# Exploration and Analysis of the Influencing Factors on Sina Weibo Hot Topics

Siqi Zhang*, Xi Shi

School of Economics and Management, Beijing University of Technology, Beijing, China

*Corresponding author Email zsq200010@163.com,
shixi20000925@126.com

**Abstract.** Weibo is a social network platform and its hot topic, as a function that is high participated among users, is of high research significance. In this paper, representative factors were sorted out and summarized as the main factors that are able to turn Weibo topics into the trending ones under factor analysis from the angle of empirical test, and were evaluated through regression analysis. At the end, this paper gave out some suggestions for the people highly-related with Weibo to create a better environment that all side can share benefits.

**Keywords:** Weibo topics, factor analysis, regression analysis

## 1    Introduction

As one of the biggest social media in China, Weibo is a platform where short real-time information is shared without the limit of time and space. Keywords will be turned into Weibo topic words with a # before and after them, allowing not only users to browse relevant microblogs and get involved into the discussions, but also Weibo itself to quickly respond to an issue outbreak. Therefore, the way to efficiently and precisely unearth hot topics from a mass of short texts has become an essential research hot spot.

In this paper, the topics that have been on the Sina Weibo hot topic list from Nov. 4th to 6th, 2021 were selected as the research object, with a total of 150 samples.

## 2    The Establish of Evaluation Indexes

### 2.1    Index Establishing

This research is consisted of two levels of indexes based on Weibo topic index description, including two primary indexes and thirteen secondary indexes. "Figures" mentioned in the following are all represent "figures who were referred to in the topic", and they can either be celebrities or ordinary people.

**Table 1.** The Indexes Based on Weibo Topic Index Description

| Primary Index | Secondary Index |
|---|---|
| (P) Index of participation | P1 repost, comment and like from topic hosts |
| | P2 popularity of topic hosts |
| | P3 activity level of topic hosts |
| | P4 influence of topic opinion leader |
| | P5 repost, comment and like from figures |
| | P6 popularity of figures |
| | P7 activity level of figures |
| (T) Index of topics | T1 cumulative activity level |
| | T2 Netizens' attention |
| | T3 sensitivity of topic |
| | T4 information volume of microblog text |
| | T5 content of topic video |
| | T6 originality of topic |

Table credit: Original

## 2.2 Secondary Indexes Quantification Methods

In this section, all secondary indexes were transferred into available statistic data from Sina Weibo (see Table 2).

**Table 2.** Secondary Indexes Quantification Methods

| Quantification Methods |
|---|
| P1= number of repost, comment and like of topic host |
| P2= number of followers of the topic host |
| P3= number of microblogs of the topic host |
| P4= number of the media that release topic-related news |
| P5=$\sum_{m=1}^{n} p_m$; pm = the mth topic's number of repost, comment and like of figures, n is the number of them |
| P6=$\sum_{i=1}^{n} p_i$; pi = the ith figure's number of followers, n is the number of figures |
| P7=$\sum_{j=1}^{n} p_j$; pj = the jth topic's number of microblogs of users, n is the number of figures |
| T1= number of discussions of that topic within 72 hours on Sina Weibo |
| T2= times of view of that topic within 72 hours on Sina Weibo |
| T3= The length of time between the first relevant Weibo post of the topic and the peak number of discussions on the topic within 72 hours |
| T4=The text word count of microblog on hot topic list |
| T5=the length of the video posted by the host (in seconds) |
| T6= the number of users who originate microblogs using the topic within 72 hours |

Table credit: Original

# 3    Empirical Test Analysis

## 3.1    Factor Calculation

### 3.2.1 Applicability Test.

Before conducting factor analysis, we used R to perform KMO test and Barlett sphericity test on the collected data (see Table 3). The KMO value is greater than 0.5 and the Bartlett significance level is less than 0.01, which indicate that the selected data are suitable for factor analysis.

**Table 3.** KMO Test and Bartlett Sphericity Test

| KMO sampling suitability | Bartlett sphericity test | | |
|---|---|---|---|
| | Approximate chi-square | df | Significance |
| 0.656 | 671.616 | 78 | .000 |

Table credit: Original

### 3.2.2 Common Factors Extraction.

**Table 4.** Initial Eigenvalue

| Component | Total | % of Variance | Cumulative % |
|---|---|---|---|
| 1 | 3.3324161 | 25.63397 | 25.63397 |
| 2 | 2.0849411 | 16.03801 | 41.67198 |
| 3 | 1.2780098 | 9.83084 | 51.50282 |
| 4 | 1.2190159 | 9.37705 | 60.87987 |
| 5 | 1.0750054 | 8.26927 | 69.14914 |
| 6 | 0.8647435 | 6.65187 | 75.80101 |
| 7 | 0.7854627 | 6.04202 | 81.84303 |
| 8 | 0.7353282 | 5.65637 | 87.4994 |
| 9 | 0.5685108 | 4.37316 | 91.87256 |
| 10 | 0.439395 | 3.37996 | 95.25252 |
| 11 | 0.3388679 | 2.60667 | 97.85919 |
| 12 | 0.1564946 | 1.20382 | 99.06301 |
| 13 | 0.121809 | 0.93699 | 100.0000 |

Table credit: Original

We use R to perform factor analysis on these indexes (see Table 4). Take 80% as the lowest standard, then the first 7 eigenvalues were chosen for the accumulative variance rate is already 81.84%. A factor loading matrix was established based on the principal component estimation method, and the maximum variance orthogonal rotation was performed on the factor loading matrix (see Table 5).

**Table 5.** Composition Matrix After Rotation

| Factor / Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $X_1$ | **0.82** | 0.34 | 0.09 | 0.01 | 0.11 | -0.06 | 0.01 |
| $X_2$ | **0.93** | 0.12 | -0.11 | -0.04 | -0.06 | -0.00 | -0.01 |
| $X_3$ | **0.67** | 0.12 | -0.30 | -0.22 | -0.34 | 0.16 | -0.02 |
| $X_4$ | 0.02 | -0.10 | 0.22 | 0.01 | 0.16 | **0.87** | -0.04 |
| $X_5$ | 0.01 | -0.24 | 0.19 | **0.70** | 0.12 | -0.31 | 0.02 |
| $X_6$ | -0.14 | 0.02 | -0.14 | **0.84** | -0.11 | 0.21 | -0.04 |
| $X_7$ | 0.34 | **0.85** | -0.05 | -0.11 | -0.02 | -0.02 | 0.08 |
| $X_8$ | 0.21 | **0.90** | 0.02 | -0.11 | -0.02 | 0.09 | 0.05 |
| $X_9$ | 0.01 | 0.07 | **0.73** | -0.08 | -0.07 | 0.17 | 0.03 |
| $X_{10}$ | -0.14 | -0.17 | **0.81** | 0.03 | 0.01 | 0.00 | -0.10 |
| $X_{11}$ | -0.01 | 0.08 | 0.06 | -0.02 | -0.01 | -0.03 | **0.98** |
| $X_{12}$ | -0.07 | 0.00 | -0.04 | -0.03 | **0.94** | 0.15 | -0.01 |
| $X_{13}$ | -0.07 | 0.53 | **0.65** | 0.15 | 0.10 | 0.07 | -0.05 |

Table credit: Original

### 3.2.3 Factor Naming.

Only a few indexes have large factor loadings per factor, then the 13 indexes can be divided into 7 categories as main interpretive indexes.

**Table 6.** The Classification Based on High Factor Loadings

| | Indexes with High Factor Loading | Naming |
|---|---|---|
| 1 | $X_1$: repost, comment and like from topic hosts<br>$X_2$: popularity of topic hosts<br>$X_3$: activity level of topic hosts | Leadership |
| 2 | $X_7$: influence of topic opinion leader<br>$X_8$: netizens' attention | Dissemina-tion |
| 3 | $X_9$: cumulative activity level<br>$X_{10}$: sensitivity of topic<br>$X_{13}$: originality of topic | Engagement |
| 4 | $X_5$: popularity of figures<br>$X_6$: activity level of figures | Attractiveness |
| 5 | $X_{12}$: content of topic video | Visualization |
| 6 | $X_4$: total repost, comment and like received by figures | Stickiness |
| 7 | $X_{11}$: information volume of microblog text | Guidance |

Table credit: Original

### 3.2.4 Factor Score Calculation.

Factor score matrix and the rank of top 15 topics were gained using Excel and R (see Table 7). They can be categorized into four main groups, which are celebrity topics, national events topics, social positive energy topics and current trend topics, and they can provide instructive information for the way Weibo-related people can follow.

**Table 7.** Factor Scores of the Top 15 Topics

| Rank | Topic name | Total Scores |
|------|------------|--------------|
| 1 | How do you view the sensation of metaverse | 1.47 |
| 2 | What dose Jay Chou means to you | 1.43 |
| 3 | A fire fighter fell to his death for protecting a girl | 1.21 |
| 4 | A community which had epidemic recently lifted the lockdown | 1.10 |
| 5 | Yang Yang sued Dr. of Pox for infringement | 1.05 |
| 6 | Four main reasons why young people suffer from hair loss | 0.95 |
| 7 | How do you think of Allen Guo being expelled | 0.90 |
| 8 | The acting ability of Wen Zhengrong worth an S card | 0.79 |
| 9 | First snow in Beijing | 0.78 |
| 10 | What are the signs of a couple before they break up | 0.77 |
| 11 | It has been 10,000 days since He Jiong made his debut | 0.65 |
| 12 | Yan Xujia broke the record of the single lap Kart racing | 0.56 |
| 13 | Song Yuqi can be both cool and cute | 0.55 |
| 14 | Pay tribute to the heroic fire fighters | 0.53 |
| 15 | The eggs sent by dad have 5 protective layers each | 0.49 |

Table credit: Original

### 3.2    Regression Analysis

### 3.3.1 Variable Selection and Processing.

We chose the seven variables that have the highest correlation coefficient with the common factors as the independent variables, which are "popularity of topic hosts" ($X4$), "activity level of figures" ($X5$), "repost, comment and like received by figures" ($X3$), "netizens' attention"($X1$), "sensitivity of topic"($X2$), "content of topic video" ($X6$) and "information volume of microblog text" ($X7$). The term u is the residual, and "originality of topic" (Y) is the dependent variable. The expression is as follow:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + u \qquad (1)$$

### 3.3.2 Parameter Estimation.

We expurgated X4, X5, X6 and X7 from the initial models for they are not significant at the level of 0.1, and the final results show that netizens' attention, sensitivity of topic and repost, comment and like from figures have associatively significant influence to the originality of hot topic.

**Table 8.** The Results of Multiple Linear Regression

| Variables | Regression Coefficients | P-value | Significance |
|---|---|---|---|
| Intercept Term | -1655.38 | 0.001 | *** |
| X1 | 984.91 | 0.000 | *** |
| X2 | 307.83 | 0.000 | *** |
| X3 | 1.65 | 0.030 | ** |
| F-Test | P-value=0.00<0.05 | Adjusted $R^2$ | 0.4014 |

Note: ***, ** and * represent significance at the level of 1%, 5%, and 10% respectively.
Table credit: Original

## 4      Conclusions

We extracted 7 factors from 13 influencing factors of Weibo hot topics, which more clearly present the effect of expediting the topics to become popular. Therefore, we are able to give out some recommendations for Weibo-related users. For amateurs, it's difficult to become famous on the Internet as easily as in their fantasy, when lacking of sufficient support and enough desire from users to participate in discussion. For Weibo platform manager, it's reasonable to keep an eye on the topics in the four categories and control them in time because of their higher possibility of fermentation. For Weibo bloggers, they will reap benefit if adding high-quality pictures and videos when presenting information in their topic page.

## References

1. Wang Jing and Sheng Junquan. Research on financial performance evaluation of agricultural listed companies based on DEA method [J], Green Finance and Accounting, 2019(06): 15-22.
2. Zhang seek. Based on DEA listed agricultural company's financial performance evaluation method [D], Wuhan: Wuhan Polytechnic University, 2015.
3. Zhu Jiaming. either Xiao quill, Based AHP-analysis of smart metering growth of entropy method [J], Journal of Anhui University, 2017,41(6): 61-67.
4. Rosnow R L, Eric K F. Rumor and Gossip Research[J]. Psychological Science Agenda，2005(19): 4:1-4.
5. Grant M，Suchitra M. Rumour and Gossip in Organisations: A Conceptual Study[J]. Management Decision, 2000, 38(5): 339-346.
6. Meng Xiaokai. Research on the performance evaluation of listed agricultural companies based on DEA method [D], Beijing: Capital University of Economics and Business, 2018.
7. Limayem M, Hirt S G, Cheung C M K. How habit limits the predictive power of intention: The case of information systems continuance[J].MIS Quarterly,2007,31(4):705-737.