An Information Dissemination Process Study Based on Data Mining Methods

Tianhao Cao[1,#], Jiaxin Wang[2,#], Shengxi Liu[3], Meizhu Chang[4], Yong Liu[5,*]

[1]Software College, Heilongjiang University, Harbin, Heilongjiang, China

[2] School of Information Management, Heilongjiang University, Harbin, Heilongjiang, China

[3] School of Electronic Engineering, Heilongjiang University, Harbin, Heilongjiang, China

[4] School of Economics and Business Administration, Heilongjiang University, Harbin, Heilongjiang, China

[5] Heilongjiang University and Nantong Xikerui Intelligent Technology Co., Ltd, Harbin, Heilongjiang, China

[*]Corresponding author: Yong Liu

[#]These authors contributed equally.

Keywords: S-type model, Hierarchical analysis method, Correlation analysis model, Linear regression fitting model, Deviation model

Abstract: Under the new information dissemination pattern, how to break the “screaming effect” and “Echo effect” and walk out of the “information cocoon” is a real problem that is urgently needed to be solved. This article mainly the formation mechanism of the three effect mechanisms has been studied. This article explores the law of information dissemination with S-type models, and explores the formation mechanism of the above three effects with mathematical analysis methods such as correlation analysis and linear regression fitting.

1. Introduction

In the era of rapid network development, various information is flooded in our daily lives. While facilitating our lives, some problems need to be solved.[1] “The screaming effect” and “Echocho effect” are two of the important and extensive phenomena, and these two are likely to lead to the formation of “Information Cocoon”. The so-called “information cocoon” refers to the fact that people’s own information needs are not comprehensive in the dissemination of information, but only choose the information they want or can make themselves happy[2]. The more limited, and eventually shackled himself in the “cocoon room” like a silkworm cocoon,Lose the ability to understand and access other different information.

Under the new pattern of information communication, how to break “screaming effect” and “the echo chamber effect”, out of the “cocoon information”, is the current urgent need to solve practical problems, this paper mainly for three effect mechanism for the formation of the mechanism of the research, this article use S quantitative analysis model and correlation analysis describes the propagation process of the subject (or information), Explore the formation mechanism of this, and discuss the influence of topic attraction, user activity, user psychology, interaction between different users, platform recommendation algorithm and other factors on the formation of these phenomena[3].

2. Methods

First of all, we should take two different types of words as data entry point: one point of view tends to be the same (neutral consensus), the other point of view is polarized (polarization). By crawling relevant data on Weibo, Twitter and other social media, we should take the amount of reading, discussion and retweet as the influencing factors of information dissemination[4]. Through quantitative amount per unit time of reading, discussion, forward the quantity and other indicators, quantitative analysis describes the propagation process of the subject (or information), through the analytic hierarchy process (ahp), the quantitative index of information communication, build a

comprehensive evaluation system, to analyze its influencing factors, analysis of the rule of information changes over time, the first stage can be called the incubation period, followed by the exploding. After a period of time, it enters a plateau period, and finally comes to the extinction period[5].

For the “scream effect”, when the topic spreads quickly in a short period of time, the number of users' comments on the topic will increase rapidly, and the spread speed of the topic will also increase rapidly. Compare the screaming effects of different themes by collecting the propagation speed of different themes at different stages. Using the above established related model, that is, the chi-square test, the relationship between the scream effect and various factors is obtained. After reviewing the relevant material, we concluded that for the “echo chamber effect”, the number of similar opinions is pushed repeatedly over time, creating an echo chamber effect. For the “information cocoon room”, the degree of deviation between the user's selected topic and the user's preference is observed over time, and the formation mechanism of the information cocoon room is explored [6,7].

3. Results

3.1 The Law of Information Dissemination and Influencing Factors

We chose Two topics:

Topic No.	Topic Name	Topic type
Topic 1	Is it true that love cannot escape novelty?	Neutral consensus
Topic 2	Does Time really kill love?	View polarization

Here is two data sheets of the influencing factors of the topics below (see Table 1 and Table 2):

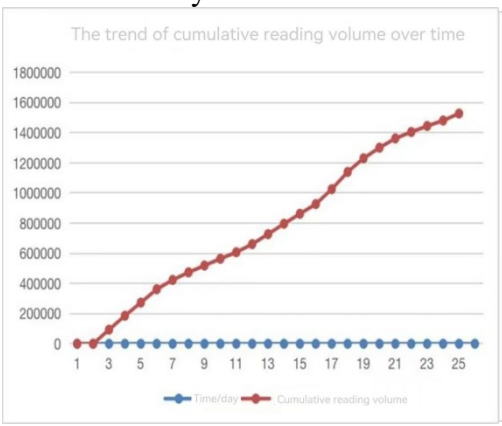
Table 1: -#Love Really Can't Escape Freshness#Subject Data

Time/day	Discuss trends	Discuss changes	Reading trends	Reading change	Forward trend	Transmission variable
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	93510	93510	14	14	10	10
4	92440	-1070	16	2	12	2
5	87219	-5221	10	-6	8	-4
6	88560	1341	10	0	10	2
7	61812	-26748	6	-4	6	-4
8	50262	-11550	4	-2	4	-2
9	45231	-5031	1	-3	2	-2
10	44965	-266	2	1	3	1
11	43381	-1584	2	0	4	1
12	53659	10278	5	3	4	0
13	66252	12593	8	3	6	2
14	69254	3002	12	4	15	9
15	65687	-3567	9	-3	8	-7
16	64631	-1056	28	19	18	10
17	98651	34020	9	-19	10	-8
18	114017	15366	34	25	30	20
19	90896	-23121	15	-19	14	-16
20	70321	-20575	15	0	12	-2
21	61572	-8749	6	-9	5	-7
22	42461	-19111	6	0	2	-3
23	39829	-2632	9	3	8	6
24	36437	-3392	2	-7	4	-4
25	46312	9875	2	0	4	0
26	42869	-3443	5	3	2	-2

Table 2: -#Time Will Really Kill Love? #Subject Data

Time/day	Discuss trends	Discuss changes	Reading trends	Reading change	Forward trend	Transmission variable
1	188131	188131	52	52	40	40
2	166053	-22078	26	-26	20	-20
3	124356	-41697	8	-18	4	-16
4	118596	-5760	7	-1	5	1
5	91589	-27007	2	-5	2	-3
6	112381	20792	2	0	2	0
7	101402	-10979	3	1	3	1
8	99910	-1492	8	5	8	5
9	125351	25441	6	-2	6	-2
10	146706	21355	12	6	10	4
11	156739	10033	25	13	24	14
12	149237	-7502	52	27	35	11
13	168485	19248	24	-28	20	-15
14	137811	-30674	32	8	30	10
15	91510	-46301	24	-8	24	-6
16	92524	1014	20	-4	15	-9
17	113561	21037	15	-5	12	-3
18	104417	-9144	13	-2	10	-2
19	89080	-15337	13	0	10	0
20	82646	-6434	20	7	18	8
21	93284	10638	20	0	18	0
22	106522	13238	11	-9	10	-8
23	72693	-33829	11	0	10	0
24	71802	-891	6	-5	5	-5

In order to verify whether it conforms to the model and its propagation law, the cumulative reading volume, cumulative discussion volume and cumulative forwarding volume of each topic are analyzed in the unit of days:



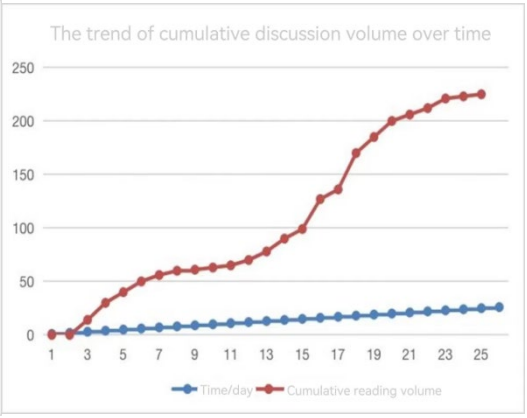


Fig.1 (Left). Topic 1 the Relationship between Cumulative Reading Volume and Time

Fig.2 (Right) Topic 1 the Relationship between Cumulative Discussion Volume and Time

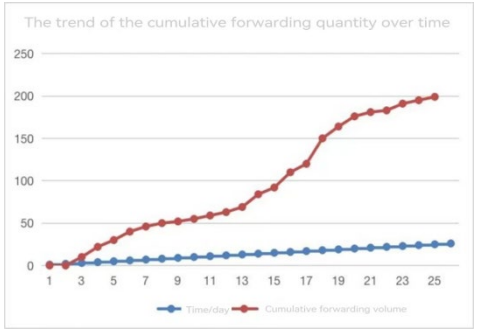


Fig.3 Topic 1 Relationship between Cumulative Forwarding Quantity and Time

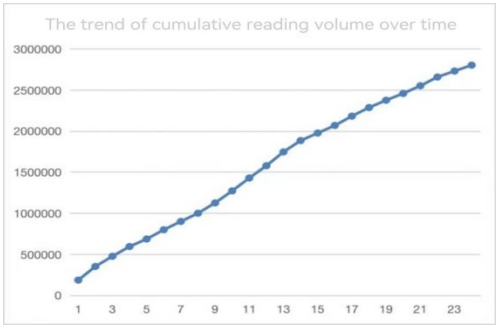


Fig.4 (Left). Topic 2 the Relationship between Cumulative Reading Volume and Time (Horizontal axis is time, vertical axis is cumulative reading)

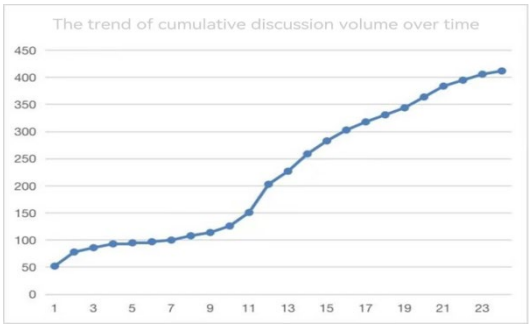


Fig.5 (Right). Topic 2 the Relationship between Cumulative Discussion Volume and Time (Horizontal axis is time, vertical axis is cumulative discussion volume)

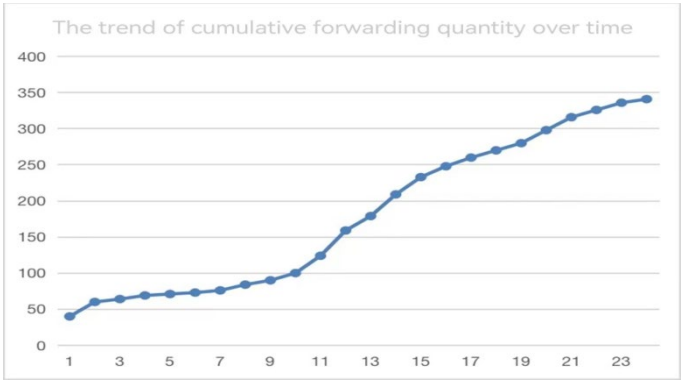


Fig.6 Topic 2 the Relationship between the Cumulative Forwarding Quantity and Time

(Horizontal axis is time, vertical axis is cumulative forwarding volume)

As shown in the figures above (see Fig. 1-6), the propagation stages of the above two topics have been directly reflected, and the basic compound S-shaped propagation law.

3.2 The Influencing Factors in the Process of Information Dissemination

(1) Number of user followers

The number of user fans plays an influential role in the process of information dissemination. It has a certain influence on the number of fans of the topic manager. The number of fans is used to measure the important role of the level of topic management on the topic impact index. The higher the number of fans of the topic manager, the greater the possibility of the continuous impact of the topic[8].

(2)Comment on the amount

The number of comments on a topic plays an important role in the spread of the topic. The number of comments on a hot topic can also affect the spread of the topic to different degrees. A large number of comments will make more users participate in the topic and spread the topic more widely. Fewer comments mean fewer reads, and fewer users engage with the topic.

(3) The reading quantity

The amount of reading about a topic will affect the spread of the topic. The more reading, the wider the information spread.

(4) Forwarding quantity

The number of retweets plays a certain role in the spread of the topic. Different retweets of hot topics can affect the spread of the topic to different degrees. If the number of retweets is large, the reading quantity will be higher and the spread of the topic will be more extensive. The less retweets, the less reads, and the less users participate in the topic.

By using the data crawler algorithm, the data of influencing factors were statistically analyzed for Topic 1# Does love really escape novelty # and topic 2# Does time really kill Love #.

The comprehensive evaluation model is built as shown in the figure below:

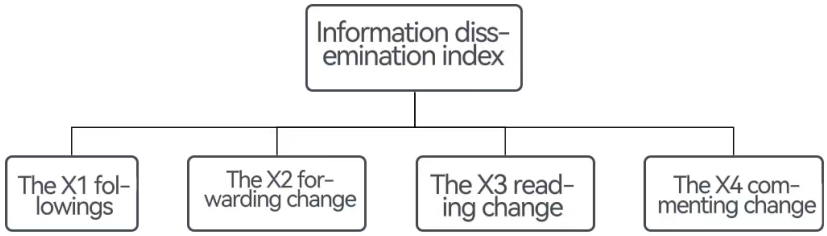


Fig.7 The Comprehensive Evaluation Model

From Fig. 7 we can see that in the first part, the comprehensive evaluation model and fuzzy comprehensive evaluation model based on AHP are established, and four indexes are established in this model.

Fuzzy comprehensive evaluation model[9]

Topic 1: Performing matrix synthesis Operations:

$$B = A \bullet R = \begin{bmatrix} 0.54 & 0.28 & 0.10 & 0.08 \end{bmatrix} \bullet \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.1 & 0.7 & 0.2 \\ 0.7 & 0.3 & 0 \\ 0.4 & 0.6 & 0 \end{bmatrix}$$

The final result:B=[0.238,0.49,0.272]

It can be obtained that the general influence of topic 1 accounts for 23.8%, the comparative influence accounts for 49%, and the extraordinary influence accounts for 27.2%. The comment with the largest value is taken as the comprehensive evaluation result, so the evaluation result is “comparative influence” (see Fig.8).

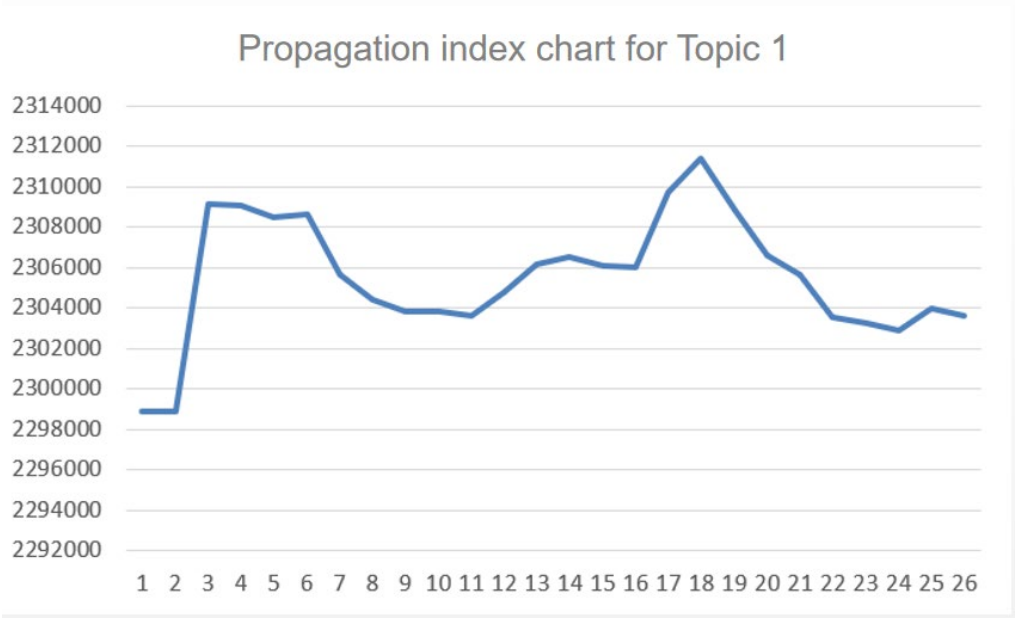


Fig.8 Propagation Index Chart for Topic 1

Topic 2: Performing matrix synthesis Operations:

$$B = A \bullet R = \begin{bmatrix} 0.29 & 0.43 & 0.13 & 0.15 \end{bmatrix} \bullet \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.2 & 0.8 & 0 \\ 0.3 & 0.7 & 0 \end{bmatrix}$$

The final result:B=[0.345,0.425,0.23]

From Fig. 9, it can be obtained that the general influence of topic 2 accounts for 34.5%, the comparative influence accounts for 42.5%, and the extreme influence accounts for 23%. The comment with the largest value is taken as the comprehensive evaluation result, so the evaluation result is “comparative influence”.

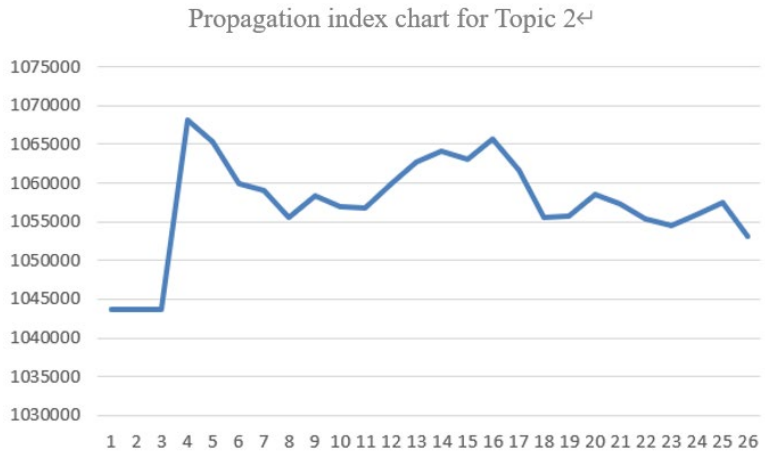


Fig.9 Propagation Index Chart for Topic 2

In summary, the communication process of the two topics is basically in line with the S-shaped model, which is divided into incubation period, outbreak period, stable period, and extinction period, and has the laws of periodic transmission and trend transmission. In the process of communication, for the topic of topic 1, which tends to have the same point of view, the largest influencing factor is the number of fans of a certain opinion user, followed by the number of topic forwards, the weakest is the amount of topic comments, and for topics where the second point of view tends to be polarized, the largest influencing factor is the amount of topic retweets, followed by a certain point of view, the number of fans of users, and the weakest is the amount of topic comments.

3.3 Modeling and Analysis of the Generation Mechanism of “Scream Effect”, “Echo Chamber Effect” and “Information Cocoon Room”

3.3.1 The Mechanism and Influencing Factors of “Scream Effect” Were Modeled and Analyzed

The establishment and solution of correlation analysis model

(1) Data preparation

For the screaming effect, the data of three topics were collected respectively through the above analysis and are listed in the appendix. The average increment of the JTH topic (average propagation speed of the topic), $\overline{y_j}$, and the increment of the average growth rate (the trend of the average propagation speed of the topic), $\overline{y_j'}$, can be obtained by using the following equation. Among them:

$$\overline{y_j} = \frac{x_t - x_0}{\Delta t}$$

$$\overline{y_j'} = \frac{y_t - y_0}{\Delta t}$$

Here are two sheets of the data for the two topics (See Table 3 and Table 4), which are topic propagation speed and propagation speed trend data tables

Table 3: Topic 1 Data Sheet

$\overline{y_j}$	$\overline{y_j'}$	User activity	Impact between different users	Platform to recommend
0.00	0.00	0	0	0
0.00	0.00	0	0	0
336.00	8064.00	87219	2	196
48.00	-6912.00	88560	5	186
-144.00	-4608.00	61812	9	173
0.00	-150.26	50262	15	172
-96.00	-2304.00	45231	48	164
-48.00	1152.00	44965	70	134
-72.00	-576.00	43381	76	129
24.00	2304.00	53659	81	111
0.00	-576.00	66252	67	106
72.00	1728.00	69254	65	103
72.00	0.00	65687	64	100
96.00	576.00	64631	60	97
-72.00	-4032.00	98651	103	93
456.00	12672.00	114017	423	74
-456.00	-21888.00	90896	349	75
600.00	25344.00	70321	452	76
-456.00	-25344.00	61572	200	92
0.00	10944.00	42461	106	93
-216.00	-5184.00	39829	98	120
0.00	5184.00	36437	68	157
72.00	1728.00	46312	67	182
-168.00	-5760.00	42869	20	200

Table 4: Topic 2 Data Sheet

$\overline{y_j}$	$\overline{y_j'}$	User activity	Impact between different users	Platform to recommend
0.00	0.00	0	0	0
0.00	0.00	0	0	0
0.00	0.00	0	0	0
56.73	61.88	188131	43	200
27.13	30.88	166053	58	193
-432.00	-11019.13	124356	67	192
-24.00	9792.00	118596	60	196
-120.00	-2304.00	91589	29	184
0.00	2880.00	112381	34	150
24.00	576.00	101402	64	129
120.00	2304.00	99910	104	101
-48.00	-4032.00	125351	302	94
144.00	4608.00	146706	345	85
312.00	4032.00	156739	689	63
648.00	8064.00	149237	1273	57
-672.00	-31680.00	168485	945	47
192.00	20736.00	137811	724	89
-192.00	-9216.00	91510	548	92
-96.00	2304.00	92524	572	104
-120.00	-576.00	113561	357	152
-48.00	1728.00	104417	359	163
0.00	1152.00	89080	201	187
168.00	4032.00	82646	187	191
0.00	-4032.00	93284	37	200

(2) Calculate the correlation coefficient

The Person coefficient method is commonly used to measure the correlation between fixed distance variables. Set the average increment (average topic propagation speed), $\overline{y_j}$,and the increment (average topic propagation speed trend) , $\overline{y_j'}$, of average growth speed as y_i , and set the user activity, influence among different users and platform recommendation as m, then the correlation coefficient is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}}$$

Where: the closer the value of correlation coefficient r is to 1 (-1 or +1), the stronger the correlation coefficient is; The closer it is to 0, the weaker the correlation coefficient.

Significance test of correlation coefficient:

Table 5: The Average Spread Speed of the Topic is Different from the User Activity, the Influence among Users and the Platform Recommendation Test Results

	Average topic spread speed * User activity		Average propagation speed of topics * Influence among different users		Average propagation speed of topics * platform recommendation	
	P	Phi	P	Phi	P	Phi
Topic 1	0.028	3.926	0.028	3.926	0.028	3.926
Topic 2	0.026	4.153	0.036	4	0.028	3.926

Table 6: The Average Spread Speed Trend of the Topic is Different from the User Activity, and the Influence among Users and the Platform Recommendation Test Results

	Average topic spread speed * User activity		Average propagation speed of topics * Influence among different users		Average propagation speed of topics * platform recommendation	
	P	Phi	P	Phi	P	Phi
Topic 1	0.036	4	0.029	4.243	0.026	4.153
Topic 2	0.036	4	0.026	4.153	0.029	4.243

By analyzing the data shown in Table 5 and Table 6, the detection value of Pearson's chi-square in chi-square test P< 0.05, indicating significant correlation. Phi &gt in symmetric measurement; 0.5 indicates a high correlation.

To sum up:

By analyzing Phi, it was found that the screaming effect in topic 1 was most affected by different users.

In topic 2, the screaming effect was most affected by platform recommendation.

3.3.2 The Mechanism of "Echo Chamber Effect" is Modeled and Analyzed

Based on the above data, the information propagation law data can be obtained. Linear regression is used to linear fit the average propagation rate of two kinds of comments on two topics, and the functional relationship is obtained as follows. The results are shown in Fig. 10 and Fig. 11:

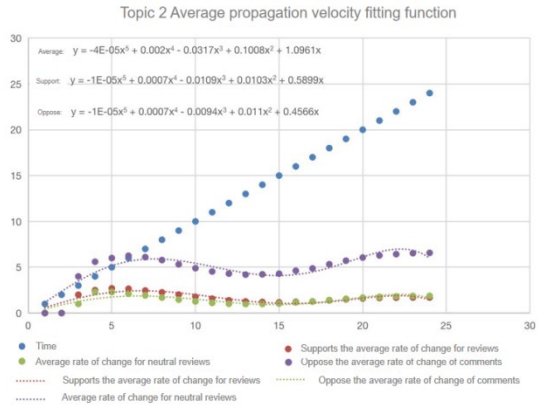


Fig.10. Topic 1 Average Propagation Velocity Fitting Function

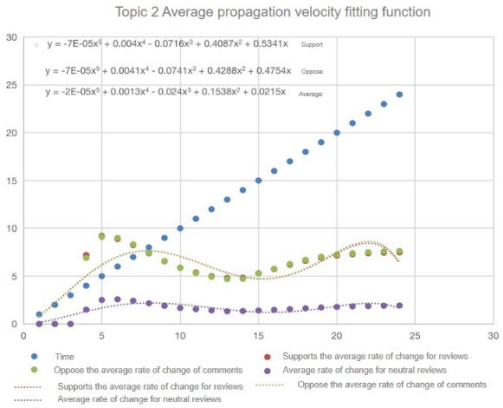


Fig.11. Topic 2 Average Propagation Velocity Fitting Function

Then, at time t, the amount of information transmission h(t):

$$H(t) = \int_0^t y(t)dt$$

Due to the limitation of the platform, the value of the user is a fixed value Y, then the number of users who do not receive information,$h^{'}$ (t) ,at time t is:

$$h^{'}(t) = Y - \int_0^t z(t)dt$$

When t value approaches infinity and $h^{'}$ (t) approaches 0, the platform has limitations in recommending topics.

It can be seen that with the continuous increase of time t, the information is constantly transmitted back and forth in the platform, and more and more users receive the information. When t value approaches infinity, all users receive the information, thus forming an "echo chamber effect". However, the Y value changes with the topic attraction, topic popularity, the base of the preference group and other factors. For example, the higher the topic attraction and topic popularity, the larger the preference group base, the more users, and the larger the Y value.

Secondly, according to the HK model[10]:Individuals update their own views only according to the views of their neighbors within the trust threshold, and the rules are as follows:

$$x_i(t+1) = \begin{cases} \dfrac{1}{|N_i(t)|}\sum_{j\in N_i(t)} x_j(t), & N_i(t) \neq \varnothing \\ x_i(t), & N_i(t) = \varnothing \end{cases}$$

$$N_i(t) = \left\{ j \in V \left| x_i(t) - x_j(t) \right| \leq \varepsilon_i \right\}$$

Where, $N_i(t)$ is the neighbor set within the trust threshold ε_i of the difference between an individual's opinion and the individual's opinion; $|N_i(t)|$ is the number of neighbors whose opinion interacts; $\dfrac{1}{|N_i(t)|}\sum_{j\in N_i(t)} x_j(t)$ is the opinion after the interaction between an individual and the individual in the neighbor set $N_i(t)$. If $N_i(t)$ is empty, then the individual opinion remains $x_i(t)$. Thus, we can learn that the "echo chamber effect" occurs when similar opinions are repeated, and the trust value of similar opinions is in the threshold. Users interact with neighbors' opinions, are "surrounded" by users with the same opinions, and thus "blinded".

3.3.3 The Mechanism of "Information Cocoon Room" is Modeled and Solved

Suppose there are n topics, let the topic type, topic, word count, title... The list of alternative

factors is $x_{u,1}, x_{u,2}, x_{u,3}, x_{u,4}, \cdots, x_{u,n}$, For individuals, a series of factors $x_1, x_2, x_3, x_4, \cdots, x_n$ are selected according to the preference degree of each factor. Then the deviation degree of information preference is μ_1 :

$$\mu_1 = \sqrt{(x_{u,1}-x_1)^2+(x_{u,2}-x_2)^2+\cdots+(x_{u,n}-x_n)^2}$$

Within a fixed period of time, individuals have the same probability of being exposed to each topic, and each topic grows at the same rate.

Let the daily growth rate of n topics of the same type be q, then the total number of topics after t days is H(t):

$$H(t) = n \cdot (1+q)^{t-1}$$

The number of page views per day is Q, and whether a user pays attention to H(t) topic is m_i: 0 means no attention, and 1 means attention. Then the deviation degree of daily information preference is

$$\mu_2 = \sum_{u=1}^{u=H(t)} q_u \sqrt{\sum_{v=1}^{v=m} (n_{v,u} - n_v)^2}$$

v=m means there are m factors, $n_{v,u}$ means the uth topic, the v factor. And $Q = \sum q_u$.

To sum up,

$$\min \mu = \sum_{u=1}^{u=H(t)} q_u \sqrt{\sum_{v=1}^{v=m} (n_{v,u} - n_v)^2}$$

The constraint conditions are:

$Q = \sum q_u$ $q_u (u = 1,2,3\cdots)$, $H(t) = n \cdot (1+q)^{t-1}$.

Therefore, when a user likes a certain type of topic, the deviation of this topic to the user's preference is the smallest, and the deviation of related topics to the user's preference is also small. As time goes by, the types of topics that users watch every day gradually tend to be the same.

At the same time, we set up the user suitability model, assuming that some information of users is as follows: User information preferences, user information needs, user values, and user social capital are $f_1, f_2, f_3, f_4, \cdots f_k,$ if a certain "cocoon room user" is assumed to be $\overline{f_1}, \overline{f_2}, \overline{f_3}, \overline{f_4}, \cdots \overline{f_k},$ then the user's cognition is $F = \sqrt{\sum_{i=1}^{i=k} (f_i - \overline{f_i})^2}$. Therefore, a certain type of user will form the same cognition with a certain type of "cocoon room user" because the deviation is not large, and thus produce the same choice behavior. This awareness, in turn, keeps users at that level of awareness, consistently generating choices and creating an "information cocoon."

In addition, due to the existence of "echo chamber effect", users' views are easily surrounded by similar views through the analysis of HK model, thus accelerating the formation of "cocoon room".

4. Conclusion

(1) Through the analysis of the data information of the three topics, such as reading volume and discussion volume, it is found that the information transmission law conforms to the distribution law of S-type model, and the influence of reading volume, discussion volume and forwarding volume on it is explored through the hierarchical analysis method.

(2) By means of correlation analysis and analysis of data such as transmission speed, this paper explores the generation mechanism of scream effect, echo chamber effect and information cocoon room, as well as the influence of discussion topic attraction, user activity, user psychology, interaction between different users, platform recommendation algorithm and other factors on these phenomena.

Acknowledgment

Fund project support: the central government guides the local science and technology development special project “Smart Irrigation Equipment Achievement Industrialization Project” (ZY21CBL001); Nantong Basic Scientific Research and Social People's Livelihood Science and Technology Plan Project “Key Technologies of Smart Water and Fertilizer Integration”; National Key Area Support Project of College Students Innovation and Entrepreneurship Training Program(202210212193S)

References

[1] WANG Chaoyang. Strategies for Improving the Dissemination of Current Political News in TV in the Era of Media Convergence[J]. Communication & Copyright, 2022, (03):1-3.

[2] E-Zhi.” Screaming Effect” and “Information Cocoon Room”[J]. Reporter Observation, 2018, (05):1.

[3] LIU Rong. Research on Analysis and Prediction Method of Information Propagation Characteristics in Social Network[D]. University of Electronic Science and Technology of China, 2022, (06):16-20.

[4] ZHANG Xiaotian. Research on influencing factors and communication prediction of government microblog communication[D].Shandong University of Finance and Economics, 2021. DOI:10.27274/d.cnki.gsdjc.2021.000074.

[5] Hu Yi. Application of fuzzy analytic hierarchy method in determining index weighting: A case study of the development model of emerging architectural ceramics industry cluster[J]. Light Industry Science and Technology ,2022, 38(05):38-41.

[6] Wang Di. Communication and Copyright, 2018, (10):13-14+17.DOI:10.16852/j.cnki.45-1390/g2.2018.10.005.

[7] YANG Yujiao,YUAN Qinjian. The Hidden Worries of Personalized Recommendation: An Analysis of Information Cocoon Room and Its Causes and Consequences Based on Grounded Theory. [J/OL]Intelligence Theory & Practice, 2022, (09):1-13.

[8] GUO Qiuju. Research on the dissemination of high-reading texts on WeChat subscription account[D]. Nanjing University of Science and Technology, 2019, (06):46-49.

[9] CHEN Chen. Application analysis of fuzzy comprehensive evaluation method based on AHP in performance evaluation of urban construction archives website[D]. Heilongjiang University, 2017, (07): 24-27.

[10] Xu Yuanyuan. Study on Echo chamber Effect in the Evolution of public Opinion [D]. Huazhong University of Science and Technology, 2020, (05):12-19.