

# Sinophobia was popular in Chinese language communities on Twitter during the early COVID-19 pandemic

Yongjun Zhang<sup>1\*</sup>, Hao Lin<sup>1</sup>, Yi Wang<sup>2</sup>, Xinguang Fan<sup>3</sup>

<sup>1</sup> Department of Sociology and Institute for Advanced Computational Science, Stony Brook University, USA

<sup>2</sup> Department of Asian and Asian American Studies, Stony Brook University, USA

<sup>3</sup> Department of Sociology, Peking University, China

\* Direct correspondence to Yongjun.Zhang@stonybrook.edu

## Abstract

The COVID-19 pandemic has led to a global surge in Sinophobia. We examine how Chinese language users responded to COVID-19 on western social media by compiling a unique database (CNTweets) with over 25 million Chinese tweets mentioning any Chinese characters related to China, the Chinese Communist Party (CCP), Chinese, and Asians from December 2019 to April 2021. Our analysis of Twitter users' self-reported geographic information shows that most Chinese language users on Twitter originated from Mainland China, Hong Kong, Taiwan, and the United States. We then adopt the Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa) and structural topic modeling to further analyze the sentiments, content, and topics of Chinese tweets during the COVID-19 pandemic. Our results suggest that 61.8% of tweets in our database were contributed by only 1% of Twitter users and 62.2% of tweets were negative toward China. Despite the prevalence of anti-China sentiments, the target entity analysis shows that these negative sentiments were more likely to target the Chinese government and CCP than the Chinese people. Our findings also show that the most popular topics were politics (e.g., Hong Kong protests and Taiwan issues), COVID-19, and the United States (e.g., the US-China relations and domestic issues). Anti-China users focused relatively more on political issues such as democracy and freedom, while pro-China users mentioned cultural and economic topics more. Our social network analysis reveals that these pro-China and anti-China Twitter users lacked in-depth engagement in China-related conversations and were highly segregated from each other. We conclude by discussing our contributions to China and social media studies and possible policy implications.

## Introduction

The COVID-19 pandemic has led to increased xenophobia and racism towards Chinese communities (Lee & Huang, 2021; D. Zhang, 2021). China, being the first country to report cases of the coronavirus, has been the subject of misinformation regarding the origin of COVID-19, which has fueled a global surge in Sinophobia (Cook, Huang, & Xie, 2021). Recent scholarship has examined public sentiment towards China and the Chinese government, with one

strand of research analyzing how social media users and media outlets framed China during the early stages of the pandemic. For instance, Cook et al. (2021) found that the pandemic led to a sharp rise in anti-China attitudes in the United States, based on an analysis of English-language tweets. Similarly, Fan and Zhang (2023), analyzing web news on China from media outlets worldwide, found a significant increase in racial slurs targeting China during the early pandemic that persisted even after the World Health Organization warned against misinformation about COVID-19. Meanwhile, there is another strand of literature assessing how Chinese citizens responded to COVID-19 on domestic social media platforms. For example, Lu, Pan, and Xu (2021) found that Sina Weibo users in China were more supportive than critical due to the effective COVID-19 responses by the Chinese government. It is unclear, however, how Chinese language users on western social media platforms like Twitter discussed COVID-19 and their sentiments towards China, as previous studies have focused either on English social media users or Chinese domestic users in a censored environment.

Recent research has shown that the COVID-19 crisis increased censorship circumvention and access to international news and political content on websites blocked in China (K.-C. Chang, Hobbs, Roberts, & Steinert-Threlkeld, 2022). When individuals seek crisis-related information, they may also come across unrelated information or misinformation that has long been censored by the government. These users may actively engage in social media conversations and increasingly influence public opinion in international society. There is reason to expect positive sentiments toward China among Chinese-language users due to the potential political propaganda by the Chinese government and cyber-nationalists. However, since China has censored most international social media platforms (Hobbs & Roberts, 2018), Chinese-language users on Twitter may represent a very selective group, such as overseas Chinese, residents from Hong Kong, Taiwan, and Singapore, Mainland Chinese with VPN access, and other organizations and bots criticizing or supporting China. Such selectivity may lead to polarized sentiments toward China in Chinese-language conversations. Although scholars have presented evidence on the popularity of Sinophobia among English tweets, little is known about the sentiments within the Chinese language communities and what drives these patterns.

To fill the research gap, this article examines how Chinese language users on Twitter engaged in China-related discussions and the associated sentiments during the early COVID-19 pandemic. Specifically, in the Twitter verse, who were those Chinese language users tweeting China-related issues during the pandemic? After the COVID-19 outbreak, how did Chinese language users on Twitter discuss the pandemic and China? What were the main public sentiments toward China? Were they targeting the Chinese people or the Chinese government? Did those pro-China and anti-China users engage in each other’s debate?

To address these questions, we queried the Twitter historical database using keywords related to China, Chinese, the Chinese Communist Party (CCP), and Asians in both simplified and traditional Chinese languages to generate our Chinese Tweets (CNTweets) analytic dataset with over 25 million Tweets by 1.32 million Twitter users between December 2019 and April 2021. We then annotated a training dataset with 10,000 tweets to build a series of deep learning algorithms to classify the sentiment and topics in these tweets by fine-tuning pre-trained Chinese Robustly Optimized Bidirectional Encoder Representations from Transformers with the Whole Word Masking models (Chinese-RoBERTa-wwm-ext) (Devlin,

Chang, Lee, & Toutanova, 2018; Liu et al., 2019; Cui, Che, Liu, Qin, & Yang, 2021).

## Source of Chinese Tweets

Twitter has been blocked by the Chinese government since 2009 due to information control, so regular Mainland Chinese internet users have to rely on virtual private network (VPN) services to access Twitter (Sullivan, 2012). As a result, Mainland Chinese users on Twitter might be a very selective group of individuals, such as lawyers, journalists, and human rights activists, seeking uncensored information and discussing sensitive topics that are not allowed in China (Song, Faris, & Kelly, 2015). These anti-Chinese state users are not the only Mainland users who can circumvent the Great Fire Wall. Previous research also shows the prevalence of pro-Chinese state users, for instance, state-sponsored institutional accounts with free access to Twitter and regular pro-China internet users. China has initiated its own foreign propaganda program mainly carried out by state-run media enterprises, such as China Central Television, China Daily, Global Times, and Xinhua News. Individual pro-state users could be part of the paid 50-cent party, government employees, and other regular nationalistic internet users (Bolsover & Howard, 2019; King, Pan, & Roberts, 2017). In addition to Mainland Chinese, Chinese language users on Twitter could stem from other countries and regions with a population of Chinese language speakers, overseas Chinese, or immigrants of Chinese descent, such as Hong Kong, Macao, Taiwan, Singapore, Thailand, the US, Australia, and Canada. Twitter has been a battlefield for anti-Chinese state groups with few financial resources who are using Twitter to spread misinformation and disinformation on China and Chinese politics (Bolsover & Howard, 2019). The diversity of Chinese language users on Twitter motivates our first research question pertaining to the sources of Chinese tweets.

RQ1: Who were those Chinese Twitter users mentioning China-related issues during the early pandemic?

## Sentiment of Chinese Tweets

A large body of literature has used Twitter to gauge public sentiments and the associated impacts on political, economic, and social outcomes, such as elections (Tumasjan, Sprenger, Sandner, & Welpe, 2010; Bovet & Makse, 2019; Shmargad, 2018), stock market (Ranco, Aleksovski, Caldarelli, Grčar, & Mozetič, 2015), and public policies (Flores, 2017). Like other social media platforms (e.g., Weibo, Facebook), public sentiment on Twitter is a mix of regular internet users, opinion leaders, organizations, and social bots, and it is part of the algorithmically infused societies co-shaped by algorithmic and human behavior (Wagner et al., 2021).

Prior studies show that both pro- and anti-Chinese state groups have used Twitter as a platform to serve their propaganda purposes (Bolsover & Howard, 2019). However, these studies tend to focus on non-Chinese audiences, and limited research has examined how these groups target Chinese language users on social media platforms. For instance, Bolsover and colleagues find no evidence of pro-Chinese state computational propaganda on Twitter but strong evidence of massive tweets associated with anti-Chinese state perspectives published in simplified Mandarin (Bolsover & Howard, 2019). This is partly due to the fact that

China’s foreign propaganda has been carried out by traditional state-run media groups such as China Central Television and Global Times with massive human and monetary resources. However, these anti-Chinese state groups have used computational propaganda to promote and disseminate their messages targeting the Chinese government due to its lower operating costs. Thus, we might observe a lot of anti-Chinese state behavior on Twitter.

For pro-Chinese state groups, prior studies have shown the rise of Chinese digital nationalism (DeLisle, Goldstein, & Yang, 2016; Schneider, 2018). Cyber nationalists, especially young Chinese internet users, have defended China and the Chinese government on western social media platforms without state blessings, such as Little Pink (xiaofenhong, i.e., young Chinese nationalists on the internet) and Diba Expedition (diba chuzheng, i.e., cyber-nationalism organized by the Diba, a Chinese online community) (Han, 2019; Bi, 2021). These cyber nationalists tend to engage in conversations with their opposing groups instead of posting comments like social bots. Previous research shows that government employees have played an important role in fabricating pro-Chinese messages online (King et al., 2017) and using the click-bait strategy to gain visibility (Lu & Pan, 2021). In addition, in recent years, Beijing has initiated a series of campaigns via soft power messaging and COVID-19 diplomacy to tell China’s story well (Huang & Wang, 2019). Thus, the complexity and dynamics of pro- and anti-Chinese state groups lead us to the second set of research questions.

RQ2: What was the overall pattern of public sentiments during the early pandemic?

RQ3: Who were the main targets of positive and negative sentiments?

RQ4: Were there any conversations between pro-China and anti-China Twitter users?

## Content of Chinese Tweets

Twitter has been a public sphere since its founding. After the COVID-19 outbreak, Twitter, like other social media platforms such as Facebook and Weibo, has been one of the major online spaces where individuals seek social support, track government announcements, and monitor the spread of the coronavirus (Lu et al., 2021). We focus on any Chinese tweets mentioning China-related keywords during the pandemic. We expect that Chinese Twitter users, such as overseas students and Chinese immigrants would use Twitter to share news and seek help when COVID-19 emerged.

Twitter has also been a fierce battlefield for conspiracy theories, hate speech, misinformation, disinformation, and fake news. COVID-19 has led to a global surge of anti-Chinese sentiment (Cook et al., 2021), and racial slurs targeting Asian and Asian American communities have been widely spread on Twitter such as Chinese Virus and KungFlu (Ziems, He, Soni, & Kumar, 2020). Chinese Americans and overseas Chinese students might use Twitter as a platform to voice themselves and combat racism and anti-Asian attacks.

The increasing tension between the United States and China such as trade wars and human rights issues pertaining to Xinjiang and Tibet and the Trump administration’s strict policy on Chinese scientists might also spark overseas Chinese users to share concerns on the US-China relations, discuss immigration policies, and express anger or fear of uncertainties in the pandemic. Pro-democracy groups might use Twitter to discuss sensitive topics such as the Xinjiang re-education camp, Uyghurs, and Falungong, while pro-Chinese state users including state-sponsored organizations and the paid 50-cent party might use Twitter to

promote China’s soft power and boost China’s global image by tweeting Chinese culture, economic development, tourism, and so on.

The 2019-2020 protest cycles in Hong Kong have drawn great attention from Chinese and global societies. Protesters used Twitter as a platform to diffuse protest information, mobilize resources, and seek solidarity, while pro-Chinese state supporters might also strategically use Twitter for political propaganda by framing protests as conflict and violence, disrupting social order and economy, and destabilizing national security (M. M. Zhang, Wang, & Hu, 2021). Twitter is also an online space where Chinese state-backed media and nationalists promote the reunification between Mainland China and Taiwan (R. Chang, Lai, Chang, & Lin, 2021). Similarly, Taiwan independence supporters might use Twitter to seek for support.

Due to the diversity of Chinese Twitter users and the confluence of COVID-19 and other political and social events, this leads to our third set of research questions.

RQ5: What was the content of these Chinese Tweets during the early pandemic?

RQ6: Was the overall sentiment pattern driving by specific topics during the early pandemic?

RQ7: Was there any variation in topics among different types of Twitter users?

## Data and Methods

Next, we first introduce how we collected the Chinese tweet dataset (CNTweets). Then, we describe how we constructed our training dataset used to build deep learning algorithms to classify sentiments and topics of tweets. Given that each research question requires different methods, we elaborate more on the specific method used for each research question.

### CNTweets Data

We used Chinese keywords to retrieve all matched tweets posted in 2019-2021 from Twitter’s historical database using academic Twitter API. Section 1 in supporting information (SI) documents the detailed keywords we used in data collection. We collected over 25 million tweets by 1.32 million users mentioning any keywords in simplified and traditional Chinese characters related to China, Chinese, and CCP. Table 1 shows the descriptive statistics of our Twitter data.

Table 1: Summary of Twitter Data

Data Type	Million
# of tweets	25.30
# of tweets mentioning China (中国)	16.64
# of tweets mentioning Asians or Chinese (亚裔/华裔)	0.28
# of tweets mentioning CCP (共产党)	7.46
# of Twitter users	1.32

## Training Data

In order to extract sentiments and topics in CNTweets data, we annotated a training dataset with 10,000 tweets to build deep learning algorithms to classify CNTweets. Section 2 in SI documents the detailed process of our training data construction, and here we briefly summarize the major steps. We started with those well-known pro- and anti-China Twitter users in the Chinese Twitter community and their followers or following accounts (e.g., PDChinese, dajiyuan). We scraped all their tweets posted in the past 2 years. We also used pro- and anti-China hashtags and keywords (e.g., against CCP) to extract potential tweets that either support or criticize the Chinese government or China. We then used a stratified sampling strategy to select 7,000 tweets from these potential positive or negative tweets targeting China. To add more potential neutral tweets to our training dataset, we then randomly selected 3,000 tweets from our CNTweets data to construct the final 10,000 tweets for human annotation. We hired both graduate and undergraduate research assistants to manually annotate the sentiment and topics in these tweets. Each tweet had been labeled by at least two annotators, and if there was inconsistency, one of our authors then adjudicated the difference.

## Source of Chinese Twitter Users

To tackle the first research question on the sources of Chinese Twitter users, we rely on partial information provided by Twitter users’ self-reported locations when they signed up for a Twitter account. To extract the major countries and regions, our location analysis first used regular expressions to search full names and abbreviations of a country or region and then searched states/provinces/major cities in a country or region. For instance, to identify whether a Twitter user is from the United states, we first searched the United States, U.S., or US, and then incorporated different states, cities, and their abbreviations such as New York and NY. Readers should be cautious when interpreting the results, as we might underestimate the total number of users who reported their locations or fail to capture the difference between users’ displayed locations and their actual locations due to address changes and reporting false locations. Note that the self-reported location analysis is highly sensitive since it depends on whether Twitter users reveal their true locations. In addition, we also asked our annotators to identify whether a tweet is related to personal opinions, organizations, government announcements, or spam. This allows us to identify whether these tweets are from individual or organizational accounts.

## Sentiment of Chinese Tweets

To answer the second question about the overall pattern of public sentiments, we fine-tune the pre-trained Robustly Optimized BERT Pretraining Approach (RoBERTa) with the Whole Word Masking models (Chinese-Roberta-wwm-ext) (Liu et al., 2019). The recent development in natural language processing with deep learning techniques shows that BERT has outperformed other state-of-the-art language models (Vaswani et al., 2017; Devlin et al., 2018; Cui et al., 2021). BERT is the most state-of-the-art language representation model, which stands for Bidirectional Encoder Representation from Transformers (for more technical

details, see Devlin et al.’s work). It is trained on large-scale unlabeled texts by randomly masking some of the tokens from the input (i.e., mask language model) and taking the input’s both left and right contexts into account (i.e., bidirectional contextual embedding). We used the pre-trained Chinese-RoBERTa-www-ext model and fine-tuned the last classification layer and several hyper-parameters of the model such as learning rate and batch size. The fine-tuned RoBERTa models were then used for our specific downstream tasks (i.e., sentiment analysis and topic classification).

Table 2 shows our accuracy and F1 scores for the sentiment classifier. We classified each tweet’s sentiment toward China into three categories - positive, negative, or neutral. Note that here we broadly define China. China can be a nation as a whole, Chinese people, Chinese central/local government, CCP, State-sponsored enterprises and organizations, places, and other entities related to China. We also compared the performance of RoBERTa with the performance of other BERT models such as MacBERT and Multilingual BERT based on our annotated training datasets, but RoBERTa outperformed others consistently.

Table 2: Model Performance on Sentiment and Target Classification

Outcomes	F1 Score	Accuracy
Sentiment	0.81	0.81
Target China in general	0.70	0.86
Target the Chinese people	0.61	0.94
Target the Chinese government	0.84	0.89

To tackle the third research question about the targets of public sentiments, we build RoBERTa models to further discern the target entities: the Chinese people, the Chinese government, or China in general. If a tweet mentions anything related to ordinary Chinese people, we label it as “Chinese people”. If a tweet discusses the political system in China, we label it as “Chinese government”. Examples of entities in the category include the central/local Chinese government or CCP, general politics in China, police departments, state media, state-sponsored companies, major political figures in China, and Beijing or Zhongnanhai when they are used to refer to the government. Sometimes people mention the Chinese government such as “China” and “authoritarian regime” without using any specific term related to the government. In this case, it requires our annotators to use their own judgment to identify their targets and label those tweets. If a tweet discusses China but can’t be categorized as “Chinese people” or “Chinese government,” we label it as “China in general,” such as festivals and traveling. Table 2 shows our accuracy and F1 scores for the target classifier. It is noteworthy that a tweet can contain multiple entities as we trained three separate classifiers to identify the target entities.

To answer the fourth research question on the dynamics between pro- and anti-Chinese state groups, we rely on social network analysis techniques. We used the conversation\_id from Twitter to construct a bipartite conversation network based on whether these pro- and anti-users engaged in the same conversations. Twitter assigns a unique conversation\_id to all the reply threads, and this conversation\_id matches the original tweet that started the conversation. Thus, a conversation contains all replies to a given tweet and replies to those replies from the single original tweet. This new API feature allows scholars to retrieve

and reconstruct an entire conversation thread and understand how conversations and ideas evolve on Twitter. On top of the conversation network, we also conducted a retweet network analysis (see Section 5 in SI) and the results are consistent. But we prefer the conversation network over the retweet network in the main text because the conversation network with a series of replies signals more in-depth engagement compared to a simple retweet.

## Content of Chinese Tweets

To address the fifth research question about the content of Chinese tweets, we train a series of classifiers to identify whether a tweet is related to *COVID-19*, *politics*, *economy*, *culture*, *religion*, and *the US*. We asked annotators to label each tweet into these different topics when we were building our training datasets. These topics were selected based on the consideration of geopolitics centering on China during the early pandemic. For instance, during the early pandemic, the US-China trade war, Taiwan issues, Hong Kong protests, and the US presidential election drew great attention from journalists, policymakers, and the public. Thus, for politics, we further discerned US politics, Hong Kong politics, and Taiwan politics. Our annotators were trained to label these topics to each tweet in our training dataset. Table 3 shows our accuracy and F1 scores for each classifier. Because some classifiers have a lower F-1 score, we also supplement our topic classification results with structural topic modeling for robustness check (Roberts, Stewart, & Tingley, 2019). Structural topic model, as an unsupervised text analysis tool, has been used to retrieve information from large-scale textual data and its advantage is to allow researchers to flexibly estimate how document-level metadata shapes topic prevalence compared to the conventional latent dirichlet allocation (LDA) model (Blei, Ng, & Jordan, 2003; Roberts et al., 2014). The most straightforward understanding of topic models is to see each document as a function of themes or topics governed by some prior distribution and each theme is a distribution of words in the fixed vocabulary, and topic modeling is to find these two sets of parameters that best fit the observed data. When estimating topic models, researchers need to pre-define the number of topics in the documents. We chose structural topic model over others because of its capacity to add metadata like timing (e.g., month) into the model estimation process, and we ran a series of structural topic models with different topics (e.g.,  $K=30$ , 50, and 100). In the main text, we only present the model with  $K = 30$ .

To address the sixth question on the relationship between topics and sentiments, we run regression models with monthly fixed-effects terms to test whether some topics such as politics, economy, culture, religion, COVID-19, and US-China relation were driving the overall sentiment pattern towards China.

To address the seventh question on the topic variation by pro- and anti-China users, we focus on two types of accounts that either support or oppose China. We analyzed the differences in the content of their posted tweets in our CNTweets database.



Table 3: Model Performance on Topic Classification

Outcomes	F1 Score	Accuracy
Tweet type	0.91	0.91
COVID-19	0.93	0.97
Culture	0.16	0.98
Democracy	0.63	0.90
Economy	0.23	0.98
Politics	0.92	0.92
US Politics	0.70	0.96
Taiwan Politics	0.68	0.99
HK Politics	0.70	0.98
Religion	0.27	0.99
US	0.86	0.96
US-China Relation	0.45	0.96

## Results

### The Sources and Types of Chinese Twitter users

To recapitulate, our first research question asks about who mentioned China-related issues during the early pandemic, so we begin by describing the overall pattern of who created these Chinese tweets. The descriptive analysis shows that 1% of Twitter users generated 62% of total Chinese tweets during the early pandemic in our database from December 2019 to April 2021. Notably, 10% of Twitter users contributed to 90% of total Chinese tweets in our CNTweets database. Thus, in the Twitter verse of Chinese language users, the majority of Chinese tweets targeting China, Chinese, CCP, and Asians in either a positive or negative direction were driven by a handful of Twitter users (around 13 thousand).

We then ran a geospatial analysis of Chinese Twitter users’ self-reported locations. The majority of Chinese language Twitter users reported a location in Mainland China, the US, Taiwan, or Hong Kong. Among 1.32 million Twitter users in our dataset, 0.58 million (43.83%) of them self-reported a location on their public profiles. Among those who reported certain information in the location part of the profile, we were able to identify 0.33 million (58%) users’ countries/regions (e.g., Europe, Singapore, Indonesia, Japan). Among those users with identified countries/regions, the majority reported a location in Mainland China (31.62%), the US (18.09%), Taiwan(8.95%), and Hong Kong (8.59%). This is reasonable as these countries or regions contain a large population of individuals who speak Chinese.

We also find that the majority of Chinese language tweets were associated with personal opinions, followed by news content. We trained a RoBERTa classifier to discern the types of these tweets. Each tweet was classified into personal opinions (i.e., any personal expression such as personal opinions, comments, discussion, or emotions about any topic), news content (e.g., news related to COVID-19, China, the US, or other countries), government or any other institutions’ announcements (e.g., announcements by government officials and World Health Organization’s health advice), advertisements and spam, and others. We find that 68.4% of

tweets were related to personal opinions, 27.6% were associated with news content, 0.71% of tweets were related to governments’ or other institutions’ announcements, and 2.16% were advertisements and spam. This suggests that Chinese language users have used Twitter as a public space to express opinions towards China instead of retweeting news-like content or government announcements.

## The Overall Sentiments and Main Targets

The second research question asks about the overall pattern of public sentiments during the early pandemic. Our RoBERTa sentiment classifier shows that the sentiments in the Chinese tweets were predominantly negative toward China. As shown in Table 4, among 25 million tweets in our CNTweets database, 15.74 million were classified as negative toward China, 5.54 million were neutral, and only 4.02 million were positive. Tweets sharing negative, positive, and neutral sentiments toward China accounted for 62%, 22%, and 16%, respectively, during the early pandemic.

Table 4: Descriptive Statistics of Prediction Results on Sentiments and Targets

Outcomes	Million of Tweets	Percent
Sentiments		
Negative	15.74	62.2
Neutral	5.54	21.9
Positive	4.02	15.9
Target the Chinese Government	15.19	60.0
Target the Chinese People	2.79	11.0
Target China in General	6.32	25.0

Fig 1 shows the time series of the percentage of positive, negative, and neutral tweets. It suggests a robust pattern that the Chinese Twitter community was consistently negative toward China during the early pandemic. But there are some nuances in the sentiment pattern as we can see how sentiments in Chinese tweets resonate with major events related to the COVID-19 pandemic and the US election. After the first cases were confirmed in Wuhan in mid-January 2020, the daily number of negative tweets soared first and then declined, but this increased again after former US President Trump tweeted about the racial slur “Chinese Virus”. For instance, on March 16, 2020, Trump tweeted, *“The United States will be powerfully supporting those industries, like Airlines and others, that are particularly affected by the Chinese Virus. We will be stronger than ever before!”* It is noteworthy that the spike in the neutral trend in early-November 2020 was due to the shifted attention toward the voter fraud conspiracy in the US 2020 presidential election, while the early February spike in 2021 was due to the discussion of the World Health Organization (WHO)’s preliminary report on the origins of the COVID-19 coronavirus (i.e., it is unlikely leaked from a Wuhan Lab) by both pro- and anti-China users. Note that the time series of unique Twitter users based on these positive, negative, and neutral tweets shows a similar pattern, and there were consistently more active users tweeting negative posts over time (see Fig S1 in SI).

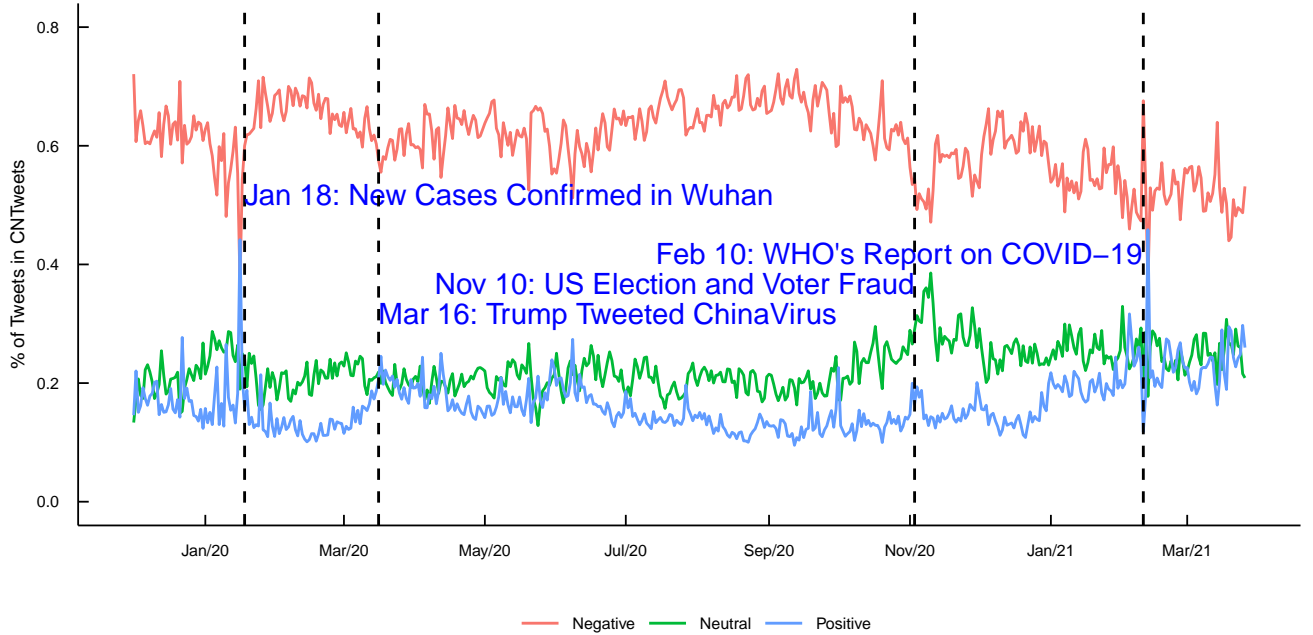


Figure 1: Daily Trends of Positive, Negative, and Neutral Sentiment toward China on Twitter. The sentiment toward China in Chinese language communities was predominantly negative during the early pandemic. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

Keywords analysis shows that China and CCP were more likely to be mentioned than people of Asian or Chinese descent. Fig 2 shows the daily trends of China, CCP, and people of Asian or Chinese descent (亚裔/华裔). It clearly shows that Chinese language Twitter users mentioned China and CCP more often than people of Asian or Chinese descent (see Fig S2 in SI for the user level analysis). Fig 2 also shows that China/CCP keywords surged during the early pandemic, peaked after former US President Donald Trump tweeted “Chinese Virus” on March 16, 2020, and then remained relatively steady. For Asian-related keywords, we have a similar pattern during the early pandemic, but these keywords also surged after March 2021 because of the tragic Atlanta SPA mass shootings. The mention of Asian or Chinese descendants was likely to be associated with these StopAsianHate movements.

Overall, the sentiment toward China was negative. But who were they targeting? To address our third research question about the main targets of positive and negative sentiments, our sentiment target analysis shows that most negative tweets were targeting the Chinese government or China in general instead of the Chinese people. Fig 3 shows the daily trends of tweets targeting different China-related entities. The majority of sentiments in the CNTweets database were directed toward the Chinese government. During the early pandemic, around 60% of tweets were targeting the Chinese government, around 11% were targeting

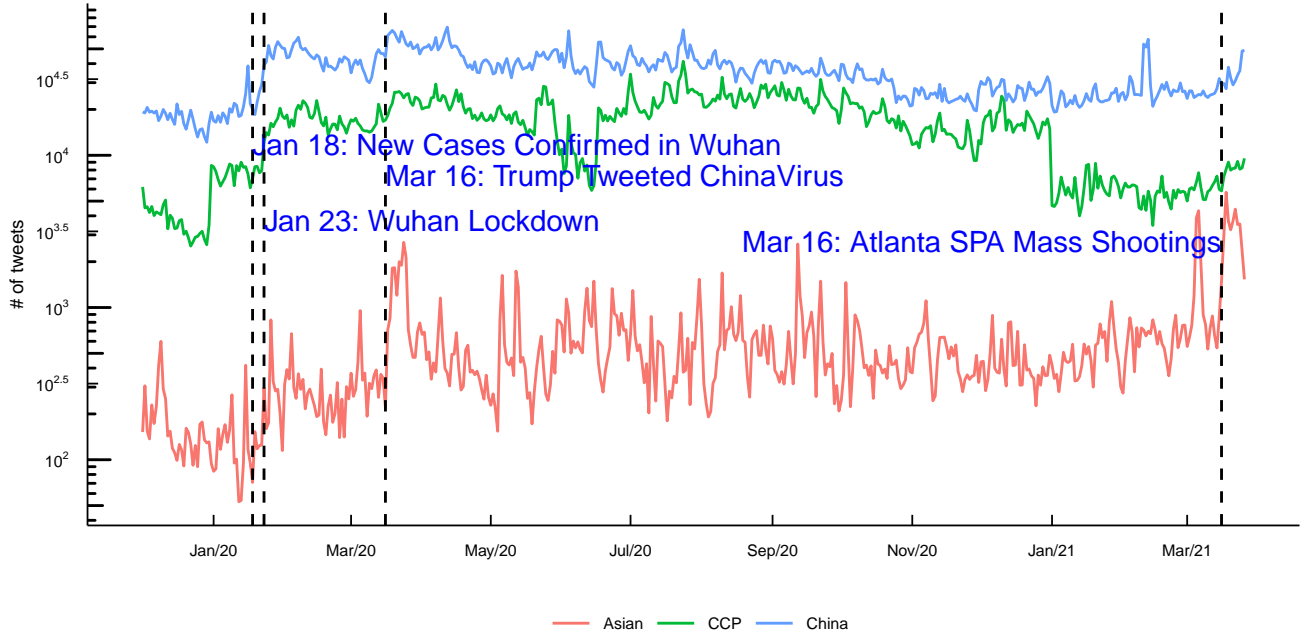


Figure 2: Daily Trends of Chinese Tweets mentioning China, Asians/Chinese, and Chinese Communist Party (CCP). Keyword analysis of Chinese tweets shows a consistent pattern that Chinese language users mentioned China and CCP more compared to People of Asian or Chinese descent. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

Chinese as an ethnic group, and around 25% were targeting China in general. Similarly, if we focus on active Twitter users instead of tweets, we find a similar pattern, as there were consistently more active Twitter users targeting the Chinese government during the early pandemic (see Fig S3 in SI).

For those tweets with negative sentiments, as shown in Table 5, 80% were targeting the Chinese government, 11% were targeting the Chinese people, and 19% were targeting China in general. For those tweets with positive sentiments, the proportions associated with the Chinese government, the Chinese people, and broad China were 20%, 34%, and 46%, respectively. These results suggest that negative tweets were more likely to target the Chinese government and positive tweets were more likely to support China in general.

## A Network Analysis of Pro- and Anti-China Twitter Users

To address the fourth research question on the engagement between pro- and anti-China Twitter users, we use the results from sentiment analysis to classify Twitter users into pro-China and anti-China users based on the rate of positive tweets. If a user's positive rate

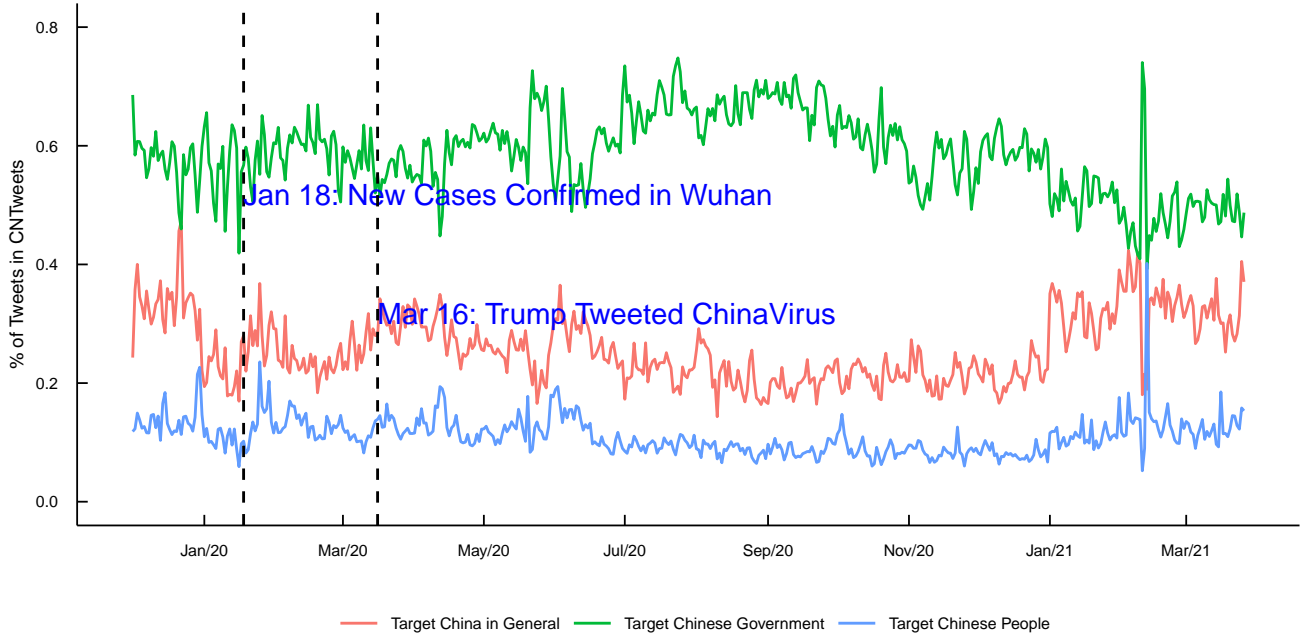


Figure 3: Daily Trends of Chinese Tweets Targeting Different Entities. Chinese language communities on Twitter were consistently targeting the Chinese government and China in general instead of the Chinese people. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

Table 5: Proportion of Target Entities by Different Sentiments

Sentiment	Chinese People	Government	Broad China
Negative	0.11	0.80	0.19
Neutral	0.06	0.22	0.27
Positive	0.20	0.34	0.46

is larger than 0.6, we label it as a pro-China user; if it is less than 0.4, we label it as an anti-China user. We have 459,821 anti-China users and 496,504 pro-China users.

Then we constructed a conversation network for these pro- and anti-China users in our database based on whether these users engaged in the same conversations using Twitter’s conversation\_id. Twitter assigns a unique conversation\_id to each tweet if they engage in the same conversation thread. Typically, the conversation\_id is identical to the tweet id posted by the first user and other replies to this post or its replies share the same conversation\_id. For these identified pro- or anti-China users, we observed 19.82 million unique conversations in our database. We also find that 96.4% of these tweets contained no replies or engagement

with others. Among these tweets with no replies, pro-China users contributed 1.78 million while anti-China users contributed 17.32 million. Notably, 0.83 million conversations had at least one pro-China user and one anti-China user. Thus, conversations between pro- and anti-China users only accounted for 4.7% of total conversations that occurred among identified pro- or anti-China users in our database. To further quantify the segregation level between pro- and anti-China users, we computed the E-I index, a measure of homophily in social networks, to capture the difference between between-group and within-group ties (Krackhardt & Stern, 1988; Bojanowski & Corten, 2014). The E-I index will take +1 if all ties fall into between groups and -1 if all ties are within groups. For more technical definitions, see Section 3 in SI. The E-I index based on our conversation network was -0.33. This clearly shows that pro- and anti-China users were more likely to engage within their own groups and lacked in-depth cross-boundary engagement with each other. We also conducted an additional retweet network analysis (see Section 5 in SI) and the results are similar. The E-I index for the retweet network was -0.906, suggesting an even more segregated pattern between pro- and anti-China users on retweeting behavior.

Fig 4 visualizes the conversation network among Twitter users. We only plot Twitter users with at least 10 conversations for ease of illustration. Red dots indicate pro-China users, while blue dots denote anti-China users. It clearly shows the polarized pattern that pro-China and anti-China users were clustered into their own groups, but pro- and anti-China users did engage in some dialogues that might support or criticize China. There are some nodes in Fig 4 that were attracting attacks from the other side. For 219,985 conversations with at least one pro and anti-China user, we find that 23% only had one pro-China and one anti-China participant, and the majority (74%) of these conversations had less than 10 pro- or anti-China users. Only 26% of these conversations involved over 10 pro-China or anti-China users. Taken together, these findings suggest a polarized pattern that pro- and anti-China users lacked in-depth engagement in China-related conversations. Given that we focus solely on Chinese Tweets centering on China, our results cannot be extended to other conversations not pertaining to China, Chinese, or CCP.

## The Content of Chinese Tweets

Next, we move to the fifth research question on the content of these Chinese tweets. Overall, the majority of tweets were related to politics, followed by democracy and freedom, US issues, and COVID-19 topics. As mentioned earlier, we trained a series of RoBERTa classifiers to identify potential topics in these tweets. Our RoBERTa topic classifiers show that 73% of tweets were broadly related to politics. Politics could be any topic related to ideology, democracy, policy, major figures in China or other countries, geopolitics, etc. More specifically, 31% were associated with discussions on democracy and freedom, 22% were discussing US politics such as domestic issues and elections, 9% were discussing Hong Kong protest issues, and 6% were mentioning Taiwan politics. 27% of these tweets were related to US topics. Note that 14% were related to the US-China relation. This is reasonable as the trade war between China and the United States. 20% of tweets discussed COVID-19 issues, while culture, economy, and religion-related topics only accounted for 6%, 5%, and 2%, respectively. Note that we define economy topics as any economic issues such as infrastructure investment, economic growth, and the development of industrialization and modernization.

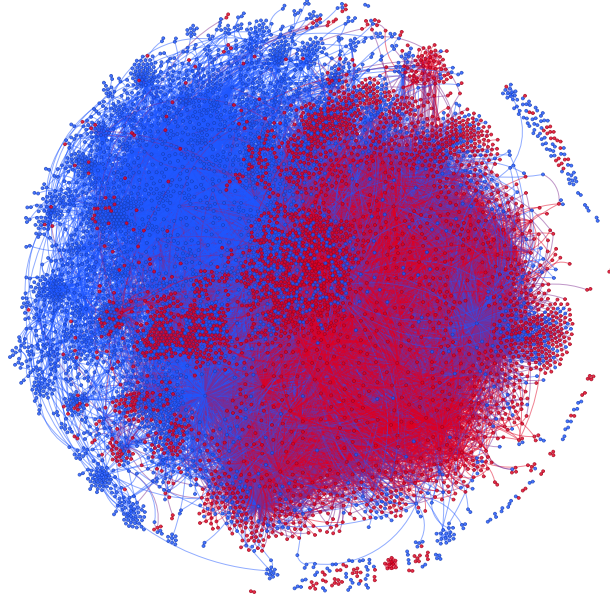


Figure 4: Conversation Network Visualization of Pro- and Anti-China Users. Nodes are pro- (Red) and anti- (Blue) China Twitter users and edges indicate at least 10 conversations between two nodes. Pro- and anti-China users were more likely to be segregated from each other, even though there was some cross-boundary engagement. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

Culture-related topics include travel, food, sports, art, entertainment, etc. Religion topic focuses on religious freedom or other religious issues.

Table 6: Proportion of Main Content

Outcomes	Proportion
Politics	0.73
Democracy	0.31
US	0.27
US Politics	0.22
COVID19	0.20
US-China Relation	0.14
HK Politics	0.09
Taiwan Politics	0.06
Culture	0.06
Economy	0.05
Religion	0.02

The keyword analysis shows that COVID-related keywords quickly peaked in Chinese language communities after the outbreak, but US and Hong Kong-related topics prevailed during the early pandemic. Fig 5 shows the daily trend of some keywords of interest, includ-

ing COVID-19, Taiwan, the US, Hong Kong, Tibet, and Xinjiang. These were the major issues targeted by anti-China Twitter users. Unsurprisingly, COVID-related Chinese keywords increased rapidly in the Twitter community after the outbreak, peaked after March, and declined after April 2020. However, the US and Hong Kong-related topics were often discussed in the community as the US-China trade war and Hong Kong protests were dominating the issue attention cycle during the early pandemic, followed by Taiwan, Xinjiang, and Tibet issues.

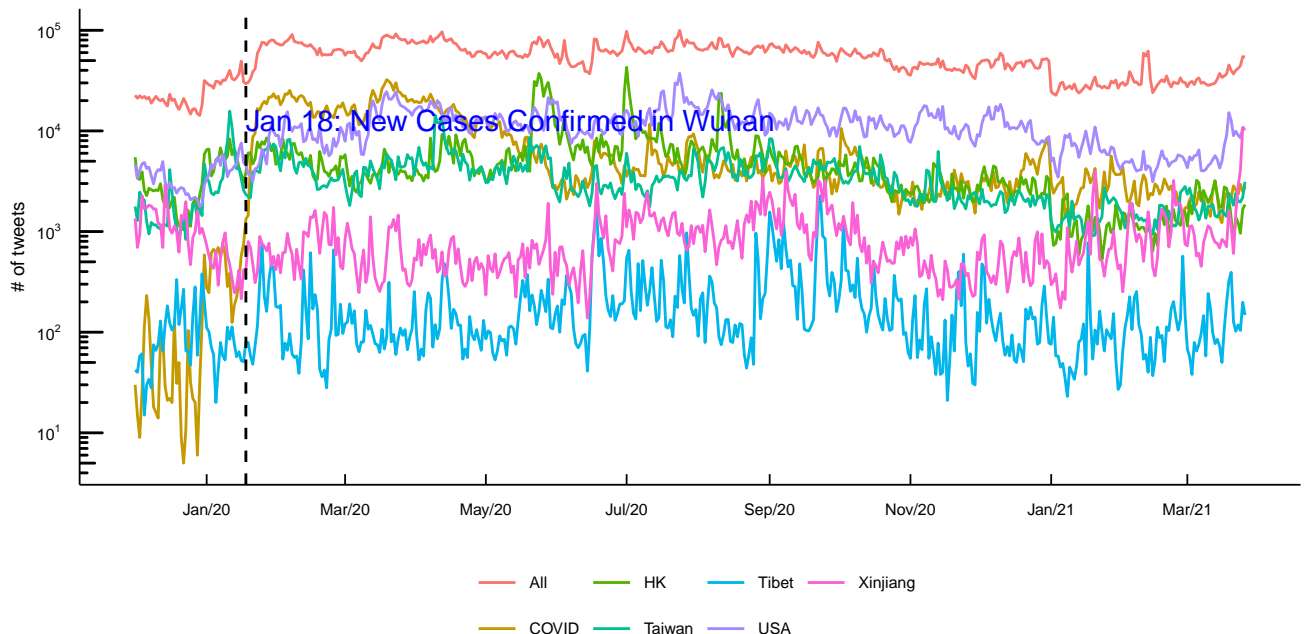


Figure 5: Daily Trends of Chinese Tweets mentioning Hong Kong, Taiwan, Xinjiang, Tibet, the US, and COVID-19. Keyword analysis shows that COVID-19, Hong Kong, and the US were frequently mentioned by Twitter users in Chinese language communities during the early pandemic. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

Structural topic modeling also shows that the most popular themes in CNTweets were China’s domestic politics, COVID-19, US politics, and Hong Kong and Taiwan issues. Fig 6 plots the distribution of themes extracted from our CNTweets data. We estimated 30 topics using the structural topic model. Results suggest that democracy-freedom (8%), US election (6.9%), global issues (6%), the 50-cent party (i.e., supporting CCP, 5.4%), culture-education (5.1%), COVID-19 (4.9%), Hong Kong-National Security Law (4.8%), Wuhan outbreak(4.8%), human rights (e.g., Xinjiang, 3.7%), and the US-China Initiative (3.6%) were the top 10 themes during the early pandemic on Twitter. Other prevalent topics include COVID origin (made in a Wuhan lab), Huawei Ban, Chinese policing, the Chinese



economy, anti-CCP, etc.

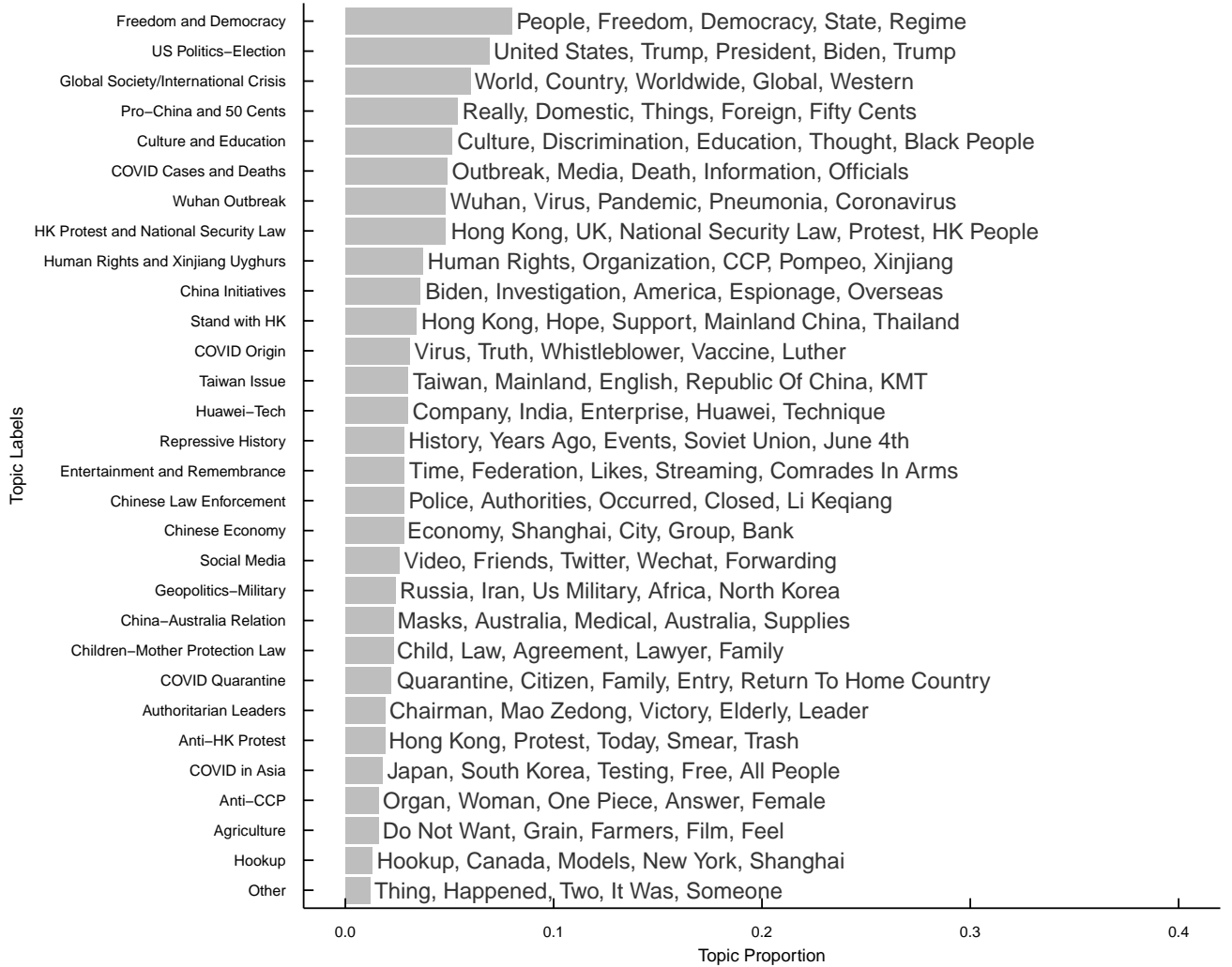


Figure 6: Structural topic model output, K=30. The most popular themes from CNTweets were related to freedom and democracy, the US election, geopolitics, and COVID-19. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

## The Dynamics between Topics and Sentiments

To address the sixth question on the relationship between topics and sentiments, we ran the logistic regression model using topics to predict whether a tweet's sentiment is positive towards China. Table 7 reports logit coefficients from the model focusing on different topics.

Model 1 explores which topics among COVID-19, politics, religion, culture, economy, and US-China relation were more likely to be positive towards China. We show that a tweet is less likely to be positive towards China when it pertains to COVID-19, politics, religion, and US-China relation but more likely to be positive if it relates to cultural or economic issues. We further plot the daily trends of the fraction of negative tweets by topics in Fig 7. It suggests that the negativity towards China among Chinese language communities during the early pandemic was mostly driven by the discussion on politics, followed by COVID-19 and US-China relation topics. After the outbreak of the coronavirus, the percentage of negative COVID-19 tweets in our database increased rapidly but then declined and was surpassed by the negative tweets on US-China relation after June 2020.

Table 7: Logistic Regression Results Predicting Positive Sentiment at the Tweet Level

	Model 1
COVID-19	−0.558*** (0.002)
Politics	−0.877*** (0.001)
Religion	−0.906*** (0.005)
Culture	0.225*** (0.002)
Economy	0.112*** (0.002)
US-China Relation	−0.093*** (0.002)
Observations	25 299 856
BIC	21 178 919.2
RMSE	0.36
FE:Month	X

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## The Topic Variation between Pro- and Anti-China Users

To examine the final research question about whether different types of users engaged in distinct topics, we ran an additional analysis to compare topic proportions between 459,821 anti-China users and 496,504 pro-China users. Table 8 reports the average number of tweets and overall proportions for each topic within all tweets posted by these pro- or anti-China users. Both sides were heavily engaged in topics including politics, the US, and COVID-19 issues. Over 30% of pro- or anti-China users’ tweets involved some aspect of politics.

Pro-China users were more likely than anti-China users to tweet about economy, culture, COVID-19, and US issues, compared to topics like politics. For an average pro-China user in our CNTweets database, as shown in Table 8, they were less active in terms of the average number of posts compared to anti-China users. For instance, on average, a pro-China user had 2.88 tweets discussing politics, while an anti-China user had 35.07 tweets. But in terms of the topic shares for all tweets made by these users, pro-China users focused more on economy, culture, COVID-19, and US issues, while anti-China users focused more on politics,

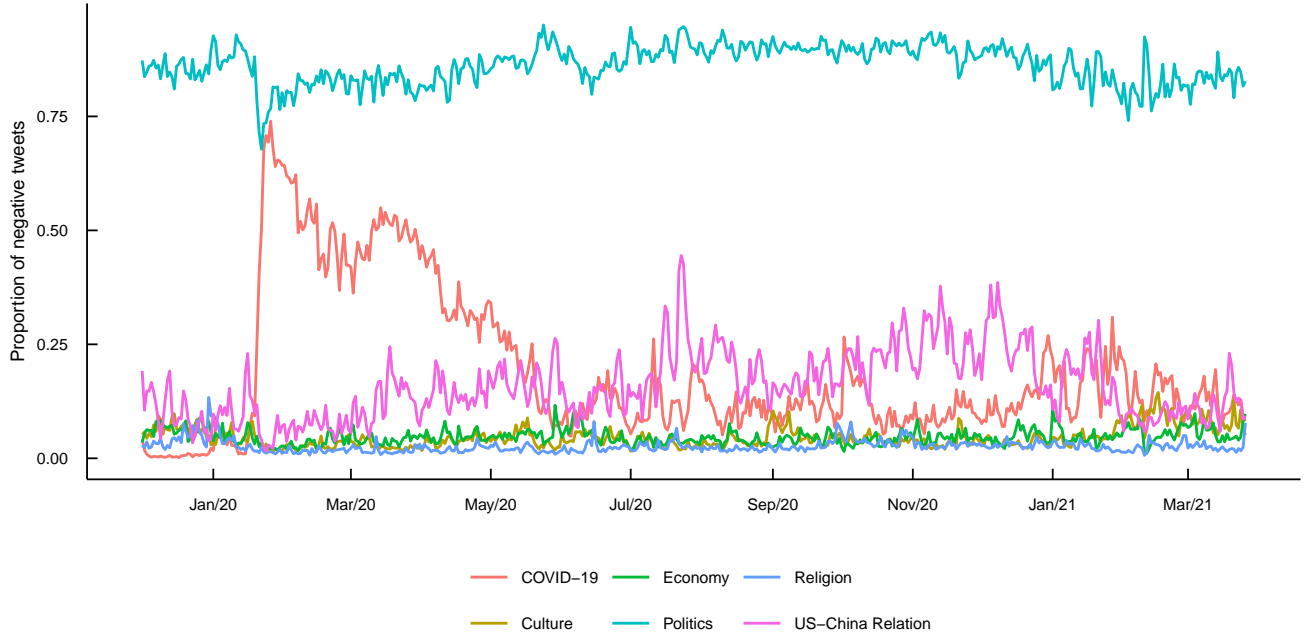


Figure 7: Daily Trends of Proportion of Negative Tweets by Topics. The negativity in the database was driven by the surging negative COVID-19 and political topics. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

Table 8: Average Tweets by Pro- and Anti-China Users

Outcomes	Pro-China (prop.)	Anti-China (prop.)
Politics	2.88(0.32)	35.07(0.34)
Democracy	0.52(0.06)	15.99(0.16)
HK Politics	0.3(0.03)	4.53(0.04)
Taiwan Politics	0.35(0.04)	2.85(0.03)
US Politics	0.94(0.1)	10.32(0.1)
US	1.28(0.14)	12.67(0.12)
US-China Relation	0.47(0.05)	6.75(0.07)
COVID19	1.02(0.11)	9.41(0.09)
Economy	0.35(0.04)	2.03(0.02)
Culture	0.98(0.11)	2.03(0.02)
Religion	0.07(0.01)	1(0.01)

particularly related to democracy and freedom and Hong Kong politics. The variation in topics reflects the different agendas of these pro- and anti-China users on Twitter.

## Discussion and Conclusion

This paper used multi-modal supervised and unsupervised machine learning tools to examine anti-China sentiments and topics in Chinese language communities on Twitter during the early COVID-19 pandemic. Since the outbreak, scholars have shown a global surge of anti-China sentiments. Our work was the first to systematically examine the dynamics of sentiments in Chinese language communities on a major western social media platform. Compared to other Chinese media platforms like Sina Weibo, Twitter is a public space that attracts users who intend to express their criticism or support toward China. Thus, it affords scholars a window to examine the relationship between pro- and anti-China users online. But readers should note that Chinese language communities on Twitter are not a representative group of Chinese communities or the Chinese diaspora.

Based on the analysis of over 25 million Chinese tweets from December 2019 to April 2021, we find that the majority of these China-related tweets were generated by only 1% of Twitter users. These Chinese language users, who reported a location in Mainland China, the US, Hong Kong, and Taiwan, were more likely to mention China or CCP instead of people of Asian or Chinese descent. The majority of these tweets were personal opinion-oriented, followed by news-like content and government or institutional announcements. These results suggest that tweets targeting Chinese language communities might be a very selective group of users as a handful of Twitter users contributed to the majority of content related to China topics.

We also find that the majority of tweets in our CNTweets database were negative toward China, although these sentiments were more likely to target the Chinese government or China in general instead of the Chinese people. These pro- and anti-China Twitter users were predominantly segregated as they were more likely to engage in conversations on their own side, and only a small size of Twitter users engaged in conversations on the other side. Note that we find evidence that pro-China users contributed 1.78 million tweets with no replies by others while anti-China users contributed 17.32 million in our database. These results suggest that Twitter has been used as a major platform by anti-China users to criticize the Chinese government and CCP. Prior research has shown the lack of evidence related to computational propaganda by CCP but strong evidence of computational propaganda by anti-China groups on Twitter due to low operating costs (Bolsover & Howard, 2019). Our results show that anti-China users were indeed more active on Twitter than pro-China users during the early pandemic, but given the large volume of tweets from both sides, our work adds evidence to the existence of potential computational propaganda by both pro- and anti-China users (see Fig S5 in SI). Since we focus solely on Chinese tweets, we cannot extend this conclusion to the entire Twitter verse as CCP might be more likely to target English language communities instead of Chinese language communities.

The most common topics discussed by these anti-China Twitter users were politics, such as democracy and freedom, Hong Kong protests, Taiwan politics, Xinjiang, and Tibet issues. Even though both pro- and anti-China users were heavily engaged in the discussions of politics, pro-China users were more likely to discuss topics related to economy, COVID-19, US issues, and culture, while anti-China users were more likely to focus on topics of democracy, freedom, and Hong Kong politics. Our regression analysis shows that tweets related to culture and economy were more likely to be positive towards China, while tweets

associated with COVID-19, politics, religion, and US-China relation were less likely to be positive towards China. These findings echo that pro-democracy activists tend to take advantage of these social media platforms to promote democracy and criticize the Chinese government, while pro-China Twitter users tend to use economy and culture topics to boost China’s international image.

Taken together, our findings show that Sinophobia was ubiquitous among the Chinese language communities on Twitter during the early pandemic, and the Twitter verse is a battlefield that attracts both pro- and anti-China users for their own potential propaganda agenda. Previous studies often focus on the English language communities on social media platforms and overlook non-English communities. The potential propaganda by both parties targeting Chinese ethnic groups might have negative consequences in the community. Many social media platforms have developed policies and tools to mitigate these negative consequences such as blocking hateful terms and suspending controversial accounts, but very few resources have been devoted to communities of minorities. Recent research has shown that the COVID-19 crisis increased censorship circumvention and access to international news and political content blocked in China, but when individuals sought crisis-related information, they were also exposed to misinformation and anti-China racism online simultaneously in their own online language communities.

Readers should note that our research has some limitations. For instance, some classifiers have a relatively low F1 score (e.g., culture, religion, and economy). One of the future directions is to use semi-supervised machine learning methods to improve the predictive power by adding more positive cases. In addition, our location analyses were self-reported by Twitter users instead of their actual geo-locations. Finally, we only obtained tweets during the early pandemic using keywords instead of the whole Twitter verse. We leave these to future research.

## Ethical Approval

The Stony Brook University Office of Research Compliance (IRB2022-00006) determines that this study does not meet the definition of human subjects research according to the Common Rule (45 CFR 46 subpart A). There are no human participants in this study.

## Informed Consent

All data were publicly available and collected via Twitter academic API. No private information or data will be published or can be seen in this article.

## Acknowledgments

We wish to thank the Institute for Advanced Computational Science for access to the Seawulf and Ookami high-performance computing systems at Stony Brook University and its generous seed grant support. For valuable feedback we wish to thank participants in the

CBSM CSS Small Group including Thomas Davidson, Daniel Karell, Laura Nelson, and Eunkyung Song.

## Data Availability

All aggregated data and codes used to replicate our main figures and regression table are available through <https://dx.doi.org/10.17605/OSF.IO/R7DE5>.

## Competing Interests

The authors have declared that no competing interests exist.

## List of Figure Legends

1. Fig 1. Daily Trends of Positive, Negative, and Neutral Sentiment toward China on Twitter. The sentiment toward China in Chinese language communities was predominantly negative during the early pandemic. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.
2. Fig 2. Daily Trends of Chinese Tweets mentioning China, Asians/Chinese, and the Chinese Communist Party (CCP). Keyword analysis of Chinese tweets shows a consistent pattern that Chinese language Twitter users mentioned China and CCP more compared to People of Asian or Chinese descent. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.
3. Fig 3. Daily Trends of Chinese Tweets Targeting Different Entities. Chinese language communities on Twitter were consistently targeting the Chinese government and China in general instead of the Chinese people. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.
4. Fig 4. Conversation Network Visualization of Pro- and Anti-China Users. Nodes are pro- (Red) and anti- (Blue) China Twitter users and edges indicate at least 10 conversations between two nodes. Pro- and anti-China users were more likely to be segregated from each other, even though there was some cross-boundary engagement. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.
5. Fig 5. Daily Trends of Chinese Tweets Mentioning Hong Kong, Taiwan, Xinjiang, Tibet, the US, and COVID-19. Keyword analysis shows that COVID-19, Hong Kong, and the US were frequently mentioned by Twitter users in Chinese language communities during the early pandemic. This figure is covered by the Creative Commons

Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

6. Fig 6. Structural topic model output,  $K=30$ . The most popular themes from CNTweets were related to freedom and democracy, the US elections, geopolitics, and COVID-19. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.
7. Fig 7. Daily Trends of Negative Tweets. The negativity in the database was driven by the surging negative COVID-19 and political topics. This figure is covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of the Authors; copyright © the Authors, all rights reserved.

## References

- Bi, W. (2021). Playing politics digitally: young chinese people’s political feelings on social media platforms. *Cultural Studies*, 1–20.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Bojanowski, M., & Corten, R. (2014). Measuring segregation in social networks. *Social networks*, 39, 14–32.
- Bolsover, G., & Howard, P. (2019). Chinese computational propaganda: automation, algorithms and the manipulation of information about chinese politics on twitter and weibo. *Information, communication & society*, 22(14), 2063–2080.
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1), 1–14.
- Chang, K.-C., Hobbs, W. R., Roberts, M. E., & Steinert-Threlkeld, Z. C. (2022). Covid-19 increased censorship circumvention and access to sensitive topics in china. *Proceedings of the National Academy of Sciences*, 119(4), e2102818119.
- Chang, R., Lai, C., Chang, K., & Lin, C. (2021). Dataset of propaganda techniques of the state-sponsored information operation of the people’s republic of china. *CoRR*, abs/2106.07544.
- Cook, G., Huang, J., & Xie, Y. (2021). How covid-19 has impacted american attitudes toward china: A study on twitter. *arXiv preprint arXiv:2108.11040*.
- Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2021). Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3504–3514. Retrieved from <http://dx.doi.org/10.1109/TASLP.2021.3124365> doi: 10.1109/taslp.2021.3124365
- DeLisle, J., Goldstein, A., & Yang, G. (2016). *The internet, social media, and a changing china*. University of Pennsylvania Press.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805. Retrieved from <http://arxiv.org/abs/1810.04805>
- Fan, X., & Zhang, Y. (2023). “just a virus” or politicized virus? global media reporting of china on covid-19. *Chinese Sociological Review*, 55(1), 38–65.
- Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? a study of arizona’s sb 1070 using twitter data. *American Journal of Sociology*, 123(2), 333–384.
- Han, R. (2019). Patriotism without state blessing: Chinese cyber nationalists in a predicament. In *Handbook of protest and resistance in china*. Edward Elgar Publishing.
- Hobbs, W. R., & Roberts, M. E. (2018). How sudden censorship can increase access to information. *American Political Science Review*, 112(3), 621–636.
- Huang, Z. A., & Wang, R. (2019). Building a network to “tell china stories well”: Chinese diplomatic communication strategies on twitter. *International Journal of Communication*, 13, 24.
- King, G., Pan, J., & Roberts, M. E. (2017). How the chinese government fabricates social media posts for strategic distraction, not engaged argument. *American political science review*, 111(3), 484–501.



- Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, 123–140.
- Lee, J., & Huang, T. J. (2021). *Reckoning with asian america* (Vol. 372) (No. 6537). American Association for the Advancement of Science.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. *CoRR*, *abs/1907.11692*.
- Lu, Y., & Pan, J. (2021). Capturing clicks: How the chinese government uses clickbait to compete for visibility. *Political Communication*, 38(1-2), 23–54.
- Lu, Y., Pan, J., & Xu, Y. (2021). Public sentiment on chinese social media during the emergence of covid19. *Journal of Quantitative Description: Digital Media*, 1.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PloS one*, 10(9), e0138441.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, 91(1), 1–40.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082.
- Schneider, F. (2018). *China’s digital nationalism*. Oxford University Press.
- Shmargad, Y. (2018). Twitter influencers in the 2016 us congressional races. *Journal of Political Marketing*, 1–18.
- Song, S. Y., Faris, R., & Kelly, J. (2015). Beyond the wall: Mapping twitter in china. *Berkman Center Research Publication*(2015-14).
- Sullivan, J. (2012). A tale of two microblogs in china. *Media, culture & society*, 34(6), 773–783.
- Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the international aaai conference on web and social media* (Vol. 4).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc.
- Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021). Measuring algorithmically infused societies. *Nature*, 595(7866), 197–204.
- Zhang, D. (2021). Sinophobic epidemics in america: Historical discontinuity in disease-related yellow peril imaginaries of the past and present. *Journal of Medical Humanities*, 42(1), 63–80.
- Zhang, M. M., Wang, X., & Hu, Y. (2021). Strategic framing matters but varies: A structural topic modeling approach to analyzing china’s foreign propaganda about the 2019 hong kong protests on twitter. *Social Science Computer Review*, 08944393211042575.
- Ziems, C., He, B., Soni, S., & Kumar, S. (2020). Racism is a virus: Anti-asian hate and counterhate in social media during the covid-19 crisis. *arXiv preprint arXiv:2005.12423*.

## Supporting information

### Section 1. Data Collection Process.

Section 2. Pro or Anti-China Training Dataset Collection Strategies.

Section 3. Additional Results at the Twitter User Level.

Section 4. Quantifying Segregation of Conversation Network.

Section 5. Quantifying Segregation of Retweet and Quote Network.