

Analysis of Topic Influence and Post Features of Sina-Weibo

Borong Lyu, Xinhui Shao*, Yinbo Huang and Yuyang Xie

Department of Mathematics, Northeastern University, Shenyang 110819, P.R.China

*Corresponding author

Abstract—Microblog serves as an outlet for expressing ideas of events and plays a significant role in propagating news. This paper focused on the most popular microblog in China - Sina-Weibo and discussed topic influences. Firstly, this paper decomposed topic structure into Average Fundamental Popularity and Information Tipping Points. This paper classified topics into three patterns. Next, we collected 10455 available Information Tipping Points and obtained all the posts on each one by Python. We processed these posts and got 20 features, including features within the social network and features in the whole network, such as Ratio of Original Posts and Forward Hierarchy. Furthermore, K-means and EM algorithm were applied to cluster subtopics. Statistical methods such as Spearman Rank Correlation Coefficient Method, the Levene Homogeneity of Variance Test, and One-way Analysis of Variance were used to analyze the relationship among some features. We find that eight of all the features have strong relationships with the Number of Weibo Posts. Furthermore, the total number of posts on one topic (the Number of Weibo Posts) can represent topic impact. This paper established a predictive model via the regression method to predict the Number of Weibo Posts. We also find that whether a topic can become influential can be predicted by its features in the whole network. Finally, this paper applied SVM algorithm to determine which subtopic can become a trending issue.

Keywords—Sina-Weibo; social network; features; information tipping points; predictive model

I. INTRODUCTION

Social networks are becoming major sources of commentaries on a wide range of topics [1]. When a user posts something on the social network, his followers can leave a comment, give a like, or even forward it. Some posts are reposted by millions of users in the social network, changing public opinions significantly. Thus, predictive analysis of the social network is of great value. Since some trending topics are caused by exclusive links which cannot be described or predicted by information diffusion models but can be analyzed by data analysis methods [2, 3], data analyzing of social networks is a prospective field with wide implications.

Studies focus on predicting popular posts based on post features in the social network from a data analyzing perspective have just begun. In [4, 5, 6], a number of features that might affect retweetability of tweets were examined. Cao, Wu, Shi, Liu, Zheng, & Luo. (2014). base on a large number of online data from Sina-Weibo [7], analyzed a variety of possible factors that can affect users' retweeting behavior. A

dataset of Twitter conversations was used in [8] and found that users could entertain a maximum of 100-200 stable relationships. In [9], celebrity and ordinary users' behaviors on Sina-Weibo were investigated and compared. Hong, Dan, & Davison (2011) predicted the popularity of message according to a number of features, such as temporal information, users, as well as structural properties of users' social graph on a large scale dataset [10]. In [11], researchers purposed an algorithm to predict the number of retweets according to a set of features. Researchers in [12] took Sina-Weibo as a case to study whether structural features of contents can be used to predict the popularity of short messages. Researchers in [13] investigated the attributes and relative influence of 1.6M Twitter users by tracking 74 million diffusion events.

However, most of the predictive works focus on features of a single blog or a whole topic. Very few researchers have analyzed the topic structure deeply and classified the topic patterns before applying data analyzing algorithms. Gruhl, Guha, Liben-Nowell & Tomkins. (2004). formalized the notion of long-running "chatter" topics consisting recursively of "spike" topics [14]. In [15], Topical Affinity Propagation (TAP) model was proposed to describe the social impact of topic hierarchy on large networks.

This paper takes Sina-Weibo, a microblog which enjoys a great popularity in China as a case and focuses on the structure of topics as well as analyzes the topic patterns by decomposing topic structures into average fundamental popularity and information tipping points. Furthermore, this paper clears out features about information tipping points to analyze topic popularity thoroughly through data mining algorithms and statistical methods.

The remainder of this article is arranged as follows: In section 2, this paper decomposes topic components and classifies topic patterns into three categories. Data background is contained in section 3. The relationship among posts features is put forward in section 4. In the end of the paper, conclusions are drawn in Section 5.

II. TOPIC STRUCTURE ANALYSIS

A. Topic Components Analysis

This section discusses basic components of a topic. Up to now, only researchers in [9] have broken down the topic structure. They formed the notion of *chatter* and *spike*: There is a possibility that bloggers express their new thoughts on any

given day about any topic which might have nothing to do with currently trending topics. This constitutes the *chatter* on that topic. Occasionally, an event occurring in the real world raises a great deal of dissension. This constitutes the *spike* on that topic. According to the data analysis result in [9], spikes do not typically propagate through Blogspace. Thus, a basic topic composed of *chatter* and recursively of *spikes*. This paper adapts the idea in [9] and improves it.

Some topics in the social network include a basic numbers of blogs, such as *Presidential Election*, *Microsoft*, and *Smart Phone*. On any given day, some bloggers express their thoughts on the topic or response to other postings. These constitute the basic popularity of that topic. The variations of these topics are apparently relevant to their social background. For example, the presidential election in 2016 facilitates *#Trump* to be the most popular topic in the past few years. The basic popularity is assumed to remain fixed during a short period. This paper now renames it appropriately as *average fundamental popularity*. For instance, topic “South China Sea” has an average fundamental popularity.

Occasionally, an event occurring in the real world leads to a heated discussion. Information thrives suddenly like a virus, such as “South China Sea arbitration” and “Establish San-Sha City in the South China Sea”. These subtopics constitute variation. This paper now renames it as *Information tipping point*[16].

To the best of our knowledge, there has not been any model describing wax and wane of a topic yet. A Model which aims to analyze and predict the variation of *average fundamental popularity* (*social background*) of a long-term should take sociology factors into consideration, such as news from mass media, government policies, GDP, and economic crisis, etc. This paper leaves it as an open problem. This paper devotes attention to short-term topics, where its *average fundamental popularity* is assumed to remain fixed and unchanged.

B. Topic Pattern Classification

Topics are classified into three patterns in this section according to the 10900 trending news we choose. First, we write a Python web crawler to obtain 10900 trending news events from 00:00:00 May 31st, 2016 to 23:59:59 December 31st, 2016 from a famous news website in China-Tencent News [12]. Next, for each news event, we obtain the number of posts of this topic per day. The detail about how the dataset is obtained is stated in Section 3.

We manually find that topic can be classified into three patterns. Most topics have several tipping points. Moreover, each tipping point obviously has a reason. Some of these tipping points are related to news events, others exist because of their content reach *resonance*. They have three patterns according to the number of *information tipping points* and their *average fundamental popularity*, which is familiar to that the researcher in [9] found on Blogspace.

- Topics with no much talk during common time with only few tipping points. These topics have only few posts before and after their tipping points. Such as

“Alipay + withdraw + charging”. There are 42387 blogs about this topic, and the most frequent posting is 27277 posts on September 12th. As is shown in fig. 1. The red line represents posts that are original and the green line represents reposts. Posts posted by news media are represented by yellow. Blogs about this topic first appeared in the social network on September 12th. Sina-Weibo reacted fast and calmed down fast too. On September 20th, the number of posts decreased to 209. Indeed, Alipay announced their withdraw charge policy on September 12th. The second information tipping point exits on October 12th mainly because it is the day when the withdraw policy became effective. The posts before September 12th exists because Wechat carried out new withdraw charge policy in March, and users on microblog compared these two applications. Furthermore, it is obvious that this tipping point is caused by exclusive links.

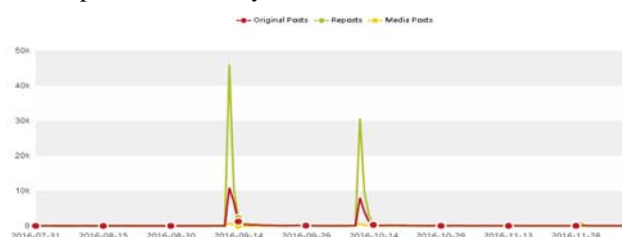


FIGURE I. THE NUMBER OF WEIBO POSTS ABOUT “ALIPAAAY WITHDRAW CHARGE” PER DAY

- Topics which their posts per day fluctuate within a narrow range with no obvious tipping point, such as “Chinese and Vietnam strategic cooperate”. We use keywords “(China+ Vietnam) + strategic+ cooperate”, and obtain the number of posts which mention these words per day. As is shown in Figure 2, there are 621 posts in total and 24 posts maximum. Although there is a fluctuating profile, the total number is low. Moreover, the topic has no tipping point and its’ number of posts per day is fluctuating in a narrow range. The average fundamental popularity remains the same. From the viewpoint of news events, there is no striking news about this topic.

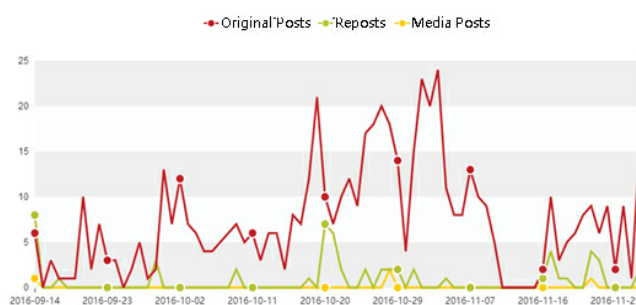


FIGURE II. THE NUMBER OF WEIBO POSTS ABOUT “CHINA AND VIETNAM STRATGIC COOPERATE” PER DAY

- Topics which have a significant *Average fundamental popularity* level and quite sensitive to external world events. They have many *Information Tipping Points*. Such as the topic *Brexit vote*. The Brexit vote

happened on June 24th. As is shown in Figure 3(a), the tipping point on the left represents the striking news-Brexit vote. However, because of its high popularity, there are small tipping points even after one month as is shown in Figure 3(b). One of the tipping points in October appears because some influential users compared this vote with the presidential election in America, pointing out how democracy and voting system sucks. A striking news like this is always accompanied by a few tipping points and any little item can result in a tipping point.

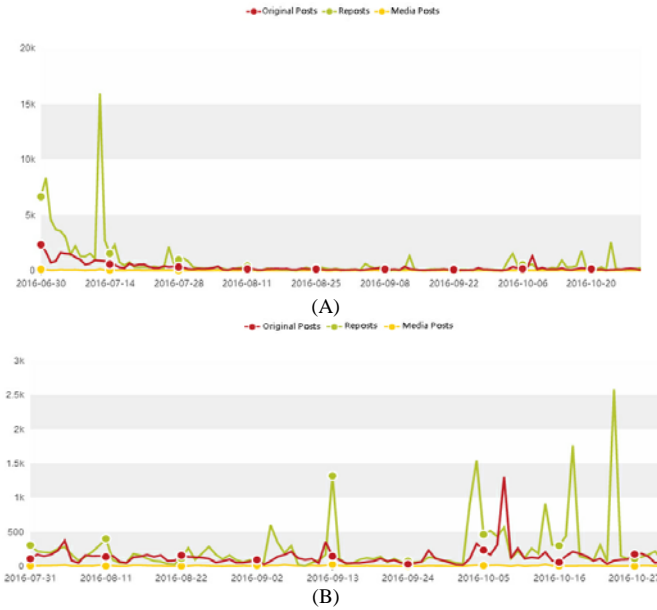


FIGURE III. (A). THE NUMBER OF WEIBO POSTS ABOUT “BREXIT VOTE” PER DAY FROM JUNE 30TH TO OCTOBER 31TH. (B). THE NUMBER OF WEIBO POSTS ABOUT “BREXIT VOTE” PER DAY FROM JULY 31ST TO OCTOBER 31TH

III. DATA BACKGROUND

This section illustrates how this paper clears out information tipping points.

It is interesting to figure out problems such as: Whether more media accounts along with hot discussions; whether hot discussions along with more original posts or reposts; and whether hot discussions have a correlation with certified accounts. This paper focuses on information tipping points of topics of type 1 and type 3 and discusses these problems

In order to focus on information tipping points, unknown social effect have to be eliminated. We focus on topics in a short-term so that social effect is assumed to remain fixed. Indeed, each tipping point is connected to a news event. Firstly, we choose 10900 trending news events between 00:00:00 May 31st, 2016 and 23:59:59 December 31st, 2016 from a famous news website- China Tencent News [17].

Next, for each news event, we obtain the number of posts on this topic per day.

Then, we need to clear out keywords connected to different news events. On each news event, to make sure the keywords

we obtain are about this news event and all the blogs related to the news event are included, we have to check repetitively. Indeed, only keywords are not enough, sometimes more exclusive words are needed.

If the profile of the number of posts per day has more than one peak, we would manually add exclusive words to this news event. If it has more than 3 peaks, we would delete this news in our database. Up to now, we have 10565 news events. Meanwhile, the date when each event happens and the date when each event comes to an end can also be determined (Usually a few days).

Moreover, we judge whether it has more than 100 relevant posts maximum. If it does, this news is regarded as a tipping point. Finally, we obtain 10507 news events.

The whole process is shown in Figure 4.

In the end, for each tipping point, this paper obtains The Number of Weibo Posts (total relevant posts) (denoted as F); 17 features within the social network: see Table 1, and (b) 3 features in the whole network as is shown in Table 2.

TABLE I. FEATURES WITHIN THE SOCIAL NETWORK

Symbol	Meaning
A	Median of Opinion Leaders' Fan Number (opinion leaders here means blog users of top 10 forwards)
X₁	Median of Forwarded Number of Core Communication Agencies (Core communication agencies means agency accounts of top 10 forwards, so as follows)
X₂	Median of Forwarded Number of Enthusiastic Media
X₃	Median of Forwarded Number of Enthusiastic Media People
X₄	Median of Forwarded Number of Core Internet Users
B	Forwarded Number of Popular Original Posts
C	Forwarded Number of Popular Forwards
D	Ratio of Original Posts
E	Forward Hierarchy (including Initial Forward, Second Forward, Third Forward, Fourth Forward)
G	Certified Account

Among these features, Original Posts refer to posts that are originally created by blog user. A post which is not originally created but forwarded is called repost. Enthusiastic Media refer to media accounts of top 10 forwards. Enthusiastic media people refer to media people's accounts of top 10 forwards. Core Internet Users means common users' accounts (which have no correlation with mass media) of top 10 forwards. When calculating Median of Forwarded Number, this paper refers to the forwarded number of popular original posts only.

Certified account: Including Common User, Orange V, Talent, and Blue V. When a user is certified by Sina-Weibo, a

logo will appear next to this users' pseudonym (a colored V for orange V and blue V; a red star for talent). Blue V means the same as Orange V but is designed for agencies, medium, and enterprises. Having an Orange V or a Blue V certification needs to: follow more than 30 accounts; possess more than 100 followers; and have more than two mutual fans with an Orange/Blue V mark. Talent mark is acquired by accumulating scores, which are obtained by login, posting, etc. Since the certified logo is obvious and it represents influential people, this might affect users' behavior on the social network.

TABLE II. FEATURES IN THE WHOLE NETWORK

Symbol	Meaning
H	The Number of Reports in the Whole Network.
I	Ratio of Blogs to Reports in the Whole Network
J	Median of Forwarded Number of Popular Reports in the Whole Network

When confronted with impossible observations in the data collected, relevant news event is deleted. In the end, there are 10455 news events left.

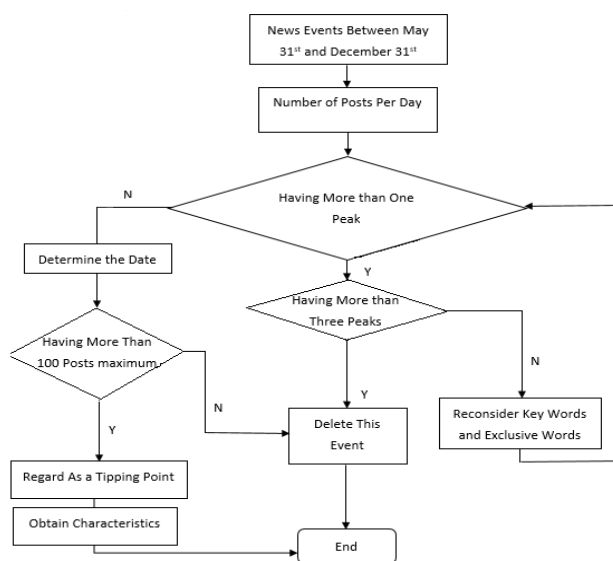


FIGURE IV. FEATURES IN THE WHOLE NETWORK

IV. FEATURES CORRELATED WITH INFORMATION TIPPING POINTS

A. Features Within the Social Network

The factors correlated with information tipping points are discussed in this section

1) Clusters

K-means Cluster algorithm is applied to classify all these tipping points into 2 clusters according to their features. It is obvious that one *tipping point* cluster is made up of news events which received little attention (Cluster 1) and the other

one is made up of news events with lively discussions (Cluster 2) [18].

Then Expectation-maximization algorithm is applied to cluster the Number of Weibo Posts. The number of categories increases to 4, but still can be manually classified into two clusters, which is the same as K-means Cluster algorithm. Therefore, they can be classified into 2 categories according to only one variable-The Number of Weibo Posts, which can represent topic's influence in the social network.

2) Ratio of Original Posts

The relationship between clusters and Ratio of Original Posts is analyzed in this section. As is shown in the box plot in Figure 5.

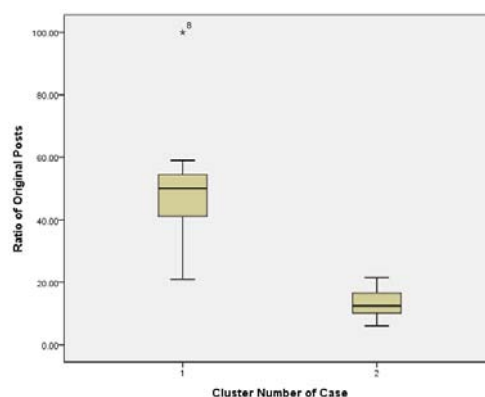


FIGURE V. RELATIONSHIP BETWEEN CLUSTERS AND RATIO OF ORIGINAL POSTS

It is obvious that most posts about lively discussed news events are reposts. In a word, more posts in total, more reposts in ratio.

3) Forwarded Number of Popular Original Posts, Forwarded Number of Popular Forwards.

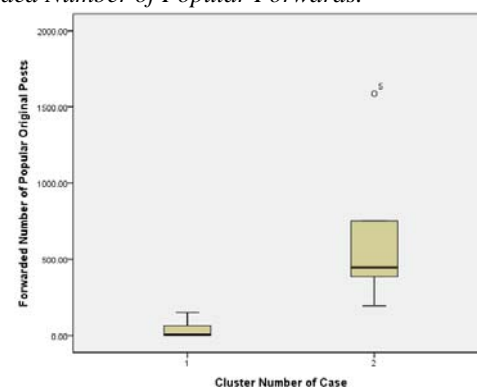


FIGURE VI. RELATIONSHIP BETWEEN FORWARDED NUMBER OF POPULAR ORIGINAL POSTS AND CLUSTERS

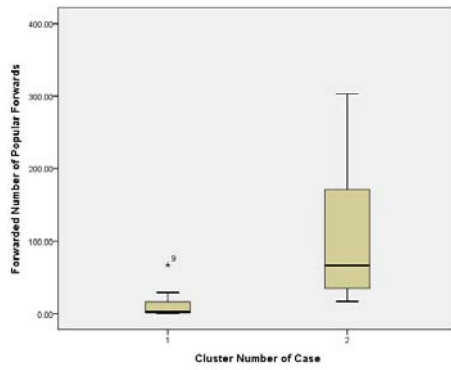


FIGURE VII. RELATIONSHIP BETWEEN FORWARDED NUMBER OF POPULAR FORWARDS AND CLUSTERS

As is shown in Figure 6 and 7, it is obvious that with more intense discussions, higher the Forwarded Number of Popular Original Posts and Forwarded Number of Popular Forwards will be.

4) Median of Opinion Leader's Fan Number

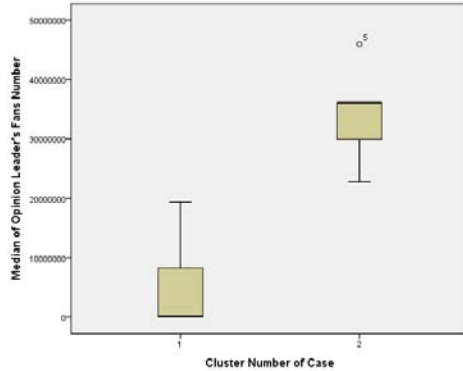


FIGURE VIII. RELATIONSHIP BETWEEN MEDIAN OF OPINION LEADER'S FAN NUMBER AND CLUSTERS.

Lively discussed topics are always related to influential users.

5) Median of Forwarded Number of Core Communication Agencies/Enthusiastic Media/Enthusiastic Media People/Core Internet Users

According to the data we have, hot topics usually have more Median of Forwarded Number of Communication Agencies/Media/Media People/ Internet Users. Moreover, Forwarded Number of Enthusiastic Media \geq Enthusiastic Media People \geq Core Internet Users \geq Core Communication Agencies under most of the circumstances.

Here we discuss the relationship between these four accounts.

This paper calculates their spearman correlation coefficient, the result is shown in Table 3.

TABLE III. SPEARMAN CORRELATION COEFFICIENT AMONG MEDIAN OF FORWARDED NUMBER OF CORE COMMUNICATION AGENCIES/ENTHUSIASTIC MEDIA/ENTHUSIASTIC MEDIA PEOPLE/CORE INTERNET USERS

		x_1	x_2	x_3	x_4
x_1	Correlation Coefficient	1.000	0.749**	0.736**	0.742**
	Sig. (2-tailed)	0	0.003	0.004	0.004
	N	13	13	13	13
x_2	Correlation Coefficient	0.749**	1.000	0.782**	0.823**
	Sig. (2-tailed)	0.003	0	0.002	0.001
	N	13	13	13	13
x_3	Correlation Coefficient	0.736**	0.782**	1.000	0.956**
	Sig. (2-tailed)	0.004	0.002	0	0
	N	13	13	13	13
x_4	Correlation Coefficient	0.742**	0.823**	0.956**	1.000
	Sig. (2-tailed)	0.004	0.001	0	0
	N	13	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

P-Value is less than significant level 0.05, thus, we reject the original hypothesis. These four variables have a strong linear correlation. Therefore, a linear regression analysis model can be set.

For example, we can use any three of these four variations to predict the other one.

The linear regression analysis result applying enters regression method is stated as follows:

$$x_4 = -8.722 + 0.094x_2 - 0.395x_1 + 0.978x_3$$

Then its' residual normal P-P plot is drawn in Figure 9.

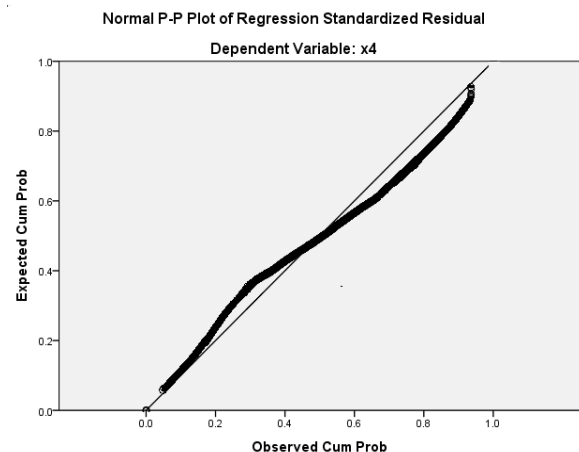


FIGURE IX. NORMAL P-P PLOT OF REGRESSION STANDARDIZED RESIDUAL

In light of our data set is impossible to be large, the linear regression model can be regarded as a satisfying result.

6) Common Account, Certified Account.

Before one-way analysis of variance, Homogeneity of variance test is needed. The P-value of Levene homogeneity of variance test is 0.135. This P-value is bigger than the significance level, so the original hypothesis should be accepted. Thus, one-way analysis of variance can be preceded. As is shown in Table 4.

TABLE IV. ANOVA COMMON ACCOUNT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.882	5235	13.882	0.128	0.727
Within Groups	1188.604	5220	108.055		
Total	1202.486	10455			

Since the P-value of one way analysis of variance is 0.727, the original hypothesis should be accepted. As a result, Common Users and topic classification are irrelevant. If we ignore a few extreme points in case1, Common Users change in the same range, and so is Certified Accounts.

Here we conclude that Certified Accounts have nothing to do with users' behavior. This is to say, no matter what kind of account this user possesses, the possibility of getting involved in an issue is the same. Moreover, a few extreme points were removed and we calculated their average: Common Users take up 82.55%, Orange V takes up 4.16%, Talent takes up 9.05%, and Blue V takes up 4.25%. Unfortunately, the official number is not available in concern of information disclosure, we cannot validate this assumption.

7) Forward Hierarchy

The analyzing process is the same as Certified Account, and so is the conclusion. For Initial Forward, Homogeneity of Variance Test is preceded first. The P-value of Levene Homogeneity of variance test is 0.112, which is bigger than the significance level, so the original hypothesis should be accepted. Thus, one-way analysis of variance can be preceded. The result is shown in Table 5.

TABLE V. ANOVA INITIAL FORWARD

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1723.331	5235	1723.331	3.963	0.072
Within Groups	4783.107	5220	434.828		
Total	6506.438	10455			

Because the P-value of one way analysis of variance is 0.072, the original hypothesis should be accepted. As a result, Initial Forward and topic classification is irrelevant. For the Second/Third/Fourth Forward, the result is the same. Here we conclude that Forward Hierarchy has nothing to do with users' behavior, no matter how many times a post being forwarded, so the possibility of being forwarded again still remains the same.

8) Predictive Model

Up to now, 8 relevant features have been recognized, and their linear relationship between The Number of Weibo Posts is obvious. Thus, we proceed linear regression analysis applying enter regression method.

The result is as follows:

$$F=174.905X_1+7.762X_2-1041.811X_3-340.076X_4-114.523B+1755.961C+78.950D-6135.466$$

Its residual normal P-P plot is shown in Figure 10.

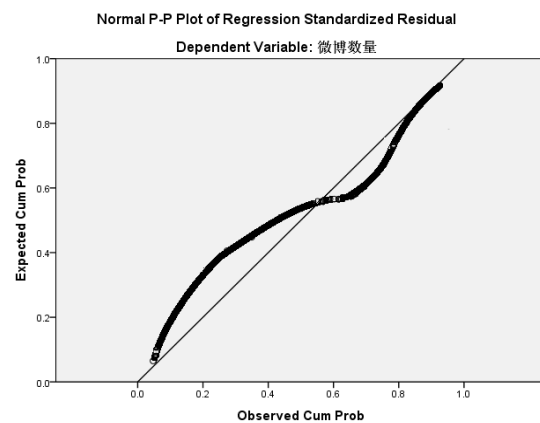


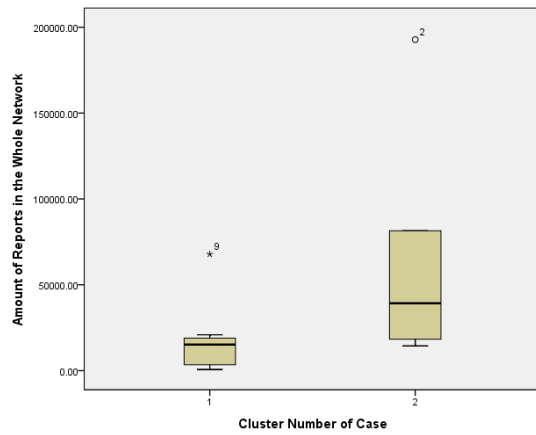
FIGURE X. NORMAL P-P PLOT OF REGRESSION STANDARDIZED RESIDUAL

In light of our dataset is not large, the standardized residual is considered to comply normal distribution, this result is reliable. This linear model can be used to predict The Number of Weibo Posts. Furthermore, this linear model can be used to predict how influential this news event can be.

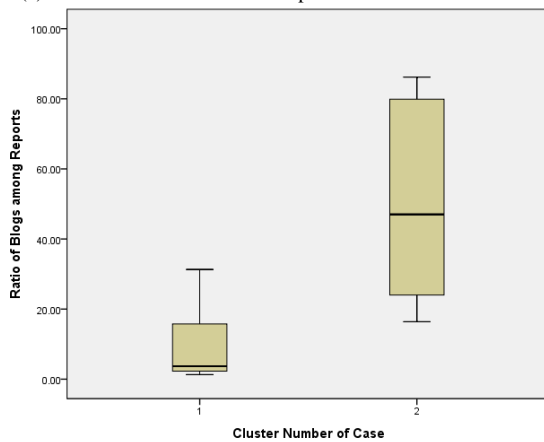
B. Features of the Whole Network

In this section, we discuss the relationship between features in the whole network and its popularity in Sina-Weibo, this section continues to use the clustering result in section 4.1.1.

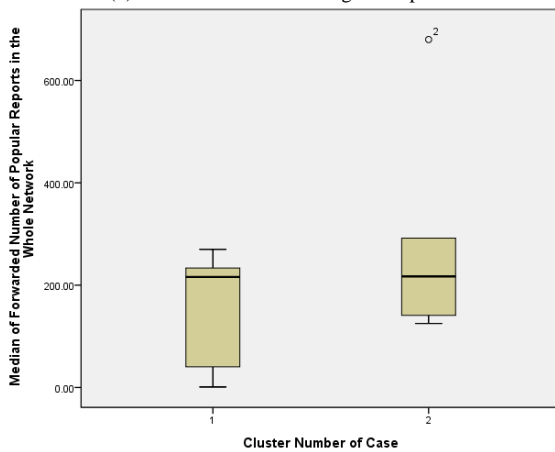
Reports Number in the whole network. The Ratio of Blogs to Reports and Median of Forwarded Number of Popular Reports in the Whole Network cannot determine clusters alone.



(a) Box Plot of the Number of Reports in the Whole Network



(b) Box Plot of Ratio of Blogs to Reports



(c) Box Plot of Median of Forwarded Number of Popular Reports in the Whole Network

FIGURE XI. BOX PLOT OF FEATURES IN THE WHOLE NETWORK

None of these features can determine subtopics clusters alone. However, as is analyzed in section 4, the Number of Reports in the Whole Network multiple The Ratio of Blogs to Reports represents the Number of Weibo Posts. Furthermore, these two features can represent clusters. Moreover, the Number of Reports in the Whole Network and Median of Forwarded Number of Popular Reports in the Whole Network

can be used to predict the subtopics' cluster. Here Support Vector Machine algorithm (SVM) is applied [13].

This paper implemented this algorithm with MATLAB. The discriminant function is as follows:

$$c(\tilde{x}) = \sum_i \beta_i K(b_i, \tilde{x}) + b = 0.3484K(b_1, \tilde{x}) + 0.1200K(b_{205}, \tilde{x}) + 0.0703K(b_{478}, \tilde{x}) - 0.0130K(b_{521}, \tilde{x}) - 0.5256K(b_{10036}, \tilde{x}) + 0.0772$$

The error rate is 0. Thus, the Number of Reports in the Whole Network and Median of Forwarded Number of Popular Reports in the Whole Network can be used to predict whether a subtopic can become a hot topic.

V. CONCLUSIONS

This paper first divided *information tipping points* into two clusters by using the K-means algorithm and found out that one cluster is composed of news events of few attentions; the other one is made up of news events with lively discussions. Moreover, with MATLAB, implementation of the Expectation-Maximization algorithm to cluster the Number of Weibo Posts was discussed and found that it can totally represent topic influence.

This paper, with SPSS, concluded that more posts in total, more reposts in ratio. Higher the forwarded number of popular original posts is, higher the forwarded number of popular forwards will be and more influential users are involved in. Furthermore, Median of Forwarded Number of Core Communication Agencies/Enthusiastic Media/Enthusiastic Media People/Core Internet Users has strong correlation relation with each other.

Next, one-way analysis of variance revealed that common account, certified account, and forward hierarchy have nothing to do with the popularity of subtopics. Also, we could deduce that the average ratio of certified account in our database equals the ratio of certified account to all blog users in the social network. Finally, we established a predictive model to predict The Number of Weibo Posts using 8 relevant features.

For features in the whole network, this paper applied Support Vector Machine algorithm to obtain a discriminant function to determine whether this topic can become a trending issue.

The main contributions of this paper are:

- Discussed topic structure more thoroughly. Topic components were decomposed into *Average fundamental popularity* and *information tipping points*.
- This paper focused on topic level and analyzed features of *information tipping points* from a data mining perspective.
- This paper is the first to take features in the whole network into consideration. Since social network is now interrelated with the whole network closely, our research is deep and profound.

ACKNOWLEDGMENT

This paper acknowledges Sina WYQ for the data they provide [19]. Sina WYQ is an official public opinion analysis website; it can provide data on Website, Microblog, WeChat, News Media, and Newspapers to analysis public sentiment. We also acknowledge National Natural Science Foundation of China (No. 11371081) for their support.

[19] Sina-WYQ. Retrieved on December 16, 2016, from <http://wyq.sina.com>.

REFERENCES

- [1] Wikipedia: Microblogging in China. Retrieved on February 20, 2017, from https://en.wikipedia.org/wiki/Microblogging_in_China.
- [2] Cha, M., Mislove, A., Adams, B., & Gummadi, K. P. (2008). Characterizing social cascades in flickr. ACM. The Workshop on Online Social Networks. (pp. 13-18). Seattle, Washington, USA.
- [3] Liben-Nowell, D., & Kleinberg, J. (2008). Tracing information flow on a global scale using internet chain-letter data. In Proceedings of the National Academy of Sciences of the United States of America (pp. 4633-4638). USA.
- [4] Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In Proceedings of the IEEE International Conference on Social Computing. (pp. 177-184). Pala Alto. USA.
- [5] Zaman, T. R., Herbrich, R., van, Gael, J., & Stern, D. (2010). Predicting information spreading in Twitter. In Proceedings of the Neural Information Processing Systems. (pp. 598-601) Vancouver, Canada.
- [6] Zarrella, D., Science of Retweets. Retrieved on February 15, 2017, from <http://danzarrella.com/the-science-of-retweets-report.html>.
- [7] Cao, J., Wu, J., Shi, W., Liu, B., Zheng, X., & Luo, J. (2014). Sina Microblog information diffusion analysis and prediction. Chinese Journal of Computers. 37(4), 779-790.
- [8] Zhang, B., Guan, X., Khan, M. J., & Zhou, Y. (2012). A Time-Varying Propagation Model of Hot Topic on BBS Sites and Blog Networks. Information Science. 187(1), 15-32.
- [9] Liu, X., Wang, W., & Wu, J. (2015). Celebrity and ordinary users: A Comparative Study of Microblog User Behaviors on Sina-Weibo. Journal of Data and Information Science. 32(9), 83-95.
- [10] Hong, L.J., Dan, O., Davison, B.D. (2011). Predicting Popular Messages in Twitter. Proceedings of the 20th international conference companion on World Wide Web. (pp. 57-58). New York, USA.
- [11] Kupavskii, A., Ostroumova, L., Umnov, A., Usachev, S., Serdyukov, P., Gusev, G., Kustarev, A. (2012). Prediction of Retweet Cascade Size over Time. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management. (pp. 2335-2338). Maui, Hawaii, USA.
- [12] Bao, P., Shen, H., Huang, J., Cheng, X. (2013). Popularity Prediction in Microblogging Network: A Case Study on Sina Weibo. In Proceedings of the 22nd International Conference on World Wide Web. (pp. 177-178). New York, USA.
- [13] Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J. (2011). Everyone's an influencer: Quantifying influence on Twitter. Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. (pp. 65-74). New York, USA.
- [14] Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information Diffusion through Blogspace. International Conference on World Wide Web (pp.491-501). New York, USA.
- [15] Tang, J., Sun, J., Wang, C., & Yang, Z. (2009). Social Influence Analysis in Large-scale Networks. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 807-816). Paris, France.
- [16] Gladwell, M. (2000). The Tipping Point: How Little Things Can Make a Big Difference. Little Brown.
- [17] Tencent News. Retrieved on December 1, 2016, from http://news.qq.com/china_index.shtml.
- [18] Wu, X., & Kumar, V. (2013). The Top Ten Algorithm in Data Mining. Tsinghua University Press.