

Events Influence Computation on Social Media

Zhang Yu^{1, a}, Yu Min^{1, b}, He Yueying^{2, c}, Zhao Zhonghua^{2, d}, Qiao Yang^{1, e} Zhang Hua-Ping^{1, f}, Shang Jianyun^{3, g}

¹ Beijing Institute of Technology School of Computer Science, Beijing, 100081, China

² CNNCERT/CC, Beijing, 100031, China

³ Beijing Institute of Technology School of Software, Beijing, 100081, China

^azhangyu2014@nlpir.org, ^byumin2014@nlpir.org, ^chy@cert.org.cn, ^dzhaozh@cert.org.cn, ^eqiaoyang2014@nlpir.org, ^fkevinzhang@bit.edu.cn, ^gshangjia@bit.edu.cn

Keywords: Sina Weibo, Specific Events, Influence, Ranking List.

Abstract.

Up to now, as the most popular and widely used domestic social media mobile application with the most registered users, Sina Weibo has become the main platform of growing numbers of hot events occurring and spreading. This paper aims at the demand for analyzing specific events on Sina Weibo for public opinions, proposing a model for computing influence of Weibo users, Weibo messages, and whole Weibo events, utilizing content attributes, dissemination attributes and sentiment attributes of Weibo comprehensively. Based on dissemination features of public opinions and research experience of Weibo events, scientific and comprehensive features for building the model are selected. Being the main outcome of the model, ranking lists of influential users, Weibo messages, and hot events' influence provide technical support for public opinion surveillance on Weibo.

1 Introduction

During the first season of 2016, the monthly active users of Sina Weibo reached 261 million, with year-on-year growth of 32 percent, and the daily active users reached 120 million, with year-on-year growth of 35 percent. Being the largest domestic social media platform, Sina Weibo has become the main front in revealing and discussing hot social events, more agile and dynamic than traditional media. It's considered a more commonsensible way of data mining to process Weibo data in units of specific events. Within the range of specific events, in order to calculate the influence of the events, analyzing the value and influence of Weibo messages and posters (users) possesses more optimal soundness and stronger application value^[1].

At present, most of the domestic and abroad research on Weibo or Twitter focus on forming topological relation network of social network, using the following relations on social media, in order to calculate the general and constant influence of users or messages. However, not only does the complete data of following relations of the whole network offer limited help for analyzing specific events, but also cause rapid increase in computational complexity^[2]. The execution falls in jeopardy when only simple retweeting relations can be obtained, which is often the case. While most researches are based on multi-layer relation network or statistical figures, it's rarely seen that analysis of text content taken into account. Methods like PageRank are used to form topological relation network through following relations to measure the long-term influence of users in online communities^[3] ^[4]; the relevance between specific messages of specific groups in emergency events and follow-up evolution is studied^[5]; retweeting models are proposed to predict the numbers of retweeting of some tweet, leading to its future impact^[6]^[7]. Dataset of credible rumors are crawled to train models in order to identify rumors on social media^[8]; research on the transmission model of information on social network based on nodes and information features^[9]^[10]; detection of hot events on Weibo through distribution of emotion^[11]; key nodes searching of social network is achieved through following relations, namely high-influence users finding^[12].

The systems for social media influence analysis which are put into actual application are as follows: the public opinion monitoring room of People.com.cn^[13] and IndexMedia^[14]. The first system computes and ranks influence and service comments of government service Wechat official accounts for influence and public opinions of hot online social events in long periods of time. It collects and clusters data of many websites, computing the influence index of events through building a multi-dimensional model with websites weights. The Micro-blog Communication Index of IndexMedia uses activity and transmission of micro-blogs to describe transmission capability of its user accounts. The evaluation of original micro-blogs' transmission is stressed to encourage high-quality original work. These two systems mainly utilize statistical transmission attributes, such as number of original weibo, number of comments, number of retweets, number of clicks, number of users.

The model proposed in the paper integrates content analysis of weibo text and principles and characteristics of events transmission, and adopts a more comprehensive attributes system with specific events being as the basic unit, for the purpose of computing the event influence index. It offers effective contributions to public opinion analysis on social media, which presents all-round information of events accurately, and can be applied for enterprises' and Internet media's references. To sum up, the model possesses practical value and realistic significance.

2 Eight Elements of Public Opinion and Framework of Influence Model

The essence of studying specific events on Weibo can be summarized as public opinion to certain extent, the formation of which is described as follows: social change leads to social issues concerning people, then people come up with their own comments and opinions, later opinion leaders stand out in the process of people expressing, and similar opinions start fusing, transforming individual opinions into the opinion or mentality of the majority, which forms public opinion at the end.^[15]

The essential theories of communication science includes the formation and development of information, the generation and interpretation of meaning, the structure and organization of information text, social interaction in transmission, and social dynamics of transmission. As the foundation of this subject, the essential theories originated in studying the universal rules of all forms of transmission and communication in human society, providing deeper understanding. When modern communication science extends to public opinion, the related knowledge is concluded as eight elements of public opinion, integrated with which the structure of attributes needed to compute the influence of specific events is established, as depicted in Fig. 1.

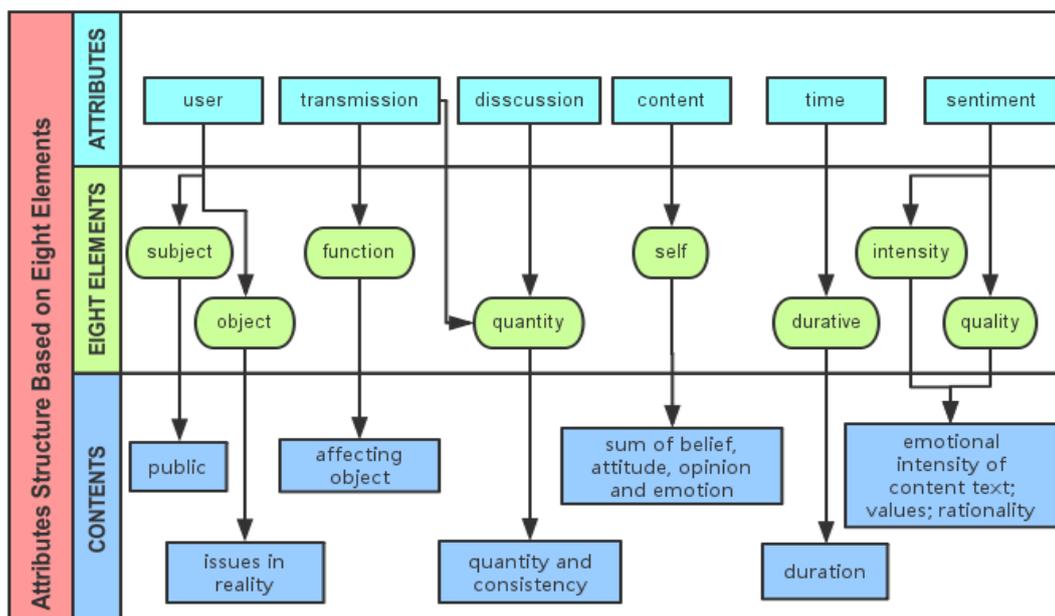


Fig. 1 Attributes Structure Based on Eight Elements of Public Opinion

As discussed above, extend of transmission, influence and participant opinion all depict a specific event from different aspects, and comprehensive and real analysis of an event is to be obtained through integrating all attributes. Besides influence index of events, this paper discusses as well the computation of users and messages.

3 Attributes Analysis and Computation of Influence of Specific Events

3.1 Influence of Complete Events

3.1.1 Attributes Analysis of Influence of Complete Events

Based on the eight elements of communication and understanding of events developing, we build a model of attributes describing influence of events from different angles, including user, transmission, discussion, content, time, and sentiment.

Attribute user is incarnated by participant users' characteristics, attribute transmission by media's involvement, attribute discussion by the quantity of related weibo, attribute content by the type and intensity of events, attribute time by the duration or time intensity, and attribute sentiment by emotional tendency of related weibo text. Taking the study of weibo text into account is an innovative approach of this model, computing the emotional score of each text with natural language processing tools, implementing the attributes of influence from the perspective of content.

3.1.2 Computation of Influence of Complete Events

Based on the listed attributes above, the concrete attribute content, computational formula and parameter description are extracted from dataset, as depicted in Table 3.1.

Table 3.1. Attributes and Computational Formulas of Events Influence

| Concrete Attributes Content | Computational Formula | Parameters Description |
|--|--|---|
| users' average followers(x_1) | $S_1 \cdot \min\left(1, \frac{x_1}{F}\right)$ | S_i : each assigned value; F (average followers), V (verified users), M (media): threshold for full credit of each attribute |
| notable verified users(x_2) | $S_2 \cdot \min\left(1, \frac{x_2}{V}\right)$ | |
| media users(x_3) | $S_3 \cdot \min\left(1, \frac{x_3}{M}\right)$ | |
| weight of event type and intensity(x_4) | see Table 2. | S_4 : assigned value |
| number of original weibo(x_5) | $S_5 \cdot \min\left(1, \frac{x_5 + x_6 + x_7}{x_8 \times W}\right)$ | S_5 : shared assigned value of four attributes; W (number of weibo) : threshold for full credit |
| number of comment(x_6) | | |
| number of retweet(x_7) | | |
| duration day(x_8) | | |
| positive emotion average score($x_{9,1}$), percentage($x_{9,2}$) | $S_6 \cdot \frac{x_{9,1} \times x_{9,2}}{L_p}$ | S_6 : assigned value; L_p : threshold for full credit |
| negative emotion average score($x_{10,1}$), percentage($x_{10,2}$) | $S_7 \cdot \frac{x_{10,1} \times x_{10,2}}{L_n}$ | S_7 : assigned value; L_n : threshold for full credit |

Table 3.2. Mapped Score of Type and Intensity of Events

| Type of Events | Intensity of Events | Mapped Score |
|---|---------------------|-----------------|
| Politic, Diplomacy, Disaster, Economic, Security, Legislation | Max | $1.0 \cdot S_4$ |
| | Middle | $0.8 \cdot S_4$ |
| | Min | $0.6 \cdot S_4$ |
| People's livelihood, Society, Education, Transportation | Max | $0.7 \cdot S_4$ |
| | Middle | $0.5 \cdot S_4$ |
| | Min | $0.3 \cdot S_4$ |
| Entertainment, Technology, Sports, Other | Max | $0.4 \cdot S_4$ |
| | Middle | $0.2 \cdot S_4$ |
| | Min | $0.1 \cdot S_4$ |

The combinations of type and intensity of events are mapped to specific scores determined by S_4 as shown in Table 3.2. . The ultimate score of centesimal system of events influence is limited by the condition described as the following equation, $\sum_{i=1}^7 s_i = 100$. The distribution of assigned value can be altered if needed. Suggested distribution based on practical working and experiment experience as shown in Table 3.3., thresholds for full credit in Table 3.4. .

Table 3.3. Suggested Distribution of Assigned Values

| | | | | | | | |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| Assigned Value | S_1 | S_2 | S_3 | S_4 | S_5 | S_6 | S_7 |
| Distribution | 20 | 15 | 10 | 10 | 35 | 3 | 7 |

Table 3.4. Suggested Thresholds for Full Credit

| | | | | | | |
|------------|-------|-----|-----|-------|-------|-------|
| Threshold | F | V | M | W | L_p | L_n |
| Suggestion | 30000 | 500 | 10 | 20000 | 5 | 5 |

The ultimate score of event influence from the computation above is called EII (Event Influence Index) in this paper.

3.2 Influence of Users

3.2.1 Attributes Analysis of Influence of Users

Based on summarizing and refining the eight elements, attributes of user influence are extracted from dataset, as shown in Table 3.5. .

Table 3.5. Attributes of User Influence

| | |
|---|--|
| user's current followers (X_1) | via statistics from data of weibo and users |
| user participation (X_2) | times the user post, retweet or comment in the event |
| number of comments and retweets of user's weibo (X_3) | number of retweets and comments of the user's weibo in the event |
| weights of user's followers (X_w) | corresponding weights of user's current followers |

3.2.2 Computation of Influence of Users

The computing scheme of user influence is described as follows: the three attributes, X_1 , X_2 , X_3 , are graded on a scale of 0 to 25 according to their numerical magnitude, of which the results are EX_1 , EX_2 , EX_3 , X_w , graded on a scale of 1.0 to 2.0, resulting in EX_w . Plugging the four results in the formulas below gives ultimate user influence score, moreover the ranking list.

$$Influence_{user} = EX_1 + EX_2 + (EX_2 \cdot EX_w) \quad (1)$$

The results are graded by formulas below:

$$EX_1 = \begin{cases} 25, X_1 \geq 5 \times 10^6 \\ 20, 5 \times 10^6 \geq X_1 \geq 10^6 \\ 15, 10^6 \geq X_1 \geq 5 \times 10^5 \\ 10, 5 \times 10^5 \geq X_1 \geq 10^5 \\ 0.0001 \cdot X_1 + 3.13, 10^5 \geq X_1 \geq 10^3 \\ 1, 10^3 \geq X_1 \geq 1 \\ 0, X_1 = 0 \end{cases} \quad (2)$$

$$EX_2 = \begin{cases} 0, X_2 = 0 \\ 2.56 \cdot X_2 - 0.49, \text{其他} \end{cases} \quad (3)$$

$$EX_2 = \begin{cases} 25, X_1 \geq 5 \times 10^5 \\ 20, 5 \times 10^5 \geq X_1 \geq 10^5 \\ 0.0002 \cdot X_2 + 3.78, 10^5 \geq X_1 \geq 5 \times 10^4 \\ 10, 5 \times 10^4 \geq X_1 \geq 10^4 \\ 0.0009 \cdot X_2 + 1.79, 10^5 \geq X_1 \geq 1 \\ 0, X_1 = 0 \end{cases} \quad (4)$$

$$EX_4 = \begin{cases} 2, X_1 \geq 5 \times 10^6 \\ 1.8, 5 \times 10^6 \geq X_1 \geq 10^6 \\ 1.6, 10^6 \geq X_1 \geq 5 \times 10^5 \\ 1.4, 5 \times 10^5 \geq X_1 \geq 10^5 \\ 1.2, 10^5 \geq X_1 \geq 5 \times 10^4 \\ 1.0, 5 \times 10^4 \geq X_1 \geq 10^4 \\ 0.8, 10^4 \geq X_1 \geq 10^3 \\ 0.6, 10^3 \geq X_1 \geq 1 \\ 0, X_1 = 0 \end{cases} \quad (5)$$

3.3 Influence of Messages

3.3.1 Attributes Analysis of Messages

The influence of a message embodies in the people it reaches and the intensity of discussion it generates, which can be interpreted as width of transmission and intensity of attention, which we think represent the personal influence of poster and the influence of the content and multiple transmission of the message. The number of followers of poster can reflect the former, namely more followers implying more readers. The comments and retweets of the message can reflect the latter, which are regarded as involved users who launch enough attention.

The momentum formula, $\mathbf{P} = \mathbf{m} \cdot \mathbf{v}$, shows momentum is decided by the mass and velocity. From the angle of analogy, we consider momentum similar to influence of message, with mass being personal influence and velocity being influence of content and transmission, and then the formula for influence of messages is constructed in reference of the momentum formula.

3.3.2 Computation of Influence of Messages

$Influence_{BLOG}$ stands for the influence score of a weibo message; m stands for the followers of poster, corresponding to mass; v stands for the number of comments and retweets of the message, corresponding to velocity.

$$Influence_{BLOG} = \ln((m + 1) \cdot (v + 1)) \quad (6)$$

The logarithmic function is added in expectation for better data characteristics, which transforms multiplying into adding, simplifying computation complexity.

4 Experiments and Evaluation

4.1 Experimental Data Set

The data set for experiment is collected from Sina Weibo in unit of event. The weibo of each event are retrieved through keywords and selected time period. The data set contains contents, comments, retweets, posting time and data type of weibo, and number of current followers of users.

The experiments are conducted on three events: "Fan Weiqi posting baby picture during triumphal review of troops" event (hereinafter referred to as "Fan" event), with 41043 weibo and 31356 users; "Bi Fujian's video leak" event (referred to as "Bi" event), with 10649 weibo and 8378 users; "Langyabang discussion" event (referred to as "Lang" event), with 39903 weibo and 11493 users.

4.2 Evaluation of Experiment Results of Event Influence

Extract attributes statistics proposed in section 3.1.2, and compute the grades with the formulas in Table 3.1., then the summation of grades generates EII. As is depicted in Table 4.1. (the grades are in parentheses).

Table 4.1. Result of Event Influence Experiment

| Attributes | "Fan" event | | "Lang" event | | "Bi" event | |
|--------------------|---------------------|------|------------------------|---------|---------------------|--------|
| X_1 | 11165.4 (7.44) | | 1942 (1.29) | | 32846.4 (20) | |
| X_2 | 390 (11.7) | | 92 (2.76) | | 189 (5.67) | |
| X_3 | 6 (6) | | 1 (1) | | 7 (7) | |
| X_4 | Politic, Max (10) | | Entertainment, Min (1) | | Politic, Max (10) | |
| X_5 | 14375 | (24) | 7509 | (34.92) | 3084 | (6.21) |
| X_6 | 15727 | | 12818 | | 2480 | |
| X_7 | 10941 | | 19576 | | 5085 | |
| X_8 | 3 | | 2 | | 3 | |
| $X_{9,1}X_{9,2}$ | 2.71,17.27% (0.28) | | 9.32,52.78% (2.95) | | 5.96,33.43% (1.2) | |
| $X_{10,1}X_{10,2}$ | -7.80,47.39% (5.17) | | -3.65,22.53% (1.15) | | -5.37,36.43% (2.74) | |
| EII | 66.59 | | 45.07 | | 55.82 | |

Events are classified according to EII score: 75 and higher means the highest level of influence, which calls for high concern; 50 to 75 means high influence; 50 and lower means rather ordinary influence.

As seen in Table 4.1. , with type of entertainment and intensity of min, the quantity and scale of "Lang" is close to "Fan" and far greater than "Bi", while its EII is much smaller than the other events, the score of which is 45.07, implying the ordinary influence of "Lang". This shows the partition degree of influence aligns highly with demands of public opinion surveillance and commonsense, namely the validity and reliability of this model.

"Fan" occurs in special time of triumphal review of troops of China, and its high grades of all attributes displays well the social attention and influence "Fan" obtains. The 66.59 EII classifies "Fan" as events of high influence, conforming to objective conditions.

Due to the incomplete data set, only part of "Bi" event's data is collected, resulting in lower grades of X_5 , X_6 , X_7 than actual circumstance. The 55.82 EII also indicates high influence, which shows the robustness of the model, namely not relying on quantity solely. This paper thinks it's safe to reason that with the increase in quantity attributes grades the EII of complete "Bi" event could reach 75, implying the highest influence.

In conclusion, the model proposed for computation of event influence possesses reasonability, validity, reliability and partition degree, satisfying its designing purpose and providing accurate analysis of hot events on social media, namely the technical support for public opinion surveillance.

4.3 Evaluation of Experiment Results of User Influence

The ranking list of user is the outcome of the model, the top 10 users of "Fan" is listed in Table 4.2. for references.

Table 4.2. Top 10 Users of "Fan" Event

| user nickname | influence |
|--------------------|-----------|
| 乱世中 | 47 |
| 撒比在看我 | 46 |
| 用户已注销 | 46 |
| 仲係咁 handsome | 46 |
| OceAnIlaX | 46 |
| 手机用户 2895409477 | 46 |
| 洛洛 boy 丶 | 46 |
| 梦一场_428 | 46 |
| 大胖子 ztp | 46 |
| Ashiiislii-Aliissi | 46 |

The process of manual distinction is described as follows using example in Table 4.2. :

- a) The number of comments and retweets of user is large, suits communication science;
- b) Their weibo is long, high-profile, sharp in words and provocative, even contains curse words;
- c) The user label akin to "music/movie/fresh", more than 30% were born after 1990;
- d) 60% of users have more than one hundred followers.

The features of event "Fan" meets with the columns above, and with more research of the user information, the manual distinction can effectively evaluate the experiment results.

After performing manual distinction on the top 50 and last 20 users on the ranking lists of three events, the precisions are in following tables.

Table 4.3. The Precision of High Influence User Recognition in "Fan"

| | | | | | |
|------------|--------|--------|--------|---------|---------|
| user range | Top 10 | Top 20 | Top 50 | Last 10 | Last 20 |
| precision | 70% | 75% | 76% | 90% | 95% |

Table 4.4. The Precision of High Influence User Recognition in "Bi"

| | | | | | |
|------------|--------|--------|--------|---------|---------|
| user range | Top 10 | Top 20 | Top 50 | Last 10 | Last 20 |
| precision | 60% | 75% | 86% | 100% | 95% |

Table 4.5. The Precision of High Influence User Recognition in "Lang"

| | | | | | |
|------------|--------|--------|--------|---------|---------|
| user range | Top 10 | Top 20 | Top 50 | Last 10 | Last 20 |
| precision | 90% | 75% | 78% | 80% | 90% |

In conclusion, the experiment results shown in tables above prove that the proposed model effectively locates and recognizes users of high influence, and maintains a high precision when the user range expands. The manual distinction of users at the bottom of lists shows the model avoids missing out influential users. The influence of users discussed is limited within the range of events, different from the constant and long-term influence in many other researches, which offers better understanding and analysis of specific events.

4.4 Evaluation of Experiment Results of Message Influence

The manual distinction above also apply to ranking lists of messages. The top 100 messages are performed manual distinction, the precision results of three events are shown in tables below.

Table 4.6. The Precision of High Influence Message Recognition in "Fan"

| | | | | |
|---------------|--------|--------|--------|---------|
| message range | Top 10 | Top 20 | Top 50 | Top 100 |
| precision | 100% | 90% | 92% | 93% |

Table 4.7. The Precision of High Influence Message Recognition in "Bi"

| | | | | |
|---------------|--------|--------|--------|---------|
| message range | Top 10 | Top 20 | Top 50 | Top 100 |
| precision | 90% | 85% | 88% | 89% |

Table 4.8. The Precision of High Influence Message Recognition in "Lang"

| | | | | |
|---------------|--------|--------|--------|---------|
| message range | Top 10 | Top 20 | Top 50 | Top 100 |
| precision | 90% | 90% | 92% | 91% |

As shown in three tables above, the proposed model filtrates and selects messages of rather high influence precisely, and maintains a steady precision when message range expands. Due to the features of messages, there is no different between constant and in-event influence, therefore the formula originated from momentum represents the preferable influence of messages. The locating of

influential messages reveals the development of events and hot point of views and direction of public opinion.

5 Summary

The model proposed in this paper aims at specific hot events on social media, typically represented by Sina Weibo, and covers the three aspects of important participants, hot messages and the complete influence of events. The model offers computational schemes for selecting and influence computation of influential users, the filtrating and locating of influential messages, and the construction and computation of attributes for events influence index, providing technical support for public opinion analysis and surveillance, as well as research on social media. The evaluation and validation of the experiment results on data set of Sina Weibo proves the reasonability, validity, reliability and partition degree of the model, also showing the fulfillment of the design purpose.

The optimization of selection and setting of parameters will be the future of the model, reducing dependency of experience values. A more effective method of evaluating and validating the experiment results with contrast experiment of similar models is also the key point of further research.

6 References

- [1] 祝华新,单学刚,胡江春.2011年中国互联网舆情分析报告[J].资治文摘:综合版,2012(8):24-29
- [2] 卢伟胜,郭躬德.基于特定话题的微博意见领袖在线检测方法[J].计算机应用与软件,2015, 32(5):70-74
- [3] 原福永,冯静,符茜茜.微博用户的影响力指数模型[J].现代图书情报技术,2012(6):60-64
- [4] F Riquelme. Measuring user influence on Twitter: A survey[J].Computer Science,2015
- [5] 陈远,袁艳红.微博信息传播的正负社会效应[J].信息资源管理学报,2012(2):48-54
- [6] Fuyong, YUAN1, Jing, FENG, Qianqian. An index model for measuring microblog users' influence[J].Chinese Journal of Library & Information Science, 2012, 5(4):67-76
- [7] Sitaram Asur , BernardoA Huberman. Predicting the Future With Social Media[J].Proc of Wiat,2010,7(2):492-499
- [8] 薛一波,鲍媛媛,易成岐.SPNR: 社交网络中的新型谣言传播模型[J].信息网络安全,2014(1):5-9
- [9] 周东浩,韩文报,王勇军.基于节点和信息特征的社会网络信息传播模型[J].计算机研究与发展,2015,52(1):156-166
- [10] Guille A,Hacid H.A predictive model for the temporal dynamics of information diffusion in online social networks[C]. International Conference Companion on World Wide Web,2012,144(1):1145-1152
- [11] 杨亮,林原,林鸿飞.基于情感分布的微博热点事件发现[J].中文信息学报,2012,26(1):84-90
- [12] Y Zhang,Y Bai.Research on the Influence of Micro-bloggers—Take Sina Celebrity Micro-blog as an Example[C].English International Conference on Semantics,2012:189-192
- [13] <http://wyq.sina.com/bd.shtml>
- [14] <http://www.gsdata.cn/>
- [15] 郭庆光.传播学教程[M].中国人民大学出版社,2011