**Analytic Techniques to Visualize Censored Content (Sina Weibo)**

Avidan Rothman

Montclair State University

Rothmana2@montclair.edu

**Abstract**

Censorship is one of the most parasitic ethical and moral problems the world is combatting today. It has infected every bit of the global social fabric makeup and we are seeing it turn entire countries upside down. The Chinese have been one of the guilty countries of this debilitating practice for quite some time, and they have become leaders in the evolution of it as well. The purpose of this study is to investigate characteristics used to identify surveillance targets on Sina Weibo. The study seeks to answer the following research questions. Are there data analytic techniques that allow for the discovery of trends within Sina Weibo data? In addition, how much does the motive of the topic play a factor in the censorship? This study would like to emphasize that idea and elaborate on the complexity of these censors. This issue is not something that can be easily broken down to a granular level and labelled as black and white. There are certainly a lot of gray areas which the purpose of the study wants to reveal through Topic Analysis, Word Embedding and Parts of Speech (POS). The goal is to shed some light on the censorship practice of Sina Weibo using these techniques to find trends that could lead to future research on this topic.

**1 Introduction**

Sina Weibo is one of the largest social media platforms in the world. It was launched in 2009 by the Sina Corporation. As of June 2021, Weibo has 566 million monthly active users. The “Great Firewall” was created by China to prevent the spread of misinformation however there is significant evidence that shows this is rather a gray area in terms of what is deemed misinformation and rather a blatant use of government power to censor content that is deemed unfit by the government. Although Sina Weibo is not the only platform that has been used for this type of censorship being that Sina Weibo is one of the largest platforms, we felt this would be a good opportunity to be able to work with a large dataset that would yield positive results. The rise of social media platforms has opened the flood gates for information to be shared and with that has created a way for people to connect like never before. Despite social media’s initial purpose being this binary bridge to bring people together. The world has unfortunately begun to witness in the last ten years the dangers of these platforms through the spread of misinformation and censorship. This has been backed up by the observations made by the human rights watch in the article “In China, the 'Great firewall’ Is Changing a Generation”. They state “The past 10 years in China have seen a combination of communications crackdown, ramped-up propaganda, and rapid expansion of surveillance efforts that—when paired with China’s rising global ambitions—have changed the public conversation in China, even among educated and younger people. It will make it harder, even in a post-Trump world, for the world’s great powers to avoid splitting further apart, perhaps dangerously.” [1] In addition, they state that the optimism of China becoming a more open and diverse country over the past decade has evaporated since Xi Xinping became the paramount leader back in 2012. When ‘Document 09’ was issued in 2013 online activism continued on an even steeper downward spiral. Fast forwarding to 2020 when Covid-19 exploded we began to see an even scarier and very real and alarming example of a global event that has through the scope of social media has raised a major red flag on how these platforms can be used in a rather malicious way. It has become especially prevalent here in the United States as well as countries like China. The growing concern is the involvement of governments and how they are leveraging these platforms for surveillance and outright censorship. Being that social media platforms have for the most part attempted to prevent the spread of misinformation and even attempts to censor specific information there is a lack of accountability being applied to these platforms. In addition to misinformation, there is a major problem of information being censored because it contradicts the views of either the government itself or the views of the platforms. One of main questions that this research is attempting to answer is through three common data analytic functions can we find any trends or patterns in how data is censored? In this research Sina Weibo was chosen for its controversial history as a social media platform to conduct our experiment to find out if these trends exist and how in the future more research could be done to find out more. Targets will include the recency of the content chosen for censorship as well as the time frame that the content remained censored. The goal is to see if there are trends based on text of the post as well as the topic of the post. The second criteria could also shed some light on virality as the more a topic is spoken about the higher the virality This will be answered by analyzing the posts scraped from Sina Weibo and put through our three data analytic tools created to perform three functions. Parts of Speech (POS), Word Embedding, and LDA Topic Analysis.

**2 Related Work**

**Object based method of censorship on Twitter**

Censorship has become a mainstay in the social media spaces across the globe. We can see from examples here in the United States that social media platforms are in some ways complacent with these practices and have been criticized for not standing up for freedom of speech. In a study performed by Tugrulcan Elmas, Rebekah Overdorf, Karl Aberer [2] they collected twitter data from twitter API and found that tweets are broken down into two parts tweet object and user object within these objects they found that tweets or retweets that were censored had an object field called “withheld\_in\_countries”. With this information they came up with a strategy to see which users had their profiles censored at some point in the past. With this idea they observed that some of the users that were censored were from retweets and not the original tweet users. Based on the findings in this study one can conclude that there may be some applicability to the users in Sina Weibo. Where one might see a difference is where the focus should be directed towards instead of focusing on users that have a lot of followers the criteria for censorship on Sina Weibo rather should be on the type of content that a person puts out that has a lot of followers. In addition, including the virality of the topic and if it could lead to the post being censored. Another interesting find from this study is that censored users are mostly based in a country other than the one that their tweet was posted in. This is another interesting idea that can have applicability to how users are censored on Sina Weibo and determine if there are any users that are censored in China when they send their message from outside the country.

**Real-time analysis of Geographics in Censorship**

Internet censorship has many different forms, wide scale disruptions that affects all users, blocking specific “hot” words or topics, or even targeting a specific user based on their status or location. In another study that focuses on the multi-perspective view of how internet censorship occurred in Myanmar [3] they used various complementary methods to obtain their data. The first method was called IODA (Internet Outage Detection and Analysis). This system monitors the internet continuously to try and identify macroscopic internet outages hitting a specific network. The second method is OONI (Open Observatory of Network Interference), this is a project develops open-source software built to measure the blocking of websites, instant messaging apps and other platforms. OONI is run by volunteer that have authorization in around 200 countries. The final method is monitoring internet traffic itself with the help of Kentik a large network observability company. Incorporating these research methods could be a good opportunity to help in helping find trends or examples of censorship on Sina Weibo that help this study gain evidence to answer the research questions. Where the methods of this study differ from the study mentioned above is the acquisition and use of the data for the study. OONI was used in the Myanmar study whereas the main source of data from our study comes from a mined database from freeweibo.com. In addition, the Myanmar study is focused on censorship as a whole in Myanmar whereas our study focuses on Sina Weibo specifically within China.

**Content deletion of Multimedia on Sina Weibo**

While acknowledging that text-based censorship is the main focus of this research it is important to not forget that multimedia censorship has a major role in online censorship and continues to grow at a rapid pace. In the study, “More than plain text: Censorship deletion in the Chinese social media” [4] they highlight the growth of multimedia censorship and shy away from text-based criteria for why a post is censored. An interesting point made specifically with the platform Sina Weibo is that “A report by Sina Data Center (2018) has acknowledged that among original posts on Weibo, 49.1% have images and 36.5% include video. A bias toward textual communication hence passes over how censors adapt their practices in response to the richness of the online practices of Chinese Internet users and the sophistication of corresponding censorship.” In the study mentioned above they primarily focus on content deletion. Based on the objective of answering the question on what constitutes a post being censored on Sina Weibo focusing only on content deletion only shows part of the story because it’s not just about what content is deleted because of a censor it is also about user suspension as well as timeline/timeframe that something stays censored. Due to the fact that they are focusing on Multimedia as the medium for censorship content deletion makes sense. When someone posts a picture that picture is unique making it easier for censor to delete because once deleted its gone and there isn’t any detriment to that. The difference between text-based and media-based censored is that text-based censors can’t delete text because then that word that is deleted is unable to be used without triggering a censor. Weiboscope was one of the primary sources that “More than plain text: Censorship deletion in the Chinese social media” [4] uses and also used the Weibo’s User Timeline API function to identify surviving posts. Using these methods poses a problem that is mentioned in the paper which is that after they identified a post that was missing, they could not distinguish whether the post was deleted by the user intentionally or the censor removed it for its content.

**Demographics and other pseudo random criteria in the application of censorship**

Demographics certainly play a part in censorship especially in China, there has been extensive research on this. In the study “User Demographics and Censorship on Sina Weibo” [5] they make the strong and compelling argument that demographics play a huge role in censorship as well as other factors like sex and verified status. The idea that demographics plays a big role in censorship has become a growing thought and through this study shows that to be a very real argument. The use of other criteria to gain knowledge of a user besides for the user ID number on Sina Weibo is where the study’s methods create a unique perspective that adds on to very scary picture of how users are labelled. The methods here focus on how these criteria affect censorship but do not focus on why they are censored. In one of the results table that showed the results of demographics based on region showed that third most demographic geographically censored was actually from outside of China. This correlation to the study “A multi-perspective view of Internet censorship in Myanmar” [3] further makes the argument that censored post coming even from outside China on Sina Weibo are being censored. Bringing these two criteria (Geography and Demographics) to be incorporated into the idea that age and virality are additional methods applied to censorship paints a very alarming and diverse picture of the censorship landscape in China.

**Patterns based censorship associated with reposted censored posts**

The idea of posting reposted censored content on Sina Weibo is a fascinating idea considering the fact that depending on how it is accomplished can shed light on the ways that people or organizations are trying to circumvent the censors found on Sina Weibo. In the study, “Understanding Patterns of Users Who Repost Censored Posts on Weibo” [6] they found the correlation between demographics and the likelihood of a censored being reposted to be high. This changes dramatically when they reviewed users that were certified or had higher social status. This finding is similar to the study “User Demographics and Censorship on Sina Weibo” [5] where there is a distinction is obviously the reposting of censored content. When bringing up the idea of virality there is something that when adding this criteria of reposting censored content creates an interesting conundrum with a few alternative questions. When a post gets censored due it going viral how does that affect the time frame that a post would be censored for? Does this affect the virality of a post? When a post is reposted does it extend the posts speed to “Post-virality” as in it is no longer considered viral and people aren’t talking about it anymore. Having this incorporation gives more possibilities to study censoring and more importantly future of censoring technology.

**Tonal spoofing used to bypass Censors on Sina Weibo**

The idea of trying to bypass censors is something that has been around since censoring began in China back in 1996. In the study “Algorithmically Bypassing Censorship on Sina Weibo with Nondeterministic Homophone Substitutions” [7] takes this idea of bypassing censors on Sina Weibo to a whole other level. “In our work, we focus on Mandarin Chinese, China’s official language. Mandarin Chinese is a tonal language: each character’s sound can be de-composed to a root sound and its tone. Some characters con-vey multiple meanings and might be associated with multiple sounds based on the meanings they convey. While the tone of a sound can change a word’s meaning, native speakers can often detect an incorrect tone by referring to its sur-rounding context.” This idea to essentially spoof Sina Weibo into thinking a post contains no harmful content by simply changing the way word is read is a really impressive method. Further examining this research, they determined that the age of a post in their context was a 48-hour period of time. This idea is an interesting concept that is applicable to the goals of determining the age and virality criteria. 48-hours seems like a short period of time especially when considering the idea of virality because depending on the topic the level at which a topic is discussed online can surely surpass the 48-hour mark. This is worth investigating further because this is where the above study does not go into further detail discussing the decision to mark the age of a post at 48-hours to determine if it was censored or not.

**3 Methodology**

In order to answer the questions proposed we developed a FreeWeibo data-scraper that scrubs freeweibo.com for posts that are no longer visible on Weibo and flagged as censored and stored in a database. From this database we applied the three programs developed to perform three different data analytic functions.

**3.1 Datasets**

The data collected in this project represents censored posts that were at one point posted on Sina Weibo and have since been censored and then stored on the website FreeWeibo.com. The dataset used is a collection of Sina Weibo posts containing information like Weibo IDs, Username, time created. The data was scrapped from August 2021 until September 2022. The date range that the data was posted on Sina Weibo is from December 2010 until September 2022. Duplicate posts were removed using the Unique ID FreeWeibo assigned giving us a clean dataset. The total number of posts collected was 20,965.

**3.2 Data collection process**

In order to be able to investigate the censorship occurring on Sina Weibo we needed to collect data from the website and create a database that we could reference in our data analytic tools to get a result. FreeWeibo.com in figure 1, is the main source for our database and, as mentioned in the datasets section contains the various data-points that are highlighted in the results.

Graphical user interface, text, email

Description automatically generated

**Figure 1: FreeWeibo Website**

Once the data is pulled from FreeWeibo.com it is stored in the database and from that database the NLP processes pull the data to be analyzed, see figure 2.

Diagram

Description automatically generated

**Figure 2: Data collection flowchart**

**4 Results**

With the data collected from FreeWeibo.com and stored on our database we can begin to build the front-end interface with Streamlit. Building the interface into Streamlit required converting the code built in our Jupyter Notebook into a raw python code. Due to the fact that Jupyter Notebook projects are saved with the suffix .ipynb the code has to been saved as a simple .py. Once the conversion is done, we needed to install on our machine the Streamlit program. After that process is complete, we can start to build in a text editor the code that will run the Streamlit code. Before the Streamlit code can be run we need to create an environment in Anaconda Navigator to run the Streamlit code. Streamlit runs automatically when opening a terminal in Anaconda and entering the code “streamlit run filename.py” The file structure has to have the code in the same directory in order to run for all files created to be run in Streamlit. After the code is initiated, the code will display the local URL and network URL as well as automatically open the browser see figure 3.

Text

Description automatically generated

**Figure 3: Running Streamlit from Terminal**

The Streamlit consists of four different parts. The first part consists of the Streamlit code that builds the interfaces main page see figure 4.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 4: Main Page Streamlit code**

Here we are introducing the first Streamlit specific code that interacts with the Python code. “Import Streamlit as st”. This is the most essential code for running any Streamlit functions within the Python code. This code allows for the use of any Streamlit code to recognized for the front-end interface. After this code we add code that will be visible in the front end like links and headers, etc see figure 5.

Graphical user interface, text

Description automatically generated

**Figure 5: Main Page from front-end**

**4.1 Parts of Speech (POS)**

Parts of Speech is the first of the Natural Language Processing (NLP) technique we used in our project. Parts of Speech is a process tagging a collection of words to a corresponding part of speech, based on the word's context and definition. The purpose of this technique in our project is to identify which words from the thousands of posts occur the most. From this we can infer that the word could potentially be a trigger for the censor and the reason that the post was censored in the first place. Implementing the parts of speech is a complex process and getting the results to display in Streamlit was a challenge see figure 6.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 6: Applying parts of speech to Segmented list of Words**

After applying the complete program with the Streamlit functions we get to the data visualization where we are able to display the results of the parts of speech function on our dataset see figure 7.

Chart, bar chart

Description automatically generated

**Figure 7: Count of full dataset broken down to parts of speech**

From this visualization we can see that the largest number of words that are censored are nouns. Now it seems obvious that nouns would be the largest part of speech censored but it is important to remember that the dataset is Chinese and with that comes a very large vocabulary. In the next figure we see why this becomes an interesting observation, see figure 8.

Chart, bar chart

Description automatically generated

**Figure 8: Top 10 nouns in the dataset**

From the above figure we can see that the most common noun is 人 which means person. When combined with the Word Embedding where we can see that the above word has different iterations of the word and therefore it has different meaning, and we can infer that if the word was change it may have allowed the censor to be bypassed.

**4.2 Word Embedding**

Word Embedding is a vector representation for a specific word in a sentence. How Word Embedding works is by classifying words based on similar context into vectors and then mathematically calculating the cosine of that angle between the vectors. The closer two words or sentences are to one the more similar the word or sentence is. The purpose of this technique is to compare posts of users and give those post the corresponding mathematical calculations associated with word embedding, see figure 9.

Text

Description automatically generated

**Figure 9: Computing the cosine for the first 10 posts**

The above code shows the cosine mathematical equation that is needed to create the word embedding. The code applies the cosine to the first 10 posts and then assign the score. In the Streamlit interface we have the user enter a Weibo ID to get the top 3 most similar IDs to the one that the user entered, see figure 10.

Graphical user interface

Description automatically generated with medium confidence

**Figure 10: Streamlit interface for Word Embedding**

In Streamlit we built a input for a user to interact with. The user is asked to enter a Weibo ID and from there we complete the comparison of that Weibo ID to 3 other IDs and give them an associated score based on the Cosine vector analysis. Below the input we have the 2D heat map which is used to visualize the data that the previous output displayed. What can be interpreted from the data is that similar words could lead to other posts being censored just by being associated with a post that was censored previously. This is something that in future work should be investigated further as this could be a real insight discovery into determining one factor that leads to posts being censored on Sina Weibo.

**4.3 LDA Topic Analysis**

LDA Topic Analysis is a topic model that is used to categorize words from a document to a specific topic. Where this model is applicable to our dataset is this model is essentially creating a “Hot Topics” analysis. Similar to the way that FreeWeibo.com lists in Hot Topics based on the recent posts it is storing here we are seeing based on the posts themselves what topics are able to be deduces from them see figure 11.

A computer screen capture

Description automatically generated with medium confidence

**Figure 11: LDA Topic Analysis Model Training**

From this code we are able to create a word cloud visualization that includes words from our dataset and display it in a unique way., see figure 12.

A picture containing text

Description automatically generated

**Figure 12: WordCloud Visualization**

Taking this analysis further we were able to create using a function called LDAvis which is an interactive tool being able to see which words are most relevant depending on the Topic and even seen which Topics may have some overlap leading one to believe that these could be some type of sub-topic, see Figure 13.

Chart

Description automatically generated

**Figure 13: LDAvis interactive visualization**

The takeaway from this analysis is that in future work someone could dig deeper into and see how these topics and the words associated with them could be bypassed. Combining the three models together can be the best way to answer the question of whether the censors on Sina Weibo can be censored and also why specific words/posts are censored in the first place especially when the topic or content of the post is necessarily assumed to be triggering.

**Conclusion**

While other studies have focused specifically on online censorship in various ways, this study presents not only an opportunity for more research to build off of what has been started. Using Streamlit to visualize the data gives a user another way to interpret the data and be able to draw new conclusions based on it. We find that in the Parts of Speech analysis that because nouns are the prominent part of language using this in a predictive manner could bring insight into how to overcome censorship. In tandem with POS Word Embedding builds a strong tool to be able to not only predict words that are triggers but could give users an ability to bypass the censor. This of course is something that has to be done in future research, but this project laid the groundwork to be able to accomplish that. LDA Topic Analysis was used in this project to visualize Topics that were shared between posts and gave some insight into what topics are the most triggering. Although further digging is necessary due to the language gap between Chinese and English.

**References**

[1] Wang, Y. (2020, October 28). *In China, the 'great firewall' is changing a generation*. Human Rights Watch. Retrieved August 7, 2022, from [https://www.hrw.org/news/2020/09/01/china-great-firewall-changing-generation](https://www.hrw.org/news/2020/09/01/china-great-firewall-changing-generation%20)

[2] Elmas, T., Overdorf, R., & Aberer, K. (2021, March 19). *A dataset of state-censored tweets*. arXiv.org. Retrieved May 18, 2022, from <https://arxiv.org/abs/2101.05919>

[3] Padmanabhan, R., Filastò, A., Xynou, M., Raman, R. S., Middleton, K., Zhang, M., Madory, D., Roberts, M., & Dainotti, A. (2021, August 23). *A multi-perspective view of internet censorship in Myanmar*. A multi-perspective view of Internet censorship in Myanmar. Retrieved May 18, 2022, from [https://dl.acm.org/doi/pdf/10.1145/3473604.3474562](https://dl.acm.org/doi/pdf/10.1145/3473604.3474562%20)

[4] Liu, J., & Zhao, J. (2020, May 30). *More than plain text: Censorship deletion in the Chinese social media*. Researchgate.net. Retrieved May 18, 2022, from [https://www.researchgate.net/profile/Jun-Liu-38/publication/342737890\_More\_than\_plain\_text\_Censorship\_deletion\_in\_the\_Chinese\_social\_media/links/5f47642c92851c6cfde5202b/More-than-plain-text-Censorship-deletion-in-the-Chinese-social-media.pdf](https://www.researchgate.net/profile/Jun-Liu-38/publication/342737890_More_than_plain_text_Censorship_deletion_in_the_Chinese_social_media/links/5f47642c92851c6cfde5202b/More-than-plain-text-Censorship-deletion-in-the-Chinese-social-media.pdf%20)

[5] Kenney, W., & Leberknight, C. S. (2021, September 1). *User Demographics and Censorship on Sina Weibo*. Aisel.asinet.org. Retrieved May 18, 2022, from <https://aisel.aisnet.org/hicss-54/dsm/data_mining/6/>

[6] Qian, Y., Yuan, F., Lyu, H., & Luo, J. (2021). (thesis). *Understanding Patterns of Users Who Repost Censored Posts on Weibo*. Retrieved from <https://arxiv.org/pdf/2101.06864.pdf>.

[7] Hiruncharoenvate, C., Lin, Z., & Gilbert, E. (2021). View of algorithmically bypassing censorship on Sina Weibo with nondeterministic homophone substitutions. Retrieved July 11, 2022, from [https://ojs.aaai.org/index.php/ICWSM/article/view/14637/14486](https://ojs.aaai.org/index.php/ICWSM/article/view/14637/14486%20)