

Mingmei Li

**Analysis of Network Rumor Dissemination and Control
Mechanisms on Chinese Social Network – Sina Weibo**

Master's thesis of mathematical information technology

December 21, 2016

University of Jyväskylä

Department of Mathematical Information Technology

Author: Mingmei Li

Contact information: ming.m.li@student.jyu.fi

Supervisors: Prof. Jari Veijanainen & Prof. Tapani Ristaniemi

Title: Analysis of Network Rumor Dissemination and Control Mechanisms on Chinese Social Network – Sina Weibo

Project: Master's thesis

Study line: Social Network Analysis

Page count: 56+2 (56 = page count without appendices; 2 = page count of appendices)

Abstract:

The social network has become a major source of information. The openness and swiftness of the network bring instant accessing to both true and false information. Sometimes the false information such as network rumors can mislead people from obtaining the true information and even create chaos. The social network platforms and the governments in different countries are making effort on enacting their own mechanisms to strive on eliminating the rumors.

The aim of this thesis is to investigate the spreading of rumors on the Chinese social networking website Sina Weibo. The censorship policy of the Chinese government and the methods used to control the spreading of rumors are discussed. The latest rumor control policy is a particular focus. It states that, if a post on Sina Weibo is tagged as a rumor and it is retweeted more than 500 times and/or viewed/clicked more than 5000 times, the originator of the post can be charged and sentenced to imprisonment for up to three years. The penalty also includes limiting access to Sina Weibo.

Previous research into rumor dissemination on social networks is reviewed, including the definition and content of the rumors, as well as rumor spreading models, such as DK (Daley-Kendal) and its variant MK (Maki-Thompson), improved ones based on SIR (susceptible-infected-removed) from the theory of epidemics such as Weighted CSR, SIHR and

SIRe. Also covered is the status of internet censorship and rumors in China and the development of Sina Weibo and its rumor refutation system.

The latest rumor control policy is examined as a case study. A dataset was collected using a web crawler system which was technically supported by a cooperative research group with the keywords in Chinese of “rumors, being retweeted 500 times”. Based on the data analysis and visualization, characteristics such as discussion of the topic over time could be observed. A supplementary lexical text analysis is also covered. The results show the attitude of people towards the case topic, although the effect of the latest rumor control policy was not obvious due to the size of the dataset. No certain result is generated to indicate the connection between people’s rumor spreading willingness and the promulgation of the latest rumor control policy.

This research addresses the following specific questions: How are rumors generated? What is the relationship between rumors and the rumor refutation system used by Sina Weibo? What is the effect of the latest rumor control policy on the dissemination of rumors and the control of public opinion?

This area of investigation would benefit from further research into rumor control mechanisms, especially for Chinese social networks, using automatic rumor detection and refutation systems based on deep learning and sentiment analysis.

Keywords: Chinese Social Network, Sina Weibo, Network Rumors, dissemination of rumors, rumor control, Social Network Analysis, Data Analysis, Data Visualization

Glossary

ASCII	American Standard Code for Information Interchange
MAU	Monthly Active Users
DAU	Daily Active Users
JSON	JavaScript Object Notation
CSV	Comma-Separated Values

List of Figures

Figure 1	The structure of Sina community content management platform	23
Figure 2	The capture of platform entry on users' homepage on Weibo	26
Figure 3	The capture of display example for the account @weibopiyao and hashtag #Weibo Rumors refutation#	27
Figure 4	The discussing tendency of the certain words from 09.09.2013 to 30.10.2013	30
Figure 5	The capture of data explanation for figure 4	31
Figure 6	The communication model of Weibo	32
Figure 7	The sample of metadata in JSON format.....	38
Figure 8	The discussing tendency indicated regarding to the time and count of comments.....	39
Figure 9	The discussing tendency indicated regarding to the time and comparison on count of comments and retweets	40
Figure 10	Word Cloud of Weibo Contents with combined keywords "Rumors, being retweeted 500 times"	41
Figure 11	Rank list of word mentioned frequency	42
Figure 12	The capture of the most commented and retweeted post regarding to the topic	43

Contents

1	INTRODUCTION.....	1
1.1	Background.....	1
1.2	Objective and significance of the research.....	2
1.3	Methodology and structure of the thesis.....	3
2	PREVIOUS RESEARCH ON RUMOR DISSEMINATION IN THE SOCIAL NETWORK.....	5
2.1	The definition of network rumor.....	5
2.2	Social network rumors.....	7
2.2.1	The content of rumors.....	7
2.2.2	Rumor spreading models.....	7
2.3	Internet censorship in China.....	11
3	THE DEVELOPMENT AND SUCCESS OF WEIBO IN CHINA.....	15
3.1	The internet, social media, microblogging and Chinese Weibo.....	15
3.2	Sina Weibo: The king of microblogging platforms in China.....	16
4	RUMORS AND RUMOR REFUTATION SYSTEM OF WEIBO.....	20
4.1	Overall situation about rumor spreading on Weibo.....	20
4.2	Rumor refutation mechanisms and strategies of Weibo.....	20
4.2.1	Real name registration.....	20
4.2.2	Community restriction and supervision.....	22
4.2.3	The national platform of rumor refutation.....	25
5	CASE STUDY.....	29
5.1	The background and introduction of the case.....	29
5.2	Official data about the case.....	30
5.3	Communication attribute and rumor spreading structure of Weibo.....	32
5.4	A Social science analysis.....	33
6	DATA ANALYSIS.....	36
6.1	Data preparation.....	36
6.2	Data analysis & visualization.....	38
7	CONCLUSION.....	45
	BIBLIOGRAPHY.....	47
	APPENDICES.....	52
A	The Sample of Case Data.....	52
B	Python code for data format transformation.....	53

1 Introduction

1.1 Background

With the rapid development of Information Technology, World Wide Web (WWW) has transited into the web 2.0 era which aims for web users' interaction more than just internet surfing based activities. This includes our daily use services such as blogs, microblogs, wiki, communities, social networks and so on. Each internet user could also be an internet wave maker instead of just being a surfer. Particularly, accompany with the popularization of mobile terminals such as smartphones and tablet PCs, which enable access to the internet via Wi-Fi and/or 3G/4G networks, the connection between the real life and network interaction activities has become more and more seamless which brings a profound impact on all the aspects of the society.

The overall internet development was revolutionary during past two decades in the mainland of China. The way of obtaining news from media for Chinese has been transformed and somehow subverted since microblogging emerged. The influence of this transformation even shakes the traditional Chinese political ecology. However, accompanied with the explosive development of microblogging service in China, the information dissemination on the internet has also been reformed.

Sina Weibo¹ (Weibo “微博” means microblog in Chinese), as a local hybrid of Twitter and Facebook (Persinos 2015), is the most popular and largest social network service in China while the Chinese government blocks Twitter and Facebook. It is also the most successful microblogging service platform in China regarding the amount of registered and active users and the market value. Further details are given in chapter 3.

Rumors, which was defined generally as “an unverified account or explanation of events circulating from person to person and pertaining to an object, event, or issue in public concern” (Peterson and Gist 1951), has been spreading widely as long as the social network

¹ www.weibo.com

been developing. Same as all the other social network service platforms, Sina Weibo faces to the problem that various rumors run rampant terribly. To control the dissemination of rumors, Sina Weibo has established its official rumor refutation system, including authoritative publishing accounts for clarifying rumors, and Weibo community convention for enhancing Weibo users' self-discipline. However, it seems not enough as the government tightens control over online speeches gradually. A new law concerns about the dissemination of internet rumors was proposed on 9th September 2013 and implemented on 1st November 2015. Verified internet rumors and baleful defamatory information to be retweeted more than 500 times and/or viewed/clicked more than 5000 times that the originator would face defamation charges and even be sentenced to up to 3 years' prison. This new policy has been causing an uproar and a heated argument on Sina Weibo for a period. People have various opinions since the definition of "rumors" is very ambiguous in China. It is possible for any piece of news or information which has not been released and/or clarified by official authorities or government channels to be considered as a rumor.

In this thesis, the analysis about the dissemination of rumors on Sina Weibo and relevant rumor control policies and mechanisms will be presented. An outline of information dissemination includes situation and problems in the Chinese social network will be indicated as well.

1.2 Objective and significance of the research

This research addresses the following more detailed questions: How are rumors generated? What is the relationship between rumors and rumor refutation? What is the effect of the latest rumor control policy on the dissemination of rumors and its impact on supervising the public opinion? In addition to the official rumor refutation system of Sina Weibo, during the rumor spreading, what else could we do from the technical aspect, to give an early warning when a rumor appears and/or to make an effective intervention?

This thesis concentrates on the analysis of network rumors and situation in mainland of China. Therefore, the rumor mentioned in this thesis is particularly defined as an unveri-

fied account or explanation of events circulating from person to person through the internet, in special of Sina Weibo and pertaining to an object, event, or issue in public concern. Weibo users' discussion about "Being retweeted 500 times" rumor control policy is analyzed as the case study.

This research will not limit and rest on the Weibo rumors themselves, but take the view on the solutions and mechanisms for controlling the rumors. The analysis of rumor dissemination can be the supplementary to help the research for rumor refutation. A dynamic and stereoscopic thinking through communication, public relations, social psychology and other disciplines on the analysis of rumor spreading picture on Weibo will be provided.

1.3 Methodology and structure of the thesis

The dissemination of rumors and the rumor refutation mechanisms will be the dual-core of the research, in which the former one is the research foundation for the latter one and the latter one will provide solutions for the former one.

The research method combines various quantitative analysis methods and qualitative ones, including theoretical studies like literature review and empirical case studies like data collection and analysis.

The whole thesis divides into 7 chapters. Chapter 2 consists of overall summary about previous research on rumor spreading in the social network. Theory from a sociological perspective like six degrees of separation will be introduced as part of the social network background.

In chapter 3 and 4, a comprehensive description of microblogging service development in China and the overall situation of rumor dissemination on Sina Weibo and its prior existed rumor refutation system will be represented. In addition, online information censorship system in China will be mentioned to help understand about the rumor refutation and controlling on Sina Weibo.

Chapter 5 is the study of the case “Being retweeted rumor over 500 times can land you in jail” rumor control policy. In this chapter, this controversial topic with a judicial interpretation in details is discussed.

The following part, which is chapter 6, is comprised of information on data collection and data analysis. The study will be based on the views of discussion popularity regarding to heat index, the tendency of propagation and some related qualitative analysis. A web crawler system collected all the data about the case since Sina Weibo’s official API has lots of limits on accessing the completed data. Total number of 4509 Weibo posts with keywords “rumors, being retweeted 500 times” was collected between 09.09.2013 - 17.09.2013. Various data analyzers such as python, Excel, BDP and HTML5 word cloud are used to analyze and visualize the data to output a direct and visual result.

Through above analysis, the dynamics of how rumors should be refuted and prevented on Sina Weibo and even on the internet will be discussed while the conclusions will lead to further discussion. The final section, conclusion chapter will draw all the general findings and thoughts as the overall conclusion for this thesis.

2 Previous Research on Rumor dissemination in the Social Network

“A lie can travel halfway around the world while the truth is putting on its shoes.”

- Mark Twain

2.1 The definition of network rumor

Rumor, as a universal phenomenon of public opinion, usually is spread by the way of oral transmission, of which the result and the range of dissemination are somehow limited. However, as the internet rapidly develops and popularizes, with the help of network, the dissemination of rumor is not like oral transmission limited geographically. The dissemination range becomes wider, the dissemination speed becomes quicker, as well as the dissemination route becomes more complicated. Apparently, rumor can be treated as a form of social communications.

The definition of rumor can be varied but it is often viewed as “an unverified account or explanation of events circulating from person to person and pertaining to an object, event, or issue in public concern” (Peterson and Gist 1951).

Rumors have been studied for several decades in social and psychological fields. In the social sciences, a rumor involves some kind of a statement whose veracity is not quickly or ever confirmed. The definitions of the rumors are various in research (Pendleton 1998). A piece of information can be considered either verified or unverified, based on the judgments made at the time of circulation. There are also various types of rumors including daydream ones which reflect public desires and wished-for outcomes, bogie or fear ones which reflect feared outcomes and wedge-driving ones that intend to undermine group loyalty or interpersonal relations. A rumor may also be part of a collective process explanation

or political communication strategies. When part of or all the above are concerned, particularly spread on the social network by the contribution of the internet, the social network rumors are shaped.

Allport and Postman have tried to formalize the rumor intensity in their frequently cited formula: $R \sim i \times a$; “the amount of rumor in circulation [R] will vary with the importance of the subject [i] to the individuals concerned times the ambiguity of the evidence [a] pertaining to the topic at issue. The relation between importance and ambiguity is not additive but multiplicative, for if either importance or ambiguity is zero, there is no rumor. Ambiguity alone does not sustain rumors. Nor does importance.” (Allport and Postman 1947)

Traditionally, the previous research on rumors is mostly based on the analytical view of psychology, sociology, and politics. In these areas, the research on rumors can be generalized into 4 paradigms, which are linear, psychology, sociology, and hermeneutics.

A piece of information can be considered either verified or unverified based on the judgments made at the time of circulation. The unverified information is commonly considered to be a rumor in social psychology field. Kwon et al. further divided the unverified information into three types which were true, false, and unknown, based on the judgment made after the time of circulation to make a research on the aspect of rumor spreading on a microblogging network like Twitter. The true type was when a piece of information that was unverified during circulation was officially confirmed as true after some time. This could be interpreted as information leakage, marketing, or prediction with enough reliable evidence. The false and unknown, which later in time were confirmed as false or remained unverified respectively, were what defined as rumors (Kwon, et al. 2013).

2.2 Social network rumors

2.2.1 The content of rumors

The content of rumors can vary from simple gossip to advanced propaganda and marketing material. Rumors with different content can be spread to different levels that leading to different kinds of impact from small public concern to large social panic.

For instance, one kind of the most popular gossip is celebrity deaths. People use social networks to chase and discuss about the entertainment news. So many celebrities have been the death hoax victims. World famous action movie star Jackie Chan was reported for “death” in 2015. His death rumor was spread widely on Twitter, Facebook and Weibo (Mehta 2015) (China Daily 2015). One theory about this kind of rumor is the “haters” of the celebrity launch a viral campaign against them by spreading their death rumors throughout the internet.

Different from the rumors concerning about personal attacks, widely spread rumors about politics and social living may bring large public concern and/or terrible social panic. For example, USA and South Korea will deploy a missile defense system (Thaad) in South Korea. China is opposed to this plan because of the border threat possibility (BBC News 2016). For a time, rumors about a Chinese government ban on South Korean entertainers and products to against Thaad spread widely on Weibo even though the Chinese Ministry of Foreign Affairs has denied. Folk worrying about the relationship between South Korea and China remains.

2.2.2 Rumor spreading models

Well known theory “six degree of separation” that describes any two individuals are connected through no more than six related individuals. This theory which was carried forward by Milgram’s “small world study” (Milgram 1967) was widely applied in the social network and communication study. Hence, the probability that two individuals would know

each other could be quantified. Accordingly, the information dissemination in the social network could be quantified as well.

There are some classic models about information spreading, such as epidemic spreading, computer viruses diffusing and so on. A standard model of rumor spreading was introduced and mathematically formalized many years ago by Daley and Kendal (Daley and Kendal, Stochastic rumors 1965). The Daley-Kendal (DK) model and its variants such as Maki-Thompson (MK) model (Maki-Thompson 1973), have been used widely in the past for quantitative studies of rumor spreading. In the DK model, a closed and homogeneously mixed population in the network are categorized into three groups: Ignorants are the individuals who have not heard the rumor and, consequently, are susceptible to being informed. Spreaders are active individuals that are spreading the rumor and the stiflers know the rumor but are no longer spreading it. These groups are denoted as I, S and S respectively. The rumor is propagated through the population by pair-wise contacts between spreaders and others in the population. The ISS (Ignorant-Spreader-Stifler) is the model where individuals are divided into two status, Susceptible and Infected. Assume that any spreader involved in a pair-wise contact attempts to “infect” the other individual with the rumor. In this case “the other individual” is an ignorant, it will become a spreader. In other two cases, either one or both of those involved in the contact know the rumor and decide not to tell it anymore, thereby turning into stiflers (Daley and Gani, Epidemic Modeling 2000). The dynamical behavior of the spreading process depends on how spreaders meet ignorants. In the DK model based variant MK model, as a supplement states the rumor is spread by directed contacts of the spreaders with others in the population. When a spreader contacts another spreader, only the initiating spreader becomes a stifler.

When an ignorant meets a spreader, it is turned into a new spreader with a probability. On the other hand, spreading decays due to a forgetting process or because spreaders learn the rumor has lost its new value. In the ISS model, the decaying process occurs when a spreader meets another spreader or a stifler and both contacts are supposed to have another probability. Piqueira presented a detailed study on ISS model by using concepts from the Dynamical Systems Theory. (Piqueira 2010)

Another widely applied theory to describe rumor spreading is from epidemiological SIR (Susceptible-Infected-Removed) model (Kermack and McKendrick 1927). The equations that model the problem using in SIR model and above ISS model are similar. The ignorant and spreader populations from ISS model are analogous to the susceptible and infected populations of the SIR model, respectively. The main difference is that, in ISS model, the stifler population plays a different role from the removed population of the SIR model. In the ISS model, the stifler does not propagate the rumor and its individuals remain in the system in a constant state (Daley and Kendal, Stochastic rumors 1965). While in SIR model, the removed individuals are either transformed into susceptible ones, creating a feedback loop, or excluded from total population (Kermack and McKendrick 1927).

However, as the social network develops along with the information technology, the original SIR model has limitation and deficiency on applying to describe the rumor spreading process. Researchers started to take the other factors such as “social reinforcement” into account to modify the SIR model. Wang et al. presented a Weighted CSR (Credulous-Spreader-Rational) model which took the weight of node into account, namely, assumed the node’s degree as the weight of this node. The weight of node could influence the probability of information acceptance. They assume that each person in networks adopts one of the three states at the initial time. Credulous people is similar to the susceptible one in SIR model that tend to easily believe a rumor and take delight in spreading the rumor. If someone like credulous hears the rumor, he or she will accept it with the acceptant probability and become another state-spreader. The spreader is the same as in SIR model to share the rumor to the others. He or she will spread the rumor until he or she consider all his or her friends have known the rumor or met the rational. Rational is analogous to the refractory state of SIR model, who could be described as rationalists and tend to lack faith in every rumor before he or she tests and verifies the rumor by himself or herself. Rational exists before rumor spreading. At the end of the process, all the spreaders will turn into rationals. The conservation equation of weighted CSR model is: $N * c(t) + N * s(t) + N * r(t) = N$ in which $c(t)$ means the densities of credulous, $s(t)$ means the densities of spreader and $r(t)$ means the densities of rational respectively. With the initial conditions: $s(0) = 1/N$

and $r(0) = n/N$, where N and n denotes the total individuals and the number of Rationals at initial time. (Wang, et al. 2012)

The other scholars thought the classical SIR, DK and MK models all had some weakness. They have not taken the topological characteristics of social networks into account and they were not suitable for describing rumor spreading mechanism on large-scale social networks. For example, when an ignorant contact with a spreader, the ignorant may maintain the state of ignorance or accept the rumor to become a spreader. This original assumption ignores the possibility that the ignorant becomes a stifler directly. This possibility lies in that some ignorants have a strong background knowledge, some ignorants have logical reasoning ability and some have little interest in the rumor. Another weakness is that they have not covered the forgetting and remembering mechanism. The individuals can spontaneously cease spreading the rumor by forgetting, also can spontaneously or stimulated restart spreading the rumor by remembering. As a result, Zhao et al. developed a new SIR based SIHR (Susceptible-Infected-Hibernator-Removed) model. The model extended the SIR one by considering a direct link from ignorants to stiflers with a certain probability and a new factor of People-Hibernators. The new Hibernators group which reflects the repeatability of rumor spreading comes from the spreaders due to forgetting mechanism and later becomes spreaders again due to remembering mechanism. (Zhao, et al. 2011)

Considered about the unique characteristics of microblog, another group of Chinese scholars (Liu, et al. 2014) presented a SIR modified SIRE model to describe microblog in pertinence. The stifler's broadcasting effect and social intimacy degree to the people were included in SIRE model. If the social intimacy is high, which means the pair-wise people have more trust in each other, hence the rumor is more likely to be transmitted when an ignorant has the possibility to become a spreader. The common friend number between the pair-wise persons could be one of the important indexes to measure the intimacy. Since the message flows in Sina microblog is mainly in the form of retweeting, so when a stifler receives a rumor from a spreader, his or her rejection to the rumor that mainly in the form of retweeting will be seen by all the stifler's friends. This leads to the transformation of spreaders and ignorants in becoming stiflers at certain possibility. In addition, a stifler will

not publish his reflection of the rumor until he or she receives a rumor. So, the rumor message becomes a trigger for stifler to publish clarifying message.

In all, compared to the DK and MK models about rumor spreading in the social network, the latter SIR based Weighted CSR, SIHR, SIRE, and the other similar research on the rumor spreading models are more up to date and advanced with the development of the social networks. The more possible to predict rumor spreading process, the more accurate the model is in describing the social networks such as microblog. These researches contributed to reveal the rumor spreading mechanism and provided the characteristics which could be involved in the rumor spreading and immunization.

2.3 Internet censorship in China

As we know, China is a very centralized developing country with one-party wielding the power. The censorship policy in China is special and strict. In 1998, the Chinese government started to build a Great Firewall (GFW) on internet to control the censorship and monitoring. GFW became the standing word for Chinese internet censorship since then. However, internet censorship is a complex concept which is not limited to the aspects such as IP and DNS filtering or access prevention to some resources.

Prior technical work in this field mainly focused on four dimensions which were network filtering, search filtering, chat censorship and blog censorship. With the help of GFW, networking filtering basically prevent the users to access the servers of the websites. The nonofficial project Hiking GFW² provides a up to date list of block websites by GFW. There is also a project called GreatFire³ that promotes to bring transparency to the GFW of China. It provides monitoring and testing service on blocked websites and keywords with its analyzer. For the search filtering, all the internet search engine providers which provide service in China should cooperate with the government censorship on the content which includes violent contents, erotic contents, and government related “sensitive words”.

² http://hikinggfw.org/blocked_sites

³ <https://en.greatfire.org/>

Google exited from the Chinese market and evacuated all the service from China in 2010 because of the uncooperative reaction on Chinese internet censorship policy. The latest news in this year mentioned that Google planned to return to China in the near future. Whether the scheme for the censorship in China is satisfied to the government will still be the key issue to Google (Waddell 2016). The service providers take the main responsibility on chat censorship and blog censorship as well. Instant message app such as WeChat which is the most popular chat app in China is censored by its provider Tencent in accordance with Chinese law. An exhaustive report from the University of Toronto's Citizen Lab on WeChat censorship subject tracked 36,000 unique WeChat "official" accounts which mainly were media outlets, brands, celebrities and small companies. Their posts between June 2014 and March 2015 were observed. The report found that during the period, 1.48% of the official user accounts were suspended, 1.55% of the posts made by active accounts were censored by the system. About 5% of the accounts had at least one post censored, meaning that majority of WeChat official accounts were not experiencing censorship at all (Ng 2015). Blog censorship usually binds with keywords filtering. (MacKinnon 2009) evaluated blog censorship practices of several providers, similar as search filtering and chat censorship in suppressed content, with most common forms of censorship being keyword filtering and deletion after posting.

Those prior researches indicated that the censorship in China was deeply fragmented and decentralized. Besides the national level network filtering, the censorship policy relies more on domestic companies to implement their own censorship measures as responsibility.

Moreover, the study of people's behavior on the social networks under the influence of active internet censorship could help understand the internet censorship policy in a country. To understand the online censorship characteristics perspective in China, the accessing available social networks in the mainland of China become the objects to the research on people's social network behavior since other world popular social networks such as Facebook, Twitter and Instagram are blocked in the mainland of China. As one kind of the most important and popular social networks in China, microblogs also follow the government control to censor the sensitive content.

A study (Bamman, O'Connor and A. Smith 2012) from the US presented a large-scale analysis of political content censorship in Chinese microblog Sina Weibo. It stated to be the first large-scale analysis on this topic. The study focused on leveraging a variety of information sources to discover and characterize censorship and deletion practices in Chinese social network via three orthogonal sources of information: message deletion patterns on Sina Weibo, the differential popularity of terms on Twitter vs. Weibo; and terms that were blocked on Weibo's search interface. The conclusions indicated that content censorship allowed users to publish politically sensitive messages but occasionally the messages would be deleted retroactively. A significant variation in the rate of posts which were deleted was very different from the provinces in the far west and north such as Tibet and Qinghai (53% deletion) to the eastern provinces and cities (ca. 12% deletion).

Sina Weibo has a complex variety of internal censorship mechanisms, including proactive and retroactive one. For the research in details, refer to (Zhu, et al. 2012). In general, proactive mechanisms may include Explicit filtering, Implicit filtering, Camouflaged posts and Surveillance. Retroactive mechanisms for removing the content that has been released may include Backwards reposts search, Backwards keyword search, Monitoring specific users, Account closures, Search filtering and Public timeline filtering. When Weibo discovers the messages, which contain sensitive content have passed through the filters listed above for proactive mechanisms, Weibo will stop the "viral" propagation of such content. Removal of sensitive content may still be included in the implementation. Besides the proactive and retroactive mechanisms, Weibo also announced a "user credit" system that users could report sensitive or rumor-based posts to the administrators. The "user credit" system is introduced in detail in chapter 4.2.2.

In addition, the scholars from The University of Hong Kong made specified research on the censorship practices of Sina Weibo (Fu, Chan and Chau 2013). They collected 111 million Weibo posts between 1- June 30, 2012, and deployed a matched case-control study which was designed to determine a list of Chinese terms to discriminate the censored and uncensored posts when they were written by the same Weibo users. The list also included some homophones and puns created by Chinese Weibo users to circumvent the censorship. Meanwhile, the research also evaluated the impact of real name registration system which

was introduced in chapter 4.2.1 on Weibo users' posting activities. The findings concluded that the real name registration policy might have stopped some users from writing about social and political subjects. Weiboscope⁴, a data collection and visualization system which uses Sina Weibo Open APIs (User Search API and User Timeline API) was developed by the research team above could be accessed (not from mainland of China) to get their research data and the latest censored messages of Weibo. User Timeline API is a stream API so the data they collected is up to date and the censorship index provided on the website is in real time.

⁴ <http://weiboscope.jmssc.hku.hk/>
<http://weiboscope.jmssc.hku.hk/webpagegrid/aboutushtml.html> (Introduction of the system)

3 The Development and Success of Weibo in China

3.1 The internet, social media, microblogging and Chinese Weibo

Following the successful development and popularities of Facebook and Twitter world widely, China has also been striding forward into a social media era. A country with such large population regarded as user base, has huge potential in the market of social networks. As the report of digital 2016 (WeAreSocial 2016) indicated, there were 668 million active internet users in China, in which 659 million users were active social network users. There were 675 million unique mobile users, in which 574 million users were active mobile social apps users. Meanwhile, the annual growth in the number of active internet users was 6% (+36 million) while the annual growth in the number of active social network users was 2% (+14 million) and annual growth in the number of active mobile social apps users was 15% (+77 million).

Among these millions active social network users, login frequency to the platforms that at least once in the most recent 30-day reporting period ranked as following: QQ, QZONE, WeChat, Baidu Tieba, Sina Weibo and YY. Sina Weibo was the only microblogging platform. In the reporting period, there was 54% of the internet users who claimed to have used Sina Weibo in the past month.

Microblog, a kind of social media as an example of Web 2.0 technology has stood for the modern communication characteristics of “anytime, anywhere, anyone and anything”. It provides a platform for each netizen to record, report, share and react with anything in people’s daily life. It also brings a worldwide revolution about the ways how people get the information instantly compared to the traditional media and/or early-staged internet web portals. So as in China.

According to the article named “Which tongues work best for microblogs?” from (The Economist 2012), Chinese was illustrated to be the most ideal language for microblogs, which typically restrict messages to 140 symbols. It gave the example, which was:

“THIS 78-character tweet in English would be only 24 characters in Chinese. = 这条七十八个字的英文推如果写成中文只有二十四个字。”

Moreover, the message could be even shortened into 16 characters in Chinese which would be 此七十八字英文推若用中文仅十六字. One Chinese character is 2 bytes while one English character is 1 byte (1 byte = 8 bits) regarding to the standard of ASCII (American Standard Code for Information Interchange). Thus, the above example Chinese sentence is total 48 bytes which are 30 bytes shorter compared to the original 78 bytes' English ones.

That principle makes Chinese ideal for microblogs, which typically restrict messages to 140 eight-bit characters. The Chinese language is succinct that most posts may never reach the limit. Most of the information that people wish to publish can be ideally expressed within 140 eight-bit characters. The language factor has laid the foundation for the explosive development of microblogging service in China. For Chinese microblogging service providers, which usually limit the post to 140 Chinese characters, people who use Chinese to tweet can even express more completely.

3.2 Sina Weibo: The king of microblogging platforms in China

Weibo which is used to name Sina Weibo is unquestionably the most popular microblogging service in China. It was officially launched in August 2009. The year of 2010 was marked as the first year of development of Chinese microblogs. Besides Sina corporation, the other three largest Chinese web portals⁵ had also launched their microblogging service platforms. However, during a few years' competition in microblogging market, Sina Weibo has held the palm while the others either shut down the services or transformed the service form or even quitted from the market competition. There were a lot of discussions and many articles in China described this business competition, as well as made the com-

⁵ Four largest web portals in China: Sina Corporation, Tencent Holdings Limited, Sohu, Inc. NetEase, Inc. according to Baidu Baike Index

parison for all the main microblog service providers in China and emphasized the successful factors of Sina Weibo. List of the main factors are summarized here:

- 1) Sina firstly established the microblogging service platform in China and got a certain users base, in which the active users' circle was shaped earlier.
- 2) Sina replicated the previous successful experience of its blogging service. For example, attracted and promoted celebrities and opinion leader users to amplify the impact on the platform.
- 3) Sina constantly improved the platform and made the innovation to enhance the user experience.
- 4) The other competitors didn't have unique selling point which could be fundamentally different from Sina Weibo.
- 5) Sina Weibo was in line with the national circumstances of China. It was open but politically not against the government.

Following the successful development of a "we-media"⁶ trend, Weibo has become an important and powerful one in China step by step.

As a social media platform and a social network service provider, Weibo is for people to create, share and discover Chinese-language content. Weibo combines the means of public self-expression in real time with a powerful platform for social interaction, as well as content aggregation and distribution. Any users can create characters and post a feed and attach multimedia or long-form content as picture format. Weibo has removed the limitation of 140 Chinese characters for posting a feed since 28th February 2016 (Millward 2016). Users can post a feed up to 2000 Chinese characters as text format. This would give the users as much freedom and convenience as they could spend to express on Weibo. Howev-

⁶ We-media is a concept introduced by Dan Gillmor in his 2004 book "We the Media," which is about how the proliferation of grassroots internet journalists (bloggers) has changed the way news is handled. Nowadays, we-media or self-media is commonly referred to the phenomenon of grassroots journalists via social media.

er, whether the word limitation affects the rumor dissemination that need further studies. This factor will not be considered in empirical part of the thesis.

Users' relationships on Weibo may be asymmetric. Any user can follow any other users and add comments to a feed while retweeting or not. This simple, asymmetric and distributed nature of Weibo allows an original feed to become a live viral conversation stream. The communication attribute is introduced in detail in chapter 5.3.

The inception of Sina Weibo was in 2010. In its first 66 days' operation, Weibo had amassed one million users. By the time of the eighth month since operation, this number had reached 10 million (Farley 2011). In December 2014, Weibo had 175,7 million monthly active users (MAUs) and 80,6 million average daily active users (DAUs), increasing from 129,1 million average MAUs and 61,4 million average DAUs in December 2013, and 96,7 million MAUs and 45,1 million average DAUs in December 2012. According to the financial report of Weibo for the first quarter 2016, in March 2016, average MAUs was 261 million with the annual growth of 32% and average DAUs was 120 million with annual grew of 35%. Meanwhile, the annual growth of mobile DAUs was 45%. Mobile DAUs reached 91% of the total DAUs, which highlighting Weibo's traffic momentum arising from high mobile usage and penetration beyond tier 1 cities. (Weibo IR n.d.) Nowadays, when you mention the social networks in China, Weibo has been a very indispensable name for the field. The fall of traditional media and the rise of new social media have made Weibo become an important force of public opinion which could not be ignored.

Like a microcosm of Chinese society, Weibo has attracted a wide range of users, including ordinary people, celebrities, and other public figures, as well as organizations, such as media outlets, business, government agencies and charities. Chinese internet development during the last few years cloud be characterized by the growth of Weibo. It has led the Chinese public communication entered a "Weibo era".

Supported and promoted by Sina, Weibo has been one of the most preferred and reliable information sources for many Chinese netizens during its development procedure. In the early-stage, the most popular daily topics on Weibo were generally entertainments news, gossip, commercial information and sports. However, people were using Weibo more and

more on getting news, shaping public opinions and political sphere since Weibo performed instantly on so many public affairs. One of the impressive successes in shaping public opinion on Weibo could be that many influential incidents were first exposed through Weibo. For example, the Wenzhou terrible train accident in 2011. It was the remarkable incident for the public and Weibo itself as well. The first news and instant report were from Weibo users. Within just a couple of hours, huge concern, discussion and reports were shaped on Weibo. People shared and commented on every related information, and suspected the official report on the number of victims and the rescue operations from the government. Weibo became the one and only media platform for the ordinary people to follow the latest news about that incident and the expectation to seek for some “truth”.

Weibo has been performing well on incidents exposure and instant reporting on latest news continually. Many Weibo exposure incidents quickly attracted the public attention and being widely discussed. Moreover, aware of the unique social media feature, rapid development of Weibo platform and its leading instant information sources for the news, more and more traditional media and other emerging media join in Weibo by opening their official accounts. As new media development report of China 2016 indicated (Chai 2016), by the end of August 2015, there were 24,259 verified media accounts, of which 17,323 ones were the traditional media such as local newspaper, magazines, television stations and radio stations etc.

Weibo also attracted the Chinese government. According to the internet public opinion report 2015 (People Yuqing 2015), by the end of September 2015, there were 150,131 verified government accounts on Weibo, including all the regional or local government at different administrative levels (ca. provinces, autonomous regions, municipalities, special administration regions and cities) in China. About 112,602 accounts were the official government agencies and 37,529 ones were the government officers.

Evidently, after the firm foundation from 2009 to 2011 and the rapid development from 2012 to 2013, Weibo has become one of the major communication platforms and a mature and leading social media in China.

4 Rumors and Rumor refutation system of Weibo

4.1 Overall situation about rumor spreading on Weibo

Tsinghua National Laboratory for Information Science and Technology (Liu, et al. 2015) has collected and analyzed the number of 9079 rumor terms on Weibo from August 2011 to May 2013. The category of rumors could be mainly divided into 5 topic types which were political rumors, economical rumors, cheat rumors, social living rumors, and commonsensible rumors. Political rumors and social living rumors took most of the ratio, totally around 70%. Even though the commonsensible rumors took little of the ratio which was 10%, they were the ones which appeared repeatedly in the same or similar formation every now and then.

Guowei Cao, who is the CEO of Sina Corporation, has been interviewed in September 2012 that indicated the containment of rumor dissemination and business model for Weibo would be the first two and the most important tasks for future development. Rumor dissemination has been one of the top problems for Weibo significantly. As Weibo develops, the huge flow of information with more instant dissemination and freer speech make a hotbed for rumor dissemination. Terrible rumors can even make huge social panic and unrest. Rumor refutation work is a long-term strategy work for Weibo.

4.2 Rumor refutation mechanisms and strategies of Weibo

4.2.1 Real name registration

In 16th March 2012, Weibo launched its real name registration system. All the old users who registered before this date should complete and submit their real name ID information (national identity number or passport number). Otherwise, they might not be possible to post, comment, and retweet but just review. All the new users would be required to submit their real name ID information when registered. Otherwise, they could not register success-

fully. This is background real name registration, which means the users' real name information will not be published on their profile but saved in Weibo system. This policy was believed as the response to the promotion of internet real name registration proposal by the National State Department of China.

However, Weibo considered the real name registration system more as a means of controlling the false information and rumors. In the internet era, everybody can be a "keyboard hero or offender" in virtual communities with just typing words from computers or smartphones, but take no responsibility for the actual life. Weibo believed the real name registration system would contribute to reduce the amounts of irresponsible messages and rumors.

(L. Wang 2013) collected 90 rumor terms from the first half of 2011 and 108 ones from the first half of 2012 to make a same-time comparison and found out the following characteristics regarding to the real name registration system.

- 1) Real name registration system has effectively controlled the tendency of rumors growth.
- 2) Originating rumors increased significantly while retweeting ones were restrained steady.
- 3) The tendency of authenticated users' intervention on rumor spreading was significant. If the rumor was posted or retweeted by the authenticated users, the rumor spreading promoted rapidly.

According to this research, the real name registration system has made some effort on constraining the rumor spreading, but has failed to restrain the emergence of rumors. In particular, the increase of originating rumors and the intervention of authenticated users on rumor spreading have propelled the emergence and dissemination of the rumors contrarily.

As mentioned in chapter 2.4, the censorship study of evaluating the impact of real name registration system of Weibo from Hong Kong scholars (Fu, Chan and Chau 2013) also found evidence to support that the real name registration might stop some specific Weibo users from continually posting, especially writing about social and political subjects. They

suspected if the real name registration system were enforced, it might exert a chilling effect selectively on some Weibo users who are characterized as stationed inside Mainland of China, with a non-authenticated identity, open for users to comment, and less online networked. This kind of users might inhibit or discourage to post rumors by the threat of legal sanction. When comes to the social and political issues, it may be more sensitive and possible to generate rumors on the internet. Real name registration system might generate some constraining force on the usage of Weibo, but the certain impact on rumor dissemination was not covered in this research.

In all, even though the real name registration system was not 100% effective in controlling the rumor dissemination on Weibo, it still made contribution and advisable outcomes which were worth to be discussed and researched on the management of online communities.

4.2.2 Community restriction and supervision

Following the real name registration system, Weibo published its community convention (trial)⁷, community management regulations (trial)⁸ and community committee system (trial)⁹ in 28th May 2012. The community treaties were drafted by Weibo and its users through an online request for proposal.

Based on the community reporting and monitoring system, Weibo introduced a community code of practice (CoP) and developed its own official rumor-busting service platform. The since-established CoP Community Center with a demerit point system has been set up to punish users who violate its rules. The rules prohibit the distribution of commercial spam, indecent affairs, false information and rumors.

⁷ <http://service.account.weibo.com/roles/gongyue>

⁸ <http://service.account.weibo.com/roles/guiding>

⁹ <http://service.account.weibo.com/roles/zhidu> (only in Chinese)

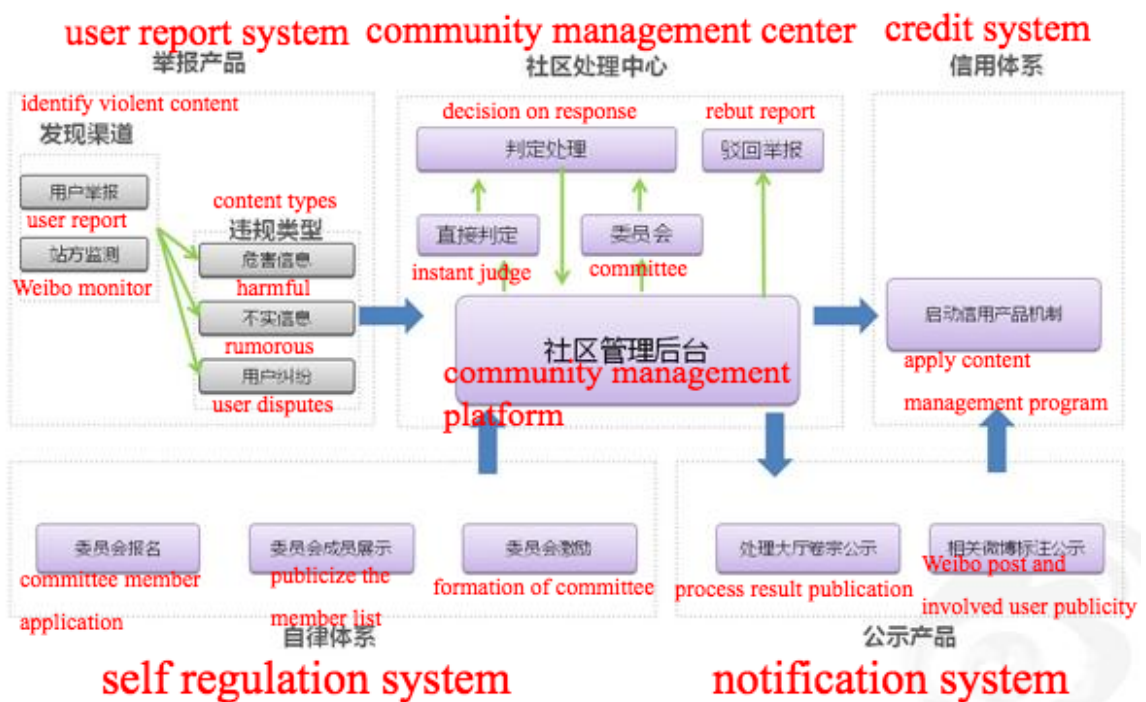


Figure 1 The structure of Sina community content management platform

Cited from the Sina news¹⁰ with English version report by Lam (Lam 2013), figure 1 displayed the structure of Sina community content management platform. The CoP community demerit point system is enforced by reports from users and processed by community management center, in which a committee review the disputed reports, and results in credit and notification systems.

In the demerit point system, each user begins with 80 credits. When the user's credits are reduced to 60, he or she will be considered as a low credit member and be subjected to certain restriction in the usage of Weibo services. Low credit members are possible to return to 80 credits if they have no anti-regulation activities for the following 2 months. When the user's credits are dropped to zero, the account will be deleted or blocked perma-

¹⁰ <http://news.sina.com.cn/c/2013-07-04/183527578723.shtml> (in Chinese)

nently. They user with real ID information will be on the blacklist that is impossible to register account with the same ID again.

The CoP community center has already recruited 19,077 people to review the disputed reports by 30th September 2014. Team member recruiting only by invitation from the CoP community center. There is “Report” button on every Weibo post and comment. Users can click the button or tag any of the official Weibo administrative accounts “@weiboxiaomishu” (@微博小秘书) or “@weiboguanliyuan” (@微博管理员) to report to the CoP community easily. Whenever a disputed report alleged to contain a rumor, the center will select a subcommittee who is usually a professional in the certain field to review the disputed case and give a judgment. If the “defendant” loses to the verdict, he or she may face to punishment such as impossible to tweet or retweet for various days, Weibo account blockade or possible legal liability besides false information deletion. Once a rumor is verified by the CoP community center, it will be deleted or marked with a label of “rumor” immediately. A notification to all the users who have involved in the dissemination of the “rumor” will be published with labelling the deleted content as a rumor. The whole dealing procedure usually takes 1 to 3 working days. Terrible disputed reports such as suspected rumors usually are processed in 1 working day.

In one year since the community management regulations have been executed, the CoP community center has received and handled more than 15 million users’ reports., in which 12 million was commercial spam, 1 million was indecent affairs and 2 million was about spreading rumors. The CoP community center has been keeping the handling rate above 95% of the users’ reports since August 2012. The daily report number had been reduced by 87% from 4,000 in May 2012 to 500 in May 2013.

According to the monthly work report of CoP community center (Weibo Community Mangament Center 2014), the false information reports kept at a low rate of averagely below 1% for all the disputed reports from November 2013 to September 2014. Referred from jury system, with each user’s mutual restriction and supervision, the CoP community system has made a potent effort on controlling the dissemination of the false information on Weibo.

4.2.3 The national platform of rumor refutation

Chinese Ministry of Public Security and Weibo jointly launched the online platform for rumor refutation in May 2016. It would observe the online rumors' related issues and release monthly report about the work on verifying and refuting rumors. The related data would be published as well.

The platform allows netizens to provide links or upload screenshots with the location of the alleged false information on any social media platforms, including Weibo and other online forums. The 189 internet police patrol accounts with verification and every local police official account will be the main authorities to accept users' report, check and publish the correct information. Besides directly submit to the platform, every single user can either report to those authorities' accounts or write feed and/or comments or retweet the certain information with hashtag “#Weibo rumors refutation#” (#微博辟谣#), will automatically connect to the official Weibo account of the platform and be reviewed by certain staffs to give a judgment. In addition, the platform will collect and publish the most spread and verified rumors and related refutation information in aperiodic time through its official Weibo account “@weibopiyao” (@微博辟谣).

The topic with the hashtag “#The National Platform of Rumor refutation#” (#全国辟谣平台#), has been read more than 620 thousand times on the press day and over 2 million mentioned discussions in 2 months. This is the first national level platform to fight against online rumors in China.



Figure 2 The capture of platform entry on users' homepage on Weibo

Figure 2 shows the entry of the platform¹¹ on the bulletin board which is placed as the sidebar on every user's homepage of Weibo. Users can find easily and access instantly to the platform.

¹¹ <http://service.account.weibo.com/derumor?rightmod=1&wvr=6&mod=noticeboard> (in Chinese)



Figure 3 The capture of display example for the account @weibopiyao and hashtag #Weibo Rumors refutation#

Figure 3 is the display example that the official Weibo account of rumor refutation platform @weibopiyao publishes the rumor refutation topic. The example is the list of serious spreading rumors published by official account of Sina Zhejiang (a local Sina media account of Zhejiang province) from 18th July 2016 to 24th July 2016. The official Weibo account of rumor refutation platform retweeted it with a hashtag of “#Weibo rumors refutation#”.

According to the first monthly work report (Weibopiyao 2016), the platform received 480 reports, verified, refuted and published 158 rumor messages from 21st April 2016 to 20th May 2016. The topic with the hashtag "#The National Platform of Rumor refutation#", has been read totally 210 million times until 20th June 2016.

Generally, it is still early to say whether this rumor refutation platform can be an effective “rumors disintegrator”. Long-term observation and data collection are needed. Similar to the community restriction and supervision, users’ participation and official refutation’s credibility are very important on building a right and effective rumor refutation platform. In case the rumor involves the polices, the ministry of public security, the politicians, and other government authorities, whether the platform can give instant, fair, transparent, reasoned and convincing judgment is the key factor affects users’ participation into the development and the reliability of the platform.

Above all are the current refutation mechanisms against rumors on Sina Weibo. Obviously, Sina has been continuously making effort on fighting against online rumors to do whatever it can as a social network platform. But as the rumors constantly emergence, there is always a long way to go.

5 Case Study

5.1 The background and introduction of the case

Besides all the mechanisms presented in the previous chapter, there was another one promoted by the government could not be ignored. China's supreme court was bound to issue a new supplementary law on 9th September 2013 and implemented in details on 1st November 2015 to regulate the speeches on the social networks. This law indicated that any defamatory or rumormongering posts or messages on the social networks would be considered "severe" breaches to the law if they were widely retweeted more than 500 times and/or viewed/clicked more than 5000 times. Those found guilty could be possible to face up to three years' imprisonment.

The law was the most serious attempt to control the dissemination of online rumors and baleful fake news and increase people's sense of responsibility about their online speeches. It also intended to crack down the "black public relationship (PR) firms" which made money from removing unflattering information and/or creating false information and/or misdirecting the public opinion on the internet. Their activities may involve huge business chain. "Black PR firms" usually control lots of verified Weibo accounts with a huge number of followers or audiences which are not hundreds or thousands but millions and even hundreds of million. With this huge scale of promotional reach, they can easily make profits that reach millions Chinese Yuans (CNY, 1CNY \approx 0.135EUR)¹².

However, the ordinary Weibo users' opinion on this new law might be diverse. In general, some users considered it as a powerful measure to control the rumor spreading and make a more trustable environment for online information dissemination. People were delighted with the net polices cracked down those "Black PR firms". The others thought it was just government's critical censorship on free speech while the rest worried it might bring false accusation if it was not exerted correctly. Certainly, there were some people who had no comments but kept looking on it.

¹² XE Live Exchange Rates www.xe.com

The following case study will go deeper into the topic about this new law and try to find out people’s opinion regarding to rumor spreading and rumor refutation mechanism by collecting and analyzing the data with the keyword of “rumors, being retweeted more than 500 times”.

5.2 Official data about the case

The official data center of Weibo provides index search engine for hot words from 2013, in which we can put words of “rumors” “conviction” “500 times” into search. Figure 4 shows the discussing tendency from 09.09.2013 to 30.10.2013, in which the blue line stands for the word “rumors” while the red and green ones stand for the word “conviction” and “500 times”. Figure 5 gives the data explanation about the tendency curve, in which the first column shows the name of words, the second column stands for monthly average mentioned time for the words and the last column stands for the monthly high peak value for the words be mentioned.

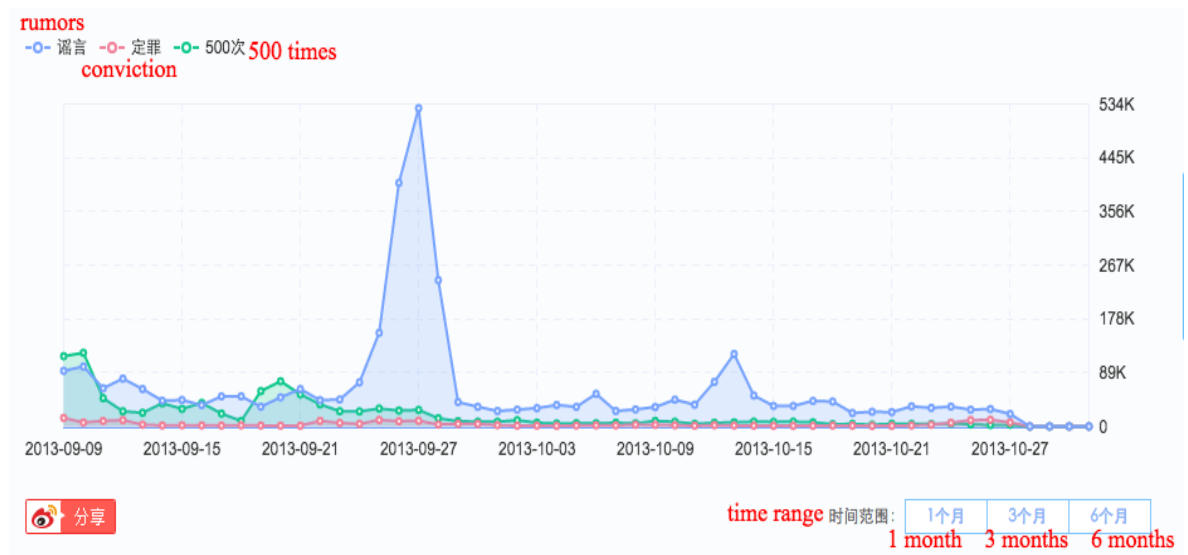


Figure 4 The discussing tendency of the certain words from 09.09.2013 to 30.10.2013

热词的相关趋势 ?	monthly average	monthly high peak
谣言 rumors	热议指数: 当月均值: 14596	当月最高: 81815
定罪 conviction	热议指数: 当月均值: 1278	当月最高: 7413
500次 500 times	热议指数: 当月均值: 428	当月最高: 1377

Figure 5 The capture of data explanation for figure 4

However, the official index search platform doesn't contain the combined words "rumors, being retweeted 500 times" in the searching engine. Those three lines are separately. We cannot simply mark the most overlapping point regarding to the "rumors, being retweeted 500 times" topic. So the initial discussing situation about this topic is not clear from the data on this index search engine. Therefore, to understand more specified impact on the "500 times" policy, with the help of web crawler system, Weibo data with combined keywords in Chinese of "rumors, being retweeted 500 time" ("谣言转发 500 次") for the first week since the policy was published were collected. The data include content texts about the topic, URL and ID number of the Weibo post, be retweeted Weibo ID, retweet time, count of comments, count of retweet and like, post sources of terminals, URL of the pictures or videos etc.

In addition, there are plenty of access ways for Weibo, including the web page www.weibo.com, the mobile page m.weibo.cn, desktop client terminal for Windows system, mobile app for Android, iOS and Windows Phone. To specialize the research, the data collection is mainly based on the access of web page www.weibo.com.

5.3 Communication attribute and rumor spreading structure of Weibo

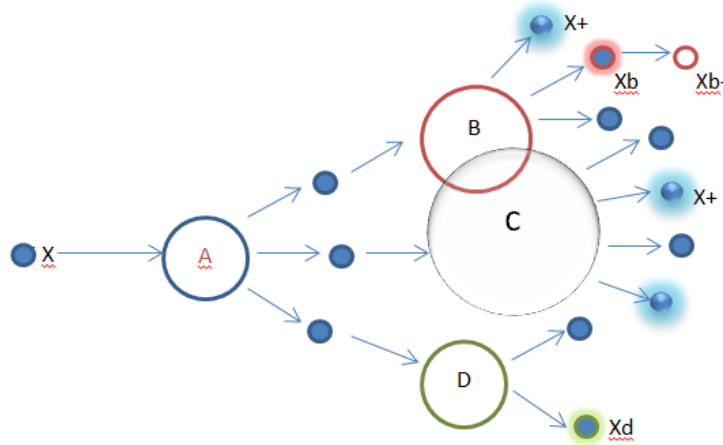


Figure 6 The communication model of Weibo
(Zhan and Gianluigi 2013)

As represents in Figure 6, each Weibo user has their own followee and follower group. The Weibo network is a directed connected structure, in which the relationship between two Weibo users (nodes) are unequal. A Weibo post X (the solid small node) origins from user A or be intaked and published by A may reach to A 's followers such as B , C and D . User C may share some followers with user B but not have any common followers with user D . Meanwhile, user B , C and D may continue spread or share some message with their followers and the followers share again to their own followers. However, during the dissemination progress, after several layers of group sharing, the message may reach the mass that could be totally different from the original one. For example, user B may retweet the original message X from user A by adding a picture or some comments so the message X becomes $X+$ or Xb . User B 's followers may receive the message $X+$ or Xb instead of original one X . In case the original message X be censored, user B 's comments remain on the plat-

form so user B's followers may continue to receive the message Xb as Xb-. Then the real information becomes more and more blurry as the process continues.

Generally, this communication attribute makes the principles of rumor spreading on Weibo. Besides the reality and property of the original message, multilayer transmission structure is also the reason for the information lost and change so that people's transmission willingness and public opinion will be affected.

In the structure of information network topology, there is the requirement on load balancing. Every single node only affords a certain amount of data. When the data load of this node exceeds its capacity, this node will share some data to the other nodes to keep the balance for information load on the whole network. In this analogy, for the information load of network rumors, when a person (single node) is unable to afford the outcomes of rumors such as anxiety and panic, he or she will unconsciously share or transmit some information to the others to reduce his or her own suffering. "The others" may refer to the multilayered followers whom can receive, understand and even comment and react to the rumor information.

5.4 A Social science analysis

There is a specific research paper on "500 times" topic (in Chinese). "Can punishment prevent rumors from spreading? -Take "being retweeted over 500 times can be sent to the jail" as example" by Yan Xiong made two times questionnaire investigation on people's reaction to the policy of "being retweeted 500 times". The investigation was last for 3 months from 05.06.2013 to 21.11.2013, with final 149 result questionnaires before the policy launching day (09.09.2013) and 151 ones after the launching day. (Xiong 2014) The research was based on hypothesizes. The Penalty Execution Variables, Incentive Variables on rumor spreading and Variables of Rumor transmission willingness were used to measure the results. The Penalty Execution Variables was set to 0 before the policy and 1 after the policy. The Variables of Rumor transmission willingness were calculated according to the question "Are you willing to share it to the others?". The Incentive Variables on rumor

spreading were calculated according to the questions such as “Is this rumor benefit for yourself?” “Is this rumor containing anti-government issues? If yes, how much do you think it contains?” “Is the rumors source professional or coming from authorities?” “Is it difficult for you to verify the rumor?” and so on. The answering scale for the questions were from 0 to 10.

By analyzing the samples which were 300 questionnaires, the results to the hypotheses were presented. The interviewees’ gender, age, and income were covered in the investigation. The research results could be generalized as following:

- 1) There was no significant relationship between the Penalty Execution Variables and Variables of Rumor transmission willingness that proved punishment was not a necessary condition nor a sufficient condition for suppressing the spread of rumors.
- 2) After the policy launched, the amount of anti-government related rumor and its terrible level were less and lower than the ones before the policy. This change would possibly reduce the rumor transmission willingness of social living related rumors towards to the predicted variable. It proved that punishment could affect to suppress the spread of rumors.
- 3) The implementation of the policy significantly enhanced the difficulties of rumor verification. This change would possibly increase the rumor transmission willingness of popularity awareness related rumors towards to the predicted variable. It also proved that punishment could affect to promote the spread of rumors. When the rumor came from “key opinion leaders” or “authorities” whom people believed, as well as widely discussed or awared, people would contrarily spread rumors actively.
- 4) The verification of the original source, the popularity awareness, benefit involvement and social living relevance were the sufficient conditions for controlling the spread of rumors. To reduce or lower down the above factors would efficiently suppress the spread of rumors.

The results of this research provided an overall investigation on the relationship of rumor dissemination and its specific punishment policy from the view of social science. It will certainly inspire the direction of data analysis on the same topic from Weibo.

However, there is always the limitation for sample investigation. So is in this paper. Firstly, the scale of the sample is small. There is a large and multi-ethnic population in China. The sample should also cover the people from different regions and ethnic groups. Secondly, education background is also a very important factor which may affect a person's rumor transmission willingness. The results may be varied if education background has been considered. Finally, as mentioned in the beginning, newly emerged media like social network has taken a significant position on the channels and sources which people get the information from. According to the researches on rumor spreading models, an ignorant may be possible to become rational. A person who usually use the social networks and get the information flow faster than from the traditional ways, and/or may view or be noticed for some rumors and rumor refutation, possibility will not spread similar rumors. Therefore, people's usual information receiving methods or sources may also affect the attitude of rumor transmission. Thus, the research on Weibo and this sample investigation will supplement each other on better understanding the case.

6 Data Analysis

6.1 Data preparation

Data from Weibo needs to be obtained via login. Every user should have an open Weibo account (could be same as normal Weibo account) to access the open (developer) service, including applying to use Application Programming Interface (API) for collecting the data from Weibo. However, there are strict limitations on the use of API. For example, for each server IP, the rate-limit for API access is 1000 times/hour for test authorization, 10000 times/hour for ordinary authorization, 20000 times/hour for middle-level authorization, 30000 times/hour for high-level authorization and 40000 times/hour for cooperated-level authorization. For each user account when using a single application, the request limit for API access is 150 times/hour for test authorization, 2000 times/hour for ordinary, middle-level and high-level authorization, unlimited for cooperated-level authorization. (Open Weibo 2013) In addition, the API for search/topics will be returned the latest 200 posts under the certain search topic with the hashtag as result. But this function is only available for high-level authorization users and above. (Open Weibo 2012)

Usually, Weibo confers ordinary authorization to ordinary developers who apply the access token for API. Middle and large-scale companies or research institutions can be issued middle to high-level authorization via application. Only cooperated-level partners with Sina group are possible to get the cooperated-level authorization. Hence, as an independent student researcher, it is difficult to get high-level authorization and almost impossible to get cooperated-level authorization to use the search/topics API. Thus, it is impossible to use API to get the data with keywords. Certainly, we can search the keywords directly on Weibo platform which provides a search interface. But only 20 pages' results with the latest posts (20 posts per page at most) will be displayed. It is also impossible to collect data from this way. Therefore, the web crawler program is the only way to get the data about this case topic from a certain period. The crawler program mimics a browser to the search interface of Weibo where we could fetch the past data on the server with keywords other than APIs.

Due to the past data requirement in 2013 and personal lack of sufficient web crawler technology, the data was collected with the help of big data open platform community (Big Data OPC¹³) which was a united research team founded by various Chinese universities and research institutes. The research group had their own developed distributing crawler system and various slave servers. So the data can be persistently crawled. A group of Weibo posts with combined keyword “rumors, being retweeted 500 times” (“谣言转发 500 次”) were collected from 09.09.2013 to 17.09.2013. The metadata was collected in March 2014. So the data were the posts still available on the Weibo server during the time 09.09.2013 - 17.09.2013 when we collected.

The results came in separate files for each hour per character string in JSON format. Figure 7 was a metadata sample in JSON format, including the items of content texts about the topic, URL and ID number of the Weibo post, be retweeted Weibo ID, retweet time, count of comments, retweet and like, post sources of terminals, URL of the pictures or videos etc. There were total 201 JSON files. Each one has the file name similar like this: “1395738762_1_500_20130909120000_20130909130000_1_x139573876310412100okws.json”. The string numbers of “20130909120000_20130909130000” (and other similar ones) indicated that this JSON file contains all the Weibo posts with keywords "rumors, being retweeted 500 times" published between 12:00 to 13:00 on 09.09.2013.

In order to be analyzed properly with the selected software, the dataset was transformed into one single file in CSV format so all the posts could be listed regarding to their publishing time. The transformation was done with Python, for which the code could be found in the appendices.

¹³ <http://bigdataopc.ihep.ac.cn> A research group formed by The Institute of High Energy Physics of the Chinese Academy of Sciences(IHEP), The Institute of Acoustics of the Chinese Academy of Sciences (IOA), Beijing University of Posts and Telecommunications, Beijing Jiaotong University, Sichuan University etc.

```

[
  {
    "beCommentWeiboId" : "",
    "beForwardWeiboId" : "3622436580427766",
    "catchTime" : "1395739167",
    "commentCount" : "0",
    "content" : "//@a没有你的城市:央视这叫揣着明白装213,无耻,颠倒黑白,散布谣言!假如转发500遍就是造谣,央视早就被关一万次了//@假面孔潘先生:央视太213了!!!!",
    "createTime" : "1379147392",
    "info1" : "",
    "info2" : "",
    "info3" : "",
    "mlevel" : "",
    "musicurl" : [],
    "pic_list" : [],
    "praiseCount" : "0",
    "reportCount" : "0",
    "source" : "新浪微博",
    "userId" : "2055888371",
    "videourl" : [
      "http://news.cntv.cn/2013/09/14/VIDE1379127361281509.shtml"
    ],
    "weiboId" : "3622469078046577",
    "weiboUrl" : "http://weibo.com/2055888371/A9qwrXLhv"
  }
]

```

Figure 7 The sample of metadata in JSON format

6.2 Data analysis & visualization

The dataset contains 4509 Weibo posts in total after cleaning and filtering in OpenRefine¹⁴. PaaS business data intelligence analysis and visualization system BDP¹⁵ is the platform used for the data analysis and visualization.

As represented in figure 8 and 9, the discussion popularity varied time by time. Figure 8 was indicated by the count of comments regarding to the time while figure 9 was indicated by the comparison for the count of comments and retweets regarding to time. The green line indicated the count of comments while the yellow one indicated the count of retweets. The first highly discussion period was between 3 pm - 4 pm on 09.09.2013. High peak was

¹⁴ <http://openrefine.org/>

¹⁵ <https://www.bdp.cn/home.html>

between 8pm - 9 pm on 10.09.2013. There were also a few high peaks during the day of 11.09.2013. (Catch time displays in unique format which needs to be checked and referred to the original JSON file name to translate into normal time format.) The first three days since the policy launched were the highly discussing period.

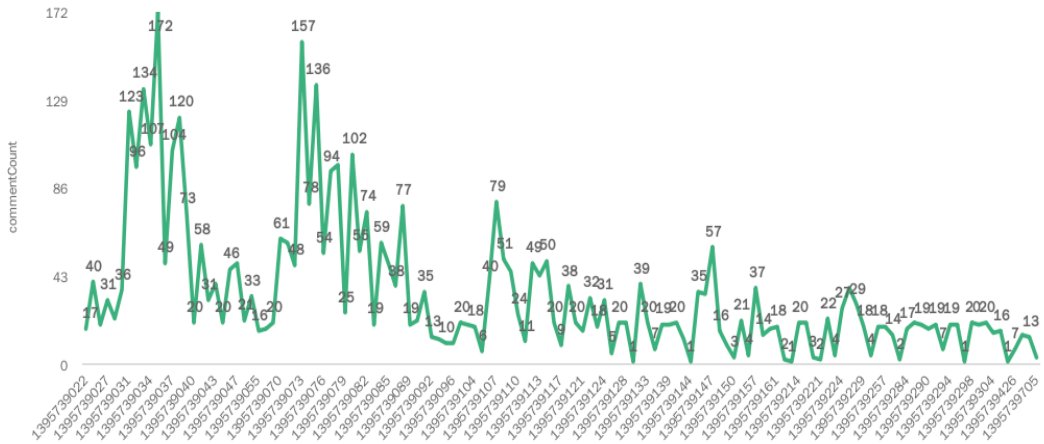


Figure 8 The discussing tendency indicated regarding to the time and count of comments

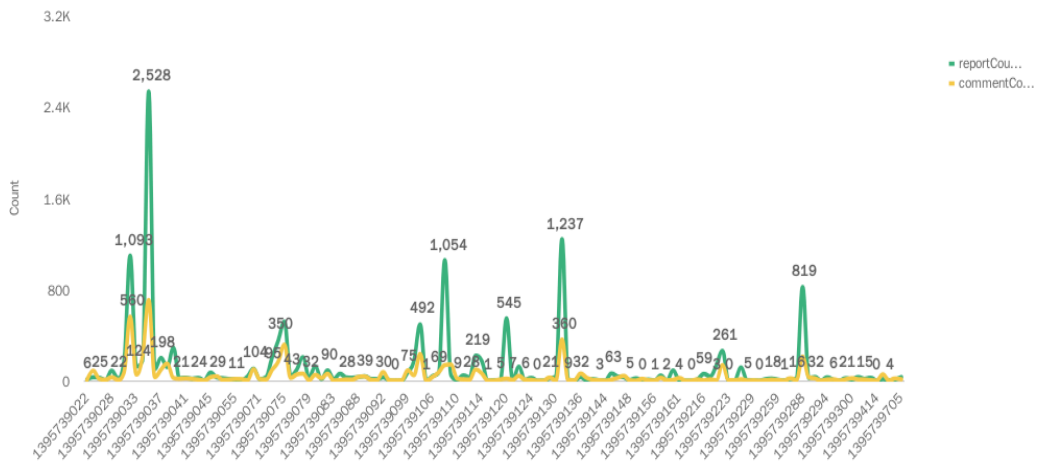


Figure 9 The discussing tendency indicated regarding to the time and comparison on count of comments and retweets

When compared to figure 4 and 5, the discussion about “Rumors, being retweeted 500 times” decayed by time generally. In addition, compared to the other hot topics on Weibo which would have billions of times of viewing, hundred thousand times of comments and retweets, the discussion index of the topic “Rumors, being retweeted 500 times” was far less.

For the discussion contents, a simple Chinese language lexical analysis applied to create the list of most mentioned words. HTML5 Word Cloud¹⁶, a browser-based HTML5 word frequency analyzer and word cloud maker which is developed by Timothy Guan-tin Chien and other contributors. It supports the lexical analysis of both English and Chinese. Word cloud of “Rumors, being retweeted 500 times” showed in figure 10, while figure 11 represented the rank list of most mentioned words. Not all the mentioned words with high frequency were displayed in the figure. “Network”, “Slander”, “Conviction”, “Rumormongering”, “Information”, “Judicial interpretation”, “Good citizen”, “Fast go retweeting”, “Arrest”, “Dangerous” and “haha (LOL)” etc. were the most mentioned words with high frequency in the discussion contents. Except the core words “Rumors” and “Retweet 500

¹⁶ <http://timdream.org/wordcloud/>

times”, as calculated in BDP, the top 5 most mentioned words were “Network(网络)”, “Good citizen(良民)”, “Slander(诽谤)”, “Conviction(判刑)” and “Rumormongering(造谣)”. It was not possible to do a sentiment analysis with this software. But from the context of Chinese network language, the most mentioned words expressed a general neutral attitude towards to the topic. However, the word “Good citizen” and its related expressions “Good citizen card” “I’m a good citizen” and other words such as “Fast go retweeting” “Arrest” “You’ll be arrested” “Haha (LOL)” described a kind of intentionally ridiculed attitude. People expressed this ridicule in sentences such as “I have good citizen card/I am a good citizen, don’t arrest me.” “It’s dangerous now. Be careful of your words otherwise, you’ll be arrested. LOL” “Let’s fast go retweeting this over 500 times and send him/her to the prison.” Obviously, a certain number of people thought the regulation was ridiculous and difficult to judge and implement, even though the supreme court gave the official judicial interpretation and explanation for establishment and implementation of the regulation. The people who had the neutral attitude were generally on the sidelines to look forward the follow-up progress.

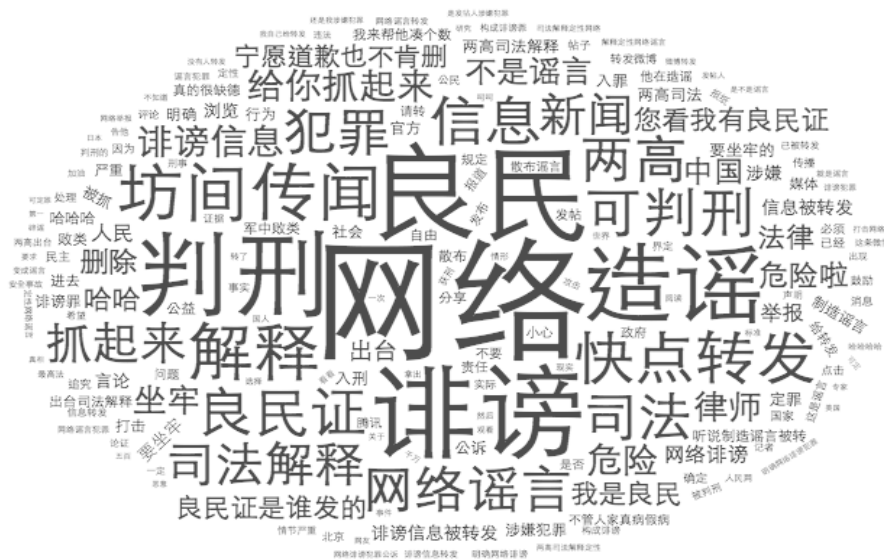


Figure 10 Word Cloud of Weibo Contents with combined key-words “Rumors, being retweeted 500 times”

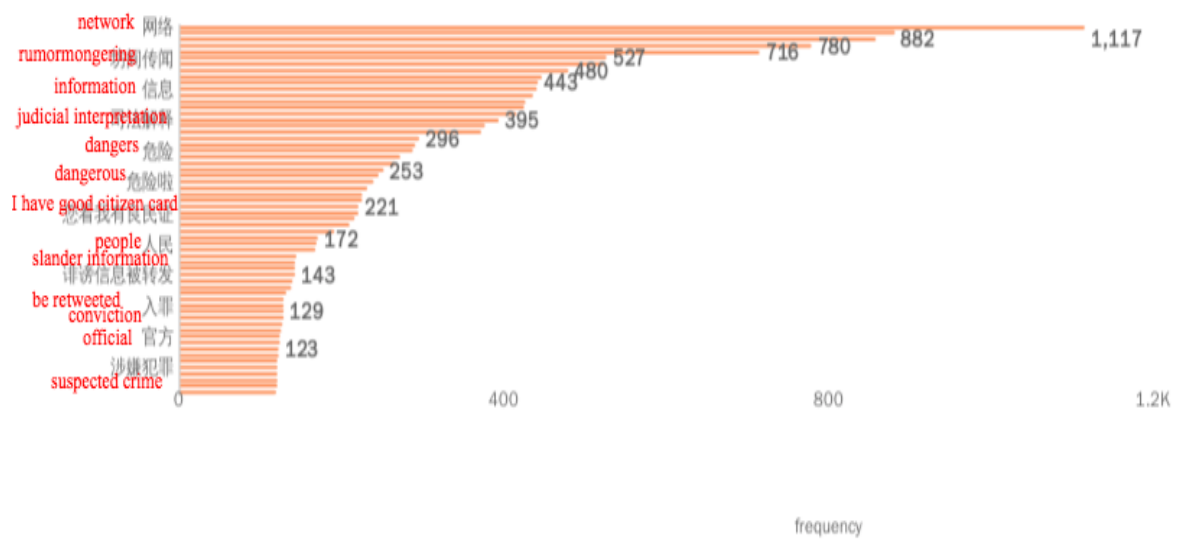


Figure 11 Rank list of word mentioned frequency

In addition, in Figure 12, it was the original Weibo post which was discussed with the largest count of comments and retweets. The post time was at 07:05 on 10.09.2013. The content was sharing a link to the report of “Being retweeted 500 times” which was emphasized the number was determined through the empirical research and professional argument. But the attached picture was describing the price of constitution of the People’s Republic of China was 8 RMB/500 g. This was an obvious satire to its linked report. It could be treated as a typical example which described the ridiculed attitude towards to the topic as mentioned above.



Figure 12 The capture of the most commented and retweeted post regarding to the topic

Overall, combined with the previous research result (Xiong 2014), in which there was no significant relationship between the Penalty Execution Variables and Variables of Rumor transmission willingness, there was also no certain result regarding to the impact on people's willingness about rumor spreading from the data on Weibo.

Certainly, the data analysis for this case was not comprehensive. If the items contain Weibo username, registration profile (age, gender, location etc.), verified situation, the count of content view and emoji, a propagation graph will be possible to generate, as well as key opinion leaders will be indicated, the topic influence and sentiment analysis will be more deeply covered.

Back to the policy “being retweeted 500 times” itself, it is essential to make long-term observation and data collection to judge its effectiveness on the containment of rumors, as well as all the rumors control policies and mechanisms which were mentioned in this thesis. Constraints, penalties, and regulations would help a community develop in order and legally. But if it is effective on controlling the spreading of false information especially rumors, specified long-term data comparison and analysis are required.

7 Conclusion

This thesis has made an overall research on rumor's generating and dissemination on Chinese Social Network, microblog Weibo. The rumor control mechanisms were covered as well. Despite the current existing rumor refutation system of Weibo, the latest "being retweeted 500 times" policy has been mentioned and investigated particularly in the research field of Chinese Social Networks. However, due to the period limitation of observation and the influence against from Weibo's anti-crawler system, a very large set of data could not be collected and analyzed. Merely the data for the case still illustrated some meaningful results.

There are rare research papers focus on the issues about "being retweeted 500 times" rumor control policy either in Chinese or in English according to the search from various academic database such as Chinese CNKI¹⁷ and Google scholar¹⁸. This thesis could inspire further research on rumor control in Chinese social networks. Besides the regulations and human intervention, automatic rumor detection and rumor refutation would be very important and popular research points in the social networks, as long as deep learning is getting popular.

There are already some researchers working on rumor issues with deep learning and natural language processing technologies. In June 2016, Tsinghua Natural Language Processing and Computational Social Science Lab launched a database platform¹⁹ for the rumors which are verified from Sina Weibo. The purpose of this platform is to provide a convenient search for ordinary internet users, as well as support academic research and personal study with free data. The users who have registered can either report or submit rumors to the platform. The staffs will check the submission. Once the rumor is verified, it will be added to the database. (Liu, et al. 2015)

¹⁷ <http://oversea.cnki.net/kns55/default.aspx>

¹⁸ <https://scholar.google.com/>

¹⁹ <http://rumor.thunlp.org/index.jsp> 2016 © THUNLP&CSS-Tsinghua Natural Language Processing and Computational Social Science Lab

This database once is consummated, will help the world researchers who are interested in the research of Chinese context analysis and social network analysis on public sentiment mining, as well as the algorithms on automatic rumor detection and refutation.

Let's assume like this: according to the characteristics comparison to the database, a newly emerged suspected post can be detected, analyzed and calculated the probability of being a rumor. Hence, the result will be added and displayed together with the suspected post until official rumor refutation system giving a stamp as YES or NO. As long as the rumor database enlarges and enriches, the algorithms will become more accurate so that the rumor detection and refutation will rely less on human but more on the system itself. Meanwhile, the rumor spreading time and influence will be reduced as less as it could be.

In all, the research on rumors is diversified and on a long way to go, especially along with the internet develops. There is no way to eliminate the emerging of rumors, but there are methods to break the dissemination chain of rumors and reduce the living time and impact of rumors. Based on the rumors formula $R \sim i \times a$, the importance of the subject $[i]$ cannot be controlled but the ambiguity of the evidence $[a]$ may be affected by the degree of information transparency. Simply the more open and transparent of the information be issued to the public, the less difficult for the rumor verification and refutation are.

Bibliography

- Allport, Gordon W., and Leo J. Postman. 1947. *The Psychology of Rumor*. New York: New York: Holt, Rinehart & Winston.
- Bamman, David, Brendan O'Connor, and Noah A. Smith. 2012. "Censorship and deletion practices in Chinese social media." *First Monday* 17: 03-05.
- BBC News. 2016. "US-South Korea missile defence system 'threatens China'." *BBC News*. 02 12. Accessed 12 05, 2016. <http://www.bbc.com/news/world-asia-35562681>.
- Chai, Yifei. 2016. ""Internet +" as the new media convergence engine (in Chinese)." *People's Daily Overseas*. 06 22. Accessed 09 01, 2016. http://paper.people.com.cn/rmrbhwb/html/2016-06/22/content_1689517.htm.
- China Daily. 2015. "Chinese stars falsely pronounced 'dead' by the media." *China Daily*. 09 17. Accessed 12 05, 2016. http://www.chinadaily.com.cn/celebrity/2015-09/17/content_21904920_2.htm.
- Daley, D.J., and D.G. Kendal. 1965. "Stochastic rumors." *J. Inst. Math. Appl.* 1: 42.
- Daley, D.J., and J. Gani. 2000. *Epidemic Modeling*. Cambridge: Cambridge University Press.
- Farley, Shane. 2011. "Sina: Great numbers for 'Chinese Twitter' service Weibo." *Seeking Alpha*. 01 07. Accessed 08 16, 2015. <http://seekingalpha.com/article/245456-sina-great-numbers-for-chinese-twitter-service-weibo>.
- Fu, King-wa, Chung-hong Chan, and Michael Chau. 2013. "Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and Impact Evaluation of the 'Real Name Registration' Policy." *IEEE Internet Computing*, 05 15: 42-50.
- Hu, Yu. 2000. *Author translated "Effects of Mass Communication"*. Beijing: Xinhua Publication.

- Kermack, W. O., and A. G. McKendrick. 1927. "Contributions of mathematical theory to epidemics." *The Royal Society Series A* 115: 700-721.
- Kwon, Sejeong, Kyomin Jung, Meeyoung Cha, Wei Chen, and Yajun Wang. 2013. "Aspects of Rumor Spreading on a Microblog Network." In *Social Information: 5th International Conference, SoInfo 2013*, by Various Authors, edited by Jatowt et al.(Eds.) Adam, 299-308. Kyoto: Springer International Publishing Switzerland.
- Lam, Oiwan. 2013. "Targeting "Rumors" on China's Sina Weibo." *GlobalVoices Advox*. 07 09. Accessed 09 09, 2015. <https://advox.globalvoices.org/2013/07/09/targeting-rumors-on-chinas-sina-weibo/>.
- Liu, Jie, Kai Niu, Zhiqiang He, and Jiaru Lin. 2014. "Analysis of Rumor Spreading in Communities Based on Modified SIR model in Microblog." Edited by G.Agre et al. *AIMSA*. Switzerland: Springer International Publishing. 69-79.
- Liu, Zhiyuan, Le Zhang, Cunchao Tu, and Maosong Sun. 2015. "Statistical and Semantic analysis of rumors in Chinese social media." *Information Science of China*, 12 09: 1536-1546.
- MacKinnon, Rebecca. 2009. "China's censorship 2.0: How companies censor bloggers." *First Monday*, 02 02.
- Maki-Thompson, D.P. 1973. *Mathematical Models and Applications, with Emphasis on Social, Life, and Management Sciences*. NJ: Prentice-Hall, Englewood Cliffs.
- Mehta, Ankita. 2015. "Jackie Chan Death Hoax: Legendary Actor Not Dead; Slams Fake Death Reports on Facebook." *International Business Times*. 05 19. Accessed 12 05, 2016. <http://www.ibtimes.co.in/jackie-chan-death-hoax-legendary-actor-not-dead-slams-fake-death-reports-facebook-632909>.
- Milgram, Stanley. 1967. "The Small World Problem." *Psychology Today* 60-67.
- Millward, Steven. 2016. "Weibo to copy Twitter, abolish 140-character limit." *Tech In Asia*. 01 20. Accessed 07 15, 2016. <https://www.techinasia.com/weibo-copy-twitter-abolish-140-character-limit>.

- Ng, Jason Q. 2015. "Politics, Rumors, and Ambiguity: Tracking Censorship on WeChat's Public Accounts Platform." *University of Toronto, Citizen Lab*. 07 20. Accessed 08 31, 2016. <https://citizenlab.org/2015/07/tracking-censorship-on-wechat-public-accounts-platform/>.
- Open Weibo. 2013. "API Access authorization (in Chinese)." *Open Weibo*. 11 26. Accessed 12 18, 2015. <http://open.weibo.com/wiki>.
- . 2012. "API Wiki Search/Topics (in Chinese)." *Open Weibo*. 11 16. Accessed 12 18, 2015. <http://open.weibo.com/wiki/2/search/topics>.
- Pendleton, S.c. 1998. "Rumor research revisited and expanded." *Language&Communication* vol.1 (no.18): 69-86.
- People Yuqing. 2015. "Internet public opinion report 2015 (in Chinese)." *People Yuqing*. 12. Accessed 07 15, 2016. <http://yuqing.people.com.cn/GB/392071/401685/>.
- Persinos, John. 2015. "This China-Based Hybrid of Facebook and Twitter Is Set to Soar in 2016." *The Street*. 11 11. Accessed 12 07, 2016. <https://www.thestreet.com/story/13361093/1/this-china-based-hybrid-of-facebook-and-twitter-is-set-to-soar-in-2016.html>.
- Peterson, Warren, and Noel Gist. 1951. "Rumor and Public Opinion." *The American Journal of Sociology*, 159-167.
- Piqueira, Roberto C., José. 2010. "Rumor Propagation Model: An Equilibrium Study." *Mathematical Problems in Engineering*, 04 02: 7.
- The Economist. 2012. "Which tongues work best for microblogs?" *The Economist*. 03 31. Accessed 08 05, 2015. <http://www.economist.com/node/21551466>.
- Waddell, Kaveh. 2016. "Why Google Quit China—and Why It's Heading Back." *The Atlantic*. 01 19. Accessed 08 31, 2016. <http://www.theatlantic.com/technology/archive/2016/01/why-google-quit-china-and-why-its-heading-back/424482/>.

- Wang, Hui, Lin Deng, Meng Zhou, Keqin Chen, and Jianghong Han. 2012. "Weighted CSR Rumor Spreading Model in Online Social Network." *Journal of Information & Computational Science*, 12 30: 5791-5798.
- Wang, Lu. 2013. "Analysis of the situation and dealing for rumors spreading under the real-name system in Weibo (in Chinese)." 06-11.
- WeAreSocial. 2016. *We Are Social*. 01 27. Accessed 07 01, 2016. <http://wearesocial.com/sg/special-reports/digital-2016>.
- Weibo Community Management Center. 2014. "Monthly Report of CoP (in Chinese)." *Weibo Community Management Center*. 09. Accessed 09 20, 2015. <http://service.account.weibo.com/monthreport>.
- Weibo IR. n.d. "Business Overview." *Weibo Investor Relations*. Accessed 08 16, 2015. <http://ir.weibo.com/phoenix.zhtml?c=253076&p=irol-homeProfile&t=&id=&>.
- Weibopiyao. 2016. "Monthly Report of Weibo Rumors Refutation(1) (in Chinese)." *Weibo*. 06 03. Accessed 08 06, 2016. <http://www.weibo.com/ttarticle/p/show?id=2309403982288034604423&mod=zwe nzhang>.
- Xiong, Yan. 2014. *Can punishment prevent rumors from spreading?-Take "Being retweeted over 500 times can be sent to the jail" as example (in Chinese)*. Vol. 2. Beijing.
- Zhan, Zhang, and Negro Gianluigi. 2013. "Weibo in China: Understanding its Development Through Communication Analysis and Cultural Studies." *Communication*, 199-216.
- Zhao, Laijun, Jiajia Wang, Yucheng Chen, Qin Wang, Jingjing Cheng, and Hongxin Cui. 2011. "SIHR rumor spreading model in social networks." *Physica A (Elsevier B.V.)* 2444-2453.

Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R. Crandall, and Dan S. Wallach. 2012.
Tracking and Quantifying Censorship on a Chinese Microblog Site. Cornell
University Library, 11 26.

Appendices

A The Sample of Case Data

content	level	musicurl	videourl	weibold	catchTime	userid	weiboUrl	commentCount	source	pic_list	beForwardV	reportCount	info1	beComme	info3	info2	createTime
已经超过500次，潘总可以投案了//@任志强 转发500次！//@潘石屹 转	3.62E+15			3.62E+15	1395739116	1897795234	http://weibo.com	616	新浪微博		0	0					1379152798
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	1196415674	http://weibo.com	349	新浪微博		0	0					1379152797
歉一帮！转！//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评	3.62E+15			3.62E+15	1395739116	2645907987	http://weibo.com	218	Android客户端		0	0					1379152795
危险了//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给	3.62E+15			3.62E+15	1395739116	1984000101	http://weibo.com	211	新浪微博		0	0					1379152782
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	3004025747	http://weibo.com	187	iPhone客户端		0	0					1379152760
你们这些坏人，就会讥讽500次。//@任志强 转发500次！//@潘石屹 接	3.62E+15			3.62E+15	1395739116	2130360654	http://weibo.com	166	三星Galaxy Note II		0	0					1379152749
过了。//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给	3.62E+15			3.62E+15	1395739116	1623853591	http://weibo.com	127	iPhone客户端		0	0					1379152748
转发500次！//@任志强 转发500次！//@潘石屹 接受道歉。//@	3.62E+15			3.62E+15	1395739116	1303139157	http://weibo.com	125	诺基亚Lumia920		0	0					1379152745
潘主要志马上道歉。像潘！//@任志强 转发500次！//@潘石屹 接受道	3.62E+15			3.62E+15	1395739116	1045291565	http://weibo.com	103	360浏览器极速版		0	0					1379152737
哈哈！我支持转一次！//@任志强 转发500次！//@潘石屹 接受道歉。//@	3.62E+15			3.62E+15	1395739116	3722476867	http://weibo.com	99	Android客户端		0	0					1379152736
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	2712418081	http://weibo.com	93	UC浏览器Android版		0	1					1379152734
转试试。估计没戏//@任志强 转发500次！//@潘石屹 接受道歉。//@名	3.62E+15			3.62E+15	1395739116	2219174713	http://weibo.com	86	Android客户端		0	0					1379152731
600了！哈哈。。//@任志强 转发500次！//@潘石屹 接受道歉。//@	3.62E+15			3.62E+15	1395739116	1450002715	http://weibo.com	78	新浪微博		0	0					1379152728
浏览量和转发量已经可以入刑//@任志强 转发500次！//@潘石屹 接受	3.62E+15			3.62E+15	1395739116	1495201983	http://weibo.com	77	新浪微博		0	0					1379152726
力提@潘石屹//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评	3.62E+15			3.62E+15	1395739116	1855796187	http://weibo.com	73	OPPO手机		0	0					1379152724
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	2396477483	http://weibo.com	71	新浪微博4G版		0	0					1379152723
到刑。//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给	3.62E+15			3.62E+15	1395739116	2095899694	http://weibo.com	70	Android客户端		0	0					1379152722
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	262477605	http://weibo.com	63	360浏览器极速版		0	0					1379152719
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	2507423973	http://weibo.com	56	新浪微博4G版		0	0					1379152716
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739116	3213078840	http://weibo.com	55	新浪微博		0	0					1379152704
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739120	1802319074	http://weibo.com	54	Weico iPhone		0	0					1379152699
对不起。在你道歉前已被转发了500次。我们将保留处理你的权利。你	3.62E+15			3.62E+15	1395739120	2688323161	http://weibo.com	53	Android客户端		0	0					1379152698
转起来//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给	3.62E+15			3.62E+15	1395739120	1707422650	http://weibo.com	51	Android客户端		0	0					1379152691
任总，我是良民，愿看我有良民证！//@任志强 转发500次！//@潘石屹	3.62E+15			3.62E+15	1395739120	2615059833	http://weibo.com	38	新浪微博		0	542					1379152686
那也是本事//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说	3.62E+15			3.62E+15	1395739120	263917432	http://weibo.com	36	iPhone客户端		0	0					1379152672
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739120	1524434715	http://weibo.com	33	iPhone客户端		0	0					1379152666
转500次//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给	3.62E+15			3.62E+15	1395739120	2718494355	http://weibo.com	33	三星GalaxyNote		0	0					1379152666
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739120	1923804707	http://weibo.com	32	iPhone客户端		0	0					1379152652
499次就够了！//@任志强 转发500次！//@潘石屹 接受道歉。//@名家	3.62E+15			3.62E+15	1395739120	2056881480	http://weibo.com	31	iPhone客户端		0	0					1379152640
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739120	1952849273	http://weibo.com	28	新浪微博		0	0					1379152639
转500//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给	3.62E+15			3.62E+15	1395739120	1134154041	http://weibo.com	28	Android客户端		0	0					1379152634
接受道歉。意味放弃诉讼。//@刑诉毛立新 涉疆诽谤//@任志强 转发50	3.62E+15			3.62E+15	1395739120	2706199953	http://weibo.com	27	新浪微博		0	1					1379152634
已超500。//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说	3.62E+15			3.62E+15	1395739120	2038762491	http://weibo.com	25	iPhone客户端		0	0					1379152628
就算真的也不会承认啊//@任志强 转发500次！//@潘石屹 接受道歉。	3.62E+15			3.62E+15	1395739120	3445774512	http://weibo.com	25	iPhone客户端		0	0					1379152627
哈。潘任美事件吧？任总必须在//@任志强 转发500次！//@潘石屹 接	3.62E+15			3.62E+15	1395739120	2134857371	http://weibo.com	24	iPhone客户端		0	0					1379152623
//@任志强 转发500次！//@潘石屹 接受道歉。//@名家评说 给潘总道	3.62E+15			3.62E+15	1395739120	2025454217	http://weibo.com	24	新浪微博		0	1					1379152618

B Python code for data format transformation

```
#!/usr/bin/env python
#-*- coding:utf-8 -*-
import os
import os.path
import sys
import json

def write_file(dict):
    #print dict
    with open('Retweet 500.csv','a') as res_file:
        resline = ""
        for key, value in dict.items():
            #print 'value type: ' + str(type(value))
            if type(value) != list:
                resline += value + ","
                #print key + ": " + value
            else:
                #print key + ":[",
                firstelem = True
                for elem in value:
                    #print elem + ' ',
                    if firstelem:
                        resline += elem
                        firstelem = False
                    else:
                        resline += '|' + elem
                #print "]"
                resline += ','
        print >> res_file, resline.encode('gbk')

if __name__ == "__main__":
    rootdir = 'D:\Data\json2csv\Retweet 500\|'
    for parent, dirnames, filenames in os.walk(rootdir):
        for filename in filenames:
            #print 'parent: ' + parent
            print 'filename: ' + filename
            with open(rootdir + filename) as file:
                lines = file.readlines()
                for line in lines:
                    try:
                        dict = json.loads(line)
                        #print dict
                        write_file(dict[0])
                        #import pdb
                        #pdb.set_trace()
                    except Exception as e:
                        print e
                        continue
```