What is social media platforms' role in constructing 'truth' around crisis events? A case study of Weibo's rumour management strategies after the 2015 Tianjin blasts

Jing Zeng, Queensland University of Technology
Chung-hong Chan, University of Hong Kong
King-wa Fu, University of Hong Kong

Abstract

This article studies the content moderation strategies used by Weibo to regulate rumour discussion in the wake of the 2015 Tianjin blasts. Over 100,000 Weibo posts were collected and categorised into three datasets, these being rumour discussion posts from the public, rumour-debunking posts from Weibo's official rumour rebuttal accounts, and posts removed from the system. We have identified two content moderation strategies on rumours, namely, rumour rebuttal and content removal. Clustering analysis and time series analysis were applied to test how these two strategies were used on different topics and how they influenced public discussion on rumour-related topics. Our findings suggest that this platform responds to rumours differently, depending upon the political sensitivity of the topic. We found that Weibo messages were refuted and censored differently based on the political sensitivity of the topics. Time series analysis identified rumour rebuttals and censorship usually associated with more general discussion about the rumour and therefore we have no evidence to support these strategies can consistently create a chilling effect.

1. Introduction

As the Internet becomes important source of news and information, people increasingly rely on online platforms for content verification and filtering. However, online news information is never presented or ordered "as is". Rather it is considerably moderated or even manipulated by online platform providers. For instance, Google, Facebook, and Twitter develop computational algorithm to moderate content on their platforms based on the information's relevance (Vaidhyanathan, 2012), authenticity (Lee, 2014), and legality (Gillespie, 2011, Travis, 2013). Alongside the public's reliance on online platforms as news source, it is a growing body of researchers who seek to examine the potentials and problems of social media's practice of content moderation (e.g. Andrejevic, 2013; Roberts, 2016; Gillespie, 2015). However, few researchers give sufficient attention to the underlying socio-

political contexts of the social media platforms. This lack of attention can be problematic because the way platforms interact with users and the subsequent impacts they have, vary depending on socio-political contexts (Bolsover, 2016; Yang et al., 2013). For instance, in an authoritarian country like China where mainstream media is tightly controlled, and where most popular western social media sites are blocked, people rely heavily on home-grown social media for alternative information or 'truth' (White & Fu, 2012; Nip & Fu, 2016). In such context, content moderation on social media may present complex dynamics that are entirely different from its western counterparts. Nevertheless, in existing literature, there is limited research studies that contextualise content moderation on non-western social media platforms.

This study aims to bridge this gap by studying China's biggest microblogging site- Weibo. Specifically, we study how this platform moderates rumour-related content during crisis events using the case of the 2015 Tianjin blasts. While most previous studies discuss the impact of social media platforms at a theoretical level (Roberts, 2016; Creemers, 2014), this study aims to test the impact of the platform's content moderation strategies empirically. To this end, we test if the rumour content moderation strategies stimulate more discussion, or lead to a decline on the discussion of relevant topics.

Over 100,000 posts from Weibo were collected to analyse how rumours have been refuted and censored during the Tianjin blasts. Statistical analysis was also used to test the chilling effect of Weibo's rumour management strategies. In the context of this study, 'chilling effect' refers to the effect of Weibo's rumour regulation strategies in silencing the public. Our findings suggest that this platform responds to rumours differently, depending upon the political sensitivity of the topic. There is also little evidence showing chilling effects from the platform's censorship of rumour.

2. Literature review

Drawing upon classic rumour theories, rumour is understood as unofficial information that results from collective uncertainty in the society when reliable information is not available (Allport & Postman, 1947; DiFonzo & Bordia, 2007; Kapferer, 2013; Rosnow, 1991;

Shibutani, 1966). Based on this understanding of rumour, what characterises rumour is not essentially falsity, i.e. rumour can be a genuine truth, but its relationship to authority or official source of information. In this case, 'authority' refers to institutions or individuals who have the power to verify the authenticity of information (Kapferer, 1990; Fine, 2007). Thus rumour's relation to power and to authority makes it an ideal subject for studies looking at institutions' (Berinski, 2015) , social groups' (Duffy, 2012), and individuals' (Manaf et al., 2013) accounts of reality. Allied with these studies, we expand this understanding of rumour to study the authoritative role of platforms in managing and defining rumour.

In the field of media and communications, there is a large and growing body of research examining rumour management on social media platforms. However, most of these studies treat rumour management as a technical matter, and little attention has been paid to the social and political aspects of rumour. For example, different automatic rumour detection methods have been proposed and discussed to help identify false rumour on social media. Castillo et al. (2011) studied information dissemination on Twitter during the 2010 Chile earthquake and found features that can be used to detect rumour on social media, such as URLs and sentiments. Procter et al.'s (2013) study of Twitter usage during the 2011 London riots discusses the potential to track rumour through tracking the lifecycles of tweets. More recently, Arif et al. (2016) used a case study of the 2014 Sydney Siege to demonstrate a mix-method approach to detect rumour, which combines the volume, exposure, and content production aspects of tweets. Studies are conducted by researchers to track rumours on Weibo. For example, based on the particularity of Weibo, Wu et al. (2015) proposed to detect rumour on Weibo through examining the propagation structure of information.

Whilst the studies mentioned above all perceive the authenticity of information as an objective value that can be measured by automated mechanisms, this research considers the authenticity of information as subjective and socially constructed. We are especially interested in the role of social media platforms in this construction process through the content moderation practices. In the context of this study, content moderation refers to the process through which social media platforms moderate user-generated content using either human labour (Roberts, 2016; Hiruncharoenvate, 2015) or algorithms (Bucher, 2012; Gillespie, 2015). In Rogers' (2014) study of information politics, he discusses how *sources* on the web compete to become *information*, and how this process is affected by 'front-end politics' and 'back-end politics'. Back-end politics involves the logics and practices in its

algorithms; whereas front-end politics can be exemplified by how an interface is designed to impact how users engage with information and each other. Rogers' analysis focuses on issues related to fairness, representation, and inclusiveness of sources on the web, and argue that both front-end politics and back-end politics contributes to the "demise of alternative accounts of reality (2014, p. 2)".

More recently, Gillespie (in press) proposed another model to analyse content moderation on social media platforms. He suggests there are two common ways through which platforms manage inappropriate content: to remove it or to mark it as such. He reasons that both approaches can be problematic. According to Gillespie (in press), removal 'opens platforms to charges of subjectivity, hypocrisy, political conservatism, and self-interest' (in press, p.21); although *marking* can be a less invasive strategy. It can also cause problems when it is used too often or when the marking process has an over-reliance on algorithms.

Both Rogers' and Gillespie's discussions highlight the significance of social media platform' content moderation strategies, and both frameworks used in their studies are useful and can be generalized to practice on Weibo. In the case of rumour management, Weibo's content censorship used to delete posts from the system is related to Rogers' back-end politics, or Gillespie's category of *marking* strategy; whereas Weibo's official posts that used to openly refute a specific rumour content (Yang, 2012) is what Rogers describes (2014) as front-end politics, and an example of *marking* strategies. Aforementioned anecdotal evidence from the Bo Xilai case and Occupy movement in Hong Kong seems to suggest the application of different content strategies depends on the topic matters. However, no previous study has investigated whether the application of the two strategies was related to the topic matters. How the two strategies were used on different topic matters might shed light on why the two strategies were used in the first place, which have not been detailed in previous studies on content moderation either. Therefore, we propose to study how Weibo's rumour management strategies are applied to different rumour topics. The first research question we propose is:

RQ 1. How were different rumour topics related to the Tianjin blasts censored and refuted in various ways?

This study also tests the impact of Weibo's rumour regulation strategies on online discussion. In existing literature on the Internet in China, there is a large number of studies examining

how content regulation works (Bamman et al., 2012; King et al., 2013; Ng & Landry, 2013; Zhu et al., 2013), but few scholars examine its impact on users' online activities empirically. Exceptionally, Fu et al. (2013) used statistical analysis to test the chilling effect of Weibo's regulatory policy. Chilling effect is based upon the psychological theory of behavioural inhibition (Carver & White, 1994), positing that an individual perceives censorship as a "punishment" or a threat to personal power of exercising freedom of expression and as a result the person might inhibit one behaviour to avoid further "punishment" or threat, leading to reduction of making posts and reposting other messages. In this context, chilling effect refers to the phenomenon that regulation policies on social media platforms leads to a decline on discussion over certain topics.

However, censorship might arouse public's interest in the censored topic. Drawing on the theory of psychological reactance (Brehm, 1966), people's reaction to censorship is directed toward the position attempting to restore the loss of freedom threatened by the censors, that is to say it results an increasing interest in making posts that are susceptible to censorship. Early psychological experiment found that censorship led to a rise in interest in the censored message and triggered a change in attitude toward the position of the message (Worchel, et al., 1975). If the theory is situated within which a Chinese microblogger responds to the message that is censored by the authority, the user may attempt to "restore" the loss of freedom of expression and the right to speak to the public by making posts that are susceptible to be censored, and/or reposting additional messages that are subject to censorship.

Although Fu et al.'s (2013) research on the chilling effect looks at the impact of real-name registration policy on online discussion exclusively, their analytical framework and methods should also be able to be applied to studies on rumour. Because few prior studies have provided empirical evidence to determine the effect of strategies used by platforms to moderate rumour content, this study aims to bridge this gap by applying Fu et al.'s (2013) analytical methods to study:

RQ 2: What was the impact of Weibo's rumour management strategies?
Do these strategies have a counter-effect that stimulates more discussion over the issue, or is there evidence of chilling effects?

As discussed in prior research, the platforms' regulatory strategies should not be discussed in isolation from its socio-political contexts. Gillespie (in press, p. 3) points out, for example, social media platforms are not just in between users, but also "in between citizens, law enforcement, policymakers and regulators." Such a mediatory role of social media between the state and the citizens is particularly important when studying social media in China, As Meng (2011) argues, the Internet in China is a result of the constant negotiation between the platform, the ruling Chinese Communist Party (CCP) and Chinese netizens. On the one hand, social media is often used by the CCP as a political instrument to suppress online discussion (Herold, 2008), but on the other it is also deployed by the grassroots citizens and opinion leaders to voice their discontent and criticism over the government (Yang, 2009; Fu & Chau, 2014).

Benney (2013, p. 2) conceptualises the relationship between Weibo and the Chinese government as a *clientelist relationship.*In this relationship, the regime approves Weibo's near monopoly of the Chinese microblogging landscape, but in return Weibo is manipulated by the Chinese government to constrain dissent, activism, and discussion over political issues in general. This conceptual understanding of the relationship between the state and social media platforms in China necessitates a contextualised discussion on how rumour is managed by Weibo, and how Chinese netizens respond to it.

3. Background

On 12 August, 2015 a series of explosions in Tianjin, northern China, devastated large areas of the region, killing at least 173 people, and leaving nearly 800 people injured (Guardian, 2015). The explosions occurred in the warehouse of a private logistics company called Ruihai that was found to have illegally stored hundreds of tons of toxic chemicals. Immediately after the blasts, footage and images of the explosions were widely shared on social media. For a long time after the explosion, only limited information about the accident was available through official media, as a result social media became the major platform where people speculated on the cause of the blasts and the damages. As rumour began to circulate on the Internet, a large number of which were critical of the authorities' response to the blasts, harsh measures were taken by the government to crackdown rumour sharing. Over 50 websites were shut down and more than 300 social media accounts in China were suspended. Twelve

people were arrested for circulating rumour online (CNBC, 2015; CYOL, 2015; Xinhua, 2015).

This was not the first time that the Chinese government has manipulated online platforms to control online discussion after massive incidents. During the 2012 Bo Xilai political scandal, the comments function on Weibo was completely shut down for a few days to stop Chinese people from participating in online discussion over this issue (Ng & Landry, 2013). A more recent instance is the 2014 Occupy Movement in Hong Kong. During the movement, non-official information relating to the incident was heavily censored on social media platforms in China (Park, 2014; Stockmann, 2015).

Both examples suggest that it is not uncommon to see aggressive approaches be used to regulate online discussion after massive incidents in China. However, these rumour-targeted regulations used after Tianjin blasts were introduced to Chinese social media only recently. In 2013, the Chinese government launched a series of campaigns to stop online rumour. For example, one such ruling states that microbloggers could be jailed for up to three years for posting rumour content if it is forwarded more than 500 times or viewed over 5,000 times (Kaiman, 2013; Chin & Mozur, 2013). Such anti-rumour policies on social media have raised concern from both commentators and scholars over Chinese people's freedom of speech (Creemers, 2014; Chin & Mozur, 2013; Jiang, 2016). It is against this background that we conducted this research to examine the impact of Weibo's content moderation over rumour-related topics.

4. Methods

4.1 Data collection

In order to answer the research questions on rumour rebuttal and censorship, three sets of Weibo messages about the Tianjin Blasts were collected and they were labelled as: messages from the web interface of the Weibo search page (dataset SR), rumour-rebuttal messages (dataset RR) and the censored messages collected from the WeiboScope project (dataset WS). The SR were used to analyse the general public's discussion about this incident but has not been rebutted and censored. The RR messages were information that has been rebutted by

Weibo's official rumour-rebuttal accounts but yet still openly available from all Weibo users. The messages of the WS dataset were censored and not accessible to Weibo users.

The methodology of data collection of SR and RR datasets was similar to a previous study and was reported elsewhere (Fung et al., 2016). In essence, a web crawler was developed to retrospectively collect search results hourly from the web interface of the Weibo Search Engine[1] by searching for a set of keywords during the entire study period (from 12-Aug-2015 to 26-Aug-2015). Deploying this method instead of retrieving data from the official Sina Weibo API is because the search API was not unrestrictedly available to non-privileged researchers and its functionality was limited2. We use the keywords 'Tianjin' (天津) and 'explosion' (爆炸) for the SR dataset collection and 'Tianjin' (天津) and 'rumour-rebuttal' (辟谣) for the RR.

The censored messages in the WS dataset were collected by the methodology illustrated in the paper by Fu et al (2013). In summary, the timeline of a group of Weibo users whose followers count was high and who were randomly selected were recorded in every time interval as defined. Their timeline was compared to the version collected at the immediately previous time point and missing messages were thus identified. The missing messages were further checked by using a Sina Weibo API call. If the error message 'Permission Denied' is returned, the message is confirmed as a censored post.

4. Data analysis

Topics of Weibo messages about Tianjin Blasts

In order to identify rumour topics, posts from dataset RR and dataset WS were manually classified by topic by the first author (JZ) using an inductive approach, i.e. open coding content analysis (Elo & Kyngäs 2008). Based on the topic list, keywords were used to select related posts under each topic from the SR dataset.

[1] s.weibo.com

[2] http://open.weibo.com/wiki/2/search/topics This topic API is the only available API related to search function. The usage of this API requires very rigid approval from Sina and only 200 messages can be request for each hashtag search.

Similarity in rumour management: clustering of topics

The similarity in management strategy between topics was examined using clustering analysis on the basis of the relative probabilities of these topic in the RR and WS set. For each topic, the probabilities of appearance in the WS and RR dataset were adjusted by the general interest in that topic as reflected by the probability of topical messages in the SR dataset. The log odds ratios (log OR) of the i^{th} topic in the j^{th} set were calculated for each topic in each set with the following formula:

$$log_e OR_{(topic=i,set=j)} = log_e \frac{(|i \cap j| + 0.5) \times (|\bar{i} \cap \bar{j}| + 0.5)}{(|i \cap \bar{j}| + 0.5) \times (|\bar{i} \cap j| + 0.5)}$$

For example, a topic x has 500 messages in the SR set of 109,099 and 50 messages in the RR set of 1,744.

Dataset	Topic x	Not Topic x
RR	50	1,694
SR	500	108,599

The $logOR_{(x, RR)}$ is equal to:
= log(((50+0.5) * (108599 + 0.5)) / ((500+0.5) * (1694+0.5))) = 1.867

A more positive value of log OR indicates the i^{th} topic is more likely to appear than expected in the j^{th} set and vice versa.
The log OR(i, SR) and log OR(i, WS) values for each topic were computed and they were used for clustering analysis. Agglomerative hierarchical clustering (AHC) algorithm was used for clustering and the optimal number of topic was determined by the elbow method (Anderberg, M. R., 1973).

'Effectiveness' of rumour management: time series analysis

To assess the effect of Weibo's rumour management strategies, we used lead-lag analyses to explore if this platform's rumour response led to a future decline in the number of discussion

in relation to the topic or otherwise, whether it actually stimulated the public to discuss more about the issue.

For each topic, the time series of daily message volume in the SR was used as the dependent time series. The time series of daily message volume in the RR and WS were considered the independent time series. The lead-lag relationships between dependent and independent time series were studied by cross-correlation function (CCF). The autocorrelation of the time series can significantly inflate the correlation and therefore we use the ARIMA model to prewhiten the time series as suggested by Cryer and Chan (2008). Due to the short study period, we can only consider 15 lag units (-7 to +7 days). The pair of dependent and independent time series was considered associated when cross-correlation was significant at the 5 percent level in any lag and lead units within the range of -7 to +7 days. For example, positive cross correlation at with a negative lag indicates the independent time series (i.e. rumour rebuttal or censorship) leads the dependent time series (i.e. general discussion).

In this analysis, only topics with at least one post per day in WS and RR were included. Finally, only six topics were included.

5. Results

In total, we collected 109,099 messages from the SR, 1,744 messages from the RR and 800 messages from the WS. Using the inductive approach, we identified 14 topics of rumour related to the blast through a content analysis (Appendix 1).

RQ1

The log OR for each topic in the WS set and RR set was plotted in Figure 1.

[Insert Figure 1 here]

Using the AHC algorithm, the optimal number of clusters was three. In the analysis, the three groups of topics were “highly refuted and maybe censored” (red topics, 7 topics), “casually refuted and casually censored” (black topics, 5 topics) and “let the public talk about them” (green topics, 2 topics), which reflected three different approaches of rumour management on

the Weibo platform. From this analysis, we found that rumour management varied according to the topical difference.

RQ2

We used time series analysis to explore if this platform's rumour response led to a decline in the amount of discussion or whether it stimulated more discussion on the issue. For each topic, the lead-lag associations between daily numbers of posts in all three datasets were evaluated. The CCFs for each topic are presented in Figure 2.

[Insert figure 2 here]

The time series analysis result reveals that the platform's efforts to debunk rumour of some topics was actually followed by a subsequent increase in the number of discussion over the issue. The RR dataset time series was positively correlated with the future level of SR dataset time series in five topics except the one of "local media". Therefore, an increase in RR activities of online rumour was usually associated with an increase in general discussion of these six rumour.

Moreover, there is little evidence suggesting that Weibo platform's censorship had chilling effects over Weibo users' sharing of rumour. A case-by-case pattern was observed. For corruption related rumours ("Ruihai" and "Officers" topics), censorship leads to more discussion over the issue. But for the rumours related to "Pollution" and "Volunteers", the future WS set time series was positively correlated with the RR set time series, which means that the censorship activities lagged behind the general discussion, indicating a pattern of information suppression. In spite of this, the general public still talked about those topics and therefore we did not observe a consistent pattern of information suppression. The chilling effect was observed only in the topic of "Casualties" in which the censorship activities were associated with a reduced level of activity in the general discussion on the same day, as indicated by the negative cross correlation at the lag unit 0.

6. Discussion

The results of this study indicate that Weibo platform's practice of *rumour refuting* and *rumour removing* are inconsistent. For instance, the platform's response to less politically sensitive rumour topics relied more on openly refuting - using the platform's official accounts to mark such information as false rumour. Rumours under this category, i.e. about public safety, include claims that the vibration of the explosions could cause organ damage in large number of residents living in the region (Topic HC), and posts requesting people to stop using certain motorways in order to make them available for ambulances (Topic TRA). By contrast, both removing and refuting strategies were used to moderate the rumours related to government officials' mismanagement of the crisis. For instance, almost all posts sharing a news that a CNN reporter was attacked by local governmental officials were removed from the system, and at the same time Weibo used its official rumour rebuttal accounts to tell the public that the attack did not happen. The result from the clustering analysis also shows that, for certain rumour topics, Weibo did not take any measure to stop the discussion. One example is the rumour that the explosions polluted the rivers in Tianjin where large numbers of fish were killed (Topic DF).

The inconsistency in Weibo's responses to rumour supports Gillespie's (in press) concern over social media platform's practice of content moderation. As he points out, it is difficult for the online platforms to apply consistent strategies to regulate content, and this inconsistency suggests subjectivity, bias, and self-interest in the platform's regulatory policies (in press, p.21). In the Chinese contexts, the forces behind social media platforms' practice of content moderation is more than simply self-interest, but also political pressure from the government (Benney, 2013; Meng, 2011). As Jiang (2016, p.140) points out, Weibo can face penalties from the party if it fails to properly handle 'public complaints'. Therefore, Weibo's content regulation over rumour content can be considered as a form of self-censorship of the platform. This self-censorship practice results from the negotiation of the platform's self-interest to maintain a high level of user traffic and the external political pressure from the party to maintain the political correctness of user's' activities.

While scholars have found evidence to show Weibo's regulation practices can chill China's microblogging environment (Creemers, 2014; Fu et al., 2013), the findings of time series analysis in the current study suggest that in some circumstances Chinese netizens can be

resilient. For instance, we found that Weibo's rumour rebuttal efforts are usually associated with more public discussion over the refuted topics. For almost all analysed topics, after Weibo began to openly refute related rumours, more posts discussing this topic were published by Weibo users. Also, there is insufficient evidence to show that *rumour removal* consistently leads to chilling effect on the Weibo platform either. We can only observe possible evidence of chilling effect on Weibo users' speculation about how many people were actually killed in this accident. This is to say, when Weibo began to remove rumour content about the casualty of the blasts, the public starts to publish fewer posts about this issue. In the case of rumour about the logistics company Ruihai's links to the government and rumour about government official's involvement in this accident, by removing related posts Weibo actually leads to an increase in the discussing over these two topics.

This counter effect of social media platform's rumour management strategies has both theoretical and practical implications. Chinese netizens' resistance to official discourse of the disaster indicates the social and political significance of rumour as a counter-power against authorities, including both the platform and the government. As discussed in literature, social media platforms are increasingly established as an *authoritative space* (Rogers, 2014) where information and user activities are shaped through both front-end and back-end regulations. However, our case study observed how users resisted social media platform's front-end strategy (official rebuttal) and back-end strategy (censorship) to regulate rumour content. In the case of China, because the government increasingly attempts to use social media platforms as a political tool, user' resistance against online platform's regulation constitutes a form of protest against the authoritarian regime.

From a practical perspective, the counter effect of Weibo's rumour management strategies suggests the importance of a more contextualised rumour rebuttal strategies. As discussed earlier, most rumour research studies in the field of media and communication often deploy automated mechanism to detect and control rumour (Castillo et al., 2011, Procter et al., 2013; Wu et al., 2015). However, our analyses the impact of Weibo's anti-rumour strategies shows that technical solutions might only solve the problem superficially in some rumour cases. Rumour management strategies should be developed based on a good understanding of the platform's geopolitical environments. In a Chinese context, as Hu (2009) points out, rumour is increasingly used by the Chinese public to challenge authority. In such circumstances, the rumour circulation is not simply the result of a lack of 'official information', but is also

caused by a 'breakdown of institutional trust' (Fine, 2007, p. 7). For this reason, re-establishing the credibility of authoritative information sources, such as the mainstream media and governmental agencies, should be on the rumour rebuttal agenda.

There are three major limitations of the current study. Firstly, the current study is observational and therefore we cannot generate any causal claim from the results. The findings should be cautiously interpreted. Further study is warranted to confirm the findings by experimental design. Large scale online experiment has been conducted to examine the Internet censorship in China, such as creating a made-up social media service (King et al. 2014). Alternative strategy is to triangulate our findings with other methodologies, for example qualitative interview.

Secondly, our approach of content analysis might be considered as not sufficiently robust and might not be replicable due to subjective coder's bias. This method is less vigor than the deductive approach (with an *a priori* fixed coding scheme) but there is no previous knowledge on topic of online discussion during the emergency in China and therefore we cannot operationalize the topical characteristics. Non-human coding approach such as unsupervised topic modelling can be used also but the topic derived might not be interpretable by human (Chang et al. 2009). Based on our findings, future study might be able to operationalize the topic characteristics of crisis-related discussion on the Chinese social media.

Thirdly, the data from the WS dataset has the problem inherited from the Weiboscope data collection methodology, such as the limitations that the subject selection is not random and the intervals of revisiting the timeline of the subject is not constant. Therefore, we cannot generalize our finding derived from the WS dataset to represent the overall Weibo user population. However, as the current study mainly focuses on comparison of content moderation strategies between different topics and all collected data suffered indiscriminately the same set of limitations as delineated above, we believe the comparison between topics should deem appropriate and thus our conclusions can be fairly considered as valid.

In conclusion, this study analysed how Weibo platform responded to rumour after the 2015 Tianjin blasts. Our findings suggest that this platform respond to rumour differently, which depends on the political sensitivity of the topics of rumour. In the second part of the research,

we examined how the platform's response to rumour might affect subsequent public discussion. Our findings suggest that in most cases the online media platform's measure to refute rumour can stimulate further discussion over the issue, whereas it is not evident to support a chilling effect as a result of the platform's censorship of rumour.

References

Allport, G. W., & Postman, L. (1947). The psychology of rumor. New York: Holt, Rinehart & Winston

Anderberg, M. R. (2014). *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks* (Vol. 19): Academic press.

Andrejevic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. New York: Routledge.

Arif, A., Shanahan, K., Chou, F.-J., Dosouto, Y., Starbird, K., & Spiro, E. S. (2016). How Information Snowballs: Exploring the Role of Exposure in Online Rumor Propagation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 466-477): ACM.

Bamman, D., O'Connor, B., & Smith, N. (2012). Censorship and deletion practices in Chinese social media. *First Monday, 17*(3).

Benney, J. (2013). The aesthetics of microblogging: How the Chinese state controls Weibo. *Tilburg paper in culture studies*. Retrieved from http://www.tilburguniversity.edu/upload/e4ef741d-2c6a-4ad9-9b17-32e103332502_TPCS_66_Benney.pdf

Berinsky, A. J. (2015). Rumors and health care reform: experiments in political misinformation. *British Journal of Political Science*, 1-22.

Bolsover, G. (2016). Harmonious communitarianism or a rational public sphere: a content analysis of the differences between comments on news stories on Weibo and Facebook. *Asian Journal of Communication*, 1-19.

Duffy, R. (2002) Hot gossip: rumor as politics. In Bondi, L. (Ed). *Subjectivities, knowledges, and feminist geographies: The subjects and ethics of social research*. Lanham, MD: Rowman & Littlefield.

Brehm, J. W. (1966). A theory of psychological reactance. New York: Academic Press.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *new media & society, 14*(7), 1164-1180.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology, 67*(2), 319.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675-684): ACM.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288-296).

Chin, J., & Mozur, P. (2013). China intensifies social-media crackdown. *Wall Street Journal (September 19)*.

China Youth Online (CYOL). (2015) Chuan bo she tian jin bao zhao shi gu yao yan yu 300 wei bo wei xin zhang hao suspended [Sharing rumor related to Tianjin blasts, over 300 Weibo and Wechat accounts were suspended]. Retrieved from http://zqb.cyol.com/html/2015-08/15/nw.D110000zgqnb_20150815_6-02.htm

CNBC (2015). Chinese police arrest 12 in Tianjin blast probe. Retrieved from http://www.cnbc.com/2015/08/26/senior-execs-at-warehouse-linked-to-tianjin-blast-arrested-report.html

Creemers, R. (2014). The privilege of speech and new media: Conceptualizing China's communications law in the internet era. *The Internet, Social Media and a Changing China, Pennsylvania, University of Pennsylvania Press, Forthcoming*.

DiFonzo, N., & Bordia, P. (2007). Rumor, gossip and urban legends. *Diogenes, 54*(1), 19-35.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing, 62*(1), 107-115.

Fine, G. A. (2007). Rumor, trust and civil society: Collective memory and cultures of judgment. *Diogenes, 54*(1), 5-18.

Fu, K. W., Chan, C. H., & Chau, M. (2013). Assessing censorship on microblogs in China: Discriminatory keyword analysis and the real-name registration policy. *IEEE Internet Computing, 17*(3), 42-50.

Fu, K. W., & Chau, M. (2014). Use of microblogs in grassroots movements in China: exploring the role of online networking in agenda setting. *Journal of Information Technology & Politics, 11*(3), 309-328.

Fung, C.I., Fu, K.W., Chan, C. H., Chan, B. S., Cheung, C. N., Abraham, T., & Tse, Z. T. H. (2016). Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors. *Public Health Reports, 131*(3).

Gillespie, T. (in press) Governance of and by platforms. In Burgess, J., Poell, T., & Marwick, A. (Eds), *SAGE handbook of social media.* Retrieved from http://culturedigitally.org/wp-content/uploads/2016/06/Gillespie-Governance-ofby-Platforms-PREPRINT.pdf

Gillespie, T. (2011). Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us. *Culture Digitally*.

Gillespie, T. (2015). Platforms intervene. *Social Media+ Society, 1*(1), 1-2.

Hiruncharoenvate, C., Lin, Z., & Gilbert, E. (2015). Algorithmically Bypassing Censorship on Sina Weibo with Nondeterministic Homophone Substitutions. In *Ninth International AAAI Conference on Web and Social Media*.

Hu, Y. (2009). Rumor as social protest. *The Chinese Journal of Communication and Society (9), 9*, 67-94.

Jiang, M. (2016). Chinese Internet Business and Human Rights. *Business and Human Rights Journal, 1*(01), 139-144.

Kaiman, J. (2013). China cracks down on social media with threat of jail for 'online rumours'. *The Guardian, 10*.

Kapferer, J. N. (2013). *Rumors: Uses, interpretations, and images*. Piscataway: Transaction Publishers.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review, 107*(02), 326-343.

King, G., Pan, J., & Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science, 345*(6199), 1251722.

Lee, N. (2014). *Facebook nation: Total information awareness*. London: Springer.

Manaf, M. M. A., Ghani, E. K., & Jais, I. R. M. (2013). Factors influencing the Conception of Rumours in Workplace. *Journal of Arts and Humanities, 2*(6), 50.

Meng, B. (2011). From steamed bun to grass mud horse: E Gao as alternative political discourse on the Chinese Internet. *Global Media and Communication, 7*(1), 33-51.

Ng, J. Q., & Landry, P. F. (2013). The Political Hierarchy of Censorship: An Analysis of Keyword Blocking of CCP Officials' Names on Sina Weibo Before and After the 2012 National Congress (S) election. In *Eleventh Chinese Internet Research Conference*.

Nip, J. Y., & Fu, K. W. (2016). Challenging Official Propaganda? Public Opinion Leaders on Sina Weibo. *The China Quarterly, 225*, 122-144.

Park, M. (2014). China's Internet firewall censors Hong Kong protest News. *CNN. com, 30*.

Procter, R., Vis, F., Voss, A., Cantijoch, M., Manykhina, Y., Thelwall, M., . . . Gray, S. (2011). Riot rumours: how misinformation spread on Twitter during a time of crisis. *Guardian, http://www. guardian. co. uk/uk/interactive/2011/dec/07/london-riots-twitter*.

Roberts, S. (2016). Commercial content moderation: Digital laborers' dirty work. In Noble, S. U. & Tynes, B. (Eds.) *The intersectional Internet. Digital formations series*. New York: Peter Lang.

Rogers, R. (2004). *Information politics on the Web*.Cambridge: MIT Press.

Rosnow, R. L. (1991). Inside rumor: A personal journey. *American Psychologist, 46*(5), 484.

Shibutani, T. (1966). *Improvised news: A sociological study of rumor.* London: Ardent Media.

Stockmann, D. (2015). Big Data from China and Its Implication for the Study of the Chinese State--A Research Report on the 2014 Hongkong Protests on Weibo. *Available at SSRN 2607998*.

The Guardian (2015). Tianjin explosion: China sets final death toll at 173, ending search for survivors. Retrieved from https://www.theguardian.com/world/2015/sep/12/tianjin-explosion-china-sets-final-death-toll-at-173-ending-search-for-survivors.

Travis, H. (2013). *Cyberspace Law: Censorship and Regulation of the Internet*. New York: Routledge.

Vaidhyanathan, S. (2012). *The Googlization of everything:(and why we should worry)*. Berkeley: Univ of California Press.

White, J. D., & Fu, K.-W. (2012). Who do you trust? Comparing people-centered communications in disaster situations in the United States and China. *Journal of Comparative Policy Analysis: Research and Practice, 14*(2), 126-142.

Worchel, S., Arnold, S., & Baker, M. (1975). The Effects of Censorship on Attitude Change: The Influence of Censor and Communication Characteristics1. *Journal of Applied Social Psychology, 5*(3), 227-239.

Wu, K., Yang, S., & Zhu, K. Q. (2015). False rumors detection on sina weibo by propagation structures. In *2015 IEEE 31st International Conference on Data Engineering* (pp. 651-662): IEEE.

Xinhua (2015) Guo jia wang xin ban yi fa cha chu 50 jia chuan bo she tian jin gang huo zai bao zha shi gu yao yan wang zhan [The Internet information office shutdown 50 websites that circulate rumour about the Tianjin balsts]. Retrieved from http://news.xinhuanet.com/newmedia/2015-08/15/c_1116265276.htm

Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on Sina Weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics* (pp. 13): ACM.

Yang, G. (2009). *The power of the Internet in China: Citizen activism online*. New York: Columbia University Press.

Zhu, T., Phipps, D., Pridgen, A., Crandall, J. R., & Wallach, D. S. (2013). The velocity of censorship: High-fidelity detection of microblog post deletions. In *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)* (pp. 227-240).

Figures

Figure 1: Log Odds ratios for each topic in the Weiboscope (WS) set and rumour-rebuttal (RR) set, with the colors of topic short names denote cluster membership.

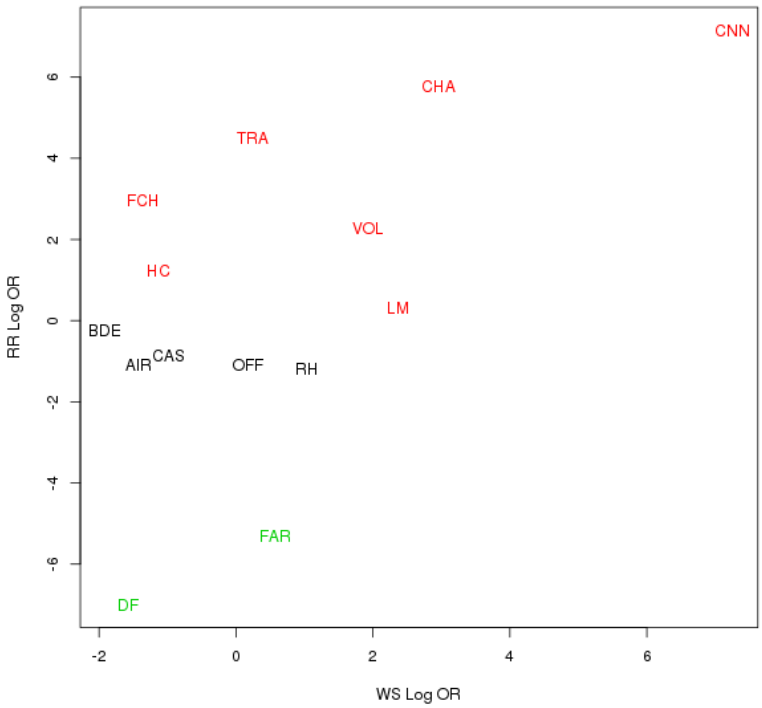
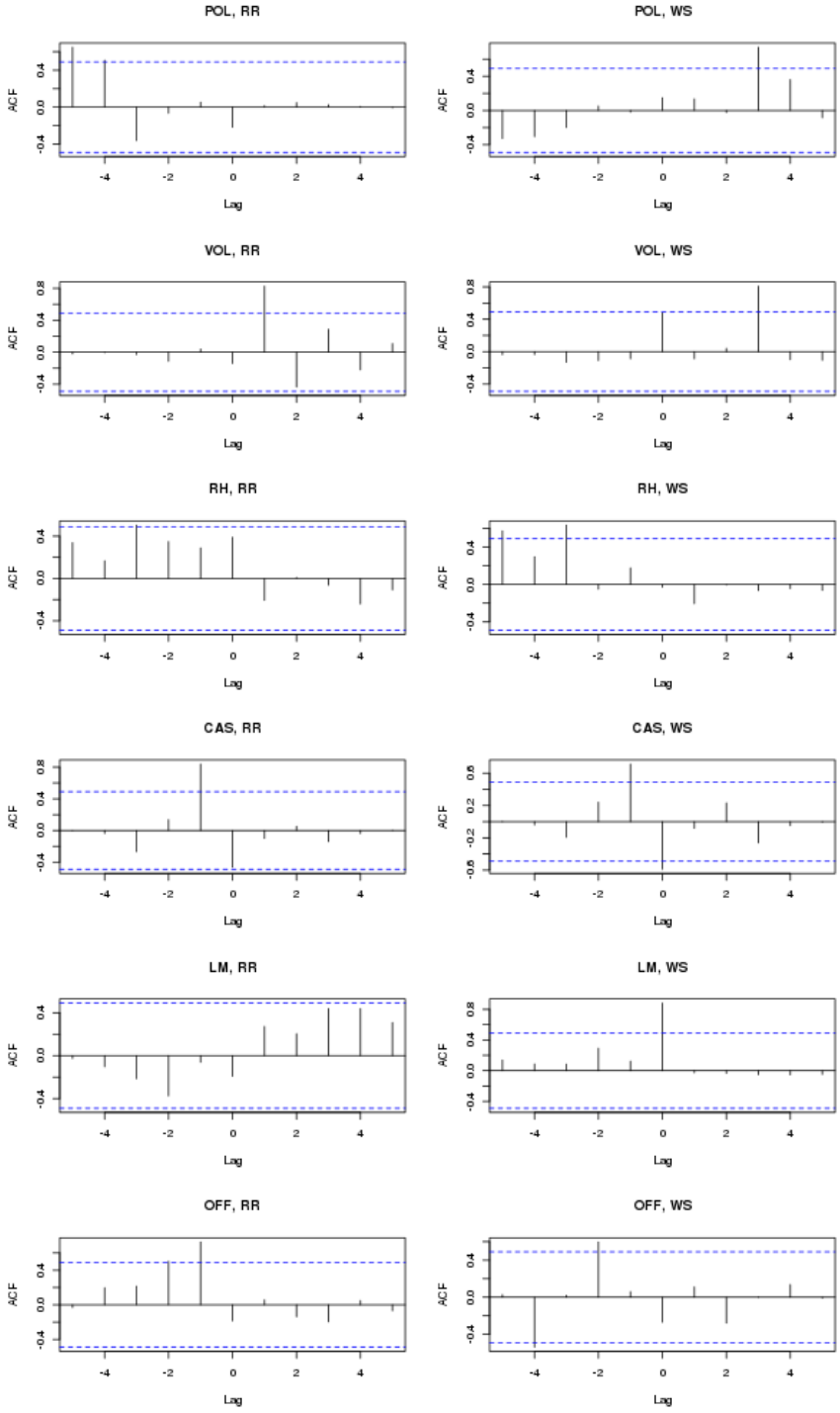


Figure 2: The lead-lag associations between Weiboscope set time series, rumour-rebuttal set time series and search set time series in six topics



Appendix 1. Basic statistics of rumour topics

Rumour topic	Abbreviation	Description	Posts in WS dataset	Posts in RR dataset	Total number of posts
Air-pollution	AIR	The blasts in Tianjin polluted the air in Beijing and Shanghai. The rumour advised people living in these two cities to close their windows.	30	809	2508
Casualty	CAS	Rumour about the actual number of people killed in this accident.	25	573	1679
Dead fish	DF	The explosions made rivers in Tianjin so polluted that large numbers of fish were killed	5	0	465
Ruihai background	RH	Claims of a link between the logistics company Ruihai (in who's warehouse the explosions occurred) and the government.	33	84	369
Officer	OFF	Rumours about specific government officials' involvement in the accident	11	66	261
Local Media	LM	The local media in Tianjin were not allowed to report on the accident.	54	184	366
Foam after rain	FAR	After the first rainfall since the explosion, Tianjin was covered in mysterious white foam, caused by chemical pollution.	8	0	100
"Burn down" principle	BDE	Firefighters was killed because they did not follow the 'burn-down principle' (which does not really exist)- firefighters should wait until everything is burnt down before accessing the site.	0	52	115
Volunteer	VOL	Volunteers were 'robbed' by local officials	14	401	457
Fake call for help	FCH	Rumour that a boy called Qile was missing	0	637	673
Health consequence	HC	The vibration of the explosions could cause organ damage and deafness in large numbers of residents living in the region.	0	103	132
Traffic	TRA	Rumour about traffic control after the explosion	0	604	611
CNN	CNN	Rumour that a CNN reporter was attacked by local government officials	31	581	612
Chaos	CHA	Rumour that the blasts lead to social chaos in the affected region.	0	162	162