vb_{lt'} + Vs_{lt'}), \tag{12}$$

where $Vb_{lt'}$ represents the buy volume of individual investors on day t' for stock l, and $Vs_{lt'}$ is the sell volume of individual investors. The consistency of investors' trading behavior between stocks l and k ($Bsi_{lk,t}$) in week t equals the Pearson correlation coefficient of $Ibm_{lt'}$ and $Ibm_{kt'}$. A higher value of $Bsi_{lk,t}$ indicates that individual investors exhibit greater consistency in their trading behavior across the two stocks. Similar to the calculation of excess co-movement, we also apply the rolling-window method to calculate $Bsi_{lk,t}$.

3.4.1. Impact of excess co-movement on information interaction behaviors

To gain insights into how changes in the level of excess co-movement of stock returns affect individual investors' information interaction behaviors on social media, we begin by examining their responses. We categorize pairs of stocks into high and low co-movement groups based on their level of excess co-movement (*Dis*) for each week. The high co-movement group consists of pairs with *Dis* values exceeding the 99th percentile value among all pairs, while the low co-movement group comprises pairs with *Dis* values falling below the 1st percentile value among all pairs. Subsequently, we calculate the average values of *Deg*, *Pst*, and *Frq* for all pairs within the high and low co-movement groups.

Additionally, we partition all pairs of stocks into positive and negative correlation groups based on their correlation coefficient (*Cor*). The positive correlation group comprises pairs with *Cor* values greater than the 99th percentile value among all pairs, while the negative correlation group consists of pairs with *Cor* values lower than the 1st percentile value among all pairs. We then calculate the average sentiment difference for all pairs within the high and low correlation groups.

Fig. 5 illustrates the dynamic evolution of various indicators reflecting co-investors' information interaction behaviors within different groups. In Fig. 5(a), we observe the average node degree of co-investors in both high (Deg^{high}) and low (Deg^{low}) co-movement groups. Notably, the average node degree of co-investors in the high co-movement group (Deg^{high}) is substantially higher than that in the low co-movement group (Deg^{low}). This discrepancy suggests that within the high co-movement group, co-investors in the two sub-forums have a greater number of neighboring nodes within the information interaction network. These nodes play a pivotal role in facilitating the dissemination of information between the sub-forums. It's worth mentioning that the

[6] The project description of Jieba: https://github.com/hiDaDeng/cnsenti.

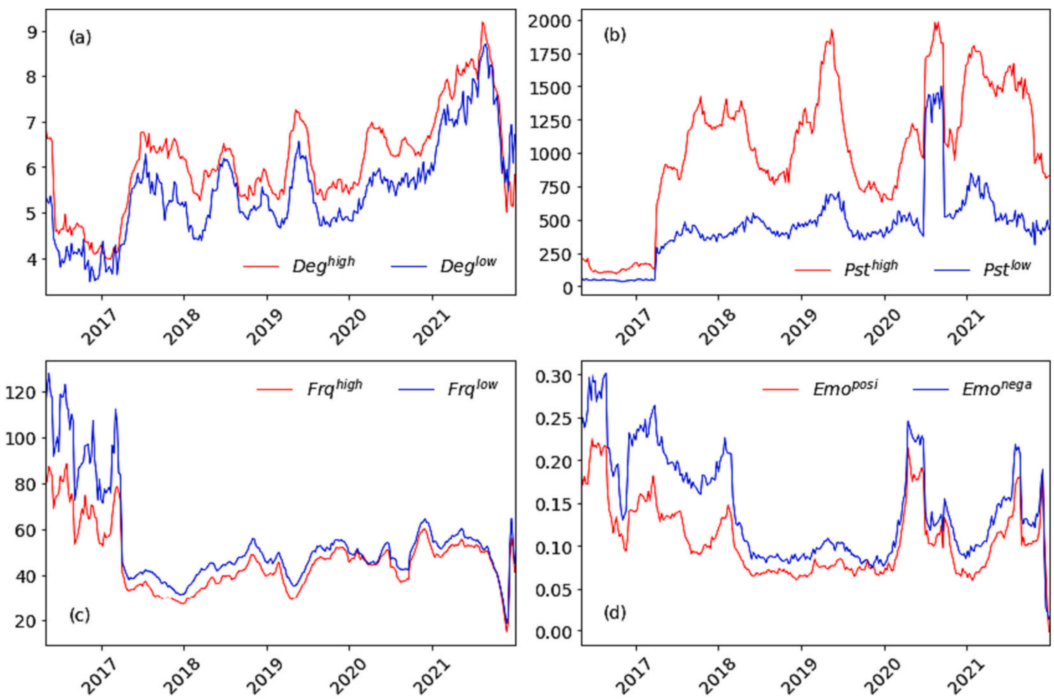

Fig. 5. Dynamic evolution of indicators reflecting co-investors' information behavior within different groups. Specifically, (a) showcases the average node degree of co-investors involved in pairs from high (Deg^{high}) and low (Deg^{low}) co-movement groups. In (b), we examine the tweet count from co-investors associated with pairs in high (Pst^{high}) and low (Pst^{low}) co-movement groups. (c) offers insight into the tweeting frequency in hours among co-investors connected to pairs in high (Frq^{high}) and low (Frq^{low}) co-movement groups. Lastly, (e) delves into the sentiment difference exhibited by co-investors within positive (Emo^{posi}) and negative (Emo^{nega}) correlation groups.

values of Deg^{high} and Deg^{low} exhibit a noticeable upward trend after February 2020, likely attributed to the shocks caused by the COVID-19 pandemic, leading to multiple stock market crashes. The heightened excess co-movement of stock returns during these market crises likely attracted investors' attention and sparked increased discussions, thus reshaping the structure of the information interaction network. These findings align with the results in Fig. 4.

In Fig. 5(b), we examine the number of tweets from co-investors in both high (Pst^{high}) and low (Pst^{low}) co-movement groups. Here, too, we find that Pst^{high} is significantly higher than Pst^{low}, with both indicators exhibiting a pronounced upward trend during the COVID-19 outbreak. Fig. 5(c) portrays the tweeting frequency in hours of co-investors in high (Pst^{high}) and low (Pst^{low}) co-movement groups. In the high co-movement group, the time interval between tweeting activities by co-investors is the shortest, fluctuating between 18 and 130 h throughout the entire sampling period. Furthermore, Frq^{high} and Frq^{low} display a clear downward trend from February 2020 to August 2022. These observations are consistent with the trends seen in Fig. 5(a) and (b), indicating that higher excess co-movement between two stocks is associated with co-investors having a larger number of neighboring nodes, increased tweet volume, and higher tweeting frequencies.

Fig. 5(e) presents the sentiment difference among co-investors in positive (Emo^{posi}) and negative (Emo^{nega}) correlation groups. Notably, the sentiment difference among co-investors in the positive correlation group is substantially lower than that in the negative correlation group. When co-investors share consistent sentiment regarding two stocks, this sentiment can propagate through information diffusion and influence other users across stock forums. This intensifies the herding effect among investors in both stocks, leading to a stronger positive correlation between the two stocks. Interestingly, there is a clear upward trend in investors' sentiment differences during the initial phase of the COVID-19 outbreak, as shown in Fig. 5(e). During this bearish market phase induced by the COVID-19 pandemic, divergent opinions among investors regarding the future trends of different stocks may have been a key factor contributing to these results.

In summary, individual investors' information interaction is influenced by the behavior of asset prices. When there is high excess co-movement in stock returns, co-investors who play a pivotal role in information diffusion tend to exhibit a higher degree of connectivity within the information interaction network, as well as increased tweet volume and frequency. Moreover, when there is a high positive correlation between stock returns, co-investors also tend to express more consistent sentiments towards both stocks.

3.4.2. *Impact of information interaction on herding behavior*

In this section, we delve deeper into whether the frequent information interaction behaviors of co-investors on social media influence individual investors' trading decisions. We categorize pairs of stocks into high and low node degree groups based on the average node degree of co-investors (*Deg*). Pairs in the high node degree group possess *Deg* values exceeding the 99th percentile value among all pairs, while those in the low node degree group have *Deg* values falling below the 1st percentile value. We subsequently compute the average consistency (*Bsi*) of investors' trading behavior towards pairs of stocks in both the high and low node degree groups. Likewise, we investigate the impact of tweet quantity, tweet frequency, and sentiment difference among co-investors on the consistency of investors' trading behavior.

Fig. 6(a) illustrates the consistency of trading behavior towards pairs in the high (Bsi_{Deg}^{high}) and low (Bsi_{Deg}^{low}) node degree groups. It becomes apparent that, in comparison to the low node degree group, the consistency of trading behavior towards pairs in the high node degree group is notably higher. These findings imply that when co-investors have a greater number of neighboring nodes within the information interaction network, individual investors tend to exhibit a heightened herding effect in their trading behavior across two stocks. This behavior may lead to an increased level of excess co-movement. Fig. 6(b) and (c) also exhibit similar trends, indicating that higher tweet volume and frequency among co-investors in two sub-forums result in larger *Bsi* values and

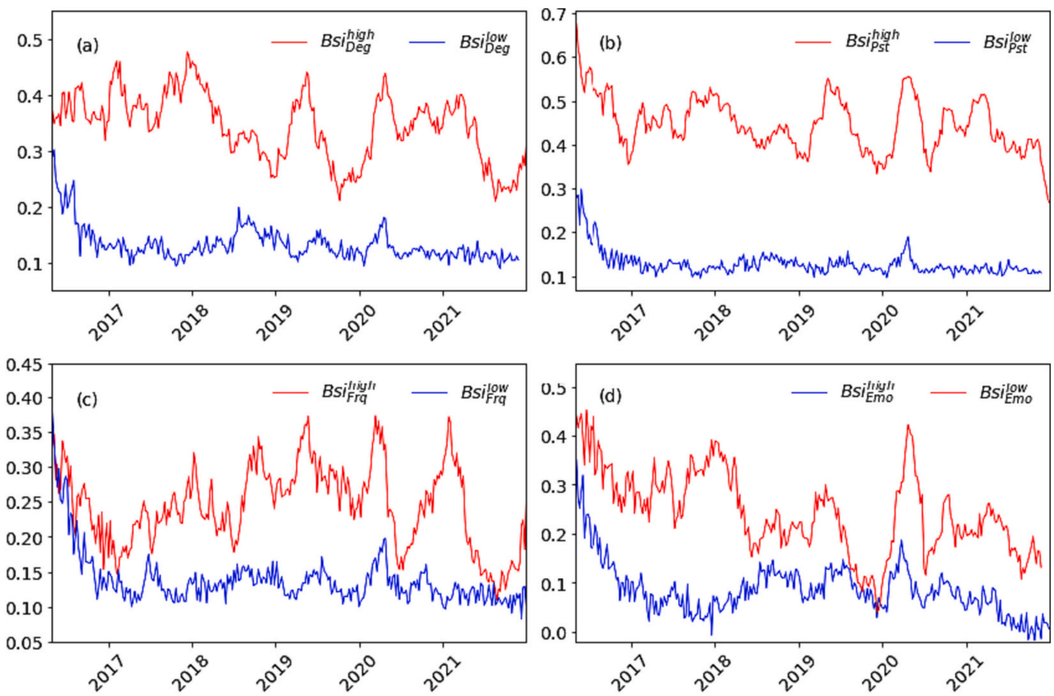

Fig. 6. Dynamic evolution of the consistency observed in individual investors' trading behavior concerning two stocks within different groups: (a) presents the consistency of trading behavior towards pairs in high (Bsi^{high}_{Deg}) and low (Bsi^{low}_{Deg}) node degree groups. (b) shows the consistency of trading behavior towards pairs in high (Bsi^{high}_{Pst}) and low (Bsi^{low}_{pst}) tweeting quantity groups. (c) depicts the consistency of trading behavior towards pairs in high (Bsi^{high}_{Frq}) and low (Bsi^{low}_{Frq}) tweeting frequency groups. (d) shows the consistency of trading behavior towards pairs in high (Bsi^{high}_{Emo}) and low (Bsi^{low}_{Emo}) sentiment difference groups.

greater consistency in their trading behavior across the two corresponding stocks.

Fig. 6(d) delves into the consistency of trading behavior towards pairs in high (Bsi^{high}_{Emo}) and low (Bsi^{low}_{Emo}) sentiment difference groups. It's evident that, in comparison to the high sentiment difference group, the *Bsi* values for pairs in the low sentiment difference group are higher. These results suggest that when there is minimal difference in sentiment among co-investors regarding two stocks, individual investors tend to exhibit more consistent trading behavior. Additionally, Fig. 6 highlights a significant increase in *Bsi* values during market crash phases, such as the early 2020s. Combining the observations from Fig. 5, we find that both co-investors' information interaction behaviors and individual investors' trading behavior are sensitive to external market shocks. This sensitivity may contribute to a more intricate interplay between social media information diffusion and excess stock returns co-movement, leading to a non-linear lead-lag relationship.

3.5. Portfolio strategies

In the era of social media's ascendancy as a potent communication channel, participants in financial markets have found themselves endowed with a veritable trove of information and an augmented ability to govern their financial decisions and trading activities (Agarwal et al., 2019). Concurrently, the diffusion of information on social media has the potential to augment market efficiency by mitigating information asymmetries (Gu, Konana, Liu, Rajagopalan, & Ghosh, 2006) and accelerating asset price responses to information, thus exerting an influence on the performance of investors' portfolio strategies (Curme, Preis, Stanley, & Moat, 2014).

In this subsection, we devise a hedging strategy for both stocks and stock index futures, predicated on the vigor of information diffusion across sub-forums. We commence by computing the information diffusion vigor for stock l using the following formula:

$$Mi_l = \frac{1}{T}\sum\nolimits_t^T \left(\frac{1}{K}\sum\nolimits_{k,k\neq l}^K Mi_{lk,t}\right), \tag{13}$$

Where $Mi_{lk,t}$ denotes the vigor of information diffusion between the two corresponding sub-forums associated with stocks l and k. To assess whether heightened information diffusion vigor enhances pricing efficiency, we adhere to the approach of Hou and Moskowitz (2005) by employing a price delay indicator (Dly_l) to quantify the speed at which stock l's prices respond to information. Smaller Dly_l values correspond to enhanced pricing efficiency.

The formulation of the hedging strategy draws upon the methodologies of Kroner and Ng (1998), Wen, Cao, Liu, and Wang (2021), and Chen, Liang, Ding, and Liu (2022). The optimal hedging weight is expressed as follows:

$$w_{lf,t} = \frac{h_{ff,t} - h_{lf,t}}{h_{ll,t} - 2h_{lf,t} + h_{ff,t}}, \tag{14}$$

where, $w_{lf,t}$ signifies the optimal weight of stock l in the hedging portfolio of two assets valued at \$1 at time t. $h_{ll,t}$ ($h_{ff,t}$) denotes the conditional variance of stock l (stock index futures f) itself, while $h_{lf,t}$ represents the conditional variance of stock l and stock index futures f. $1 - w_{lf,t}$ represents the optimal weight of stock index futures f. If $w_{lf,t} < 0$, then $w_{lf,t} = 0$; and if $w_{lf,t} > 1$, then $w_{lf,t} = 1$. Consistent with Kroner and Sultan (1993), we calculate the minimum hedge ratio between stock l and stock index futures f as $\beta_{lf,t} = h_{ll,t}/h_{ff,t}$. We also adhere to the approach of Antonakakis, Cunado, Filis, Gabauer, and Gracia (2020) to measure the hedging effectiveness (*He*) of the hedging portfolio.

To assess the hedging performance of all sampled stocks across various stock index futures, including the CSI 500 Stock Index Futures (IC500), the CSI 300 Stock Index Futures (IF300), and the Shanghai 50 Stock Index Futures (IH50), we categorize stocks into high and low information diffusion groups based on the vigor of information diffusion for each stock, denoted as Mi_l. Stocks in the high (low) information diffusion group have Mi_l values surpassing (falling below) the 90th (10th) percentile threshold of all stocks.[7]

Fig. 7(a) illustrates the average cumulative returns ($Cumret_{IC500}^{high}$) of the hedging portfolio encompassing all stocks within the high information diffusion group, utilizing the IC500 for hedging. Conversely, Fig. 7(b) portrays the average cumulative returns ($Cumret_{IC500}^{low}$) of the hedging portfolio that comprises all stocks belonging to the low information diffusion group, with the IC500 as the hedging instrument. It is evident that, in comparison to the low information diffusion group, stocks in the high information diffusion group exhibit superior performance when hedging with the IC500. The cumulative returns of the hedging portfolio display a discernible upward trajectory throughout the entire sampling period. The values of $Cumret_{IC500}^{high}$ consistently remain above zero, with fluctuations ranging from −0.07 to 0.38.

Moving on to Fig. 7(c) and (d), they showcase the average cumulative returns of the hedging portfolio for all stocks within the high ($Cumret_{IF300}^{high}$) and low ($Cumret_{IF300}^{low}$) information diffusion groups, respectively, utilizing the IF300 for hedging. Similarly, Fig. 7(e) and (f) portray the average cumulative returns of the hedging portfolio for all stocks within the high ($Cumret_{IH50}^{high}$) and low ($Cumret_{IH50}^{low}$) information diffusion groups, with the IH50 as the hedging instrument. Consistent with the observations in Fig. 7(a) and (b), stocks in the high information diffusion group exhibit enhanced performance when hedging with stock index futures. Furthermore, in comparison to the IF300 and IH50, stocks within the high information diffusion group demonstrate the most favorable performance within the hedging portfolio involving the IC500, as depicted in Fig. 7(a), (c), and (e).

Table 6 presents the average values of various measures pertaining to the hedging portfolio. Notably, the average price delay measure (Del) for stocks in the high information diffusion group stands significantly lower than that of the low information diffusion group at the 1% level, with respective values of 0.7631 and 0.831. These findings indicate that stocks within the high information diffusion group react more promptly to information and demonstrate heightened pricing efficiency, which, in turn, enhances the performance of the hedging portfolio.

The average $Cumret_{IC500}^{high}$ throughout the entire sampling period amounts to 0.1794, a figure significantly larger than the average $Cumret_{IC500}^{low}$, which registers at −0.3228. A parallel pattern is observed in the hedging portfolios involving stocks with IF300 and IH50. Notably, only the average $Cumret_{IC500}^{high}$ is the highest and positive among these observations.

Within Table 6, the average effectiveness (*He*) of the hedging portfolio involving IC500 and stocks in the high information diffusion group surpasses that of the low information diffusion group at the 10% level. Specifically, the values are 0.2762 and 0.2388, respectively. Comparable trends are evident in the hedging portfolios of stocks with IF300 and IH50. These results imply that, relative to the low information diffusion group, stocks in the high information diffusion group can achieve better risk mitigation by investing in stock index futures.

Further, Table 6 reveals that the average minimum hedging ratio (β) between stocks in the high information diffusion group and IC500 stands at 1.6870. This implies that maintaining a long position of $1 in the stock market necessitates a corresponding $1.6870 in the IC500 futures market for hedging purposes. Notably, the average minimum hedging ratio in the hedging portfolio involving stocks and IC500 is lower than those in the hedging portfolios involving stocks with IF300 and IH50.

In summary, stocks characterized by a stronger information diffusion capability exhibit heightened responsiveness to information and yield superior results within a hedging strategy involving IC500 stock index futures.

3.6. Robustness tests

In Section 3.3, we investigate the microcosmic mechanism of the interrelationship between information diffusion and excess co-movement, focusing on the information interaction behaviors of sub-forums' co-investors. To validate and fortify our primary findings, we also explored the impact of the information interaction behaviors of all users within sub-forums on this interrelationship. This approach ensures the robustness and credibility of our conclusions. We computed metrics such as *Deg, Pst, Frq*, and Emo for all users, determining the average *Deg*, average *Pst*, and average *Frq* for all pairs in the high (low) co-movement group, as well as the average Emo for all pairs in the high (low) correlation group. Additionally, we assessed how *Deg, Pst, Frq*, and *Emo* for all users influenced the consistency of investors' trading behavior.

Fig. 8 provides a dynamic overview of the indicators of all users' information behavior in different groups. It's evident that, in some trading days, the average *Deg* and average *Pst* of pairs in the high co-movement group are slightly higher than those in the low co-movement group, as depicted in Fig. 8(a) and (b). However, these differences are not statistically significant. Fig. 8(c) also illustrates that there is no significant contrast in the average *Frq* between pairs in the high and low co-movement groups. This observation aligns with the idea that co-investors between two sub-forums primarily drive information diffusion, potentially explaining why the information interaction behavior of co-investors is more responsive to excess co-movement. As shown in Fig. 8(d), in accordance with the results in Fig. 5(d), the average *Emo* of pairs in the high correlation group is consistently lower than that of pairs in the low correlation group on most trading days.

Fig. 9 illustrates the dynamic evolution of the consistency of individual investors' trading behavior towards two stocks in different groups, as based on the information interaction behaviors of all users. Mirroring the outcomes in Fig. 5, Fig. 9(a)–(c) demonstrate that higher node degrees, tweeting quantities, and tweeting frequencies of all users in two sub-forums lead to greater values of *Bsi*, indicating more consistent trading behavior across the two corresponding stocks. Furthermore, Fig. 9(d) reveals that the values of *Bsi* are higher in pairs from the low sentiment difference group. These results reaffirm that smaller differences in co-investors' sentiments towards two stocks lead to more consistent trading behavior among individual investors. To provide statistical validation for the findings in Figs. 5, 6, 8, and 9, *t*-tests were conducted, and the results are presented in Tables 7 and 8. These t-test results are largely consistent with those observed in Figs. 5, 6, 8, and 9.

These robustness tests further confirm the reliability and stability of our conclusions regarding the interplay between information diffusion, excess co-movement, and investors' trading behavior.

4. Conclusion

The rapid evolution of the Internet has profoundly transformed how investors share information, giving rise to an intricate interplay between information dissemination through social media and asset price dynamics in financial markets. In this paper, we introduce an innovative measure of information diffusion strength employing the information entropy method. We empirically investigate the intricate relationship between social media information diffusion and the excess co-movement of stock returns. We delve into the microcosmic mechanism of this relationship by examining the information interaction behaviors of individual investors on sub-forums and their subsequent trading decisions. Furthermore, we assess the performance of hedging strategies among stocks characterized by varying strengths of information diffusion and various stock index futures.

[7] We have further organized all stocks into groups using threshold values of 80% and 20%, and the outcomes are detailed in Fig. A1 and Table A1, available in the Appendix. These supplementary findings closely align with the results showcased in Fig. 7 and Table 6.

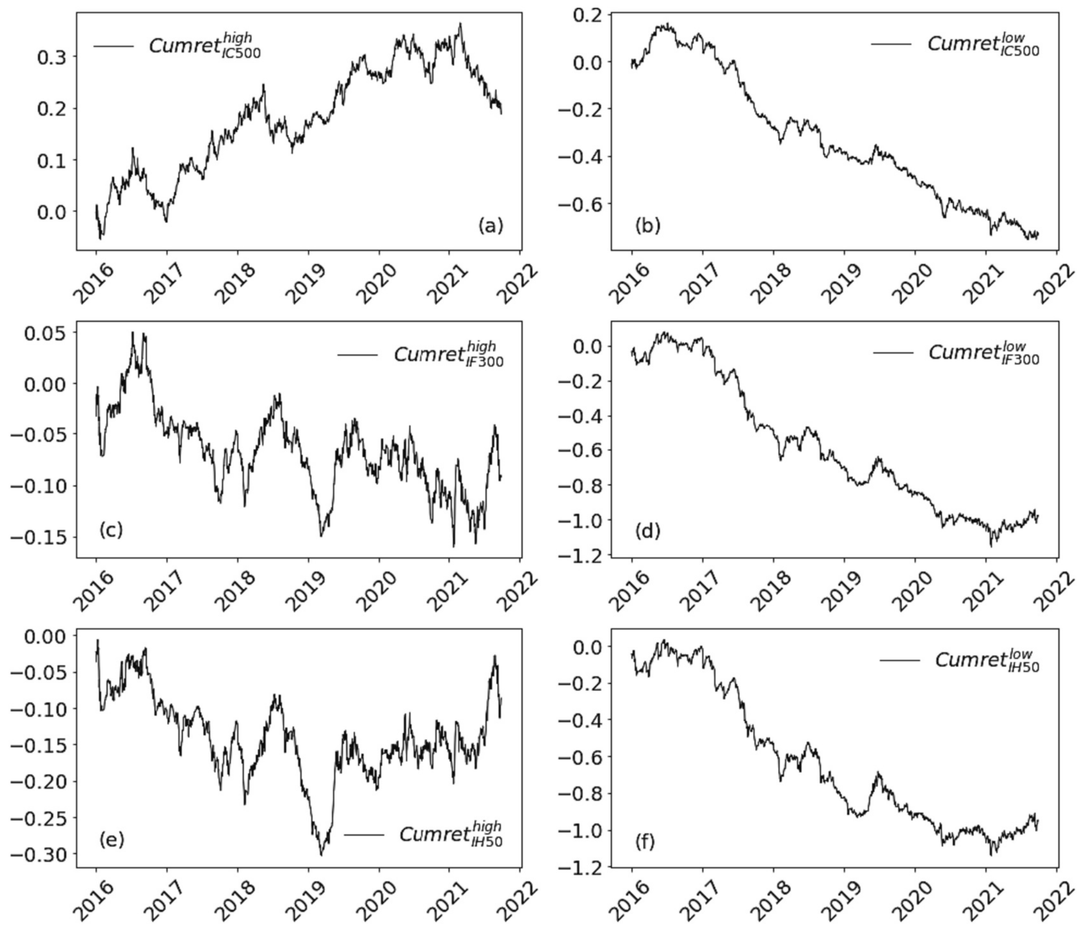

Fig. 7. Dynamic progression of cumulative returns in the hedging portfolio: (a) provides insights into the hedge portfolio's performance involving the IC500 and stocks within the high information diffusion group. (b) presents the corresponding outcomes for the low information diffusion group. The results for the hedge portfolio between the IF300 and stocks in the high information diffusion group and the low information diffusion group are presented in (c) and (d), respectively. (e) and (f) display the results for the hedge portfolio involving the IH50 and stocks within the high and low information diffusion groups, respectively.

Table 6
Average values of different measures of hedging portfolio.

	Stock l/IC500		Stock l/IF300		Stock l/IH50	
	High information diffusion group	Low information diffusion group	High information diffusion group	Low information diffusion group	High information diffusion group	Low information diffusion group
Mi	1.9171*** (22.6720)	0.4997				
Dly	0.7631*** (−2.5722)	0.8310				
Cumret	0.1794*** (5.3268)	−0.3228	−0.0663*** (5.4226)	−0.5823	−0.1411*** (5.1724)	−0.6323
He	0.2762* (1.8239)	0.2388	0.2881*** (4.7469)	0.1806	0.2802*** (5.8006)	0.1209
β	1.6870* (−1.8336)	1.8613	2.0562*** (4.1764)	1.7420	1.9739*** (7.1947)	1.4534

Note: *Mi* denotes the strength of information diffusion of stocks, *Del* denotes the price delay of stocks, *He* denotes the hedging effectiveness of hedging portfolio, *Cumret* denotes the cumulative returns of hedging portfolio, and β denotes the minimum hedge ratio between stocks and stock index futures. We perform the *t*-test with the null hypothesis that the measure of high information diffusion group is larger or smaller than that of low information diffusion group. ***, ** and * denote the 1%, 5% and 10% significance level respectively. The values in parentheses represent the *t*-statistic.

The principal findings of our study are as follows: (1) Empirical results from the PVAR model reveal a significant and lagged mutual relationship between the strength of information diffusion among corresponding sub-forums and the excess co-movement of stock returns. This relationship persists even after multiple lags. Moreover, the nonlinear connection between information diffusion and excess co-movement underscores that neither always dominates the other. (2) The topology metrics and connectedness of information diffusion and excess co-movement networks exhibit noticeable temporal fluctuations. Financial stress events can enhance the interconnectedness of these networks. Notably, the global efficiency of the information diffusion network surpasses that of the excess co-movement network. (3) In stock forums where excess co-movement between corresponding stocks is relatively high, co-investors tend to have a more extensive network of interactions, engage in higher tweet volumes, and exhibit increased tweeting frequency. Additionally, when stock returns display a high positive correlation, co-investors express more consistent sentiments towards both stocks. (4) Co-investors with larger followings in sub-forums, higher tweet volumes, greater tweeting frequency, and smaller sentiment disparities between two stocks are more likely to

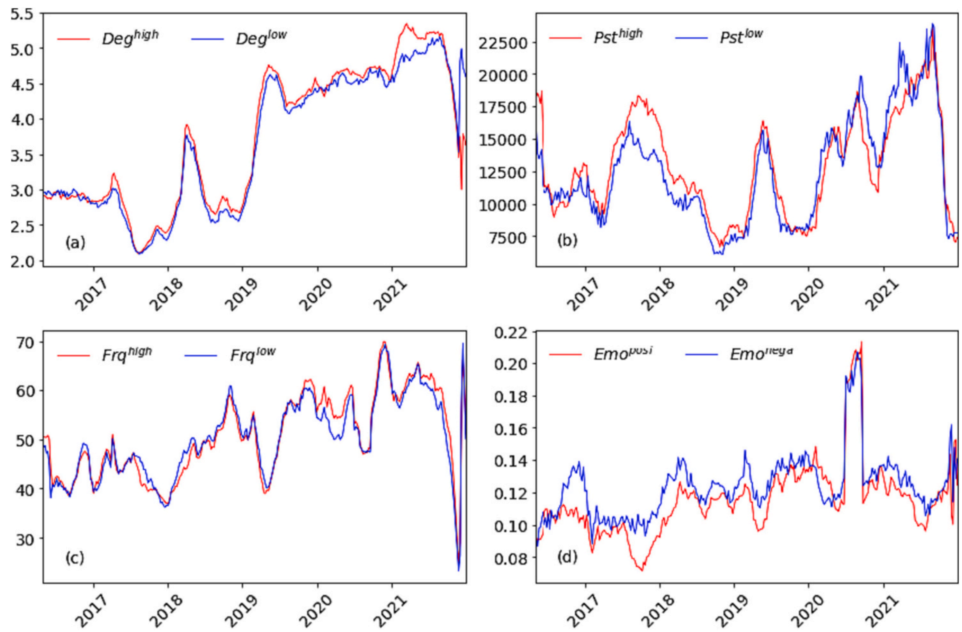

Fig. 8. Dynamic evolution of indicators related to the information behavior of all users within different groups: (a) presents the average node degree of all users participating in pairs within the high (Deg^{high}) and low (Deg^{low}) co-movement groups. (b) shows the number of tweets made by all users involved in pairs within high (Pst^{high}) and low (Pst^{low}) co-movement groups. (c) presents the tweeting frequency in hours exhibited by all users involved in pairs within high (Frq^{high}) and low (Frq^{low}) co-movement groups. (d) presents the sentiment difference among all users engaged in pairs within positive (Emo^{posi}) and negative (Emo^{nega}) correlation groups.

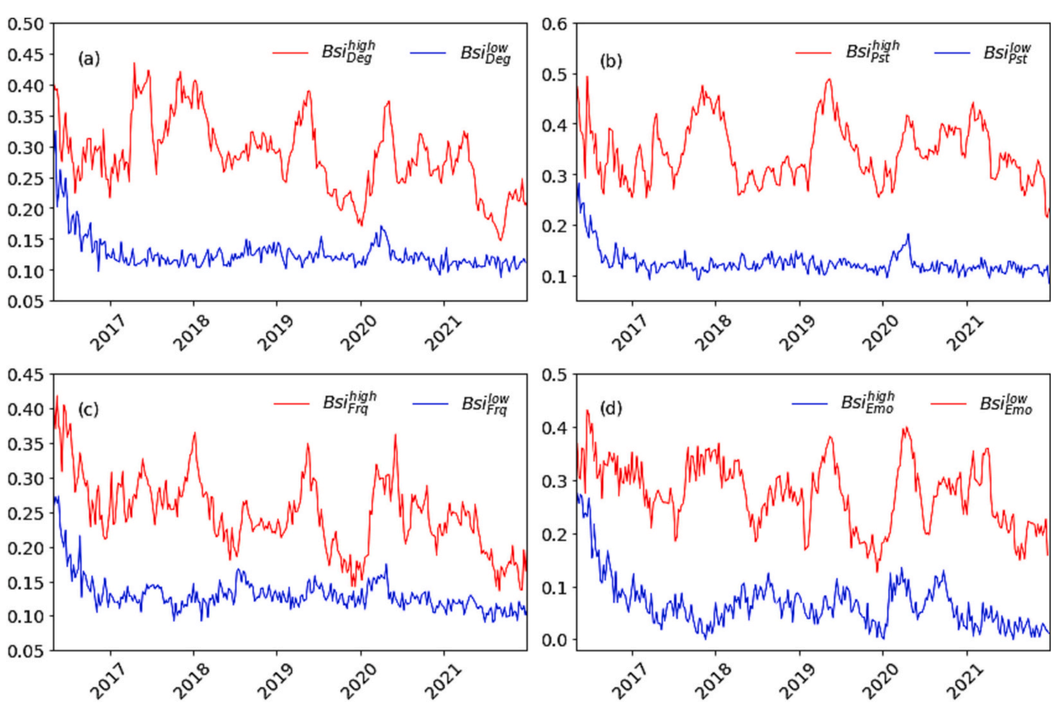

Fig. 9. Dynamic evolution of the consistency of individual investors' trading behavior towards two stocks in different groups based on the information interaction behaviors of all users: (a) illustrates the consistency of trading behavior towards pairs in high (Bsi_{Deg}^{high}) and low (Bsi_{Deg}^{low}) node degree groups. (b) shows the consistency of trading behavior towards pairs in high (Bsi_{Pst}^{high}) and low (Bsi_{pst}^{low}) tweeting quantity groups. (c) depicts the consistency of trading behavior towards pairs in high (Bsi_{Frq}^{high}) and low (Bsi_{Frq}^{low}) tweeting frequency groups. (d) shows the consistency of trading behavior towards pairs in high (Bsi_{Emo}^{high}) and low (Bsi_{Emo}^{low}) sentiment difference groups.

demonstrate heightened consistency in their trading behavior across the two stocks. (5) Stocks characterized by stronger information diffusion exhibit faster responsiveness to information and perform more effectively in hedging strategies involving the IC500 stock index futures. (6) Robustness tests confirm the consistency of our results when examining the information interaction behaviors of all users in sub-forums.

Our findings hold practical significance for market participants and regulators. Investors can promptly adjust their investment strategies

Table 7
Results of *t*-tests on indicators of information behavior in different groups.

	Co-investors			All users		
	Deg	*Pst*	*Frq*	*Deg*	*Pst*	*Frq*
High co-movement group	6.10*** (6.90)	1019.51*** (16.74)	46.50*** (−5.42)	3.74*** (8.90)	12,993.99** (2.58)	50.19 (0.56)
Low co-movement group	5.50	443.09	54.70	3.65	12,459.85	50.60

	Co-investors	All users
	Emo	*Emo*
Positive correlation group	0.11*** (−4.69)	0.10** (−8.60)
Negative correlation group	0.13	0.15

Note: The average values of the indicators during the entire sampling period are given in this table. We perform the *t*-test with the null hypothesis that the indicator of high co-movement group (positive correlation group) is larger or smaller than that of low co-movement group (negative correlation group). ***, ** and * denote the 1%, 5% and 10% significance level respectively. The values in parentheses represent the *t*-statistic.

Table 8
Results of *t*-tests on the consistency of individual investors' trading behavior towards two stocks in different groups.

	Co-investors				All users			
	node degree	tweeting quantity	tweeting frequency	sentiment difference	node degree	tweeting quantity	tweeting frequency	sentiment difference
High	0.3454*** (52.0361)	0.4488*** (81.1631)	0.2434*** (25.1299)	0.0865*** (−26.0984)	0.2862*** (51.8780)	0.3461*** (81.1631)	0.2483*** (32.4374)	0.0685*** (−43.2944)
Low	0.1332	0.1286	0.1379	0.2946	0.1264	0.1255	0.1309	0.2764

Note: The average values of the indicators during the entire sampling period are given in this table. We perform the *t*-test with the null hypothesis that the indicator of high group is larger or smaller than that of low group. ***, ** and * denote the 1%, 5% and 10% significance level respectively. The values in parentheses represent the *t*-statistic.

based on the strength of information diffusion in corresponding subforums of stocks, thus optimizing their hedging and arbitrage opportunities. By leveraging our hedging strategies, investors can further fine-tune their investment portfolios in response to varying market conditions. For regulators, our results underscore the substantial influence of social media information diffusion on the excess co-movement of stock returns, particularly in response to exogenous shocks such as the COVID-19 outbreak. This highlights the importance of establishing a risk warning mechanism for the stock market based on the activities of individual investors on social media platforms and prevailing market conditions.

Future research endeavors will likely focus on predicting the interrelationship between social media information diffusion and the excess co-movement of stock returns, providing further insights into the dynamics of financial markets in the digital age.

Ethical statements

This article does not contain any studies with human participants.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Declaration of Competing Interest

The author(s) declare no competing interests.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

This work was partially supported by the National Natural Science Foundation (No. 72201003), the Anhui Provincial Philosophy and Social Science Planning Project (No. AHSKQ2022D027), and the Anhui Province Social Science Innovation and Development Research Project (No. 2022CX031). All authors contributed equally to this manuscript.

Appendix

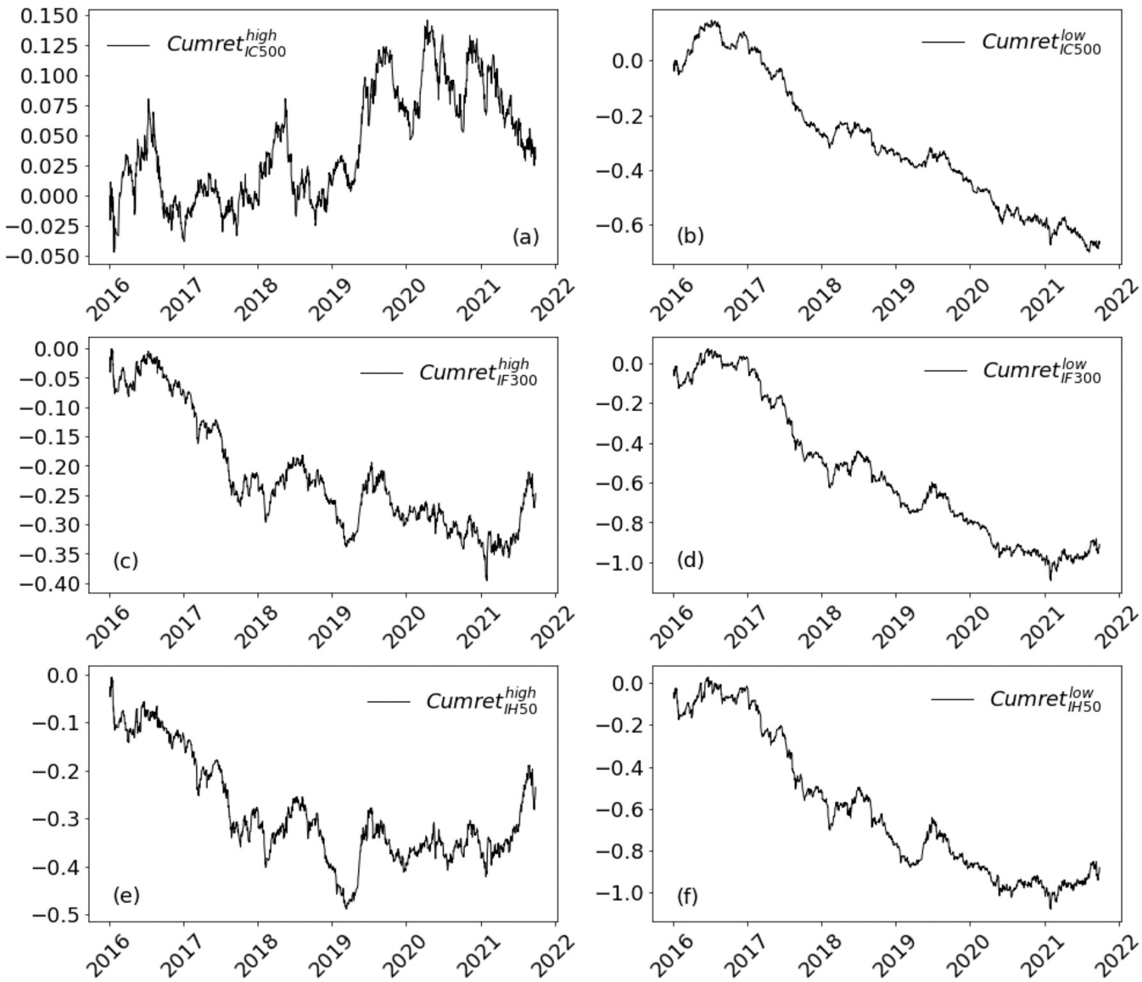

Fig. A1. Dynamic evolution of the cumulative returns of hedging portfolio based on threshold values of 80% and 10%: (a) presents the results of hedge portfolio between the IC500 and stocks in the high information diffusion group, and (b) presents the results of hedge portfolio between the IC500 and stocks in the low information diffusion group. (c) presents the results of hedge portfolio between the IF300 and stocks in the high information diffusion group, and (d) presents the results of hedge portfolio between the IF300 and stocks in the low information diffusion group. (e) presents the results of hedge portfolio between the IH50 and stocks in the high information diffusion group, and (f) presents the results of hedge portfolio between the IH50 and stocks in the low information diffusion group.

Table A1
Average values of different measures of hedging portfolio based on threshold values of 80% and 10%.

	Stock *i*/IC500		Stock *i*/IF300		Stock *i*/IH50	
	High	Low	High	Low	High	Low
Mi	1.6930*** (26.7481)	0.5804				
Del	0.8461*** (−3.5970)	0.8237				
He	0.2414*** (4.7258)	−0.2969	0.2157*** (4.6027)	−0.5505	−0.2858*** (4.3386)	−0.6008
β	0.2727*** (2.3359)	0.2384	0.2501*** (4.2746)	0.1854	0.2260*** (5.6285)	0.1267
Cumret	1.7859 (−0.4048)	1.8153	2.0376*** (5.4461)	1.7239	1.8901*** (8.5687)	1.4528

Note: *Mi* denotes the strength of information diffusion of stocks, *Del* denotes the price delay of stocks, *He* denotes the hedging effectiveness of hedging portfolio, *Cumret* denotes the cumulative returns of hedging portfolio, and β denotes the minimum hedge ratio between stocks and stock index futures. We perform the *t*-test with the null hypothesis that the measure of high information diffusion group is larger or smaller than that of low information diffusion group. ***, ** and * denote the 1%, 5% and 10% significance level respectively. The values in parentheses represent the *t*-statistic.

References

Abrigo, M. R. M., & Love, I. (2016). Estimation of panel vector autoregression in stata. *The Stata Journal, 16*, 778–804.
Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management, 45*, 118–131.
Aldasoro, I., & Alves, I. (2018). Multiplex interbank networks and systemic importance: An application to European data. *Journal of Financial Stability, 35*, 17–37.
Anton, M., & Polk, C. (2014). Connected stocks. *The Journal of Finance, 69*, 1099–1127.
Antonakakis, N., Cunado, J., Filis, G., Gabauer, D., & Gracia, F. P. D. (2020). Oil and asset classes implied volatilities: Investment strategies and hedging effectiveness. *Energy Economics, 91*, Article 104762.
Badrinath, S. G., Kale, J. R., & Noe, T. H. (1995). Of shepherds, sheep, and the cross-autocorrelations in equity returns. *The Review of Financial Studies, 8*, 401–430.

Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *The Review of Financial Studies, 22*, 151–186.
Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics, 6*(02), 161–199.
Bartesaghi, P., Clemente, G. P., & Grassi, R. (2022). A tensor-based unified approach for clustering coefficients in financial multiplex networks. *Information Sciences, 601*, 268–286.
Bianconi, G. (2013). Statistical mechanics of multiplex networks: Entropy and overlap. *Physical Review E, 87*, Article 062806.
Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives, 12*, 151–170.
Campbell, J. Y., & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management, 24*, 11–26.
Challet, D., Chicheportiche, R., Lallouache, M., & Kassibrakis, S. (2018). Statistically validated lead-lag networks and inventory prediction in the foreign exchange market. *Advances in Complex Systems, 21*, 1850019.
Chen, J. Y., Liang, Z. P., Ding, Q., & Liu, Z. H. (2022). Quantile connectedness between energy, metal, and carbon markets. *International Review of Financial Analysis, 83*, Article 102282.
Chen, Z. H. J., Ren, F., Yang, M. Y., Lu, F. Z., & Li, S. P. (2023). Dynamic lead-lag relationship between Chinese carbon emission trading and stock markets under exogenous shocks. *International Review of Economics and Finance, 85*, 295–305.
Christos, P., & Wang, B. (2017). Shareholder coordination, information diffusion and stock returns. *Financial Review, 52*, 563–595.
Colla, P., & Mele, A. (2010). Information linkages and correlated trading. *The Review of Financial Studies, 23*, 203–246.
Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences, 111*, 11600–11605.
Ding, R., Zhou, H., & Li, Y. F. (2020). Social media, financial reporting opacity, and return comovement: evidence from seeking alpha. *Journal of Financial Markets, 50*, Article 100511.
Drake, M. S., Jennings, J., Roulstone, D. T., & Thornock, J. R. (2017). The comovement of investor attention. *Management Science, 63*, 2847–2867.
Egeth, H., & Kahneman, D. (1973). Attention and effort. *American Journal of Psychology, 88*, 339.
Fama, E., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*, 1–22.
Fang, Y. B., Jia, J. W., Li, J. Z., & Liu, S. Y. (2022). Analysis of public information demand during the covid-19 pandemic based on four-stage crisis model. *Frontiers in Physics, 10*, Article 964142.
Francisco, M. (2017). Analyst coverage network and stock return comovement in emerging markets. *Emerging Markets Review, 32*, 1–27.
Fricke, D., & Savoie, C. (2017). *Common asset holdings and stock return comovement*. SSRN. Available at SSRN: https://ssrn.com/abstract=2949853.
Gao, Y., Zhao, K., Wang, C., & Liu, C. (2020). The dynamic relationship between internet attention and stock market liquidity: A thermal optimal path method. *Physica A: Statistical Mechanics and its Applications, 550*, Article 124180.
Gong, X. L., & Du, Z. Q. (2020). Institutional investor information sharing, stock market extreme risk, and financial systemic risk. *Complexity*, 1–10.
Granovetter, M. (1985). Economic action and social structure: The problem of embeddedeness. *U.S. Journal of Sociology, 91*(03), 481–510.
Green, T. C., & Hwang, B. H. (2009). Price-based return comovement. *Journal of Financial Economics, 93*, 37–50.
Gu, B., Konana, P., Liu, A., Rajagopalan, B., & Ghosh, J. (2006). Identifying information in stock message boards and its implications for stock market efficiency. In *Workshop on information systems and economics, Los Angeles, CA*.
Hameed, A., Morck, R., Shen, J. F., & Yeung, B. (2015). Information, analysts, and stock return comovement. *The Review of Financial Studies, 28*, 3153–3187.
Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information, the. *Review of Asset Pricing Studies, 1*, 35–73.
Holden, C. W., & Subrahmanyam, A. (1992). Long-lived private information and imperfect competition. *The Journal of Finance, 47*(01), 247–270.
Hong, H., Jeffrey, D. K., & Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, the. *Journal of Finance, 60*, 2801–2824.
Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *The Review of Financial Studies, 18*, 981–1020.
Huang, H., Sun, Y., & Chu, Q. (2022). Can we-media information disclosure drive listed companies' innovation? From the perspective of financing constraints. *China Finance Review International, 12*(03), 477–495.
Ivkovich, Z., & Weisbenner, S. J. (2007). Information diffusion effects in individual investors, common stock purchases covet thy Neighbors' Investment choices. *The Review of Financial Studies, 20*(04), 1327–1357.
Jiang, L., Liu, J. Y., & Yang, B. Z. (2019). Communication and comovement: Evidence from online stock forums. *Financial Management, 48*, 805–847.
Kallberg, J., & Pasquariello, P. (2008). Time-series and cross-sectional excess comovement in stock indexes. *Journal of Empirical Finance, 15*, 481–502.
Khanna, T., & Thomas, C. (2009). Synchronicity and firm interlocks in an emerging market. *Journal of Financial Economics, 92*, 182–204.
Kim, H. J., Lee, Y., Kahng, B., & Kim, I. M. (2002). Weighted scale-free network in financial correlations. *Journal of the Physical Society of Japan, 71*, 2133–2136.
Kimberly, F. L., & Michael, J. S. (2013). The effect of exogenous information signal strength on herding. *Review of Behavioral Finance, 5*, 153–174.
Koch, A., Ruenzi, S., & Starks, L. T. (2016). Commonality in liquidity: A demand-side explanation. *Review of Financial Studies, 29*, 1943–1974.
Kroner, K. F., & Ng, V. K. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies, 11*, 817–844.
Kroner, K. F., & Sultan, J. (1993). Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis, 28*, 535–551.
Kumar, A., & Lee, C. M. C. (2006). Retail investor sentiment and return comovement. *Journal of Finance, 61*, 2451–2486.
Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters, 87*, Article 198701.
Levin, A., Lin, C. F., & Chu, C. S. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics, 108*, 1–24.
Li, J., Zhang, Y. J., Feng, X., & An, Y. H. (2019). Which kind of investor causes comovement? *Journal of International Financial Markets Institutions and Money, 61*, 1–15.
Li, Y. C., Cai, W. H., Li, Y., & Du, X. (2019). Key node ranking in complex networks: A novel entropy and mutual information-based approach. *Entropy, 22*, 52.
Li, Y. S., Zhuang, X. T., Wang, J., & Zhang, W. P. (2020). Analysis of the impact of Sino-US trade friction on china's stock market based on complex networks. *The North American Journal of Economics and Finance, 52*, Article 101185.
Liu, L., Wu, J., Li, P., & Li, Q. (2015). A social-media-based approach to predicting stock comovement. *Expert Systems with Applications, 42*, 3893–3901.
Love, I., & Zicchino, L. (2006). Financial development and dynamic investment behavior: Evidence from panel VAR. *The Quarterly Review of Economics and Finance, 46*(2), 190–210.
Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., & Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging, 16*, 187–198.
Musciotto, F., Marotta, L., Piilo, J., & Mantegna, R. N. (2018). Long-term ecology of investors in a financial market. *Humanities and Social Sciences Communications, 4*, 1–12.
Muslu, V., Rebello, M., & Xu, Y. X. (2014). Sell-side analyst research and stock comovement. *Journal of Accounting Research, 52*, 911–954.
Musmeci, N., Nicosia, V., Aste, T., Matteo, T. D., & Latora, V. (2017). The multiplex dependency structure of financial markets. *Complexity, 9586064*.
Pareek, A. (2012). Information networks: Implications for mutual fund trading behavior and stock returns. In *AFA 2010 Atlanta Meetings Paper*.
Pasquariello, P. (2007). Imperfect competition, information heterogeneity, and financial contagion. *The Review of Financial Studies, 20*, 391–426.
Peng, L., Xiong, W., & Bollerslev, T. (2007). Investor attention and time-varying comovements. *European Financial Management, 13*, 394–422.
Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics, 22*, 265–312.
Rantala, V. (2019). How do investment ideas spread through social interaction? Evidence from a ponzi scheme. *The Journal of Finance, 74*, 2349–2389.
Ren, F., Cai, M. L., Li, S. P., Xiong, X., & Chen, Z. H. J. (2023). A multi-market comparison of the intraday lead-lag relations among stock index-based spot, futures and options. *Computational Economics, 62*, 1–28.
Rizkiana, A., Hasrini, S., Hardjomidjojo, P., Prihartono, B., Sunaryo, I., & Prasetyo, I. R. (2018). Lead-lag relationship between investor sentiment in social media investor attention in google, and stock return. In *2018 Thirteenth International Conference on Digital Information Management (ICDIM)* (pp. 204–209).
Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences, 107*, 13636–13641.
Tumminello, M., Di Matteo, T., Aste, T., & Mantegna, R. N. (2007). Correlation based networks of equity returns sampled at different time horizons. *The European Physical Journal B, 55*, 209–217.
Wahal, S., & Yavuz, M. D. (2013). Style investing, comovement and return predictability. *Journal of Financial Economics, 107*(01), 136–154.
Wang, D. H., Tu, J. Q., Chang, X. H., & Li, S. P. (2017). The lead-lag structure between the spot and futures markets in China. *Quantitative Finance, 02*, 1447–1456.
Wang, G. J., Xie, C., Zhao, L. F., & Jiang, Z. Q. (2018). Volatility connectedness in the Chinese banking system: Do state-owned commercial banks contribute more? *Journal of International Financial Markets Institutions and Money, 57*, 205–230.
Wen, F. H., Cao, J. H., Liu, Z., & Wang, X. (2021). Dynamic volatility spillovers and investment strategies between the Chinese stock market and commodity markets. *International Review of Financial Analysis, 76*, Article 101772.
Werner, A., & Murray, Z. F. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance, 59*, 1259–1294.
Yang, Y. H., & Shao, Y. H. (2020). Time-dependent lead-lag structures between the VIX and VIX futures markets. *North American Journal of Economics and Finance, 53*, Article 101196.
Yao, C. Z., & Li, H. Y. (2020). Time-varying lead-lag structure between investor sentiment and stock market. *North American Journal of Economics and Finance, 52*, 10114.
Zhang, F., Zhang, Y., Xu, Y., & Chen, Y. (2023). Dynamic relationship between volume and volatility in the Chinese stock market: Evidence from the MS-VAR model. *Data Science and Management*. https://doi.org/10.1016/j.dsm.2023.09.003
Zhou, W. X., & Sornette, D. (2006). Non-parametric determination of real-time lag structure between two time series: The "optimal thermal causal path" method with applications to economic data. *Journal of Macroeconomics, 28*, 195–224.