**1. Project Title and Team**

Detect China Online Censorship - The impact of major political events on Weibo posts

Team members: Yulun Han (12332015), Xin Tang (12331956), Hazel Chui (12325007), Peihan Gao (12336506)

**2. Project Summary**

The increasing use of social media platforms in China has led to a growing need to understand the impact of censorship on online public opinion. As the Chinese government reinforces censorship measures before major political events to ensure their success, it is essential to investigate how these measures affect the public's posts on social media platforms such as Sina Weibo (Weibo).

This paper focuses on analyzing the public's posts on Weibo before and after the Twentieth National Congress of the Communist Party of China (the 20th CPC National Congress) to investigate the impact of censorship on public opinion. In this paper, we use computational methods to analyze public posts on Weibo before and after the 20th CPC National Congress. Our findings could provide insight into the effectiveness of censorship measures in controlling public opinion during major political events in China.

Our work differs from previous studies in its use of computational methods to analyze a large dataset of social media posts. Additionally, our focus on a specific major political event allows for a detailed examination of the impact of censorship on public opinion during a specific time period.

**3. Related Work**

Research on online censorship in China has focused on analyzing the content of censored social media posts and identifying keywords that are likely to be deleted by the government. King, Pan, and Roberts (2013) found that the censorship program aims to silence collective expression, but not criticism, by analyzing censored social media posts. This conclusion was supported by a separate study in which they created accounts and randomly submitted texts. Similarly, Bamman, O’Conner, and Smith (2012) uncovered keywords that are likely to be censored by the Chinese government through an analysis of deleted Weibo posts. Additionally, Tai and Fu (2020) found that even articles on seemingly non-sensitive topics and those containing pro-regime messages are also removed by censors on WeChat.

Several studies have found that the level of censorship in China may not remain constant over time. Han and Shao (2022) show that the Chinese state scales up control over citizenry complaints in response to a series of important political events. They find that the government tightens information control before important political events. Ruan et al. (2020) have demonstrated that by studying WeChat, a popular Chinese communication app, the government alters censorship patterns before, during, and after significant events. The results suggest that censorship in China may be dynamic and responsive to political developments.

Our study aims to expand upon previous research on censorship in China in two key ways. First, instead of analyzing deleted posts, our study concentrates on Weibo posts that have not been censored. This approach enables us to examine the atmosphere that the government desires to create before and after symbolic political events, as well as to evaluate users' self-censorship or platforms' filtering effects. Second, to the best of our knowledge, no researchers have studied the impact of major political events on censorship in Weibo, which is the most widely used microblog in China. Weibo posts are an important source for understanding the outcome of the interaction between censorship and critical political events.

**4. Project Preparation and Prerequisites**

In this project, we want to find out how the Chinese government censors’ social media and figure out how much it does it. In most cases, more social stability and speech censorship might be required to keep things running smoothly when significant political events are held in China. Weibo is China’s most often-used social media platform for people to share their thoughts and emotions towards a specific topic. Before December 2022, China put out a set of strict rules to stop pandemics during COVID-19. Some of these rules raised public outrage.

We will scrape Weibo posts mentioning keywords including “疫情“ (”covid“), ”隔离“ (”quarantine”), and “抗疫“ (”prevention and control of covid-19”) spanning one month before and after the 20th CPC National Congress (October 16th, 2022) as our main observational sample, and analyze the change of sentiments, keywords, and topics of these posts to measure the potential censorship from the Chinese government.

We have the following hypotheses: Our research will demonstrate that, during significant political events, it is much more difficult for people to disagree with policies on social media due to censorship. Our analysis will highlight the significant decrease in posts containing politically sensitive keywords that occurred after the conference. Our findings may demonstrate how successfully censorship is used in China to control public opinion during significant political events.

We'll use sentiment analysis to compare the evolution of posts' sentiments and display the trend of sentiment scores in a line graph. The keywords for posts will be displayed as word clouds, with the size of the words indicating how frequently they occur. In order to analyze content changes, we will then use topic modeling. Our expected results will show that censorship measures significantly impact the public's ability to express dissenting opinions on social media during major political events. Specifically, our analysis will show a significant decrease in posts containing keywords related to sensitive political topics after the conference.

**5. Evaluation**

We have decided to evaluate this project by 1. comparing sentiment scores before and after (the 20th CPC National Congress) to see if the difference is significant or if the scores were positive before the conference and negative after that; 2. comparing the word clouds for the two time periods and analyzing the relatively bigger words in each of the graphs; 3. analyzing the topic modeling results for the two time periods and determining if the topics the models identified make sense and if the topics can be informative. 4. deciding if the methods are feasible. For example, if we can find a sentiment measuring function to reasonably analyze the sentiment for this specific dataset.

**6. Ethics**

There would be biases in our project. For one thing, our result might be influenced by “解封” (the lift of lockdown), which happened approximately two months after the conference. There is a chance that during the month after the conference, there have already been some changes in the policies that China chose to implement. For another, we might not be able to detect significant results if the sample size that we get is not big enough due to different reasons e.g. unable to scrape a sufficient number of Weibo posts because of rate limits.

**References**

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

Han, R., & Shao, L. (2022). Scaling authoritarian information control: How China adjusts the level of online censorship. Political Research Quarterly, 75(4), 1345-1359.

King, G., Pan, J., & Roberts, M. E. (2013). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. American Political Science Review, 107(2), 326-343.

Ruan, L., Crete-Nishihata, M., Knockel, J., Xiong, R., & Dalek, J. (2021). The intermingling of state and private companies: analysing censorship of the 19th National Communist Party Congress on WeChat. The China Quarterly, 246, 497-526.

Tai, Y., & Fu, K. W. (2020). Specificity, conflict, and focal point: A systematic investigation into social media censorship in China. Journal of Communication, 70(6), 842-867.