



SSLE: A framework for evaluating the “Filter Bubble” effect on the news aggregator and recommenders

Han Han¹ · Can Wang² · Yunwei Zhao¹ · Min Shu¹ · Wenlei Wang¹ · Yong Min³

Received: 1 June 2021 / Revised: 1 December 2021 / Accepted: 24 February 2022 /
Published online: 8 March 2022
© The Author(s) 2022

Abstract

Recommendation algorithms are data filtering tools that make use of algorithms and data to recommend the most relevant items to a particular user. The algorithm-driven recommenders become indispensable and supersede search engines as the most important information dissemination channel. On one hand, it becomes an integral component in the existing social media, e.g. Weibo, Twitter, etc. On the other hand, news aggregators and recommenders have proliferated and gained an increasing market share. As a result, the previous studies usually study the “filter bubbles” phenomenon in the context where the social filtering dominates the dissemination of information. However, less attention is paid to the news aggregators and recommenders where algorithm-driven technological filtering dominates. Therefore, in the previous research, “filter bubbles” are usually equated with the community structure, but lack of the detailed analysis of the content agglomeration through the users’ interaction with the platforms. Based on these concerns, we propose a four-phase (“Selection”, “Setup”, “Link”, and “Evaluation”) skeletal solution framework targeted at exploiting the filter bubble effect of the personalized news aggregation and recommendation system. Furthermore, we illustrate the effectiveness of the proposed framework with a case study in three top Chinese news aggregators, i.e. Toutiao, Baidu News, and Tencent News. The results show that the users are narrowed into one or a limited number of topics over time. The phenomenon of the narrowed topics is deemed as the emergence of the “filter bubbles”. We also observe that the filter bubbles demonstrate different convergence degrees as user’s individual preference varies.

Keywords Personalized recommendation · Computational social science · Sociotechnical systems · Polarization · Filter bubble

This article belongs to the Topical Collection: *Special Issue on Web Intelligence = Artificial Intelligence in the Connected World*
Guest Editors: Yuefeng Li, Amit Sheth, Athena Vakali, and Xiaohui Tao

✉ Yunwei Zhao
zhaoyw@cert.org.cn

Extended author information available on the last page of the article

1 Introduction

The emergence of “self-media” has resulted in a flood of information generation, and thus the main body of information communication becomes more complex [41]. To help the end users deal with the information overload problem, the news aggregators and recommenders use recommendation algorithms to learn the user’s preferences and behavioral patterns, and recommend the content catering for the user’s preference. The news aggregators and recommenders built upon recommendation algorithms can achieve a fine recommendation result based on user segmentation and neighbor discovery on a hyper-dimensional basis [7, 20, 24], and learn with cross-references to what each user is interested in. Despite the ease of access to information, there are mounting concerns about the recommendation algorithms providing the audience with the filtered channels of information [3, 9, 15]. The goal of the platform is to design those algorithms which maximize the time that users spend on the platforms in order to maximize the profits generated through advertising [13, 37]. This predefined optimization goal is usually criticized as “bias” embedded in those information filtering algorithms [5, 12], and those biases may have unintended consequences. Reliance on the homogeneity signals may lead to an entrenchment of the established sources at the expense of novel sources [4, 28]. This “homogeneity bias” is also termed as “filter bubble” in the media theories [27]. The filter bubble has become one of today’s most urgent issues for social media. Extensive research efforts have been devoted to this area [6, 10, 11, 17, 26, 27, 32]. However, existing research suffers from the following limitations.

Firstly, the previous studies usually study the “filter bubbles” phenomenon in the context where the social filtering mechanism and the algorithm-driven technological filtering effect co-exist [40]. This is in line with the fact that the recommendation engine comes to be an integral component in the existing social media, e.g. Weibo, Twitter, and etc. On the other hand, we note that due to the exponential increase in the available contents online and the technological development in the field of recommendation technologies, recommendation algorithm-driven news aggregators and recommenders have played an increasing important role. However, less attention is paid to the news aggregators and recommenders where the information dissemination is dominated by the algorithm-driven technological filtering. Secondly, the previous studies usually equate the filter bubbles with community structures and emphasize this exogenous isolation effect. There is however a lack of the full discussion of the internal organization of filter bubbles [25]. In [25], the authors studied the endogenous structure of filter bubbles in social networks. However, a detailed understanding of the endogenous news composition, w.r.t topic distribution, is missing.

In the industrial real-world recommenders, such as Toutiao¹, the recommendation system is designed in a way where tens of millions of customers and millions of items are discretized into the one-hot representation [7, 20, 24]. The feature size after discretization and the feature crossing can easily reach millions or even billions [7, 20]. The internal hyperparameter mechanics of the algorithm design suffers from a significant lack of explainability, which makes it a major analysis barrier to evaluate the impact of recommenders in a “white-box” way (i.e. the algorithm design is given) through directly analyzing the algorithmic design.

¹ <https://www.toutiao.com/>

In this paper, we consider the filter bubble effect problem of the personalized recommenders and develop a four-phase Select-Setup-Link-Evaluate (SSLE) framework to exploit in a “black-box” way (i.e. the algorithm design is unknown) for the problem.

We outline the procedures in applying SSLE to address the filter bubble effect evaluation problem. Further, we propose the impact evaluation metrics of news aggregation and recommendation systems, and expand the traditional user satisfaction and the accuracy-oriented evaluation metrics of recommenders. We illustrate the effectiveness of the SSLE framework with empirical studies on Toutiao, Baidu News, and Tencent News, which are the top three news aggregators and recommenders.

Our main contributions are listed as follows:

1. A four-phase skeletal solution framework is proposed to target at exploiting the filter bubble effect on the information receiver’s exposure to the recommended news at the joint influence of individual filter and technological filter. This solution framework does not rely on the design of the recommendation function, and thus, is a black-box and lightweight solution.
2. We present an empirical analysis of the filter bubbles on three real-world news recommenders, including Toutiao, Baidu News, and Tencent News, which are the top three news aggregator apps in China, by illustrating the effectiveness of the proposed solution framework.
3. We find that the news recommenders interactively filter out the users’ uninterested content and filter in the content that caters for the user’s interests, and thus, the users are narrowed into one or a limited number of topics over about 7-day period. The phenomenon of the narrowed topics is deemed as the emergence of the “filter bubbles”.
4. We also observe that the filter bubbles demonstrate different convergence degrees as user’s individual preference varies, with users interested in “Society” facing with several converged topics as the users are also interested in several other topics as well, while users interested in “Entertainment” topic facing with one big converged topic. With the same reason, the converged topics differ with different user groups. Furthermore, the technological filter reflects and strengthens the individual filter. With the proliferation of “self-media” and users’ growth of interest of UGC over PGC, there is a corresponding increment of UGC over PGC recommended by the news aggregator and recommender, with the highest percentage of UGC recommended reaching around 83%, and the lowest reaching about 50%.

We survey related work in Section 2. This is followed by a concrete outline of the problem statement and the solution framework in Sections 3.1 and 3.2, respectively. Section 4 details our experiment processes using the Select-Setup-Link-Evaluate (SSLE) solution framework. This followed by our empirical evaluation in Section 5 and conclusions in Section 6.

2 Related work

In this Section, we present previous works related to our study. First the concept of filter bubble and the empirical investigation on social media is briefly presented in order to introduce the reader into the world of network evolution. After that recommendation engines

that enable the information spread in an algorithm-driven way are discussed. Finally, we present the state-of-the-arts of the evaluation metrics of recommenders.

2.1 Filter bubble and social media

The filter bubble is an intermediate structure formed in online digital space constraining an individual's exposure to a full spectrum of news and other information on the internet. Online media providers, such as Google or Facebook, compete for user attention. The goal is to maximize the time users spend on their respective sites, thus to maximize profits generated through advertising. Therefore, the information provided on online media is selectively filtered through the available flood of information, according to the algorithmically assumed individual users' wants and needs. In [13], the authors propose the triple filter theory, namely, individual filter, social filter, technological filter [13] in the formation of filter bubbles. Individual filter refers to the filtering of content caused by users' own preferences, tastes, interests. Social filter refers to the filtering of content by pushing the content that is posted, retweeted, and favored by the user's online social connections, whether it is bilateral friendship relation as in Facebook, or unilateral following relation as in Twitter. Technological filter refers to the filtering of content caused by algorithmically selecting and tailoring the content that suits the user's tastes or excluding content that does not, examples include pagerank, recommendation algorithms [4], etc.

Previous studies usually study the "filter bubbles" phenomenon in the context where social filtering mechanism and algorithm-driven technological filtering effect coexist. This is in line with the fact that the recommendation engine comes to be an integral component in existing social media, e.g. Weibo, Twitter, etc. On the other hand, we note that due to the exponential increase in available content online and technological development in the field of recommendation technologies, recommendation algorithm-driven news aggregators and recommenders have proliferated and gained an increasing market share. However, less attention is paid to the news aggregators and recommenders where the information dissemination is dominated by algorithm-driven technological filtering.

2.2 Recommendation engines and its filtering mechanisms

Recommendation engines (REs) enable the web platforms to elicit the interests or preferences of individual users for the items, either explicitly or implicitly. Recommendation engines have been used in different areas, including e-commerce, news media, and education since its invention [35]. To the end users, the degree to which the recommendation engines understand them and are personalized for them are key factors in the adoption of the web platforms [36]. Therefore, a vast amount of research [7, 20, 23, 38] has been devoted to design and develop underlying algorithms to generate the personalized recommendations.

Recommendation engines essentially are data filtering tools designed to understand the individual needs of particular users and fetch the most relevant items to a particular user. Broadly speaking, there are three categories of the filtering mechanisms: (i) collaborative filtering, where recommendation engines recommend content based on user neighbourhood information and give recommendations by combing the neighbors' previously selected content; (ii) content-based filtering, where recommendation engines recommend content based on similarities of properties and characteristics between previously chosen and available content; (iii) hybrid filtering, where recommendation

engines combine both approaches [36]. Of these three mechanisms, we can see that the “relevancy” of the items to the end users are determined based on “similarity”, whether it is similarity between users or item or a combination of both. The technological filtering effect driven by recommendation is criticized as “homogeneity bias” [4, 31], with the concern that people are exposed to similar minds and discuss similar views with like-minded peers [4, 31].

2.3 Evaluation metrics of recommenders

Prior research on REs evaluations [1, 18, 21, 22, 29] has focused mostly on accuracy metrics, such as accuracy, Root Mean Square Error, roc curve, ranking score, and Precision@ N and recall@ N for top- k recommendations, where N represents the size of the recommendation list [16]. Silveira’s work [34] categorizes the evaluation metrics of recommendation quality into two groups: (i) base metrics, including “utility” (i.e. accuracy metrics), “novelty” (the extent to which the items are unknown to a user), “diversity” (variety of items in recommendation lists), “coverage” (the extent of items that a recommender is capable of working with), and (ii) compound metrics, including “unexpectedness” (divergence from the history of a user, related to “diversity”, “coverage”), “serendipity” (a lucky finding or a satisfying surprise, related to “accuracy”, “novelty”, and “diversity”). Other work [33] further adds “context compatibility” (referring to whether or not the recommendations consider general or personal context requirements, e.g. users’ current mood, different occasions for watching the movie, etc.) to the category of recommendation quality.

However, these evaluation metrics largely focus on the satisfaction or improving the user experience. Little is known about the social impact that these systems have on the audiences. Take “novelty” as an example, the concept of “novelty” involves three levels: (i) life level novelty, the extent to which the items are never consumed or known in the users’ lifetime, (ii) system level novelty, the popularity of items in the list, (iii) recommendation list level novelty, the extent to which the items are not repeatedly recommended [34]. To put it briefly, the optimization goal of recommenders is to improve the customer retention rate, thus it will inevitably lead to the negligence of its potential cocoon effect and other impacts on the audiences/information receivers [2].

Here, we approach the impact evaluation problem from an information receiver’s perspective, while existing research approaches from an information sender’s perspective, e.g. opinion leader [40] and his/her social status (usually in terms of follower count and the position in the network structure). In our previous work [39], we proposed the impact evaluation metrics of the personalized news recommenders that approach from an information receiver’s perspective. In this paper, we further generalize the evaluation process into a four-phase framework, and and dive into the differences of the filter bubbles formed under different topics.

Note that there are certain evaluation aspects of recommender systems that are out of the scope of this paper. In particular, we have decided to avoid the general usability evaluation of the interface, including the perceived usefulness and perceived ease of use of recommender systems, and the large area of marketing-inspired evaluation, as these are well covered in research and practitioner literature [19, 33]. We also do not consider the emerging robustness and explainability discussions of recommender algorithms, as substantially more solid work needs to occur.

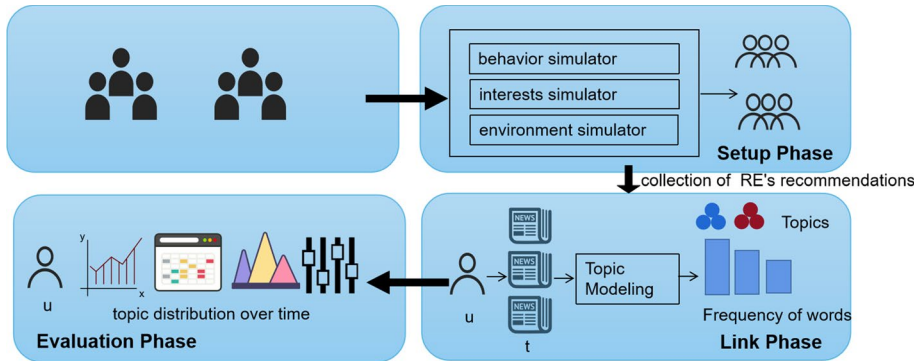


Fig. 1 Solution framework

3 Problem statement and solution framework

We now outline the problem statement more formally and introduce the solution framework.

3.1 Problem statement

Given a recommender engine RE and the demographic profile of a user group UG that is of interest to the analysts, the problem is defined as to evaluate to what extent a user $u \in UG$ receives content from a narrow set of information sources. The recommendation function itself is external to the task. It may be noted that we do not presume availability of usage data (e.g., query logs) in addressing the RE evaluation tasks. However, apriori-knowledge $A = \langle F_B, F_I, F_E \rangle$ comprising of behavioral feature F_B , interests feature F_I and environmental feature F_E is presumed as available.

3.2 Solution framework

We now outline our four-phase skeletal framework for receiver-centric impact evaluation of personalized news aggregator and recommender. An overview is given in Figure 1. The four phases are as follows:

1. **Selection:** Different user groups bear different interests and preferences, thus may lead to different topic convergence distributions over time. To shed light on the influence of users' preferences on the technological algorithm-driven filters of the news aggregator and recommender, we select four user groups that are of interest, that is, white-collar, blue-collar, cyber-celebrity, and students.
2. **Setup:** This phase sets up the bots that mimic the selected user group's online behavior based on the apriori-knowledge $A = \langle F_I, F_B, F_E \rangle$. The bots simulation outline is shown in Fig. 2. To be more specific, the simulation comprises of the following steps:

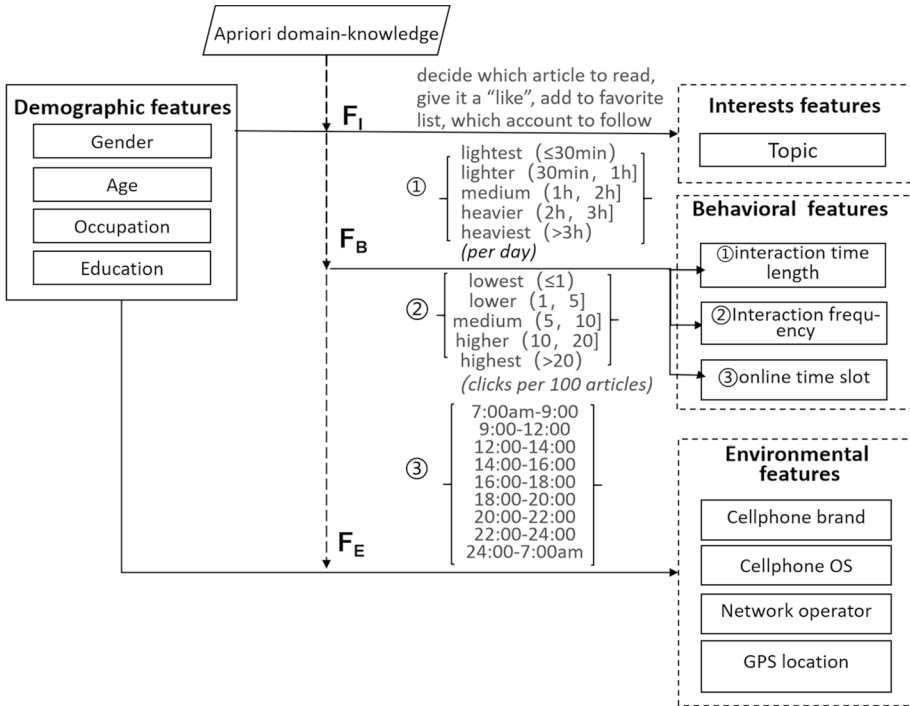


Fig. 2 Simulation Outline

- (a) **Log-on Behavior Simulation based on F_B .** The behavioral feature $F_B = \langle F_{IF}, F_{ITL}, F_{OTS} \rangle$ comprises of interaction frequency (clicks per 100 articles), interaction time length per day, and online timeslots. For each avatar user u , first we get the randomized online time length tl within the intervals specified in the apriori domain knowledge based F_B template, and the randomized log-on times l . When $l=1$, it indicates a continuous activity within time slot i . When $l>1$, it indicates consecutive activities, l is no more than the number of hours of time slot i . Then, we get the randomized online time lengths tl'_i within each subdivided online time slot i' ($\sum tl'_i = tl_i, \sum tl_i = tl$). Let ts'_i denote the length of time slot i' , the user will log on within the subdivided time slot i' with starting time no later than $ts'_i - tl'_i$.
- (b) **News Reading Behavior Simulation based on F_I .** The initial topic features $F_I = \langle F_T, F_P \rangle$, $F_T = \langle T, W, D \rangle$ is the topic-keyword feature set, where $T = \{t_j\}$ is the topic set, $W = \bigcup W_j$ is the keyword set, and $D = f(T, W)$ is the topic-key distribution feature, $f(T, W)$ is the mapping function between T and W , F_P is the topic preference feature. Thus, using the top few terms in D relevant to topics in F_P are capable of eliciting news relevant to users' interests on RE . To simulate a user u 's news reading behavior, firstly, we get the randomized values of the simulated interaction frequency f . Then, the simulator decide on whether the

avatar clicks to read the news or not based on the topic preference distribution F_p , and the keywords in the keywords topic distribution feature D . If the article falls within W_p^u (the set of keywords of the preferred topics)², the articles will be clicked at the simulated clicked frequency within the subdivided online time slot i' . The recommendation page is simulated to slide till the ts'_i ends.

- (c) **Log-on Environment Simulation based on F_E .** The environmental feature $F_E = \langle F_{CellBrand}, F_{CellOS}, F_{NetworkOperator}, F_{GPS}, F_{IP} \rangle$, of which $F_{CellBrand}$ is defined as $F_{CellBrand} = \langle F_{AccessDistribution(CellCategory)}, F_{SalesDistribution(CellBrand)} \rangle$, where $CellBrand = \{\text{Huawei, Apple, Oppo, Vivo, Xiaomi, Oneplus, ...}\}$, and $CellCategory = \{\text{“Economy } (\leq 1500 \text{ RMB)”, “Normal (1500-5000 RMB)”, “Flagship } (\geq 5000 \text{ RMB)”, “Special (limited editions, and hunger marketing)”}\}$ is a partition of $CellBrand$, $F_{AccessDistribution(CellCategory)}$ is the social media access distribution w.r.t cell categories in $CellCategory$, and $F_{SalesDistribution(CellBrand)}$ is the sales distribution of cell brands within each category. We simulated cellbrand feature based on $F_{CellBrand}$ and then F_{CellOS} can be directly derived from $F_{CellBrand}$.
- (d) **Update User Profile and Model Adjustment.** For each topic, it may happen that topic keyword distribution shifts over a certain time period. That is, F_t of the simulation model is adjusted. Let W_t denote the set of the top m ($m=3$ in the experiment) keywords extracted for each article using TF-IDF method for the time interval t . Considering the recency effect, we segment the keyword set W_t (t is set weekly in the experiment) into recent past keyword set W_t^R (keywords of articles read in the past 7 days) and further past keyword set W_t^F (keywords of articles read in more than 7 days ago), and we have $W = W_t^R \cup W_t^F, W_t^R \cap W_t^F \neq \emptyset$. W_{t+1} is adjusted as follows,
 - For the keyword $w_j \notin W_t^R$ and $w_j \notin W_t^F, w_j \in W_{t+1}$
 - For the keyword $w_j \notin W_t^R$ and $w_j \in W_t^F$ and $w_j \notin W_t^H = \{w_j | w_j \notin W_{t-1}^R \text{ and } w_j \in W_{t-1}^F\}, w_j \in W_{t+1}$
 - For the keyword $w_j \in W_t^R, w_j \in W_{t+1}$

In addition, we note that users’ interests shift may also happen as time goes by, the interest shift is adjusted as follows. First, we apply CNN-based neural network to learn and predict to which topics the articles fall within³. Then, for each user, we sort the topics by its count of articles, and get the top n_t^u topic categories. If it is not identical to the set of topic categories defined in F_p , update F_p^u and W_p^u . Initially, the keywords are assigned with weights w_k as defined in D . As the simulator runs, we assign weights $w_k = w_k + w_c$ to the keywords belonging to the articles clicked to read, and weights $w_k = w_k + w_r$ to the keywords simply belonging to the articles recommended (in the experiment, $w_c=2, w_r=0.5$). The weights are then normalized to $[0, 1]$.

3. **Link:** This phase involves linking the news recommended to the receiver to topics. To achieve this, we make use of Latent Dirichlet Allocation based topic modeling (LDA)⁴.

² Note that $|W_p| = a * n_t^u$, n_t^u is the number of topics a user is interested in, $a = 3$ indicating that for each topic the top 3 most weighted keywords are kept).

³ In the experiment, we use K -fold ($K=10$) cross validation for the hyperparameter optimization. That is, Layer number= 2^n ($n=3,4,5,6$), Layer size= 2^m ($m=6,7,8$), filter size: $L * H$ ($L=3,4,5, H=1,2,3$), with the best parameters Layer number = 32 ($n=5$), Layer size = 128 ($m=7$), filter size = 9 ($L=3, H=3$).

⁴ In the experiment, the initial parameters are set as $\alpha = 50/k$, k is the topic number ($k=26$ in the experiment), $\beta = 0.01$. In hyperparameter tuning, $\alpha = 50/k \pm n*(50/k)*0.2$ ($n=1,2,3,4,5$), $\beta = 0.01 \pm n*0.01*0.2$ ($n=1,2,3,4,5$).

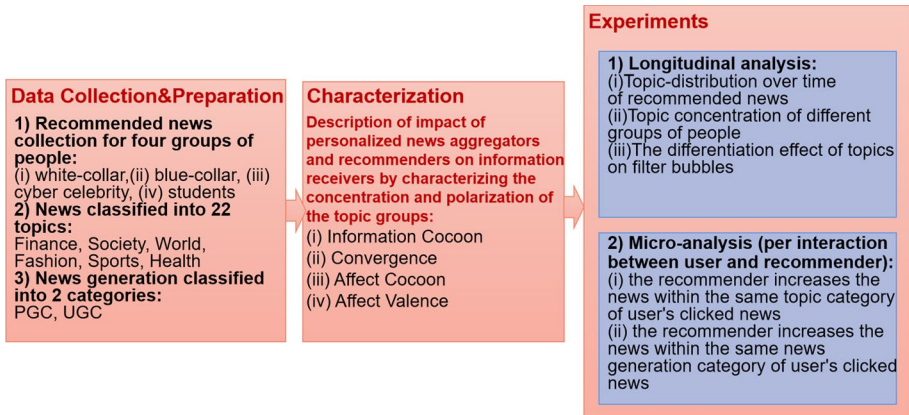


Fig. 3 Research Methodology

4. Evaluation: This phase involves analyzing the topic distribution over time of the news recommended to the receiver and evaluate whether the topic distribution tends to form filter bubbles. The topic distribution might vary with different user groups due to user's preference and interaction. Thus, differences in the previous phase across the user group entail consequent differences in the number and composition of the filter bubbles. The analysis consists of:
 - (a) longitudinal analysis. If the topics change over time from an evenly-distributed state to a skewed distributed state, the head several topics that take up the most percentage of the topic distribution is deemed as the filter bubbles for the users on the personalized news aggregator and recommender. Also the differentiation effect of topics on filter bubbles was investigated in order to find out the characteristics of filter bubbles formed under the joint influence of individual filter and technological filter.
 - (b) microscopic per interaction analysis. The technological filtering w.r.t user's individual preference is analysed at a per interaction micro-scopic view. Firstly, the recommender's topic-wide reaction after a user's click was analyzed for all the users, whether it increases the news within the same topic category of user's clicked news. Secondly, the recommender's reaction w.r.t news generation type (user generated content(UGC) and professionally generated content(PGC)) after a user's click was analyzed for all the users, whether it increases the news within the same topic generation type category of user's clicked news.

4 Experiments

The research methodology employed in this work is outlined in Figure 3. The building blocks of the methodology are: (i) data collection and preparation process, (ii) characteristics of recommended news distribution at the receiver's end.

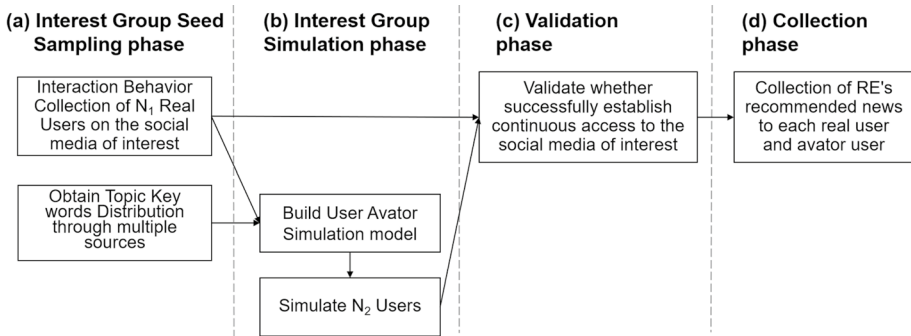


Fig. 4 Overview of data collection

4.1 Data collection

An overview of data collection is given in Figure 4. We have distributed 15 cellphones and deployed 100 bots (that is, $N_1 = 15, N_2 = 100$ in Figure 4) in 2019.12~2020.05 to collect the news recommended to the users (white-collar, blue collar, cyber celebrity, students) on Toutiao, Baidu News, and Tencent News. The bot mimics the user’s reading behavior by clicking the news containing information of relevance to the topics of interest specified in the user group’s profile Section 3.2. The parameters setting are as follows: For behavior features F_B , examples are given in Table 1. For interests features F_I , top 20 key words distribution of selective topic categories are shown in Table 2. For environmental feature F_E , the $F_{AccessDistribution(CellCategory)}$ and $F_{SalesDistribution(CellBrand)}$ are given in Table 3 and 4. The GPS locations and IP addresses are set in Nanyang in Henan province, China. The responses of the recommender per user’s click was collected. The behavioral statistics is shown in Table 5.

Bot validation To perform the effectiveness evaluation of the setup bot avator, we utilize two metrics to evaluate the compatibility of the avatar simulator and real users, with corrected conditional entropy (CCE) for temporality compatibility and Kullback-Leibler divergence [14] for interests compatibility. Note that the lower value for these metrics indicates the better performance.

Temporality compatibility The corrected conditional entropy (CCE) [8] evaluates whether the complexity of the avatar users’ online behavior is compatible with that of the real users’.

$$CCE = CE(X_m|X_1, \dots, X_{m-1}) + perc(X_m)dotEN(X_1) \tag{1}$$

where the first part is the conditional term CE , and the second part is the corrective term, X_m represents the online click times sequence within the time slots. A lower value of CCE indicates periodic or regular timing of tweeting behavior, while a higher value implies irregular behavior. Simply put, a lower value of CCE can be seen as a sign of automation, while a higher value can be seen as a sign of human participation.

Interests compatibility Kullback-Leibler divergence evaluates the difference between the probability distributions of topics read (on a weekly basis) for real user u_r and avatar user u_a , and is defined as:

$$KL(u_r, u_a) = H(u_r, u_a) - H(u_r) \tag{2}$$

where i denotes topic index, $H(u_r, u_a)$ is the cross-entropy, and $H(u_r)$ is the entropy of real user. A smaller value of Kullback-Leibler divergence indicates a closer interest distribution.

Table 1 Example Apriori Behavioral Features

Demographic features	F_p	F_{InterL} (per day)	F_{InterF} (per 100 articles)	F_{OnSlot}
(male, 16, blue-collar worker, third-tier city)	games, manga, video	(2h,3h]	(1,5]	18:00-20:00, 20:00-22:00
(male, 18, college student, first-tier city)	games, science and technology, video	(2h,3h]	(5,10]	7:00-9:00, 12:00-14:00, 18:00-20:00, 20:00-22:00
(female, 25, white-collar worker, first-tier city)	finance, tourism, society, entertainment, pet	(30min, 1h)	(1,5]	20:00-22:00, 22:00-24:00
...

Table 2 Topic-Keyword distribution matrix with top 20 keywords for selective topic categories

#	Category			Category			Category					
	Key Words	Degree	“Manga”	Key Words	Degree	“Finance”	Key Words	Degree	“Game”	Key Words	Degree	“Tourism”
1	iphone	0.9918	cosplay	bicycle	0.9807	bicycle	lpl	0.9788	bicycling	bicycling	0.9984	bicycling
2	epi	0.9983	dinosaur	credit card	0.9513	credit card	edg crew	0.9738	Self-drive tour	Self-drive tour	0.9985	Self-drive tour
3	provident fund	0.9718	exotic	Xiaomi phone	0.9437	Xiaomi phone	World of Warcraft	0.9738	seafood	seafood	0.9492	seafood
4	credit card	0.9747	Naruto	Pak	0.9419	Pak	League of Legends	0.9586	wedding	wedding	0.9484	wedding
5	diet	0.9639	Miyazaki Hayao	entrepreneurship	0.9417	entrepreneurship	diamond	0.9490	swimming	swimming	0.9462	swimming
6	entrepreneurship	0.9500	Inuyasha	bio-technology	0.9398	bio-technology	Ye Luo Li	0.9484	snack	snack	0.9413	snack
7	wedding	0.9507	toro (Japanese cartoon)	cloud computing	0.9390	cloud computing	Nintendo	0.9483	typhoon	typhoon	0.9406	typhoon
8	luxury goods	0.9490	arts	Maker	0.9381	Maker	Minecraft	0.9482	photography	photography	0.9388	photography
9	e-commerce	0.9498	Attack on Titan	coal	0.9381	coal	Sun Shangxiang	0.9433	folk custom	folk custom	0.9384	folk custom
10	graduate school exam	0.9488	Elsa	philosophy	0.9379	philosophy	Overwatch	0.9431	visa	visa	0.9381	visa
11	silver	0.9487	ACG	oil price	0.9375	oil price	soccer	0.9423	bougain-villea spectabilis	bougain-villea spectabilis	0.9381	bougain-villea spectabilis
12	university	0.9485	Cosmo	Xerox	0.9374	Xerox	ps5	0.9435	aspen	aspen	0.9374	aspen
13	kinder-garten	0.9483	Nezha	salary guideline	0.9373	salary guideline	skt	0.9394	the Palace Museum	the Palace Museum	0.9371	the Palace Museum
14	mental health	0.9481	Hisugaya Toushirou	Teamax	0.9359	Teamax	FC	0.9390	Mount Huang	Mount Huang	0.9368	Mount Huang
15	cellphone	0.9480	Jiang Ziya	tea seed oil	0.9349	tea seed oil	Zelda	0.9388	Shangri-La	Shangri-La	0.9367	Shangri-La
16	square dancing	0.9470	Slam Dunk	spinning	0.9349	spinning	Devil May Cry	0.9387	Tarim Basin	Tarim Basin	0.9333	Tarim Basin
17	Ren Chang-xia	0.9393	Disney	rice	0.9348	rice	Tang clan	0.9387	Han Chinese clothing	Han Chinese clothing	0.9331	Han Chinese clothing
18	second-hand housing	0.9388	Kindaichi	agricultural products	0.9347	agricultural products	Netease	0.9385	Qimling Mountains	Qimling Mountains	0.9300	Qimling Mountains
19	Shiba Inu	0.9387	Nobita	acquisition and reorganization	0.9347	acquisition and reorganization	Sony	0.9373	red maple leaf	red maple leaf	0.9288	red maple leaf
20	the Drung	0.9386	Photo stickers	Evergrande	0.9347	Evergrande	Luigi	0.9372	stalactite	stalactite	0.9295	stalactite

Table 3 Access Distribution of Cellphone Categories—“Economy (≤ 1500 RMB)”, “Normal (1500-5000 RMB)”, “Flagship (≥ 5000 RMB)”, “Special (limited editions, and hunger marketing)”

News Apps	Economy(≤ 1500)	Normal(1500-5000)	Flagship(≥ 5000)	Special(limited editions, and hunger marketing, etc.)
Toutiao	34.30%	51.60%	9.80%	4.30%
Baidu News	32.30%	50.50%	12.30%	4.90%
Tencent News	28.50%	57.90%	8.40%	5.20%
Average	31.70%	53.40%	10.10%	4.80%

* The access distribution was computed based on mainstream application store (Apple App store, Huawei App Gallery, PeaPod, AppChina) downloads of Toutiao, Baidu News, and Tencent News within 2020.01-2020.02

The results of behavior temporality verification is shown in Table 6. The results of interests compatibility verification is shown in Table 7. For each real user u_r , we get the minimum value of $KL(u_r, u_a)$, i.e. the closest interests distribution of avatar users to u_r . Note that the average topic read on Toutiao per week is 10.5, while the number for Baidu News is 8.4.

4.2 Data preparation

In preparing the data for evaluation analysis, the affect valence of an article is obtained using Naive Bayes classification models, see Figure 5. In the training phase, the news and comments are labelled as “positive” and “negative”, with examples shown in Table 8 and 9. The model learns the affect valence of each tokenized word based on the probability of its appearance in the positive and negative corpus, and predict the affect valence of testing news and comments.

4.3 Characteristics of the recommended news to the information receivers

An overview is given in Table 10. The evaluation of the filter bubble effect is based on the concentration of the topic groups reflected in the recommended news w.r.t each individual user, corresponding to the “Information Cocoon” and “Convergence Factor” in Table 10. If a user reads a large number of articles belonging to one single topic category, the risk of falling into the information cocoon is relatively large; otherwise, if a user’s reading interest is relatively scattered in the topics categories, the risk of falling into the information cocoon house is relatively low. As given in Table 10, the metric can be calculated w.r.t time interval or number of clicks. The latter enables a finer view as the number of news a user reads per day differs with each person. Further, to shed light on the filter bubble effect and polarization, we zoom into the concentration of the affects reflected in the recommended news w.r.t polarized groups of users, corresponding to the “Affect Cocoon” and “Affect Valence” in Table 10.

Table 4 Sales Distribution of main cell brands w.r.t the Cellphone Categories—“Economy (≤ 1500 RMB)”, “Normal (1500-5000 RMB)”, “Flagship (≥ 5000 RMB)”, “Special (limited editions, and hunger marketing)”

Economy (≤ 1500)	Sales (%)	Normal (1500-5000)	Sales (%)	Flagship (≥ 5000)	Sales (%)	Special (limited editions, etc.)	Sales (%)
Huawei Enjoy 10 Plus	7.91%	Apple 8 Plus	0.39%	Apple iPhone 11 Pro	23%	Samsung Galaxy Z Flip Thom Browne	4.69%
Huawei Honor 20 Youth	5.16%	Apple iPhone x	3.71%	Apple iPhone XS Max	18%	Huawei matexs	23.98%
Vivo y93	8.86%	Apple iPhone 7 Plus	2.78%	Huawei Mate 10	25%	Meizu E3 with Limited J-20 Edition	15.42%
Redmi Note8Pro	0.20%	Apple iPhone XR	8.44%	Huawei P40 Pro	3%	Xiaomi Mi 8 Explorer	31.12%
Redmi Note8	0.23%	Huawei Mate 30	1.36%	OnePlus 8 Pro	3%	OPPO Find X Lamborghini	4.69%
OPPO A5	9.81%	Huawei P30	0.37%	Huawei Mate 30 Pro	4%	Huawei Mate RS Porsche	8.21%
Xiaomi cc9e	0.89%	Xiaomi 10	0.86%	Apple iPhone 11	12%	Redmi Note4X Hatsune Miku Limited Edition	1.53%
Huawei Enjoy 9S	0.14%	OPPO R17	17.98%	Xiaomi 10 Pro	1%	OPPO R11 FC Barcelona Edition	1.99%
OPPO A7	2.86%	Samsung Galaxy S10	2.64%	Huawei Mate 30 RS	2%	Vivo Xplay6 Stephen Curry Edition	0.82%
Huawei Maimang 8	0.52%	Vivo iQOO 3	0.56%	Samsung Galaxy S20 5G	2%	Vertu Ti Ferrari	3.16%
Others	63.42%	Others	60.92%	Others	6%	Others	4.42%

* The sales data was collected within 2020.01-2020.02 from Jingdong (<https://www.jd.com/>)

Table 5 Behavioral Statistics on Toutiao, Baidu News, and Tencent News

User Type	Real User		Avatar User	
	# of Clicked Topic Categories*	# of Clicks*	# of Clicked Topic Categories*	# of Clicks*
Toutiao	(6,17,10.5)	(1,99,32.35)	(5,17,10.5)	(2,128,35.1)
Baidu News	(4,13,8.1)	(8,40,27.4)	(2,18,8.4)	(7,99,29)
Tencent News	(3,23,8.1)	(8,89,27.4)	(4,14,8.5)	(8,99,28.4)

*The (min, max, mean) statistics is calculated per day

Table 6 Behavior Temporality Verification of Simulation Model

	Toutiao			BaiduNews			TencentNews		
	minCCE	avgCCE	maxCCE	minCCE	avgCCE	maxCCE	minCCE	avgCCE	maxCCE
Real User	0.1	0.2	0.32	0.03	0.18	0.35	0.05	0.2	0.33
Avatar User	0.1	0.18	0.39	0.01	0.14	0.038	0.01	0.15	0.29

Table 7 Interests Compatibility Verification of Simulation Model

	Toutiao			BaiduNews			Tencent News		
	$H(u_r, u_a)$	$H(u_r)$	$KL(u_r, u_a)$	$H(u_r, u_a)$	$H(u_r)$	$KL(u_r, u_a)$	$H(u_r, u_a)$	$H(u_r)$	$KL(u_r, u_a)$
u_1	0.54	0.33	0.21	0.79	0.57	0.22	0.22	0.15	0.07
u_2	0.28	0.25	0.03	0.69	0.57	0.12	0.31	0.17	0.14
u_3	1.38	0.86	0.86	0.38	0.31	0.07	0.24	0.21	0.03
u_4	0.73	0.45	0.28	0.51	0.42	0.09	0.2	0.05	0.15
u_5	0.43	0.23	0.2	0.54	0.42	0.12	0.18	0.09	0.09
u_6	0.54	0.34	0.20	0.67	0.57	0.10	0.31	0.23	0.08
u_7	0.27	0.13	0.14	0.53	0.29	0.24	0.29	0.07	0.22
u_8	0.31	0.19	0.12	0.38	0.26	0.12	0.34	0.18	0.16
u_9	0.28	0.27	0.01	0.62	0.28	0.34	0.41	0.3	0.11
u_{10}	0.50	0.28	0.22	0.77	0.52	0.25	0.25	0.14	0.11
u_{11}	0.20	0.17	0.03	0.90	0.46	0.44	0.63	0.31	0.32
u_{12}	0.94	0.60	0.34	0.70	0.47	0.23	0.25	0.22	0.03
u_{13}	0.28	0.24	0.04	0.79	0.46	0.33	0.32	0.26	0.06
u_{14}	0.87	0.41	0.46	0.82	0.46	0.36	0.17	0.15	0.02
u_{15}	1.12	0.47	0.65	0.51	0.42	0.09	0.25	0.15	0.1

5 Experiment analysis and discussion

5.1 Longitudinal analysis

In this section, to evaluate the filter bubbles effect on information receivers, we analyze (i) the topic convergence trend over time, (ii) the time length it takes for the topics to get

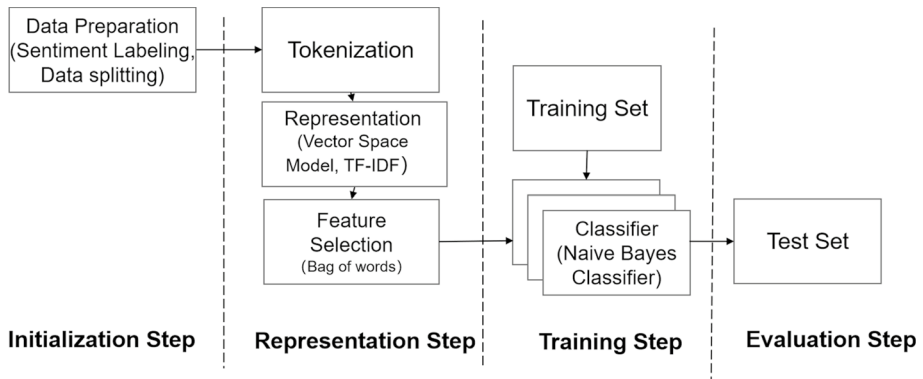


Fig. 5 Sentiment Classification

Table 8 Examples of positive news and comments

#	Example positive content
1	Comments of Zhejiang Daily The post-90s, well done
2	Dandelion Volunteer Association spreads love to the old and the young
3	Beijing Branch of Xinhua Insurance Company received the silk banner from the orphanage “365 Twilight Baby’s Home”
4	#The 20th anniversary of Da Zhang Wei’s debut# [love] # Da Zhang Wei, sunshine, rainbow, little white hors#So looking forward to it!!
5	The drama “Little veteran actor-Dream of the Red Chamber”, Lin Daiyu(character in the drama)’s crying and acting are heartbreaking. She did a fanstastic job in playing Lin Daiyu’s smile and frown. These young actors are amazing!
6	“The Wandering Earth” outdid the novels I read at high school [tear][tear][tear] wish that it can be equally excellent when they bring “The Three-Body Problem” alive on screen
7	The latest tour points of the National Table Tennis Federation came out, with Sun Yingsha at tYhe top for two consecutive months, and Chen Xingtong’s ranking soared
8	Zhu Yilong is super handsome!I am so dying!
9	Xihe Town: Carry forward the spirit of volunteerism and inspire the power of benevolence
10	#Lu Han [Chaohua]# checked in just in time, clicked in and exited by accident[Ha Husky]. Otherwise, he will get even higher on the rank. May August be a happy month

converged.

5.1.1 Impact evaluation of personalized news aggregators & recommenders

Figure 6 reports the evaluation of new recommenders w.r.t the constructs given in Table 10⁵. The measurement of “information cocoon” concerns about the concentration of the topics covered in the recommended news as a user’s online time (or a user’s clicks) increases,

⁵ The continuous affect valence within [0,1] is segmented into A categories, in the experiment we have A=3, with [0-0.35] as “Negative”, [0.35-0.75] as “Neutral”, and [0.75-1] as “Positive”.

Table 9 Examples of negative news and comments

#	Example negative news and comments
1	it is suspected that a billboard fell off in Huanshan road in Jinan, a passer-by was injured
2	Detailed rules of bank financial management was disclosed and sadphishing companies got punished Financial Daily review
3	What happened yesterday? It is clear that the stock is rising, but the earnings of the day are negative???
4	Huatai New Energy EV160r steering gear stuck leads to oil leakage of gearbox
5	25 civilian casualties in a roadside bomb attack in Eastern Yemen
6	Reply: the big rappers in Hunan province all came out this time, but still failed
7	Disasters caused by continuous rainfall in Myanmar resulted in at least 90 deaths
8	I gave out at least eight bad comments to DiDi drivers these two days
9	Tong Ren Tang (Traditional Chinese Medicine company) apologized for the incident. Netizens shouted: “don’t smash the signboard by yourself”
10	Shanghai police reported that the arrest of Huang Yiqing had been submitted for rastification on a charge of drug trafficking and drug use

see the first and the third curve in Figure 6 A. In addition, we are also interested in its convergence factor, i.e. the time (or the clicks) it takes for the news recommended to get concentrated, see the second and the fourth curve in Figure 6 A. The cocoon factor reflects the convergence degree of the topics of the news recommended to the end user. if the recommender recommends a large number of articles belonging to one single topic category to the end user, the risk of falling into the information cocoon is relatively large; otherwise, if the news recommended to the user is relatively scattered in the topics categories, the risk of falling into the information cocoon house is relatively low. As mentioned earlier, the metric can be calculated w.r.t time interval or number of clicks. The latter enables a finer view as the number of news a user reads per day differs with each person. The convergence factor characterizes the length of the interval that the news recommended to a user gets converged. We can observe from Figure 6 that it takes around 7 days or 300 clicks for the recommender to get converged.

Table 11 shows the highest information cocoon factor that the four group of users could achieve. We can see that the information cocoon factor the students and blue-collar group are the highest, the information cocoon factor for the white-collar group is the lowest. The cyber celebrity group is second to the lowest. This is in line with the observation that the capability in information retrieval and the diversity of information acquisition channels for the white-collar group is better, and the interests for the cyber celebrity group are relatively more diverse than the others.

As for the affect cocoon evaluation, the affect cocoon result is given in the top part in Figure 6 B. We can see that there is no significant trend of the affect cocoon curve. Therefore, to shed light on the affect distribution within the news recommended to the end users, we classify the users into three groups, positive, negative, and neutral⁶. The results of affect distribution within each user group is shown as the bottom part in Figure 6 B. We observe that unlike user information cocoon, the affect distribution is distinctly related with the

⁶ The classification for positive/negative group is with average percentage of positive/neutral clicked articles > 50%, the remaining group of users is neutral group

Table 10 Characteristics of recommended news to the information receivers

Variables	Definition	Measurement
Information Cocoon	the concentration of the topic groups reflected in the recommended news w.r.t each individual user.	
· information cocoon over time	the degree to which the topic distribution over the articles recommended to user u by recommender r gets concentrated as the user's online time increases.	$InformationCocoon_r(r, u, D_t) = norm(\sum_k^K ((p_k^u)^2))$, $p_k^u \in D_t$, where D_t is the topic distribution of the recommended articles read by user u within the t^{th} day, $p_k^u \in D_r$, $p_k^u = Count_k^u / Count_{Total}^u$ is the proportion of the articles recommended to user u belonging to topic k , K is the number of topics recommended by recommender r
· information cocoon over clicks	the degree to which the topic distribution over the articles recommended to user u by recommender r gets concentrated as the user's number of clicks increases.	$InformationCocoon_c(r, u, D_c) = norm(\sum_k^K ((p_k^u)^2))$, $p_k^u \in D_c$, where D_v is the topic distribution of the recommended articles read by user u within the $((c - 1) * s, (c - 1) * s + 1)$ clicks, s is slot size, in the experiment, $s = 500$, $p_k^u \in D_{cs} p_k^u = Count_k^u / Count_{Total}^u$ is the proportion of the articles recommended to user u belonging to topic k , K is the number of topics recommended by recommender r
Convergence Factor	the movement of the topic distribution of a particular user from the lowest point towards the highest point.	
· convergence factor over time	the number of days that it takes for the number of articles recommended within the topic category k that constitutes the largest percentage within an observing period (per week/per 500 clicks) to reach its highest point.	$Convergence_r(r, u) = \{d\}max\{Count_k^r(d) / Count_{Total}^r(d)\}$, where $Convergence_r(r, u)$ represents the convergence index of recommender r w.r.t user's click/news read online, $Count_k^r(t) / Count_{Total}^r(t)$ is the proportion of the articles belong to topic k at day d .
· convergence factor over clicks	the number of clicks that it takes for the number of articles recommended within the topic category k that constitutes the largest percentage within an observing period (per week/per 500 clicks) to reach its highest point.	$Convergence_c(r, u) = \{c\}max\{Count_k^r(c) / Count_{Total}^r(c)\}$, where $Convergence_c(r, u)$ represents the convergence index of recommender r w.r.t user's clicks, $Count_k^r(c) / Count_{Total}^r(c)$ is the proportion of the articles belonging to topic k at the c^{th} slot of clicks (news read online).
Affect Cocoon	the concentration of the affects reflected in the recommended news	$AffectCocoon_r(r, u) = norm(\sum_a^A ((p_a^u)^2))$ where A is the number of affect categories, p_a^u is the proportion of the articles belonging to category a .
Affect Valence	the valence of the affects reflected in the recommended news	see the description in Section 4.2

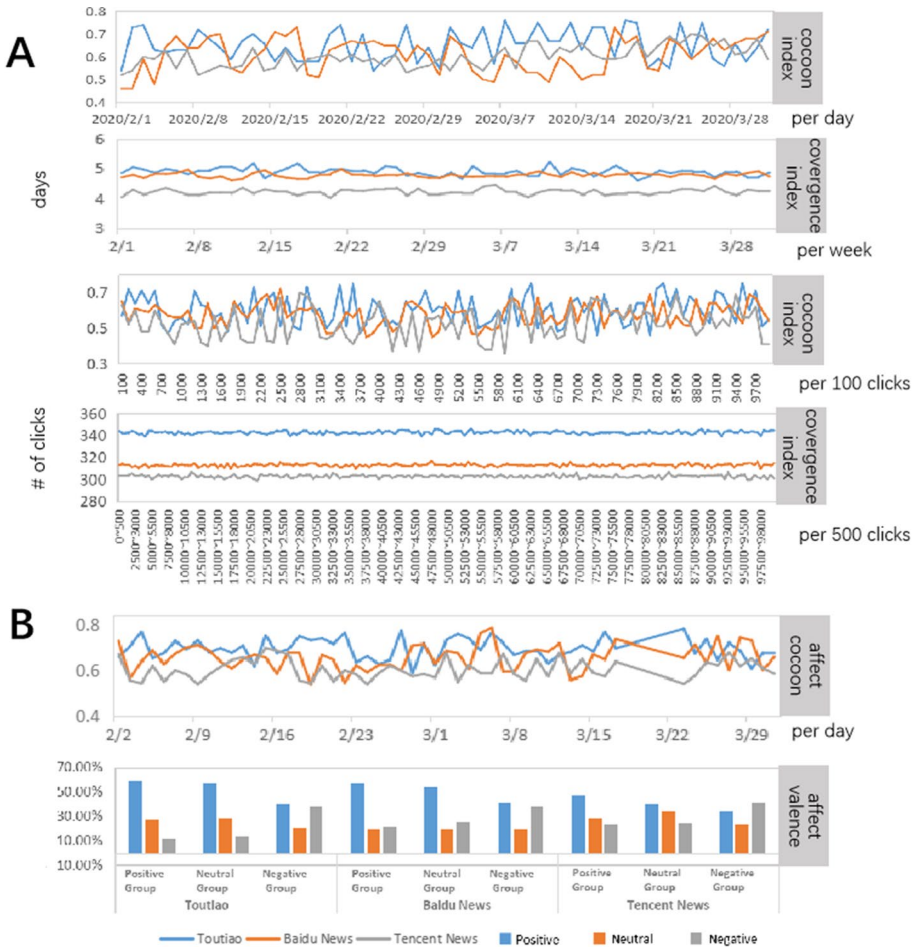


Fig. 6 Impact evaluation of Toutiao, Baidu News and Tencent News over time. A. Information cocoon and convergence evaluation. B. Affect cocoon and affect valence evaluation. The users are classified into three groups, positive, negative, and neutral. The classification for positive/negative group is with average percentage of positive/neutral clicked articles > 50%, the remaining group of users is neutral group

Table 11 The highest information cocoon factor w.r.t the four groups on Toutiao, Baidu News, and Tencent News

	White-Collar	Blue Collar	Cyber Celebrity	Students
Toutiao	0.6	0.67	0.63	0.65
Baidu News	0.68	0.71	0.69	0.72
Tencent News	0.57	0.63	0.59	0.66

user group. The affect distribution for the positive and neutral user group is similar, with a skewed distribution towards the positive news category. The affect for the negative user group is more evenly distributed.

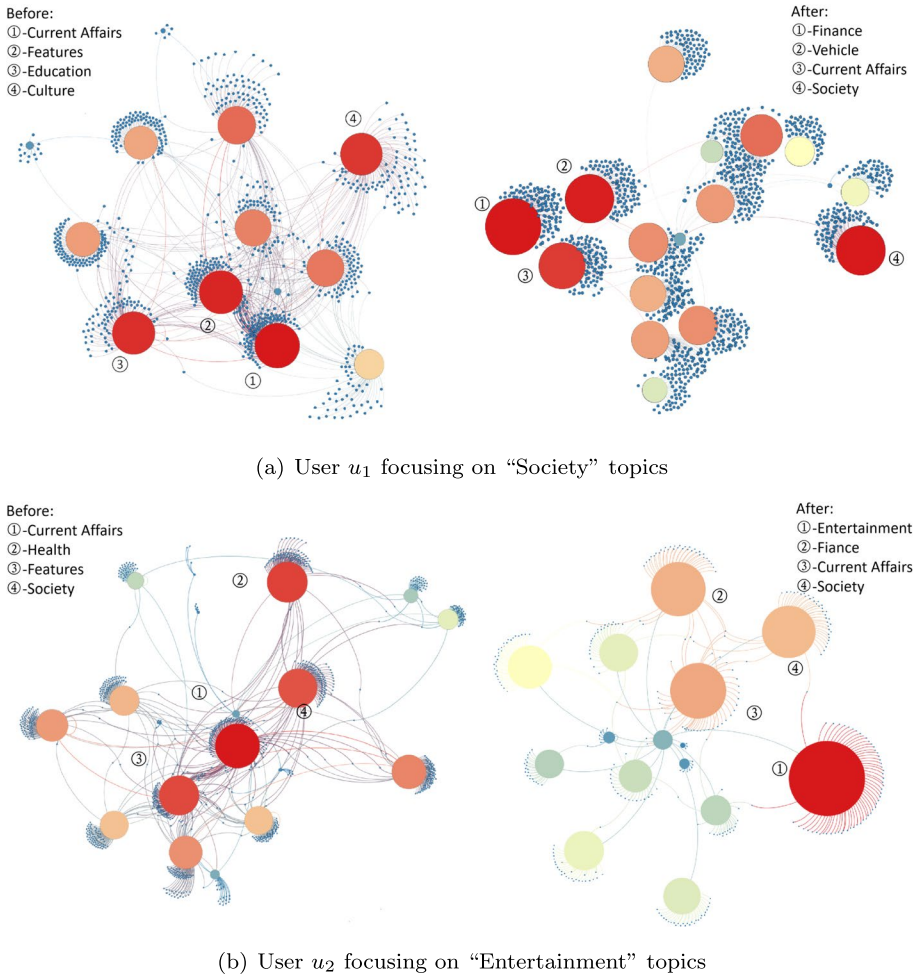


Fig. 7 Filter Bubbles – the converged topics as a user’s online time and interaction with the recommender increases

5.1.2 Differentiation effect of topics on filter bubbles

Figure 7 shows the topic distribution of the news recommended to a particular user, where the node size is correlated with the number of news within a topic, the color of the node varies from warm color to cold color with the node size as well, with a warmer color indicating a larger topic size, the blue nodes surrounding the topic node represent the keywords, a keyword may connect to multiple topics. Figure 7 (a) illustrates the changes in the topic distribution of the news recommended to user u_1 focusing on “Society” topics before and after the topic converges. The two points are selected at the beginning and the end of the experiment period. Figure 7 (b) illustrates the changes in the topic distribution of the news recommended to user u_2 focusing on “Entertainment” topics.

The results show that the topic distribution demonstrates different convergence degrees as user’s preference varies. For the users interested in “Society” topics, the news will turn

Table 12 The converged topics distribution w.r.t the four groups on Toutiao

Topics	White-Collar	Blue Collar	Cyber Celebrity	Students
Finance	18.78%	7.10%	5.85%	3.76%
Society	16.11%	27.63%	15.18%	26.44%
World	14.52%	14.97%	7.13%	5.65%
Current Affairs	11.26%	18.49%	9.92%	9.82%
Culture	8.63%	0.61%	0.76%	2.69%
Sports	7.77%	2.96%	2.85%	3.76%
Features	6.68%	6.29%	5.31%	6.26%
Health	3.49%	6.58%	2.63%	0.81%
Education	2.88%	2.65%	2.76%	22.71%
Vehicle	2.42%	0.29%	0.33%	0.27%
Entertainment	2.29%	2.21%	12.07%	5.16%
Parenting	1.25%	1.05%	1.28%	–
Tech	0.97%	1.09%	0.90%	0.90%
Food	0.45%	0.53%	4.36%	0.37%
Sentiment	0.44%	0.41%	3.38%	0.35%
Travel	0.40%	0.41%	2.49%	0.50%
Music	0.34%	0.29%	0.21%	0.30%
Manga	0.34%	0.38%	0.35%	2.32%
Horoscope	0.11%	0.08%	0.07%	0.08%
Household	0.10%	1.11%	0.11%	0.06%
History	0.08%	2.08%	0.04%	0.08%
Fashion	0.04%	0.07%	16.03%	0.03%
Games	0.01%	0.02%	5.02%	7.01%
Pet	–	–	0.20%	0.01%

into several converged topics (see Figure 7 (a)). For the users interested in “Entertainment”, the news will turn into one big converged topic (see Figure 7 (b)). This indicates the users interested in “Society” topics may also be interested in other topics, e.g. “Finance”, etc. While the users interested in “Entertainment” topic are solely interested in “Entertainment” topic, which indicates a certain degree of addictiveness. Table 12 shows the converged topics distribution w.r.t the four groups on Toutiao. We can see that the topics converged for white-collar group are “Finance”, “Society”, “World”, and “Current Affairs”, the topics converged for blue-collar group are “Society”, “World”, and “Current Affairs”, the topics converged for cyber celebrities are “Society”, “Current Affairs”, “Fashion”, the topics converged for students are “Society”, and “Education”. The results on Baidu News and Tencent News demonstrate similar convergence distribution.

5.2 Micro analysis

5.2.1 Interactive adjustment of recommenders w.r.t topic preference

Figure 8 shows the topic distribution of the recommended news and the topic distribution of the read news. We can see that for the topics that take up the most percentage, there

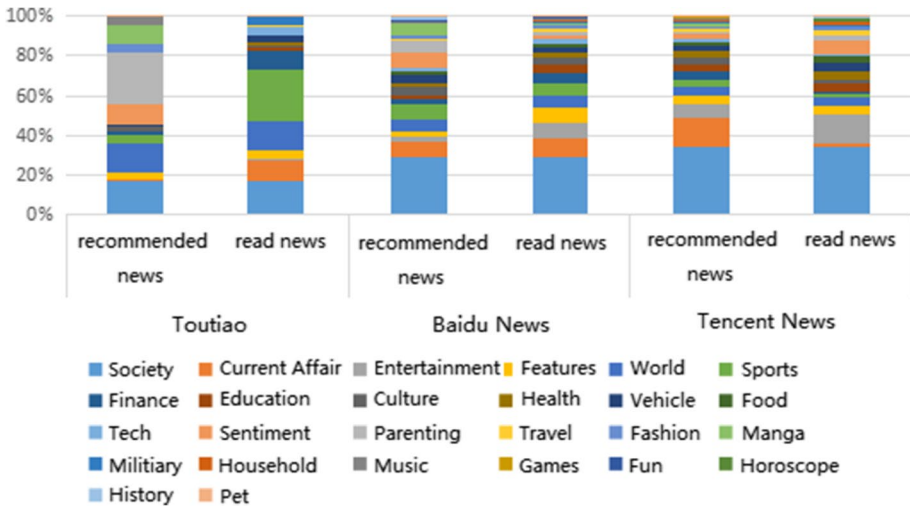


Fig. 8 Topic distribution of recommended news and read news on Toutiao, Baidu News, and Tencent News

Table 13 Topic-wide News Increment w.r.t Top 5 Keywords of Users’ Read News

Week Index	Toutiao		Baidu News		Tencent News	
	# of News Increment	% of News Increment	# of News Increment	% of News Increment	# of News Increment	% of News Increment
2019.12.6~13	39	39.0%	56	54.9%	45	44.1%
2019.12.13~20	6	4.32%	-52	-32.91%	-23	-14.9%
2019.12.20~27	-7	-4.83%	24	22.64%	20	21.33%
2019.12.27~1.03	34	24.64%	16	12.31%	15	11.9%

exists a high overlapping degree. Furthermore, we zoom into the topic-wide news increment w.r.t top 5 keywords of users’ read news at each interaction (see Table 13). We can see that the keywords belonging to the topics that a user previously clicked on the recommenders tend to bring an increment in the recommender’s responses. This explains the previous observation that the topics tend to converge in the long-run. Note that the decrement indicates a shift of keywords. This is in line with our daily experience and the emergence of new hot topics from time to time.

5.2.2 Interactive adjustment of recommenders w.r.t news generation type preference

Table 14 illustrates the cultivation between recommender and users w.r.t news generation type, i.e PGC and UGC. We can clearly see from the table that there is a significant decrement trend in the recommended PGC, in particular, there are three notably significant periods, i.e. 2020.01.05-2020.01.09, 2020.01.13-2020.01.20, and 2020.01.21-2020.01.27, with the highest reaching around 40%, and the lowest reaching about 16%. Correspondingly, there is a significant increment trend in the recommended UGC, with the highest percentage of UGC recommended reaching around 83%, and the lowest reaching about 50%.

Table 14 The distribution of recommended and read news w.r.t topic generation type

date	Percentage Recommended PGC (%)	Percentage Recommended UGC (%)	Percentage Read PGC (%)	Percentage Read UGC (%)
2020-01-02	31.47	68.53	40.00	60.00
2020-01-03	18.97	81.03	16.22	83.78
2020-01-04	20.65	79.35	23.33	76.67
2020-01-05	28.16	71.84	15.56	84.44
2020-01-06	26.50	73.50	34.38	65.63
2020-01-07	26.15	73.85	25.53	74.47
2020-01-08	28.02	71.98	22.86	77.14
2020-01-09	16.18	83.82	20.00	80.00
2020-01-10	29.48	70.52	22.22	77.78
2020-01-11	24.44	75.56	27.03	72.97
2020-01-12	29.67	70.33	20.93	79.07
2020-01-13	40.00	60.00	28.26	71.74
2020-01-14	36.64	63.36	45.24	54.76
2020-01-15	36.91	63.09	36.84	63.16
2020-01-16	35.96	64.04	26.83	73.17
2020-01-17	30.50	69.50	34.21	65.79
2020-01-18	32.17	67.83	38.71	61.29
2020-01-19	39.60	60.40	40.00	60.00
2020-01-20	29.33	70.67	24.00	76.00
2020-01-21	44.44	55.56	40.00	60.00
2020-01-22	41.33	58.67	40.63	59.38
2020-01-23	44.34	55.66	48.48	51.52
2020-01-24	38.51	61.49	33.33	66.67
2020-01-25	36.58	63.42	28.13	71.88
2020-01-26	32.37	67.63	11.11	88.89
2020-01-27	26.14	73.86	14.29	85.71
2020-01-28	39.75	60.25	17.65	82.35
2020-01-29	37.29	62.71	47.83	52.17
2020-01-30	49.32	50.68	57.14	42.86
2020-01-31	29.81	70.19	56.52	43.48

In the meanwhile, we can see that there is a similar trend in the read PGC and UGC on the recommender. However, the read news distribution w.r.t PGC and UGC is skewer than the recommended new distribution w.r.t PGC and UGC. The lowest point for the read PGC with the decrement trend could go even lower, around 11.11%, and the highest point for the read UGC with the increment trend could achieve even higher, 88.89%. This explains the increment in the recommended PGC news and decrement in the recommended UGC news.

The above experiments on Toutiao, Baidu News, and Tencent News illustrate that all three applications demonstrate a periodic topic convergence trend over about 7-day time interval, indicating the formation of “filter bubble” effect on the information receivers. Zooming into the topic-wide news increment at each interaction gives us a fine insight with the keywords belonging to the topics that a user previously clicked on the

recommenders tending to bring an increment in the recommender's responses. The topic distribution demonstrates different convergence degrees as user's preference varies. For the users interested in "Society" topics, the news will turn into several converged topics. For the users interested in "Entertainment", the news will turn into one big converged topic. The technological filter brings the individual preference to surface with the help of algorithms. The users interested in "Society" topics may also be interested in other topics. While the users interested in "Entertainment" topic are solely interested in "Entertainment" topic, and the recommender captures the preference, assimilates similar content and feeds back to the user at the per interaction, which indicates a certain degree of addictiveness over the long run. From this sense, the technological filter reflects and strengthens the individual filter. Furthermore, a similar strengthening effect can be observed that at the influence of individual preference of UGC over PGC, there is a corresponding increment of UGC over PGC recommended by the news aggregator and recommender.

6 Conclusion and future work

This paper proposes a four phase solution framework of evaluating the impact of personalized news aggregation and recommendation system on the information receivers. The introduced framework starts from an information-receiver perspective and focuses on exploiting the filter bubble effect at the joint influence of individual filter and technological filter. Further, we provide empirical evidence of this "partial information blindness" of filter bubble effects on personalized news aggregation and recommendation applications, with experiments on Toutiao, Baidu News, and Tencent News, the top three news aggregator apps in China. The results show that it demonstrates a periodic information cocoon of about 7-day time interval with the news recommenders interactively filters out the users' uninterested content and filter in the content that caters to the user's interests. Thus, the users are narrowed into one or a limited number of topics over time. The phenomenon of the narrowed topics is deemed as the emergence of the "filter bubbles". In addition, we also observe that the filter bubbles demonstrate different convergence degrees as user's individual preference varies. Although limited in scope, our findings may be of interest to those who are working to reduce political polarization in applied settings.

As for the generalization, the proposed solution framework approaches from a "black-box" way, and thus has wide applicability irrespective of the internal mechanics. One future direction is to extend the solution framework by incorporating the social filtering into the evaluation design. Another direction is to investigate the correlation between the triple-filtering mechanisms with the user retention rate, and social media addiction.

Acknowledgements The authors would like to thank all the people who have contributed to the data preparation for their selfless work. We also thank the anonymous reviewers for their valuable feedback and constructive criticism that have helped us to revise the manuscript. The work is supported by the National Key R&D Program of China (NO.2019YFC1509402 and NO.2018YFB0804204).

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Adomavicius, G., Kwon, Y.: Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering* **24**(5), 896–911 (2012)
2. Alkan, O., Daly, E. M., Botea, A.: An evaluation framework for interactive recommender systems. In: *Proceeding UMAP'19 Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization(UMAP)*, pp. 217–218 (2019)
3. Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A.: Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* **115**(37), 9216–9221 (2018)
4. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on facebook. *Science* **348**(6239), 1130–1132 (2015)
5. Beam, M.A.: Automating the news: How personalized news recommender system design choices impact news reception. *Communication Research* **14**(8), 1019–1041 (2014)
6. Borgesius, F.J.Z., Trilling, D., Möller, J., Bodó, B., de Vreese, C.H., Helberger, N.: Should we worry about filter bubbles. *Internet Policy Review* **5**(1), 1–16 (2016)
7. Chen, C., Zhao, P., Li, L., Zhou, J., Li, X., Qiu, M.: Locally connected deep learning framework for industrial-scale recommender systems. In: *The World Wide Web Conference (WWW)*, pp. 769–770 (2017)
8. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S.: Detecting automation of twitter accounts: are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing* **9**(6), 811–824 (2012)
9. Dandekar, P., Goel, A., Leec, D.T.: Biased assimilation, homophily, and the dynamics of polarization. *PNAS* **110**, 5791–5796 (2013)
10. Flaxman, S., Goel, S., Rao, J.M.: Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* **80**, 298–320 (2016)
11. Fletcher, R., Nielsen, R.K.: Are news audiences increasingly fragmented? a cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication* **67**(4), 476–498 (2017)
12. Garimella, K., Gionis, A., Parotsidis, N., Tatti, N.: Balancing information exposure in social networks. In: *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*, pp. 4663–4671 (2017)
13. Geschke, D., Lorenz, J., Holtz, P.: The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology* **58**(1), 129–149 (2019)
14. Goodfellow, I., Bengio, Y., Courville, A.: *Deep learning (Vol.1)*. Cambridge: MIT Press, pp. 71–73 (2016)
15. Guerra, P.H.C., Meira, W., Cardie, C., Kleinberg, R.: A measure of polarization on social media networks based on community boundaries. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media(AAAI)*, pp. 215–224 (2013)
16. Gunawardana, A., Shani, G.: A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research* **10**, 2935–2962 (2009)
17. Haim, M., Graefe, A., Brosius, H.: Burst of the Filter Bubble? Effects of personalization on the diversity of Google News. *Digital Journalism* **6**(3), 330–343 (2018)
18. Hayes, C., Massa, P., Avesani, P., Cunningham, P.: An on-line evaluation framework for recommender systems. In: *Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces* (2010)
19. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* **22**(1), 5–53 (2004)

20. Hu, Y., Yu, Y., Tu, W., Yang, Q., Chen, Y., Dai, W.: Multi-fidelity automatic hyper-parameter tuning via transfer series expansion. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence(AAAI), pp. 3846–3853 (2019)
21. Knijnenburg, B., Meesters, L., Marrow, P., Bouwhuis, D.: User-centric evaluation framework for multimedia recommender systems. *User Centric Media*, pp. 366–369, Springer Berlin Heidelberg (2009)
22. Krohn-Grimberghe, A., Nanopoulos, A., Schmidt-Thieme, L.: A novel multidimensional framework for evaluating recommender systems. In: Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (2010)
23. Liu, D., Bai, T., Lian, J., Sun, G., Zhao, W., Wen, J., Xie, X.: News graph: an enhanced knowledge graph for news recommendation. In: Proceedings of 2nd Workshop on Knowledge-Aware Recommender Systems (KARS) (2019)
24. Luo, Y., Wang, M., Zhou, H., Yao, Q., Tu, W., Chen, Y., Yang, Q., Dai, W.: AutoCross: automatic feature crossing for tabular data in real-world applications. In: KDD, pp. 1936–1945 (2019)
25. Min, Y., Jiang, T., Jin, C., Li, Q., Jin, X.: Endogenous structure of filter bubble in social networks. *Royal Society Open Science* **6**(11), 190868 (2019)
26. Möller, J., Trilling, D., Helberger, N., van Es, B.: Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society* **21**(7), 959–977 (2018)
27. Nikolov, D., Oliveira, D.F.M., Flammini, A., Menczer, F.: Measuring online social bubbles. *PeerJ Computer Science* **1**, e38 (2015)
28. Nikolov, D., Lalmas, M., Flammini, A., Menczer, F.: Quantifying biases in online information exposure. *JASIST* **20**(3), 218–229 (2019)
29. Olmo, F.H.D., Gaudioso, E.: Evaluation of recommender systems: A new approach. *Expert Systems with Applications* **35**(3), 790–804 (2008)
30. Paramythis, A., Weibelzahl, S., Masthoff, J.: Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adaptation Interaction* **20**(5), 383–453 (2010)
31. Pariser, E.: *The filter bubble: how the new personalized web is changing what we read and how we think*. Penguin, New York (2011)
32. Pariser, E.: *The filter bubble: What the internet is hiding from you*. Penguin UK, London, UK (2011)
33. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework of recommender systems. In: *RecSys* (2011)
34. Silveira, T., Zhang, M., Lin, X., Liu, Y., Ma, S.: How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics* **10**, 813–831 (2019)
35. Terveen, L., Hill, W. (eds.): *Beyond recommender systems: helping people help each other*. Addison-Wesley, New York (2001)
36. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: use, characteristics, and impact. *MIS Quarterly* **31**(1), 137–209 (2007)
37. Zhang, Y., Yin, H., Huang, Z., Du, X., Yang, G., Lian, D.: Discrete deep learning for fast content-aware recommendation. In: Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM), pp. 709–717 (2018)
38. Zhang, Y., Yin, H., Huang, Z., Du, X., Yang, G., Lian, D.: Discrete deep learning for fast content-aware recommendation. In: *WSDM*, pp. 709–717 (2018)
39. Zhao, Y., Wang, C., Han, H., Shu, M., Wang, W.: An impact evaluation framework of personalized news aggregation and recommendation systems. In: Proceedings of the 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 893–900 (2020)
40. Zhao, H., Xu, X., Song, Y., Lee, D.L., Chen, Z., Gao, H.: Ranking users in social networks with higher-order structures. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence(AAAI), pp. 232–239 (2018)
41. Zhou, X., Feng, H.: Research on the prevention and control of the internet rumor from the perspective of the self-media. *Journal of Computer and Communications* **7**, 1–7 (2019)

Authors and Affiliations

Han Han¹ · Can Wang² · Yunwei Zhao¹ · Min Shu¹ · Wenlei Wang¹ · Yong Min³

Han Han
hanh@cert.org.cn

Can Wang
can.wang@griffith.edu.au

Min Shu
shumin@cert.org.cn

Wenlei Wang
wwl@cert.org.cn

Yong Min
myong@zjut.edu.cn

¹ CNCERT/CC, Beijing, China

² School of ICT, Griffith University, Gold Coast, Australia

³ Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou, China