A generative model for the collective attention of the Chinese stock market investors

Jian-Guo Liu, Zhen-Hua Yang, Sheng-Nan Li, Chang-Rui Yu

PII: S0378-4371(18)30987-7
DOI: https://doi.org/10.1016/j.physa.2018.08.036
Reference: PHYSA 19922

To appear in: *Physica A*

Received date : 13 January 2018
Revised date : 27 April 2018

A Generative Model for the Collective Attention of the Chinese Stock Market Investors

Jian-Guo Liu* [a], Zhen-Hua Yang [b,c,e], Sheng-Nan Li [d], Chang-Rui Yu [c]

[a]Research Center of Fin. Tech. and Shanghai Key Lab. of Fin. Inf. Tech., Shanghai University of Finance and Economics, Shanghai 200433, PR China

[b]Business School, Huzhou University, Huzhou 313000, PR China

[c]School of Information Management Engineering, Shanghai University of Finance and Economics, Shanghai 200433, PR China

[d]Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China

[e]Warrington College of Business, University of Florida, Gainesville, Florida 32611, USA

Abstract

Collective attention of investors maps the interests and intention of investors directly in the stock market. However, the evolution mechanism of the collective attention from the viewpoint of complex system is missing. In this paper, we empirically investigate the investor collective attention mechanism based on a best-known stock trading platform between 2014 and 2016. Taking the global and recent popularity effects into account, we introduce a generative model for the collective attention of millions of investors who are deciding their trading behavior among thousands of stocks in Chinese stock market. The experimental results show that the investor attention is more closely affected by recent attention, with the optimal case, when the memory effect parameter $T = 10$ and the recent popularity parameter $\gamma = 0.1$, the model could regenerate the collective attention more accurately, say Kendall's $\tau = 0.92$ for the Shanghai Stock Exchange(SSE) and Shenzhen Stock Exchange(SZSE) simultaneously. This work may shed some lights for deeply understanding the mechanism of the investor collective attention for the financial market.

Key words: Investor collective attention, Recent attention, Cumulative attention, Stock market.

Email address: `liujg004@ustc.edu.cn` (Jian-Guo Liu*).

1 Introduction

Collective attention given by the investors in the stock market closely reflects the investors' potential interests and trading behaviors, therefore, it is of significant for understanding how investor attention to different stocks propagates and eventually fades among large population. Particularly, recent studies provide a theoretical framework in which investor limited attention can affect asset pricing statics as well as dynamics [1–3]. Heiberger *et al.* [4] argued that collective attention shifts precede structural changes in stock market networks and that this connection is mostly carried by companies that already dominate the development of the S&P 100. However, the evolution mechanism of the investor collective attention is missing, which could help us to understand the interaction between investor attention and stock market performance. In this paper, we empirically investigate the statistical properties of the collective attention for Chinese stock market, and present a generative model for regenerating the collective attention of the investors from the viewpoint of complex system.

The problem of collective attention is at the heart of decision making and the spread of ideas, and, it has been studied in a wide range of disciplines. Wu and Huberman [5] argued that novelty within groups decays with a stretched exponential law, suggesting the existence of a natural time scale over which attention fades. Lehmann *et al.* [6] focused on spikes of collective attention in Twitter and found that epidemic spreading was mostly driven by exogenous factors. Mocanu *et al.* [7] refered that the users who were prominently interacting with alternative information sources were more prone to interact with false claims. In order to understand the collective attention deeply, researchers proposed many different models to analyze it, for instance, Kyumin *et al.* [8] developed spam classifiers to detect spam messages generated by collective attention spammers. Ye *et al.* [9] presented a study of the group purchasing behavior of daily deals in Groupon and LivingSocial and formulated a predictive dynamic model of collective attention for group buying behavior. Sasahara *et al.* [10] proposed a simple method for detecting and measuring the collective attention evoked by various types of events. Bao *et al.* [11] proposed a generative probabilistic model using a self-excited Hawkes process with survival theory to model and predict the process through which individual items gain their attentions. All the researches would be a strong basics for understanding the investor collective attention.

In fact, every investor eager to capture more information about the evolution dynamic of stocks, however, each investor could only get limited information of limited number of stock items due to the time and energy. They would add the potential interested stocks into the watching list to capture the dynamic properties, and pay more attention to the attracted information to adjust their trading behavior, leading to the movement of the stock market [12–14]. Therefore, investor attention, as a key factor in the stock market, attracted more and more attention. In order to describe the investor attention as exact as possible, many researchers uncover different kinds of proxies, including extreme return [15], trading volume [16,17], turnover [18], search volume index (SVI) [19–21], social network (Twitter feeds, blogs, forum, Wikipedia etc.) [22–25], news [26,27], etc. [28,29]. In particular, the massive data sources resulting from human interaction with the Internet have offered a new perspective on the behavior of market participants besides investors in the stock market [30]. For example, Yang *et al.*[31] found that the increments of the attention volume for each stock (IAVS) from the stock trading platform positively correlated with the next day price of the corresponding stock index for the 2014-2016 three-year data, which suggests that the collective attention of the stock investors contain more important information for the investing strategy. More concretely, the measure IAVS refers to the daily increments of each stock which is chosen by investors into their watch lists in a stock trading platform, and we call it "stock attention" for ease of understanding in this paper. Figure 1 shows the evolution dynamic property of the cumulative and daily increment attention of one specific stock in 2014.

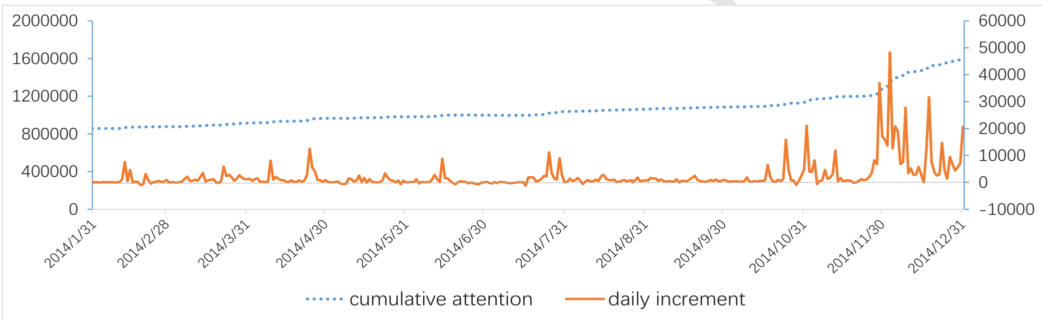

Fig. 1. (Color Online) **Illustration of the cumulative and daily increment attention for the stock in 2014**, which shows the investor collective attention of stock market from a stock trading platform named Choice, available online at http://choice.eastmoney.com/. In fact, the collective attention infers to the total attention number of each stock which is in the investors' watch lists and the daily increment for the stock infers to the increasing attention number of each day for each stock located in the investors' watch lists.

The stocks locating at the top position of the stock list may attract more investors' attention, leading to the "rich get richer" phenomenon, which has been extensively investigated in the popularity dynamics and network science [32,33]. Borghol *et al.* [34] empirically measured the popularity of videos and found that preferential selection could be used to interpret the video popularity evolution. Szabo *et al.* [35] found that the long time popularity of online content could be predicted by the early user accesses. Comparing with the rich-get-richer phenomenon, Bentley *et al.* [36] introduced a "memory" parameter defined as the number of previous steps which affects an individual's decision. Furthermore, it should be noticed that the popularity of most empirical systems is heterogeneous [37]. Liu *et al.* [38] found that the online user interests could be divided as common interests and specific interests. Li *et al.* [39] uncovered the popularity mechanisms for Facebook Apps and found the recent popularity plays more important role as the popularity of Facebook Apps increase. The heterogeneous physics of the object popularity plays an important role for the online social systems evolution [40].

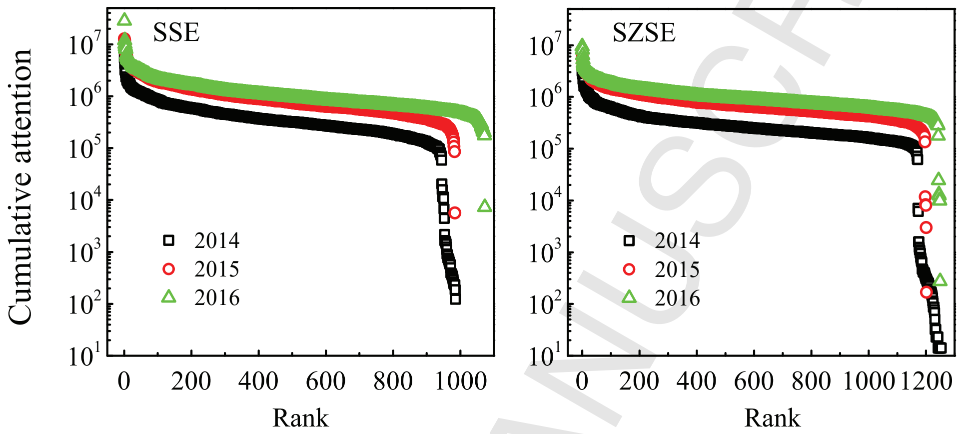

Fig. 2. (Color Online) **The rank of the total number of cumulative attention for stocks of Shanghai and Shenzhen Stock Exchange in each year from 2014 to 2016.** One can find that, for each year, the total number of cumulative attention increases from 2014 to 2016 and there are similar patterns for the cumulative attention of the SSE and SZSE respectively.

In this paper, we investigate the roles of the cumulative and recent attention in the collective attention pattern of the investors for the stock market. Based on the investor collective attention on the stock trading platform, we calculate the statistical properties of investor attention for above 2000 stocks in Shanghai and Shenzhen Stock Exchanges from 2014 to 2016. Then, taking into account the cumulative and recent number of attention, we present a model to regenerate the collective attention pattern of the empirical attention of stocks, and find that either for the SSE and the SZSE, the recent popularity of the collective attention plays more important role to regenerate the collective attention ranking list of the the investor behaviors, which suggests that the information technology, such as the App download list, online social network, affects the user behaviors more closely. Moreover, investor attention is more affected by recent attention both in Shanghai Stock Exchange(SSE) and Shenzhen Stock Exchange(SZSE). Finally, we analyze the trend of collective attention pattern of the investors for the stock market in terms of different market environments.

2 Data and model

2.1 Description of the dataset

The original dataset includes the number of cumulative attention for each stock in Shanghai Stock Exchange(SSE) and Shenzhen Stock Exchange(SZSE) in mainland China from January 1, 2014 to December 31, 2016, spanning $a_{max} = 1095$ days. All the data collected from the stock trading platform named Choice from eastmoney.com (http://choice.eastmoney.com/), which is one of the best-known financial trading platform in mainland China with approximately 1.6 million active users daily and more than one hundred million users so far. As shown in Fig. 2, for the investors' collective attention of the stock trading platforms, the smallest cumulative attention number of stocks is only six, while the largest attention number of stocks is over ten million both in Shanghai Stock Exchange and in Shenzhen Stock Exchange from 2014 to 2016. In addition, the total number of cumulative attention of the stocks in the stock market shows a trend of rising year by year and there are similar patterns for the cumulative attention of the SSE and SZSE respectively.

In this paper, we take into account the stocks which are listed in the stock trading platform at the beginning of each year and whose launch times are unknown. Finally, by removing the stocks with missing information, the amount of stocks mentioned in our following study is detailed in Table 1. As shown in Table 1, we calculate the statistical properties of the empirical data, where the measure CAVS refers the cumulative attention volume of each stock which is chosen by investors into their watch lists in the stock trading platform. In fact, the investors would like not only to choose new stocks but also to cancel some stocks. Generally speaking, when the investors pay attention to the stocks, it is possible for them to buy it, thereby causing the movement of the corresponding stocks. However, the investors would cancel their attention any time with no significant effect on the movement of the stock market, as explained by Barber *et al*. [12,15].Therefore, we only take account of the number of choosing stocks by investors in this paper.

2.2 Construction of the Regenerative model

Taking the cumulative and recent attention into account, we present a model to regenerate the collective attention behaviors for the stocks of SSE and SZSE. Since there are significant difference between the SSE and SZSE, say the SSE mainly include the heavyweights while the SZSE mainly include the SME (Small and Medium-sized Enterprises) board companies, we suppose the investor's attention behaviors depends on the stocks of SSE or SZSE separatively. Then, the model could be given in the following way. Firstly, the total number of attention for all

Table 1
The statistical properties of the empirical data for each year. CAVS refers the Cumulative Attention Volume of each Stock

Year	Exchange	# stocks	max_CAVS	avg_CAVS	med_CAVS	min_CAVS
2014	SSE	986	5,403,434	272,959	217,769	86
2015	SSE	984	12,765,074	783,611	537,831	401
2016	SSE	1073	28,339,129	1,126,740	819,464	5,621
2014	SZSE	1254	3,154,883	217,230	178,934	4
2015	SZSE	1202	8,163,248	550,947	429,314	113
2016	SZSE	1249	9,514,702	910,789	746,150	167

stocks at time a, denoted by $F(a)$, is calculated by,

$$F(a) = \sum_i \widetilde{f}_i(a), i \in \text{SSE } or \text{ SZSE}, \tag{1}$$

where $\widetilde{f}_i(a)$ is the number of attention of stock i at time a, the stock i belongs to the SSE or SZSE respectively. For the SSE or SZSE, the total increment $F(a)$ at time a is redistributed to each stock i with probability $p_i(a)$, which is given as,

$$p_i(a) = \gamma p_i^c(a) + (1 - \gamma) p_i^r(a), \tag{2}$$

where $p_i^c(a)$ is defined as the *cumulative attention* that users pay attention to stock i at time a with a probability proportional to its cumulative attention $\widetilde{n}_i(a-1)$, yielding

$$p_i^c(a) = \frac{\widetilde{n}_i(a-1)}{\sum_i \widetilde{n}_i(a-1)}, \tag{3}$$

and $p_i^r(a)$ is defined as the *recent attention* that users pay attention to stock i at time a in terms of the probability proportional to its recent received attention, which is described as

$$p_i^r(a) = \frac{\sum_{t=1}^{a-1} \frac{1}{T} e^{\frac{-(a-t)}{T}} \widetilde{f}_i(t)}{\sum_i \sum_{t=1}^{a-1} \frac{1}{T} e^{\frac{-(a-t)}{T}} \widetilde{f}_i(t)}, \tag{4}$$

where $e^{-(a-t)}$ is an exponential "memory" function which assigns weight to the age-shifted increment $\widetilde{f}_i$ from t days ago [36,41], the parameter T denotes the timeliness of memory effect. In Fig. 3, we show results for stochastic simulations using an exponential response-time distribution $\omega(a) = \frac{1}{T} e^{\frac{-a}{T}}$ to determine the weights assigned to attention from a days earlier for varying response-time parameters T. One can find that, the curve of the attenuation function $\omega(a) = \frac{1}{T} e^{\frac{-a}{T}}$ get slow down with the T increase, that is to say, the timeliness of memory effect get longer as T increase.

In the regenerating process for each time step, the $\tilde{n}_i(a)$ in Eq.(3) and $\tilde{f}_i(t)$ in Eq.(4) are obtained from the empirical data to determine the probability $p_i(a)$ of Eq.(2).

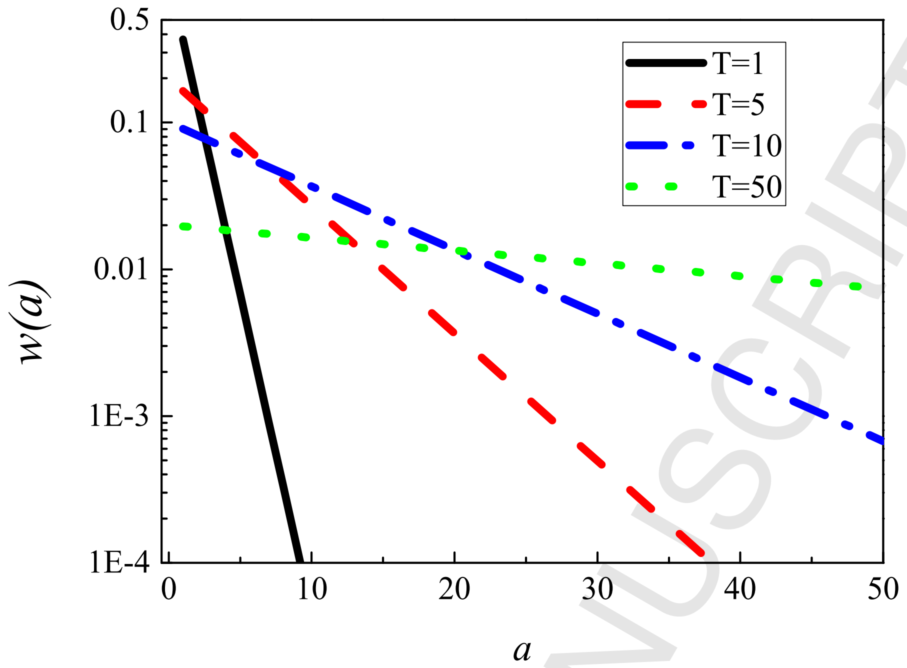

Fig. 3. (Color Online) **Illustration of attenuation function $\omega(a)$ under different memory effect parameter T.** The response-time parameter T denotes the timeliness of memory effect, as the T increase from 1 to 50, the curve of the attenuation function $\omega(a)$ begin to slow down.

The attention probability $p_i(a)$ interpolates between the extremes of $\gamma = 0$ (recent attention) and $\gamma = 1$ (cumulative attention). After distributing each total increment $F(a)$, we regenerate the age-shifted increment $\widetilde{f}_i(a)$ of each stock exchange. Specifically, the collective attention of last trading days from 2014 to 2016 is regenerated in this paper, and the stock's initial age-shifted increment $\widetilde{f}_i(1)$ is extracted from the empirical data. The results for other time parameter a are similar with this case.

3 Experimental results

3.1 Measurement

For the SSE and SZSE, to investigate the performance of cumulative and recent attention for investor attention, we introduce the Kendall's τ [42] to measure the rank correlation between the regenerated number of attention f and the empirical number of attention $\widetilde{f}$. Mathematically, it reads

$$\tau = \frac{\sum_{i \in M} \sum_{j \in M} \mathrm{sgn}[(f_i - f_j)(\widetilde{f}_i - \widetilde{f}_j)]}{|M|(|M|-1)} \tag{5}$$

where M is set of the different stock exchanges e.g. SSE and SZSE, $\mathrm{sgn}(x)$ is the sign function, which returns 1 if $x > 0$; -1 if $x < 0$; and 0 for $x = 0$. According to the definition, $\tau \in [-1, 1]$. A higher τ indicates a more accurate estimation of objects' true quality.

More generally, the Kendall's τ is a non-parametric statistic used to measure the degree of correspondence between two rankings and assessing the significance of this correspondence. In other words, it measures the strength of association of the cross tabulations. Firstly, if the agreement between the two rankings is perfect (i.e., the two rankings are the same) the coefficient has value 1. Secondly, if the disagreement between the two rankings is perfect (i.e., one ranking is the reverse of the other) the coefficient has value -1. Thirdly, for all other arrangements the value lies between -1 and 1, and increasing values imply increasing agreement between the rankings. If the rankings are completely independent, the coefficient has value 0 on average.

For understanding the meaning of τ, we give an example here. Suppose we rank a group of eight people by height and by weight where person A is tallest and third heaviest, and so on: $Persons = \{A, B, C, D, E, F, G, H\}$, $Rank\_Height = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $Rank\_Weight = \{3, 4, 1, 2, 5, 7, 8, 6\}$. We see that there is some correlation between the two rankings but the correlation is far from perfect. We can use the Kendall's τ coefficient to objectively measure the degree of correspondence, and get that $\tau = 0.57$, suggesting that, roughly speaking, there is a positive correlation between the two ranking lists.

3.2 Result Analysis

Firstly, we investigate the Kendall's τ by varying response-time parameters T in terms of the effect of recent or cumulative attention on the following investor attention of stocks for SSE from 2014 to 2016. From Fig. 4 (a)-(c), one can find that investor attention is more likely to be affected by recent attention for the SSE with long memory ($\gamma \le 0.1$, $\tau \ge 0.94$, $T = 10$) during the study time, and investor attention is also more likely to be affected by recent attention for the SSE with short memory ($\gamma \le 0.3$, $\tau \ge 0.89$, $T = 5$) during the study time. Particularly, the model parameter achieve the best condition with $\tau^* = 0.95$ while $\gamma \le 0.1$ and $T = 5$ or $T = 10$ for the SSE during the study time, that is to say, the generative model emphasizes the investor recent attention of stocks over their cumulative attention for SSE with short memory and long memory.

Secondly, we investigate the Kendall's τ with different response-time parameters

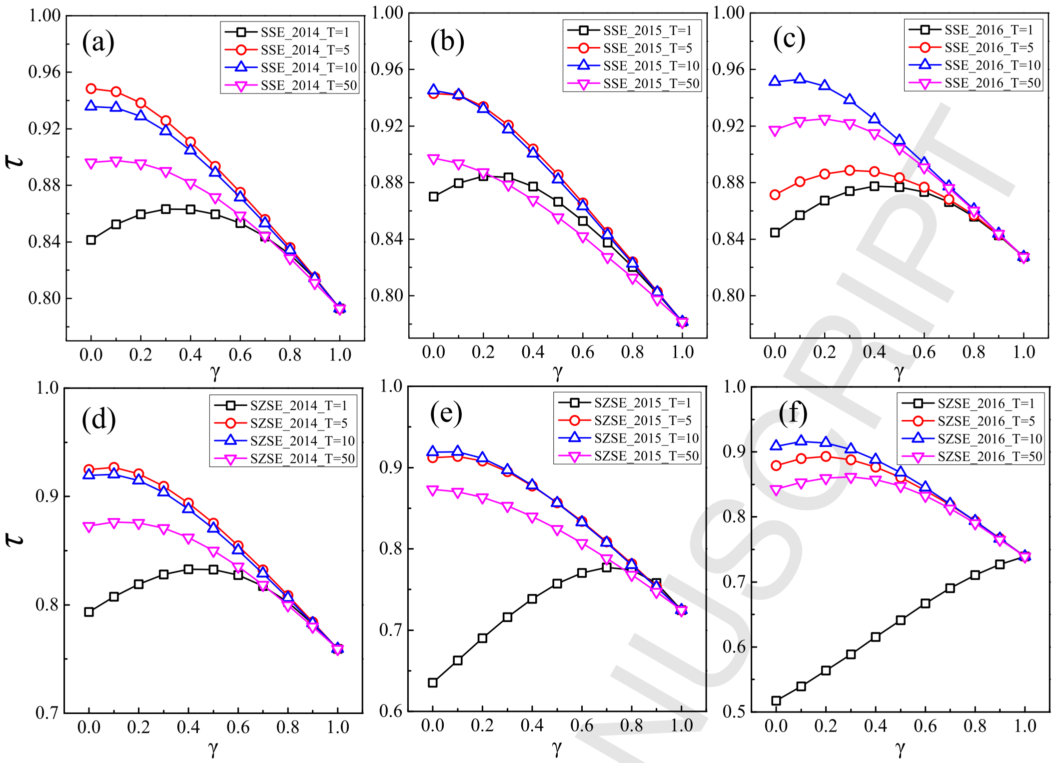

Fig. 4. (Color Online) **Distribution of the Kendall's τ of the SSE and SZSE for different parameter γ and different response-time parameters T during 2014 and 2016**, where the Kendall's τ is calculated according to the ranking lists of the generated and empirical collective attention of the last day in each year, and the parameter γ indicates the effect of the cumulative and recent popularity effects. **(a-c)** Distribution of the Kendall's τ as $T = 1$, $T = 5$, $T = 10$, $T = 50$ for SSE from 2014 to 2016. **(d-f)** Distribution of the Kendall's τ as $T = 1$, $T = 5$, $T = 10$, $T = 50$ for SZSE from 2014 to 2016. From which one can find that, for both SSE and SZSE stock exchanges, the Kendall's τ will reach its largest values when the γ is close to 0 when the decay function parameter T =5, 10, or 50, suggesting that the recent popularity plays more important role for the collective attention evolution. Technical speaking, the model could be used to measure the effect of collective attention for each day of different markets.

T in terms of the effect of recent or cumulative attention on the following investor attention of stocks for SZSE from 2014 to 2016. From Fig. 4 (d)-(f), one can find that investor attention is more likely to be affected by recent attention for the SZSE with long memory ($\gamma = 0.1, \tau = 0.92, T = 10$) during the study time. However, the effect of recent or cumulative attention on the following investor attention of stocks changed for the SZSE with different response-time parameters T in the view of short memory effect. As $T = 1$, the model parameter achieve the varying condition with $\tau = 0.83$ while $\gamma = 0.4$ in 2014, $\tau = 0.78$ while $\gamma = 0.7$ in 2015, $\tau = 0.74$ while $\gamma = 1.0$ in 2016, which shows that both the recent attention and cumulative attention affect the following investor attention of stocks for SZSE between 2014 and 2016. More over, as $T = 5$, the model parameter achieve the relative stable condition with $\tau = 0.93$ while $\gamma = 0.1$ in 2014, $\tau = 0.91$ while $\gamma = 0.1$ in 2015,

$\tau = 0.89$ while $\gamma = 0.2$ in 2016, which shows that the recent attention gives more influence on the following investor attention of stocks for SZSE between 2014 and 2016.

Generally speaking, by comparing the results with different response-time parameters T, we find that the influencing trend of cumulative or recent investor attention get stable gradually with the T increase and the most stable results is the state of $T = 10$ both in SSE and in SZSE simultaneously. For the optimal case, when the memory effect parameter $T = 10$ and the recent popularity parameter $\gamma = 0.1$, the regenerative model could identify the popularity stock list more accurately, say Kendall's $\tau = 0.92$ for the SSE and SZSE simultaneously. From the experimental results, we get a stable result that the investor attention is more closely affected by recent attention in Chinese stock market. See the detailed data in Table 2.

Table 2
The optimal Kendall's τ and the corresponding parameters of experimental results

exchange	year	$\gamma^*_{T=1}$	$\tau^*_{T=1}$	$\gamma^*_{T=10}$	$\tau^*_{T=10}$
SSE	2014	0.30	0.86	0.00	0.94
SSE	2015	0.20	0.89	0.00	0.95
SSE	2016	0.40	0.88	0.10	0.95
SZSE	2014	0.40	0.83	0.10	0.92
SZSE	2015	0.70	0.83	0.10	0.92
SZSE	2016	1.00	0.74	0.10	0.92

4 Conclusion and discussions

Collective attention of the stock market investors maps the interests and intention of investors directly. Different from the investigation of the physics of collective online behavior for online social systems [43,44], we focus on the collective attention pattern of the investors for the stock market. First, we investigate the number of cumulative attention for stocks in SSE and SZSE from 2014 to 2016 based on the investor collective attention on the stock trading platform, and find that for each year, the total number of cumulative attention increases from 2014 to 2016 and there are similar patterns for the cumulative attention of the SSE and SZSE respectively. Then, we present a model to regenerate the collective attention pattern of the empirical investor attention of stocks in term of the cumulative and recent popularity of different stocks, and measure with Kendall's τ of the regenerated watching list from 2014 to 2016. Our model suggests that investor attention guided more by recent attention both in SSE and SZSE in the view of long memory effect. More over, under different stock market environments, the effect of recent or cumulative attention on the following investor attention of stocks changed in different pattern with different stock exchanges.

We analyze the collective attention pattern of the investors for the stock market. First, give a brief review of the Chinese stock market for the years from 2014 to 2016. In 2014, A shares ushered the bull market and created a new high of nearly five years, e.g. the SSE index broke 3200 points, showing the 52.87% uplift, the SZSE index broke 11000 points, showing 35.62% uplift, and the highest volume breakthrough trillion yuan, which is the so-called "bull market". In 2015, SSE index seesawed between the highest point 5178 and lowest point 2850, and SZSE index seesawed between the highest point 18211 and lowest point 9259, and we call it as "shocked and crashed market". In 2016, both SSE index and SZSE index produce a negative income with -12.31% and -19.64%, which is the so-called "bear market". In addition, the listed companies in Shanghai Stock Exchange have the characteristics of larger financing scale in the key industries of the country and the listed companies in Shenzhen Stock Exchange almost are growth enterprises with small or medium size.

To summarize, the investor attention is more closely affected by recent attention in the view of longer memory effect both in SSE and SZSE. But, specifically, on the one hand, although the stock market environments changed greatly, the recent attention always plays an important effect on the following investor attention of stocks in Shanghai Stock Exchange both in the views of short and long memory effect, which indicates that the stocks of the listed companies with larger financing scale is more likely to be affected by the recent attention. On the other hand, the influencing trend of cumulative or recent attention on the following investor attention in SZSE is differ in SSE, and both the cumulative or recent attention play role in the investor attention in the view of short memory effect as $T = 1$ in SZSE, which indicates that the stocks of the listed companies with small or medium size is affected by both the recent attention and the cumulative attention. To our understanding, compared with the listed companies with larger financing scale, investors are more likely to overview the listed companies with small or medium size from the historical and recent perspectives. The results may shed some lights for deeply understanding attention mechanisms of the investors for the financial market.

From a theoretical perspective, our study enriches extant research by focusing on the investor attention pattern of the stock market. Specifically, we present a model to regenerate the collective attention pattern of the empirical attention of stocks to test the effect of cumulative attention and recent attention on the following investor attention of stocks. From a practical perspective, our findings may help the supervisors to better understand the investor collective attention pattern, thereby to predict the trends of stock market, based upon which they can make better investment decisions or predict risk. A next step in this research initiative considers applying the model to test the collective attention pattern of new stocks to predict the attention growth, thereby to predict the returns. For another, the path from investor attention to investor trading behavior is still incompletely clear, which also is a valuable research issue to be discussed.

5 Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 61773248, 71771152), and the Graduate Innovation Fund in Shanghai University of Finance and Economics (No. 2017110357), ZHY is supported by the Chinese Scholarship Council (No. 201706480044).

References

[1] Merton, R. C. A simple model of capital market equilibrium with incomplete information. The Journal of Finance **42**, 483 (1987).

[2] Hirshleifer, D. & Teoh, S. H. Limited attention, information disclosure, and financial reporting. Journal of Accounting and Economics **36**, 337 (2003).

[3] Peng, L. & Xiong, W. Investor attention, overconfidence and category learning. Journal of Financial Economics **80**, 563 (2006).

[4] Heiberger, R. H. Shifts in collective attention and stock networks. In International Conference on Computational Social Networks, 296–306 (Springer, 2015).

[5] Wu, F. & Huberman, B. A. Novelty and collective attention. Proceedings of the National Academy of Sciences of U.S.A. **104**, 17599 (2007).

[6] Lehmann, J., Gonçalves, B., Ramasco, J. J. & Cattuto, C. Dynamical classes of collective attention in twitter. In Proceedings of the 21st International Conference on World Wide Web, 251 (ACM, 2012).

[7] Mocanu, D., Rossi, L., Zhang, Q., Karsai, M. & Quattrociocchi, W. Collective attention in the age of (mis) information. Computers in Human Behavior **51**, 1198 (2015).

[8] Lee, K., Caverlee, J., Kamath, K. Y. & Cheng, Z. Detecting collective attention spam. In Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality, 48–55 (ACM, 2012).

[9] Ye, M., Sandholm, T., Wang, C., Aperjis, C. & Huberman, B. A. Collective attention and the dynamics of group deals. In Proceedings of the 21st International Conference on World Wide Web, 1205–1212 (ACM, 2012).

[10] Sasahara, K., Hirata, Y., Toyoda, M., Kitsuregawa, M. & Aihara, K. Quantifying collective attention from tweet stream. PLoS ONE **8**, e61823 (2013).

[11] Bao, P. & Zhang, X. Uncovering and predicting the dynamic process of collective attention with survival theory. Scientific Reports **7**, 2621 (2017).

[12] Da, Z., Engelberg, J. & Gao, P. In search of attention. The Journal of Finance **66**, 1461 (2011).

[13] Lou, D. Attracting investor attention through advertising. The Review of Financial Studies **27**, 1797 (2014).

[14] Daniel, K., Hirshleifer, D. & Teoh, S. H. Investor psychology in capital markets: Evidence and policy implications. Journal of Monetary Economics **49**, 139 (2002).

[15] Barber, B. M. & Odean, T. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The Review of Financial Studies **21**, 785 (2007).

[16] Statman, M., Thorley, S. & Vorkink, K. Investor overconfidence and trading volume. The Review of Financial Studies **19**, 1531 (2006).

[17] Kaniel, R., Ozoguz, A. & Starks, L. The high volume return premium: Cross-country evidence. Journal of Financial Economics **103**, 255 (2012).

[18] Loh, R. K. Investor inattention and the underreaction to stock recommendations. Financial Management **39**, 1223 (2010).

[19] Bordino, I. et al. Web search queries can predict stock market volumes. PLoS ONE **7**, e40014 (2012).

[20] Preis, T., Moat, H. S. & Stanley, H. E. Quantifying trading behavior in financial markets using google trends. Scientific Reports **3**, 01684 (2013).

[21] Dimpfl, T. & Jank, S. Can internet search queries help to predict stock market volatility? European Financial Management **22**, 171 (2016).

[22] Bollen, J., Mao, H. & Zeng, X. Twitter mood predicts the stock market. Journal of Computational Science **2**, 1 (2011).

[23] Ackert, L. F., Jiang, L., Lee, H. S. & Liu, J. Influential investors in online stock forums. International Review of Financial Analysis **45**, 39 (2016).

[24] Zhang, W., Li, X., Shen, D. & Teglio, A. Daily happiness and stock returns: Some international evidence. Physica A **460**, 201 (2016).

[25] Siikanen, M. et al. How facebook drives investor behavior. arXiv preprint arXiv:1709.07300 (2017).

[26] Mian, G. M. & Sankaraguruswamy, S. Investor sentiment and stock market response to corporate news. National University of Singapore Working Paper (2008).

[27] Zhang, Y., Song, W., Shen, D. & Zhang, W. Market reaction to internet news: Information diffusion and price pressure. Economic Modelling **56**, 43 (2016).

[28] Corwin, S. A. & Coughenour, J. F. Limited attention and the allocation of effort in securities trading. The Journal of Finance **63**, 3031 (2008).

[29] Lim, K. P., Hooy, C. W., Chang, K. B. & Brooks, R. Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity. The North American Journal of Economics and Finance **36**, 1 (2016).

[30] Gao, Y.-C., Zeng, Y. & Cai, S.-M. Influence network in the chinese stock market. Journal of Statistical Mechanics: Theory and Experiment **2015**, P03017 (2015).

[31] Yang, Z. H., Liu, J. G., Yu, C. R. & Han, J. T. Quantifying the effect of investors? attention on stock market. PLoS ONE **12**, e0176836 (2017).

[32] Barabási, A. L. & Albert, R. Emergence of scaling in random networks. Science **286**, 509 (1999).

[33] Bianconi, G. & Barabási, A. L. Competition and multiscaling in evolving networks. Europhysics Letters **54**, 436 (2001).

[34] Borghol, Y., Ardon, S., Carlsson, N., Eager, D. & Mahanti, A. The untold story of the clones: content-agnostic factors that impact youtube video popularity. In Proceedings of the 18th ACM SIGKDD, 1186–1194 (ACM, 2012).

[35] Szabo, G. & Huberman, B. A. Predicting the popularity of online content. Communications of the ACM **53**, 80 (2008).

[36] Bentley, R. A., Ormerod, P. & Batty, M. Evolving social influence in large populations. Behavioral Ecology and Sociobiology **65**, 537 (2011).

[37] Santos, F. C., Pacheco, J. M. & Lenaerts, T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proceedings of the National Academy of Sciences of the U. S. A. **103**, 3490 (2006).

[38] Liu, J. G., Zhou, T. & Guo, Q. Information filtering via biased heat conduction. Physical Review E **84**, 037101 (2011).

[39] Li, S. N., Guo, Q., Yang, K., Liu, J. G. & Zhang, Y. C. Uncovering the popularity mechanisms for facebook applications. Physica A **494**, 422 (2018).

[40] Zhou, T., Liu, J. G., Bai, W. J., Chen, G. & Wang, B. H. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. Physical Review E **74**, 056109 (2006).

[41] Hou, L., Pan, X., Guo, Q. & Liu, J. G. Memory effect of the online user preference. Scientific Reports **4**, 6560 (2014).

[42] Kendall, M. G. A new measure of rank correlation. Biometrika **30**, 81–93 (1938).

[43] Gleeson, J. P., Cellai, D., Onnela, J. P., Porter, M. A. & Reed Tsochas, F. A simple generative model of collective online behavior. Proceedings of the National Academy of Sciences of U.S.A. **111**, 10411 (2014).

[44] Valverde, S. Crossover from endogenous to exogenous activity in open-source software development. Europhysics Letters **77**, 20002 (2007).

- Taking into account the cumulative and recent popularity, we introduce a generative model for the collective attention of investors.
- The investor attention is more closely affected by recent attention. With the optimal case, the Kendall's τ=0.92 for two different stock exchanges.
- This work enrich the research of investor collective attention from the view of complex system.