THE UNIVERSITY OF CHICAGO

Exploring the Relationship Between Online and Offline Sinophobia Amid COVID-19

By

Yuzhou Wang

May 2024

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Jean Clipperton

Preceptor: Sabrina Nardin

Abstract

The COVID-19 pandemic has intensified global Sinophobia, leading to escalated verbal and physical assaults on the Chinese diaspora (Cabral, 2021; Reja, 2021; Salcedo, 2021; Than, 2021). Existing research indicates that online rhetoric, especially from prominent figures, can incite real-world racial animosity and aggression (Chan et al., 2016; Huang et al., 2022; Williams et al., 2020). This study first explores the overarching themes of online discourse centering on the COVID-19 news report in the United States, then focuses on the specific narratives that address China and the Chinese community to understand how Sinophobia is manifested and associated with the pandemic. Afterward, it delves into the temporal relationship between online Sinophobic sentiment and offline anti-Asian violence, accounting for the spill-over effect from targeted anti-Chinese antipathy to a broader anti-Asian hostility. Online Sinophobia is measured by comments under YouTube news videos about COVID-19 and Google searches for Sinophobic terms; offline Sinophobia is calculated by FBI hate crime statistics. This study identifies a bidirectional Granger causality between online and offline Sinophobia during the COVID-19 pandemic: anti-Asian hate crimes Granger-cause Google searches for Sinophobic terms, and conversely, these searches Granger-cause offline hate crimes, with a lag of two weeks. Additionally, it confirms a reinforcing dynamic between implicit and explicit online Sinophobia: the volume and intensity of anti-Chinese comments on YouTube news videos lead to increased Google searches for Sinophobic terms, and vice versa, with this effect observable over a ten-week lag. Furthermore, when accounting for confirmed infection cases, online Sinophobia is positively correlated with offline Sinophobia. These findings contribute to the existing literature on the relationship between online racist sentiment and real-life racial violence amidst this global crisis.

Keywords: Sinophobia; social media; hate speech; hate crime

1 Introduction

Sinophobia, or anti-Chinese xenophobia, has deep historical roots in the United States, dating back to the late 1840s when a large wave of Chinese immigrants arrived during the gold rush and was stereotyped as unhygienic and medically underdeveloped (Zhang, 2022). The 2002-2004 severe acute respiratory syndrome (SARS) outbreak witnessed a resurgence of Sinophobia, characterized as the “modern Sinophobic racialization of infectious disease” (Zhang, 2022, p. 72) by the phenomenon of the Chinese population unjustly being scapegoated for the outbreak. This historical context lays the foundation for the resurgence of Sinophobia amid the recent COVID-19 pandemic, evident in both online anti-Chinese sentiment and real-world violence against the Chinese and broader Asian community.

This research aims to uncover the intricate correlation between online Sinophobia, indicated by explicit comments under relevant YouTube news videos and implicit Google searches on pertinent keywords, and offline anti-Chinese crime incidents, measured by FBI Uniform Crime Reporting Program's (UCR) Hate Crime Statistics. It seeks to empower policymakers, community leaders, and social media platforms in their efforts to counter Sinophobic sentiment and violence, fostering racial inclusivity in the United States. Ultimately, the research intends to answer the question: What is the relationship between online Sinophobic sentiment, as influenced by COVID-19-related news, and the occurrence of offline anti-Chinese hate incidents?

In a preliminary pilot study (Y. Wang, 2023) focusing on Twitter data and hate crime statistics in 2020, the following correlations were identified:

- There were moderately positive correlations between the weekly count of tweets containing Sinophobic keywords (both generic and COVID-19-triggered) and the frequency of Google searches using the same terms. The Pearson coefficients were 0.347 for generic keywords and 0.111 for COVID-19-triggered keywords.
- There was a positive correlation between incidents of anti-Asian hate crimes and the number of tweets containing Sinophobic keywords, with a Pearson coefficient of 0.364.
- Anti-Asian hate crimes demonstrated predictive power for the use of Sinophobic slurs on Twitter and the frequency of Google searches on traditional Sinophobic terms, as indicated by Granger Causality tests. The p-values were 0.039 at lag 4-week and 0.044 at lag 2-week for tweets and Google searches, respectively.

2 Literature Review

2.1 Social Media Posts

Mirroring the pervasive depiction of the Chinese as carriers of viruses and disease, there has been a notably increasing usage of stigmatizing terms such as "Chinese virus," "China virus," "Wuhan virus," and "Kung flu" on Twitter and 4chan uttered both by political figures and by the general public during the outbreak of COVID-19 (Hswen et al., 2021; Schild et al., 2020). In a study conducted from March 9 to 23, 2020, approximately 20% of hashtags with #covid19 exhibited anti-Asian sentiment, while half of the hashtags with #chinesevirus exhibited anti-Asian sentiment (Hswen et al., 2021, p. 960). Furthermore, a significant surge in the use of #chinesevirus hashtag was observed following President Trump's tweet referring to the virus as the "Chinese virus" on March 16, 2020, resulting in a staggering increase of anti-Asian hashtags by 17,400% (p. 960). Simultaneously, there was another upsurge in the utilization of generic and historically rooted Sinophobic slurs, including "chink," "chinaman," "chyna," "gook," and "chingchong," on social media platforms (Schild et al., 2020). The usage of these slurs exhibited two notable peaks on January 23 and March 16, 2020, with the latter coinciding with Trump's reference to COVID-19 as the "Chinese virus" on Twitter, highlighting the interconnectedness between public figures' speech and public sentiment.

In historical contexts, striking parallels emerge among instances of Sinophobic sentiment. For example, during the 1875-1876 smallpox epidemic in San Francisco, the city's Health Officer suspected the Chinese population of spreading the disease, resulting in the fumigation of all houses in Chinatown (Zhang, 2022, p. 66). Similarly, during the 2002-2004 SARS outbreak, New York City's Chinatown was wrongly labeled a contagion site, leading to a significant decline in local business and tourism, despite the absence of reported SARS cases (Zhang, 2022, p. 72). These historical events share common themes of scapegoating and stigmatization of the Chinese community. More recently, COVID-19 witnesses this recurring pattern in which blame is again directed at the Chinese and China, where the accusations extend beyond disease transmission to allegations of China purposefully creating the virus as bioweapon (Schild et al., 2020, p. 8). These historical parallels emphasize the enduring nature of Sinophobic sentiment, showcasing the persistence of similar themes of scapegoating across different time periods, each manifesting in response to various disease outbreaks.

2.2 YouTube News Consumption and Racism-Charged Comment

Twitter (now X) and 4chan are among the most utilized platforms for social media sentiment analysis, with extensive literature examining these sites to study online racism. Nevertheless, YouTube has become a valuable resource for public sentiment analysis, providing insights into individuals' opinions triggered and persuaded by video content and interactions within the comments section. Breazu and Machin (2023) analyzed anti-Roma comments posted under a widely viewed YouTube news clip that reported on the lockdown of a Romanian town with a significant Roma population, and their findings revealed that anti-Roma sentiment had transformed into a form of "new racism," characterized by a tendency to avoid overtly racist language in favor of framing ethnic minorities in the context of social or cultural issues or as an economic burden. Likewise, in a study conducted by P. Wang and Catalano (2023), 2,071 comments posted under a YouTube video documenting anti-China rhetoric employed by the Trump administration were examined. Their research unveiled that YouTube, like Twitter and 4chan, served as a platform for fostering unity among racist groups while also acting as a space for expressions of resistance.

These two studies demonstrate that YouTube, particularly within its news video segment, is an under-explored yet rich platform for deepening our comprehension of racist sentiments. Moreover, the utilization of YouTube as a news source is not uncommon in the U.S.: a study shows that approximately 71% of Americans use this platform and about 26% of U.S. adults use it for news consumption. Additionally, 66% of YouTube users claim that YouTube's news videos have enhanced their comprehension of current events, and a significant majority (73%) have confidence in the accuracy of the information provided by the site's news videos (Stocking et al., 2020).

2.3 Google Search Trend

In addition to explicit expressions of racism in YouTube comments, Google search inquiries provide additional insights into people's implicit attitudes. Online search traffic data have been demonstrated to be an effective analyzer of internet behavior and Google Search Trends a reliable tool in predicting changes in human behavior with careful selection of the search terms. Google Trends data have been proven useful in measuring the public's reaction to various outbreaks or incidents, such as Swine flu (Bentley & Ormerod, 2009), the epidemic of Middle East Respiratory Syndrome (Poletto et al., 2016), the Ebola outbreak (Hossain et al., 2016), and measles (Mavragani & Ochoa, 2018). Google searches for the respective terms have also been proven to increase or peak when a public figure or celebrity is related (Bragazzi et al., 2017; Brigo et al., 2014; Pandey et al., 2014).

Evidence strongly suggests that Google Trends can uncover socially sensitive attitudes (Stephens-Davidowitz, 2014), and such implicit searches can complement the explicit opinions on social media. Stephens-Davidowitz proposed the approach of measuring racial animus by analyzing the percentage of Google search queries containing racially charged language. Specifically, he defined an area's racially charged search rate as a robust negative predictor of Obama's vote share. His findings indicated that estimates derived from Google search data are 1.5 to 3 times larger than those obtained through traditional survey-based methods (Stephens-Davidowitz, 2014, p. 27).

2.4 Anti-Chinese Hate Incidents

Public figures' use of Sinophobic language and the Sinophobic public sentiment expressed on social media have been linked to increasing hate incidents towards the Chinese and broader Asian populations (Gover et al., 2020). Hate crimes, as defined by the U.S. Department of Justice, are criminal offenses that are motivated, wholly or in part, by the offender's biases against a race, religion, disability, sexual orientation, ethnicity, gender, or gender identity. In a study conducted in London, UK, the probability of being a victim of hate crimes increased from around 3% before COVID-19 to 10% in February and further rose to over 16% in March 2020 (Gray & Hansen, 2021). According to the FBI, hate crimes against Asian Americans increased by 77% between 2019 and 2020 (Kapadia, 2022). Additionally, a survey conducted by National Public Radio (NPR) found that one in four Asian American households reported being afraid of a physical threat or attack due to their ethnicity (Fadel, 2021). Likewise, Stop Asian Americans and Pacific Islanders (AAPI) Hate reported that out of the 2,583 incidents they received, 70% involved verbal harassment, 9% were physical assaults, and 8% were potential civil rights violations since March 2020 (Tekumalla et al., 2022).

Despite the surge of anti-Chinese and anti-Asian hate crimes around the globe, a large number of these incidents go unreported. Discrepancies in percentages suggest vast under-reporting of hate crimes to the police and reveal the hidden nature of hate crimes against Asian Americans in the U.S. today (Gover et al., 2020). Specifically, the hate crime data collected by the National Crime Victimization Survey reveals that 47.6% of Asian victims do not report to the police, highlighting the issue of under-reporting in official statistics (p. 657).

2.5 Relationship Between Online Racism and Offline Hate Incidents

Online platforms can act as echo chambers that amplify hate speech and racist ideologies, emboldening individuals to carry out discriminatory actions or violence in the real world. In a study conducted in the United Kingdom, utilizing time series analysis on datasets of hateful Twitter posts and hate crime incidents reported to the Metropolitan Police Service, it was observed that offline Islamophobic hate predicts anti-Islamic hate speech, and these factors frequently reinforce a broader climate of hostility towards Islam and Muslims (Wiedlitzka et al., 2023). However, there have been contrasting findings from other studies regarding the correlation between online and offline activities concerning racism and racial issues. Numerous studies on online to offline racism revealed that hateful online content and activity occur before offline hate crimes and incidents (Benesch, 2012; Williams et al., 2020). Williams et al. (2020) also demonstrated a consistent positive correlation between Twitter hate speech targeting race and religion and offline racially and religiously aggravated offenses in London. Additionally, another research conducted in 2016 presented an integrated model based on the numerous previous studies on dangerous speech and ideological dynamics of mass atrocities, indicating the positive correlation between speeches and ideologies and the risk of genocides and mass atrocities (Leader Maynard & Benesch, 2016).

Building upon the above mentioned research endeavors, my research seeks to provide a fresh perspective on overt online Sinophobic sentiments on social media by analyzing comments posted under YouTube news videos reporting the COVID-19 pandemic. Notably, YouTube is an under-explored platform for social media analysis, particularly regarding Sinophobic attitudes. In addition to overt Sinophobia, this paper aims to investigate covert Sinophobic trends by studying Google search data. This involves identifying keywords from previous social media content analyses, such as pandemic-related slurs and broader, historically rooted slurs. By combining these two approaches, my research offers a novel and comprehensive examination of online Sinophobic sentiment triggered or shaped by news on COVID-19. Following previous studies on anti-Chinese hate crimes, I will incorporate FBI's UCR hate crime data to examine the dynamics between online and offline Sinophobia.

3 Data

This study operationalizes online Sinophobia by explicit anti-Chinese sentiment expressed by YouTube comments under COVID-19 pandemic reports and implicit anti-Chinese sentiment manifested by Google searches. Anti-Asian hate crimes reported by the FBI is used as the proxy for offline Sinophobia.

3.1 YouTube News Video Comments

To identify relevant platforms for analyzing Sinophobic sentiment on social media, I compile a list of news channels based on their political stance (Langlois, 2018) and popularity according to a survey by the Pew Research Center (Stocking et al., 2020). I then examine news videos from these channels filtered by the keyword "COVID-19". This analysis includes assessing the amount of COVID-related videos, the status of the comment sections (whether enabled or disabled), and the volume of comments. This meticulous approach enables the selection of three prominent news channels for this study: Fox News, ABC News, and MSNBC News.

In the next phase, the official YouTube Data API is employed to scrape all videos whose titles contain the keyword "covid19" between January 1st, 2020, to December 31st, 2021. To ensure continuity and diversity, three video links per month per channel are randomly selected, resulting in a curated list of 3 channels and 24 months, totaling 204 video links [1]. Adjustments are made by regenerating a small number of links when the initially chosen

[1] FOX news has no relevant video in January and February, one in August, and two in November, 2020; MSNBC news has no relevant video in January, 2020.

ones are deemed irrelevant to the target topic [2] or if the comment section is disabled [3]. Finally, leveraging the YouTube Data API again, I scrape all comments and replies from the randomly selected 204 videos, resulting total 434,715 comments for subsequent content analysis.

The final YouTube comment dataset contains information such as comment texts, video publish dates, comment publish dates, news video's channel name, and the video titles. After tokenizing and normalizing the comment corpus, I filter it by a list of Chinese related or Chinese stigmatizing terms defined by Biswas et al. (2022). These filtering terms include: "China", "Chinese", "Wuhan", "CCP", "Kungflu", "Kung Flu", "Wuflu", "ChineseVirus", "WuhanFlu", "Chinazi", "Chink", "SARS", "Communist", "Communism", "Bat", "Bioweapon". Following this filtering process, a total of 24,986 comments are identified and presumed to either reflect anti-Chinese sentiment or relate to China in some capacity.

3.2 Google Trends

Online search rate for racially charged terms can be a robust proxy for the racial animus (Stephens-Davidowitz, 2014) and can capture the sensitive attitude online, so this study leverages the Google Trends to examine implicit online Sinophobia, complementing the analysis of explicit online Sinophobia in YouTube comments. Unlike YouTube comments, Google Trends provides easy accessibility and facilitates data retrieval based on specific search terms and time periods, enhancing reproducibility for future research. Noticeably, Google Trends data represents not the actual counts of searches but the relative frequency of searches for specific terms. These frequencies are expressed as a weekly percentage of all searches that include the target terms, calculated from a representative random sample at a given location and a fixed time level. The values are then scaled relative to the highest weekly search frequency for any term within the same criteria, which in this study spans from January 2020 to December 2021.

To capture both COVID-19-triggered animus and general Sinophobia, this study uses two sets of search terms. The first keywords set comprises COVID-19 related Sinophobic terms "Chinese virus," "China virus," "Wuhan virus," and "Kung flu," which have recently gained prominence and are increasingly used by the general public and public figures. The

[2] Videos are deemed irrelevant to the study if they do not directly address topics related to COVID-19 or its socio-cultural implications. Irrelevant content includes videos that focus primarily on unrelated political events, media personalities, or other non-pandemic-related subjects, such as election analyses or reactions to non-COVID-related deaths. A total number of 9 videos have irrelevant report content, which are consequently replaced by more suitable videos.

[3] YouTube allows video uploaders to disable the comment section for individual videos when they initially post the video or after some audience have already posted comments. When comments are disabled, previously posted comments will no longer be visible to viewers, and no further comments can be made. A total number of 2 videos have closed comment section.

Table 1: Google search terms and time range

Search Terms	Time Range	Region
Chink + Chinaman + Chinkland + Gook + Chinese virus + China virus + Wuhan virus + Kung flu	1/1/20 - 12/31/21	United States

second set includes broader and historical slurs "Chink," "Chinaman," "Chinkland" and "Gook." The timeline for the Google Trends data aligns with the period covered in the YouTube comment dataset, spanning from January 1st, 2020 to December 31st, 2021. The search terms, time ranges, and regions (Table 1) are inputted into the Google Trends Explore tool, and the resulting data in CSV format is downloaded for subsequent time series analysis.

3.3 FBI Uniform Crime Report

To investigate violence targeting the Chinese population fueled by Sinophobic sentiment, I use hate crime statistics data from the FBI's Uniform Crime Reporting (UCR) Program. According to the Hate Crime Statistic Act (28 U.S.C. § 534), hate crime is defined as "crimes that manifest evidence of prejudice based on race, gender and gender identity, religion, disability, sexual orientation, or ethnicity." The FBI hate crime statistics data provides detailed information on recorded crimes, including the date, crime type, specific precincts within the United States, and descriptions of bias provided by police officers. The Hate Crime Statistics Data Collection compiles data about both single-bias and multiple-bias hate crimes. A single-bias incident is defined as an incident in which one or more offense types are motivated by the same bias. Beginning in 2013, law enforcement agencies could report up to five bias motivations per offense type. The focus of this study is anti-Asian bias, with a particular emphasis on anti-Chinese violence, so the filter mechanism is any reported incident containing "Anti-Asian" bias. As official sources for anti-Chinese hate crime statistics are limited, anti-Asian hate crime statistics serve as a proxy measure. This broader category also encompasses the spillover effect of anti-Chinese sentiments to a wider anti-Asian sentiment based on the "all Asians look the same" stereotype.

Specifically, I use the weekly proportion of anti-Asian hate crimes over all crimes, rather than absolute counts of anti-Asian hate crimes. This is because proportional data allows for a clearer understanding of the scale of anti-Asian hate crimes relative to overall criminal activity, and the subsequent normalization automatically aligns with other normalized datasets.

4 Methods

I begin my research with a comprehensive content analysis to identify key themes in the YouTube comments on news videos about the COVID-19 pandemic. This analysis utilizes Latent Dirichlet Allocation (LDA) topic modeling to determine the subjects of discussion and sentiment classification labeled by large language model to assess the emotions expressed regarding these topics. Subsequently, I focus specifically on comments that mention China and the Chinese community, exploring the prevalent themes and manifestations of hate speech within these discourses. This study progresses to a time series analysis employing the Vector Autoregression (VAR) model and Granger causality tests. This analysis investigates the temporal relationships between online and offline manifestations of Sinophobia, with online factors measured by the amount and intensity of hate speech in YouTube comments and Google searches for Sinophobic terms, and offline factors measured by FBI-reported anti-Asian hate crimes.

4.1 Content Analysis

4.1.1 LDA Topic Modeling

To understand public attitudes toward the Chinese community during the COVID-19 pandemic, I analyze the overarching themes present in the YouTube comment corpus using topic modeling. This unsupervised method identifies recurring word patterns and effectively uncovers the underlying topics that define each cluster of comments. Specifically, I employ Latent Dirichlet Allocation (LDA), a three-tier hierarchical Bayesian model, to discover abstract topics within the YouTube comment corpus. In LDA, documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA can identify each topic by its top terms and examine the complete corpus by analyzing the composition of these topics. This includes determining the most prominent topics, quantifying how many documents contain a given topic, and assessing the distribution of topics across the entire corpus. LDA assumes the following generative process for each document w in a corpus D (Blei et al., 2003, p. 996):

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dir}(\alpha)$.
3. For each of the N words w_n:
   (a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
   (b) Choose a word w_n from $p(w_n \mid z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

According to Figure 1, there are three levels to the LDA representation. The parameters α and β are corpuslevel parameters, assumed to be sampled once in the process of generating a corpus. The variables θ_d are document-level variables, sampled once per document. Finally, the variables z_{dn} and w_{dn} are word-level variables and are sampled once for each word in each document.

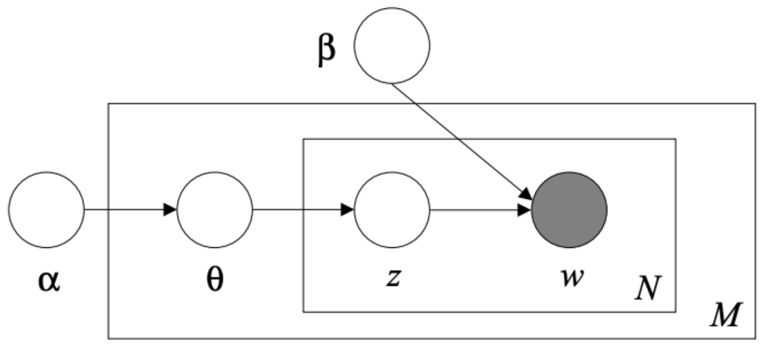

Figure 1: Graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

Given that YouTube comments typically consist of brief text inputs and constructing an effective LDA model requires a substantially large dataset, I aggregate the comment dataset by compiling all comments and replies of a given video as the unit document. This method enhances the robustness of the dataset, providing a more substantial textual basis for identifying and analyzing thematic structures with the LDA model. Afterwards, I create a vocabulary dictionary, filter the missing data and generate a corpus object as the document matrix that LDA model needs as the main input. Text feature extraction is based on vector space model, where a text is viewed as a dot in a N-dimensional space. I use the TF-IDF to weigh a keyword in any content and assign importance to that keyword based on the number of times it appears in the document, which helps enhance relevancy, reduce noise, and increase discriminative power of the topic model to provide a clearer, more differentiated view of the underlying topics.

LDA topic modeling in this study involves three stages: (1) topic modeling on the general corpus, (2) topic modeling on comments mentioning China filtered by keywords, and (3) topic modeling on the Sinophobic corpus where hate speech is detected by the roberta-hate-speech-dynabench-r4-target model.

Table 2: Hate speech score and label by RoBERTa

Comment	Hate Speech Score	Hate Speech
I'm starting to get the idea that Tucker doesn't like Chinese people	0.005	0
The CCP coronavirus	0.020	0
Chinese probably did this to hurt their greatest foe and prevent the reelection of President Donald Trump. Trump is the first president to stand up against the PRC's relentless aggression.	0.780	1
Absolutely spot on, dead perfect monolog. People really need to be ready for this.. we will certainly survive it..but its gonna be here in mass directly. Hopefully we make it a teaching moment..at least its not H7N9 or H5N1.. and we can learn to deal with china. We should have done that many years ago. CHEERS!	0.999	1
We know China doesn't care about its citizens lives or their heath. This was unleashed on us. I believe first and foremost for the economic damage, but also as retaliation for our (correct) policies towards them.	0.969	1

4.1.2 Hate Speech Detection

To further identify Sinophobic comments as mentioned above, I use the pre-trained model facebook/roberta-hate-speech-dynabench-r4-target (Vidgen et al., 2020). This pre-trained large language model is provided by Huggingface based on the RoBERTa architecture, and has been fine-tuned on a dataset of text that contains instances of hate speech. Specifically, Vidgen et al. (2020) incorporates a human-and-model-in-the-loop process, enabling iterative improvements through four rounds of data generation. In its fourth round (r4), this model version is designed to target and identify instances of hate speech more effectively, using a refined dataset that includes a wide variety of expressions and contexts. Performance is significantly improved compared to all four models evaluated by Röttger et al., 2020: the performance of facebook/roberta-hate-speech-dynabench-r4-target is consistent across both 'Hate' and 'Not Hate', achieving 95% and 93% respectively.

Table 2 exemplifies some of the hate speech scores generated by this RoBERTa model and the hate speech tag determined by a threshold of 0.34 computed by the Euclidean distance from each point on the ROC curve to the top-left corner based on a set of 50 randomly selected comments manually labeled by me. This combination of keyword-filtering and hate-speech-filtering results an ultimate Sinophobic comment dataset.

According to Table 2, comments that describe a phenomenon with flat languages generally receive lower hate speech scores, even when using stigmatizing terminology such as "CCP coronavirus." Conversely, comments with higher hate speech scores often feature more intense languages or stronger tones, irrespective of the COVID-19 context. For instance,

Table 3: Sentiment classification by GPT-3.5 Turbo Model

Comment	Labeled Sentiment
Damn crazy this is when it all started. Why tf was he in china	Anger
I had it January 12 all the symptoms doctors could not figure it out I even had a flu swapping and it came back negative and I never been to China I was passing it to my coworkers through the days and some got very sick	Fear
The vaccine has been in development for a long time. Information from the SARS vaccine that never got released was used in the COVID vaccine because of their extremely similar genetic makeup. They are both coronavirus, after all. Also, Government organizations have been donating money to campaigns so that they could research viruses more and be prepared for the next pandemic so a vaccine can be produced quickly. The vaccine has been in development way longer behind the scenes. If you ask me, it's amazing that science and technology has been able to progress this far to the point that we can have a safe vaccine within a year. Also, there has never been a vaccine that has ever had any adverse affects after 6 months of taking it. There are no "long term" affects to taking a vaccine	Joy
Let's let trump call it the Wuhan virus that's ok but let it be forever known as the "trump pandemic"; that's killing America.	Irony/Sarcasm

while "The CCP coronavirus" is associated with a lower hate speech score, the statement "We know China doesn't care about its citizens' lives or their deaths..." receives a significantly higher hate speech score.

4.1.3 Sentiment Classification

To analyze public sentiments beyond Sinophobia, this study utilizes GPT-3.5 Turbo model provided by OpenAI to classify all comments in the corpus filtered by China-related keywords. This large language model represents a significant improvement over its predecessors due to its enhanced ability to understand and generate highly context-aware text. It facilitates precise multi-class text classification through zero-shot learning, which allows for effective categorization of comments into predefined sentiment labels without the need for additional labeled training data. This method involves custom prompt engineering to tailor the model's capabilities to the specific needs of our sentiment analysis, ensuring accurate and detailed understanding of the nuanced public sentiments expressed in the comments.

I prompt the model with a specific sentiment analysis instruction: "You are an AI trained to analyze sentiment. Categorize the following comment into one of the emotions: Anger, Fear, Sadness, Irony/Sarcasm, Neutral, Joy." GPT-3.5 turbo's sentiment classification is notably precise, exceeding the supervised SVM classifier's accuracy. For example, it correctly labels “Trump Rocks! Wake up People!” as “Joy,” while SVM classifier labels it as “Anger” and VADER model labels it as “Neutral”. Table 3 demonstrates some example comments with their corresponding sentiment labeled by the model.

4.2 Time Series Analysis

To investigate the dynamics between online expressions of Sinophobia indicated by YouTube comments and Google Search Trends, and offline manifestations of anti-Asian hate crimes measured by FBI hate crime statistics, I employ a multivariate time series analysis. This analysis assesses the strength of associations among various variables over time, employing cross-correlations to explore the relationships at different time lags. These lags allow for an examination of how past values of one variable may influence future values of another, capturing temporal dependencies within the data.

4.2.1 Ordinary Least Squares Regression

Before introducing lags and autoregressive model, I first apply Ordinary Least Squares (OLS) regression to identify basic correlations and potential predictive relationships between different Sinophobia variables. This step provides a preliminary understanding of which variables have significant linear relationships, which can guide more complex analyses afterwards. OLS regression estimates the parameters of a linear regression model, aiming to find the values of the linear regression model’s parameters (i.e., the coefficients) that minimize the sum of the squared residuals. The residuals are the differences between the observed values of the dependent variable and the predicted values of the dependent variable given the independent variables. OLS algorithm assumes that the errors are normally distributed with zero mean and constant variance and that there is no multicollinearity (high correlation) among the independent variables.

4.2.2 Vector Autoregression Model

The Vector Autoregression (VAR) model is a statistical model that captures the linear interdependencies among multiple time series first introduced by Sims (1980). It operates by regressing each variable on its own lagged past values and the lagged values of all other variables in the system. It generalizes the single-variable (univariate) autoregressive model by allowing for multivariate time series. The key assumptions of the VAR model are stationarity, linearity, and a constant covariance matrix of the error terms. Additionally, the VAR model assumes that variables in the system have a contemporaneous effect on each other, capturing the dynamic interactions within the system. Specifically, VAR is a n-equation, n-variable linear model in which each variable is in turn explained by its own lagged values, plus current and past values of the remaining n-1 variables. In this study, by denoting the YouTube comment hate amount by $x_{t,1}$, hate intensity by $x_{t,2}$, Sinophobic Google searches by $x_{t,3}$, and anti-Asian hate crime rate by $x_{t,4}$, the vector autoregressive model of order 1, i.e., with a time lag of 1 week, denoted as VAR(1), can be calculated as below:

- $x_{t,1} = \alpha_1 + \phi_{11}x_{t-1,1} + \phi_{12}x_{t-1,2} + \phi_{13}x_{t-1,3} + \phi_{14}x_{t-1,4} + w_{t,1}$
- $x_{t,2} = \alpha_2 + \phi_{21}x_{t-1,1} + \phi_{22}x_{t-1,2} + \phi_{23}x_{t-1,3} + \phi_{24}x_{t-1,4} + w_{t,2}$
- $x_{t,3} = \alpha_3 + \phi_{31}x_{t-1,1} + \phi_{32}x_{t-1,2} + \phi_{33}x_{t-1,3} + \phi_{34}x_{t-1,4} + w_{t,3}$
- $x_{t,4} = \alpha_4 + \phi_{41}x_{t-1,1} + \phi_{42}x_{t-1,2} + \phi_{43}x_{t-1,3} + \phi_{44}x_{t-1,4} + w_{t,4}$

The VAR model is highly effective in capturing variations in autocorrelated data series, making it particularly valuable for this analysis. It delineates causal relationships among multiple time series and enables the prediction of future trends based on these interrelations. The efficiency of the VAR model in both explaining past interactions and forecasting future occurrences relies on the accurate specification of the model's structure and the precise estimation of its parameters (Bose et al., 2017). A key advantage of the VAR model is its lack of restrictive assumptions regarding the structure or distribution of the data, allowing it to forecast any time series variables that are either stationary or cointegrated. However, the VAR model also presents certain challenges. It requires a substantial amount of data and careful selection of lag length: Lütkepohl (2005) cautioned that using too few lags can result in autocorrelated errors whereas using too many lags results in over-fitting, causing an increase in mean-square-forecast errors of the VAR model. Lag length p refers to the number of previous observations in a time series that will be used as predictors in the VAR model. Typically, a large number of lags will be used to generate a model and then a restriction applied to select a more parsimonious model. The lag length for the VAR (p) model can be determined using model selection criteria such as Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (BIC), and the Hannan-Quinn criterion (HQ) (Lütkepohl, 2005, p. 325). Additionally, the VAR model may not adequately account for structural changes or nonlinearities in the data, which necessitates testing for stability, heteroscedasticity, and nonlinearity to ensure robustness and reliability in the model's outputs.

4.2.3 Granger Causality

To interpret the results for the VAR model, Granger-causality statistics is used to examine whether lagged values of one variable helps to predict another variable. Granger causality is a concept of causality derived from the notion that causes may not occur after effects and that if one variable is the cause of another, knowing the status on the cause at an earlier point in time can enhance prediction of the effect at a later point in time (Granger, 1969). Granger causality tests determine whether the inclusion of a second time series enhances the prediction of a first time series, implying a causal influence from the latter to the former. This is tested by fitting two autoregressive models to the first time series: one including the second time series and one without it. The effectiveness of this inclusion is assessed by comparing the variance in the error terms of both models. The null hypothesis in Granger causality states that the second time series adds no explanatory power when its lagged values are considered jointly with those of the first time series as predictors. This hypothesis is rejected if the lagged values of the second time series significantly improve the forecast, indicating a causal influence. The VAR model's flexibility facilitates this type of analysis, allowing for a robust framework to test and interpret Granger causality (Bose et al., 2017, p. 8). Through VAR modeling with Granger causality tests, I can examine whether observable patterns in online Sinophobia are predictive of subsequent shifts in hate crime rates, offering insights into the causal mechanisms at play between digital discourse and real-world outcomes.

5 Results

5.1 Content Analysis

5.1.1 Topic Modeling

The topic modeling results, illustrated by the top 10 terms for each topic shown in Figure 2, reveal a diverse range of discussion topics beyond hared toward China and the Chinese community within the complete, unfiltered YouTube comment corpus. Upon aggregating all comments from specific videos into the input documents, Figure 3 displays the distribution of dominant topics across the discourse. Notably, the most prevalent topic pertains to the global impact of the coronavirus, particularly in the United States and China. This discussion includes discussions on confirmed cases, infection rates, death rates, and speculations about the pandemic's future trajectory. The second most prominent theme revolves around political responses to the crisis, encompassing political figures' statements, policy announcements, and presidential debates. This theme also extends to the 2020 U.S. presidential election involving Donald Trump and Joe Biden. The third leading topic captures people's personal experiences and opinions regarding COVID-19 vaccinations, which often includes descriptions of the vaccination process at local pharmacies and the symptoms experienced post-vaccination. The next topic, though less prevalent compared to the prior

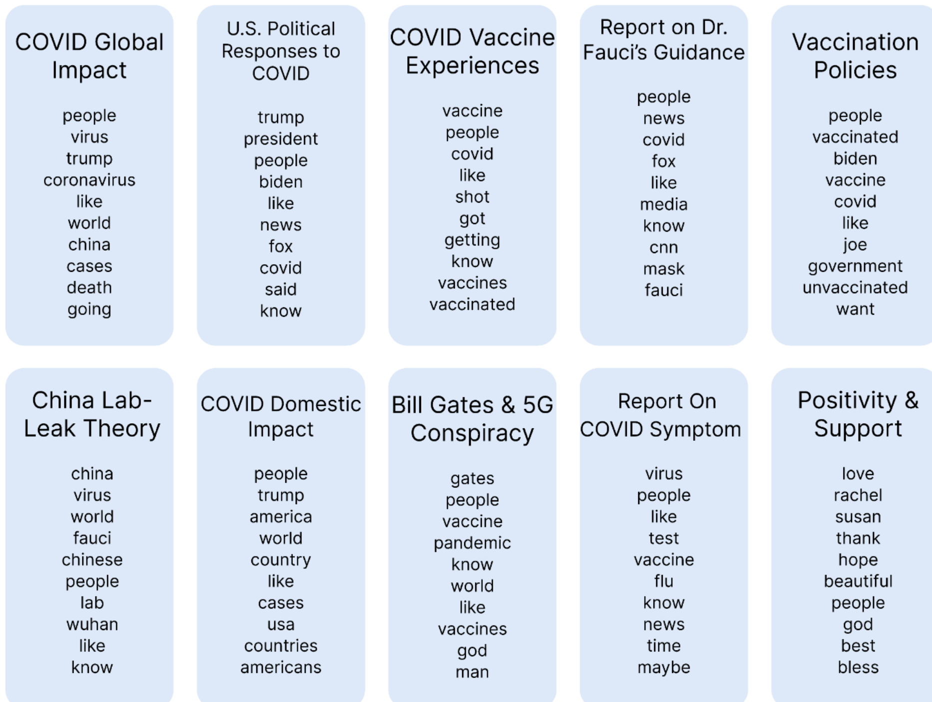

Figure 2: Top 10 terms for each topic in all YouTube comments

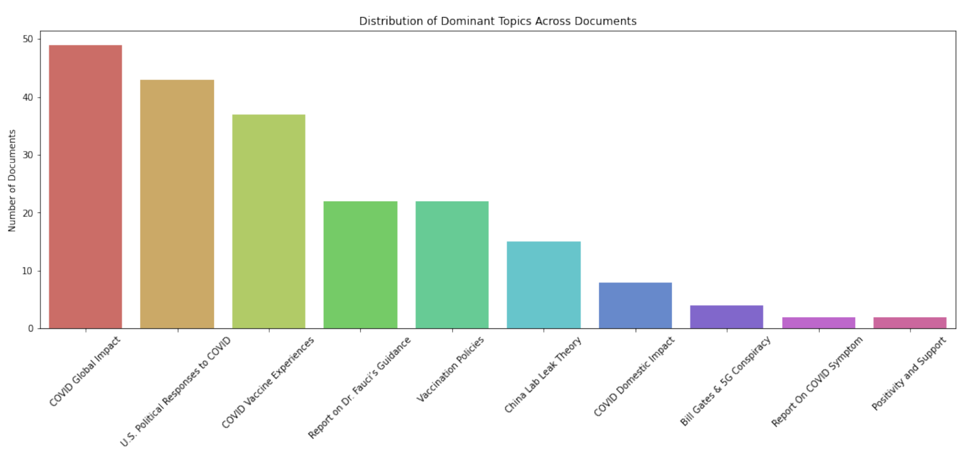

Figure 3: Distribution of dominant topics across all comments

three, focuses on Dr. Anthony Fauci, who served as a leading member of the White House Coronavirus Task Force under former President Donald Trump. Dr. Fauci's guidance during the pandemic often conflicted with Trump's statements, leading to accusations from Trump's supporters that Fauci was attempting to politically undermine Trump's reelection campaign. The comments within this topic are divided: some support Dr. Fauci's medical advice and appeal for fellow YouTube consumer to follow such advice, while others accuse him of disseminating misinformation. The fifth topic centers on vaccination policies, distinguishing itself from the third topic which primarily addresses personal experiences related to receiving or refusing the vaccine. This topic frequently includes comments that criticize President Biden's vaccination mandates, blame the unvaccinated for contributing to the uncontrolled spread of the virus, or call for vaccination in the interest of public safety. The next topic is the only predominant topic that directly attacks Chinese community by accusing China either for deliberately leaking the coronavirus into the world or engineering the virus as a bio-weapon. The next four topics, although less prominent, cover a wide array of discussions: the domestic impact of the pandemic, conspiracy theories about Bill Gates developing and distributing the vaccine as human micro-chipping, or the transmission of coronavirus using the newly-developed 5G networks, the symptoms associated with COVID-19 infection, and a distinctly positive topic arises from discussions around a specific video featuring Rachel Maddow on MSNBC, where she shares her emotional response to her partner Susan's battle with COVID-19, encapsulating support and blessings for loved ones affected by the virus.

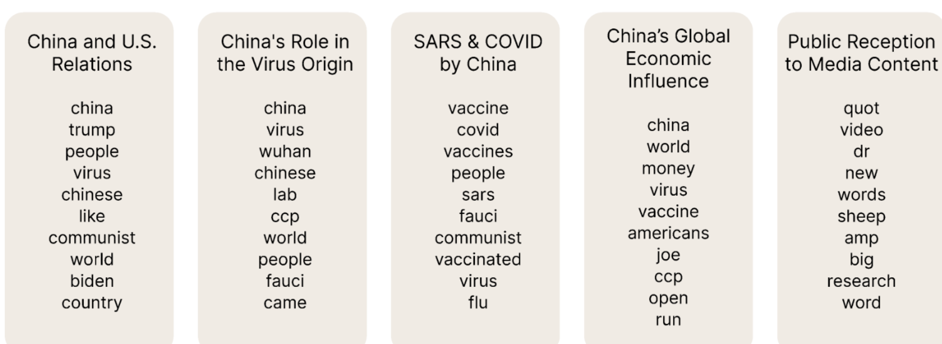

Figure 4: Top 10 terms for each topic in Chinese-keywords-filtered comments

While China and the Chinese community are not the most dominant target in discussions about the COVID-19 pandemic on YouTube, a significant number of comments still reflect Sinophobic themes and sentiments (Figure 4). A predominant theme emerging from these discussions is Sino-US relations, particularly focusing on pandemic management in both

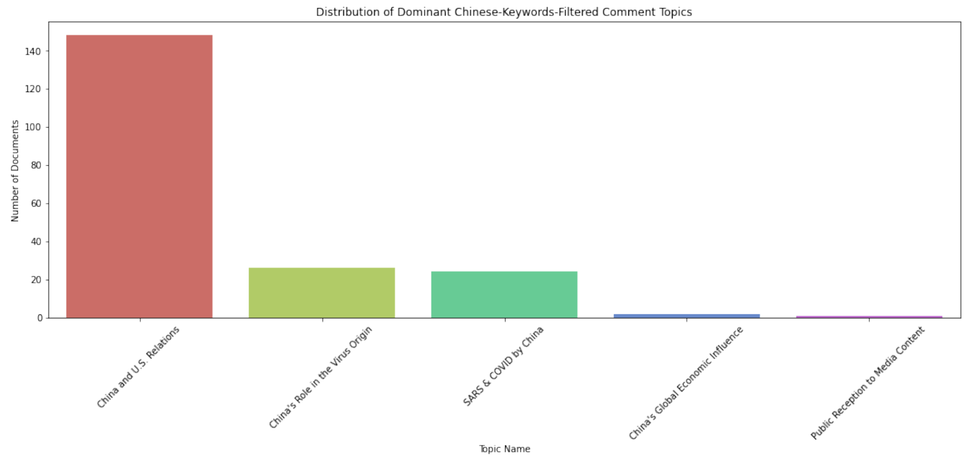

Figure 5: Distribution of dominant topics across Chinese-keywords-filtered comments

countries and the ideological differences that fuel antipathy towards China and its citizens. Common topic terms such as “China,” “Chinese,” “Trump,” “Biden,” “communist,” and “country” often appear in these discussions, highlighting the geopolitical and ideological tensions highlighted in the discourse. This antipathy is frequently connected to other topics involving global economic impact, perceptions of the virus’s origin, and the parallel between COVID and SARS, encapsulating a complex web of sentiments that go beyond health concerns to international relations and political ideologies. The next two leading topics relate to China’s role in the virus’s origin and spread, often tying back to its political party. These conversations frequently draw parallels between SARS and COVID-19, where comments reflect a “yellow fever” sentiment which critique China and the Chinese people for their alleged roles in propagating infectious diseases globally. This discourse typically emphasizes perceived political and health-related failures, illustrating a strong undercurrent of criticism aimed at China’s handling of the pandemic and its initial response. The final two topics focus on China’s impact on the global economy and public discourse. The former explores China’s significant influence through actions such as travel bans and its role in the vaccine development, emphasizing on the China’s substantial effect on worldwide economic dynamics during the pandemic. The latter centers on the critique of and engagement with media content, as YouTube users frequently reference additional source materials and tag other users in their responses to specific posts. According to the above topic modeling results on Chinese-keywords-filtered comments, an overarching negative sentiment appear without further filtering using hate score.

Using hate scores to identify Sinophobic comments, Figures 6 and 7 illustrate the topic modeling results of online Sinophobia related to the COVID-19 pandemic on YouTube.

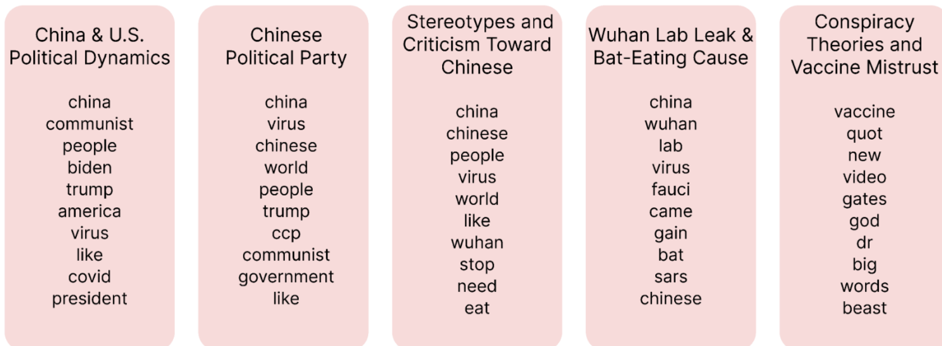

Figure 6: Top 10 terms for each topic in Sinophobic comments

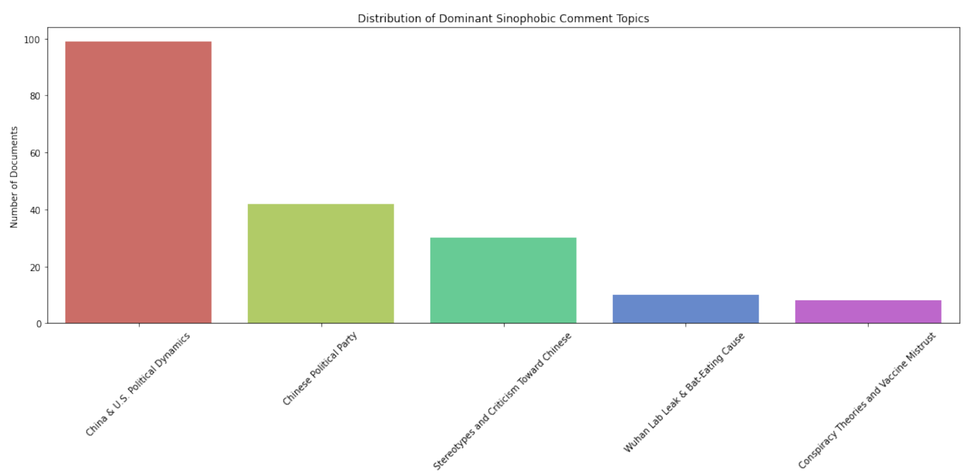

Figure 7: Distribution of dominant topics across Sinophobic comments

These discussions reveal specific criticisms against China and the Chinese people, encompassing several themes: the role of the Chinese political party, deep-rooted stereotypes about Chinese people's eating habits and hygienic underdevelopment, and conspiracy theories about the origin of the coronavirus, including speculations about a lab leak in Wuhan or transmission through the consumption of bats. These topics not only highlight long-standing "yellow fever" stereotypes portraying Chinese dietary customs as barbaric and disease-ridden, but they also reflect the "minority threat" narrative among mainstream discourse, suggesting a deliberate creation of a bio-weapon due to ideological conflicts.

5.1.2 Sentiment Classification

Figure 8 reveals the sentiment distribution of the Chinese-keywords-filtered comment corpus and captures people's emotional response triggered by COVID-19 news relating to China and the Chinese people. "Anger" emerges as the absolute predominant sentiment, significantly overshadowing others, indicating a strong frustration response towards or relates to the Chinese community. "Fear" ranks as the second most frequent sentiment, though markedly less pronounced than "Anger." "Irony/Sarcasm" and "Neutral" are observed next, sharing a comparable presence. Conversely, "Joy" and "Sadness" are the least represented emotions, highlighting their comparatively insignificant occurrence in the dataset.

Overall, discussions about China or the Chinese community in the context of the pandemic often exhibit negative emotions, particularly in those report the virus's origin or preventative actions. The notably lower frequencies of "Joy" and "Sadness" imply that comments are more reactive or critical rather than based on personal feelings or positive reflections. Additionally, the presence of "Irony" indicates a complex engagement with the news content, reflecting skepticism or a nuanced perspective on the serious topic of COVID-19.

5.2 Hate Intensity and Hate Amount

Hate intensity and hate amount are two distinct indicators for understanding hate speech in the YouTube comment corpus. Using the RoBERTa hate detection model, hate intensity is measured by the average hate speech score for comments posted within a given week, indicating the level of hate speech severity. In contrast, hate amount refers to the total number of comments identified as hateful during the same period, reflecting the volume of hate speech. To account for the potential variability in comment volumes collected, the proportion of hateful comments relative to the total number of comments is used as the indicator of hate amount. This method ensures a standardized measure that facilitates more accurate comparisons across different time period. For instance, a week characterized

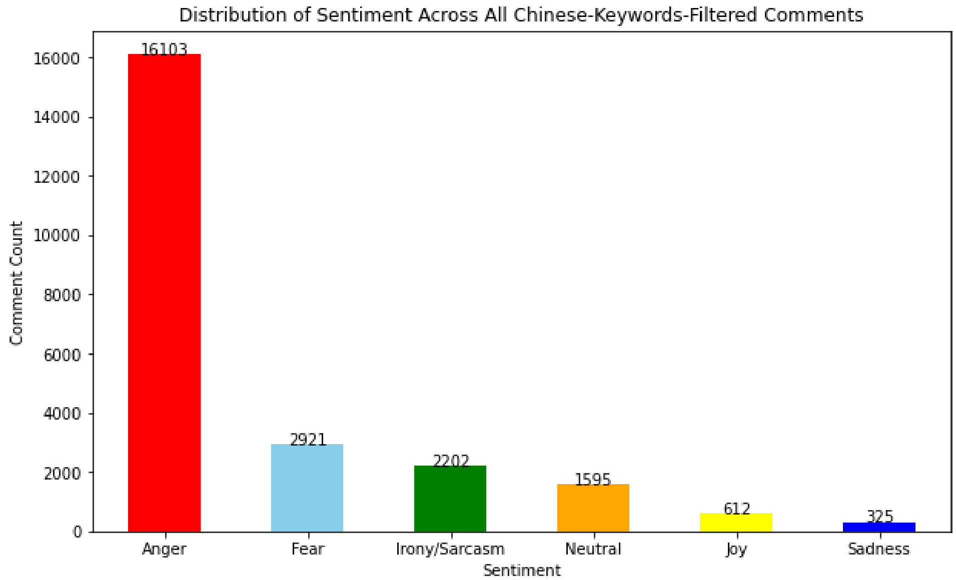

Figure 8: Sentiment distribution classified by GPT-3.5 Turbo Model across Chinese-Keywords-Filtered Comments

by a large volume of comments containing moderately hateful language would reflect a high hate amount but a low hate intensity. Conversely, another week with fewer comments, each marked by highly offensive languages, would demonstrate high hate intensity alongside a low hate amount. This differentiation allows for a nuanced understanding of hate speech dynamics against the Chinese community, capturing both the prevalence and severity of hateful expressions over time.

According to Figure 9, hate intensity largely precedes hate amount while the two share similar temporal trends. This pattern indicates a broadcasting effect where a smaller group initially expresses strong negative sentiments, setting a tone that later attracts a larger audience. As more individuals contribute, the overall intensity may stabilize or decrease due to a dilution effect, where the mix of incoming comments includes both moderate and extreme sentiments. Early in 2020, high levels of hate intensity and amount were possibly triggered by the initial COVID-19 outbreak and the ensuing lockdown in the U.S. A notable second surge occurred in October, largely driven by several key videos: a FOX News video titled "Trump says he no longer has coronavirus and is 'immune'" with 8,929 comments, "Trump's recent updates on coronavirus condition are 'encouraging'" with 2,964 comments, and "Trump joins 'Tucker' for first on-camera interview since COVID-19 diagnosis" with 16,725 comments. The high levels of hate amount and intensity continued into 2021, likely influenced by the Biden Administration's implementation of the National Strategy for the

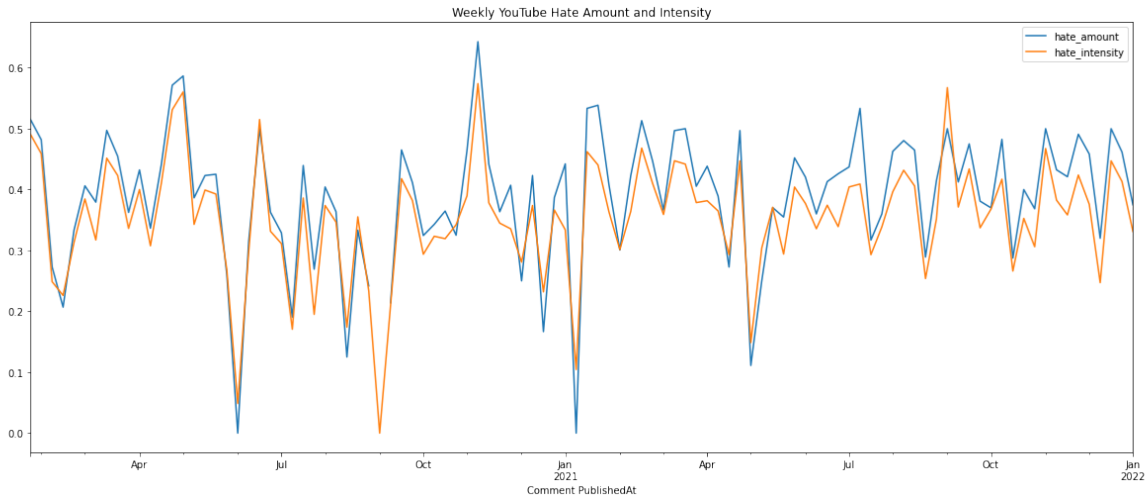

Figure 9: Weekly hate amount and hate intensity of Chinese-keywords-filtered comments

COVID-19 Response and vaccination mandates on January 21, 2021. This period is also marked by the emergence of various coronavirus variants, including the “Gamma” and “Alpha” in January, “Delta” in June, and “Omicron” in November, each contributing to the sustained intensity and volume of hateful comments in 2021.

Figure 10 (middle) shows the distribution of overall hate speech scores, which generally trends towards lower scores, indicating that most comments contain small or moderate amount hateful language. However, Figure 10 (right) shows a bimodal distribution with a noticeable uptick in the number of comments that score at the highest level of hate speech. This suggests that while overtly hateful comments are less common, they are still notable within the dataset. Figure 10 (left) breaks down the hate speech distribution across three different news channels and reveals distinct patterns of hate in these news consumption platforms. First, while ABC News consistently exhibits high hate amount levels similar to FOX News, the timing of their peaks in hate speech differs. Secondly, MSNBC News generally shows a lower volume of hateful comments compared to FOX and ABC News, with the notable exception of a spike at the end of 2021. This consistently lower hate amount and intensity potentially demonstrate MSNBC news channel as a more liberate and positive social media platform where its audience tend to express less Sinophobic sentiment. This disparity in hate speech patterns across channels could be reflective of the varying viewer demographics and editorial policies of each network, which in turn influence the nature and tone of user interactions. Indeed, Figure 11 shows that comments under ABC news and FOX news videos tend to have higher hate speech score than MSNBC news, with ABC even has a higher upper quartile than FOX news; however, the median hate speech scores are similar across news channels.

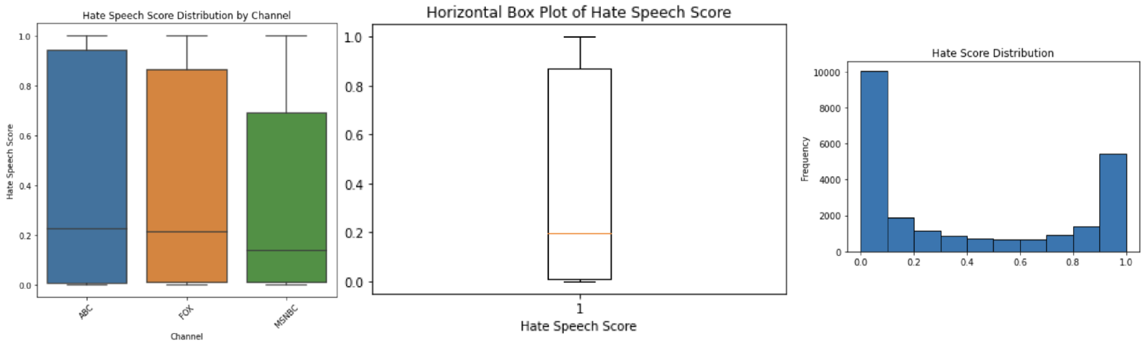

Figure 10: Distribution of the comment hate scores (left), box plots of hate score among three channels (middle) and the overall hate score distribution (right)

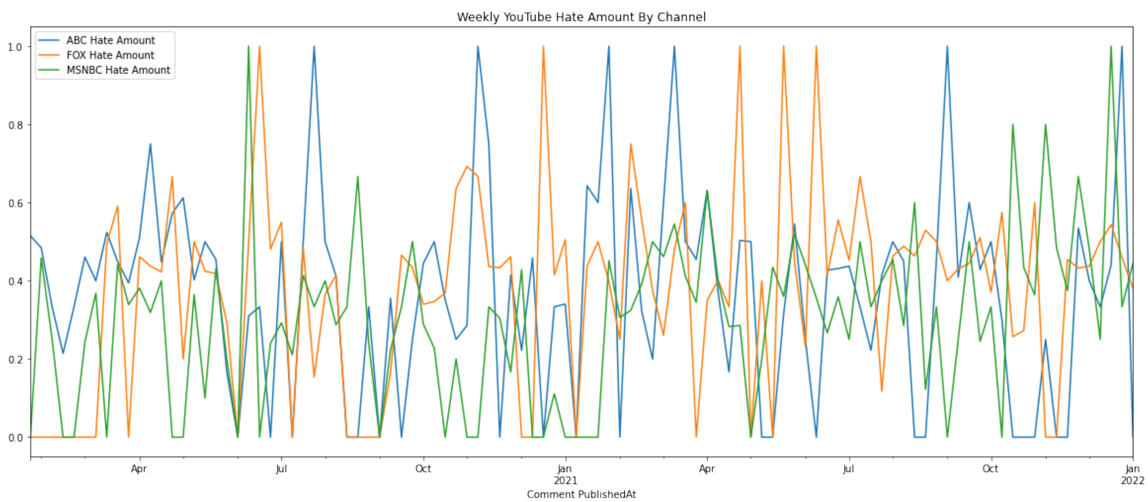

Figure 11: Weekly hate amount proportion of Chinese-keywords-filtered comments by channels

5.3 Time Series Analysis

5.3.1 Temporal Trend Analysis

March and April 2020 saw significant surges in Google searches for Sinophobic terms and incidents of anti-Asian hate crimes. These peaks corresponded with a sharp rise in confirmed COVID-19 cases and deaths in the U.S., alongside several critical events. On March 13, the Trump Administration declared a nationwide emergency and issued an additional travel ban on non-U.S. citizens. By March 15, states began implementing shutdowns to curb the spread of the virus. The situation was further exacerbated when former President Trump referred to COVID-19 as “the Chinese virus” in a tweet on March 16, which could incite increasing online Sinophobia and the Google searches of this racially derogatory term. Later, on March 28, the White House extended all social distancing measures. On March 31, Dr. Anthony Fauci and Dr. Deborah Brix projected that the U.S. could see between 100,000 and 240,000 deaths—a stark reminder of the pandemic’s potential impact.

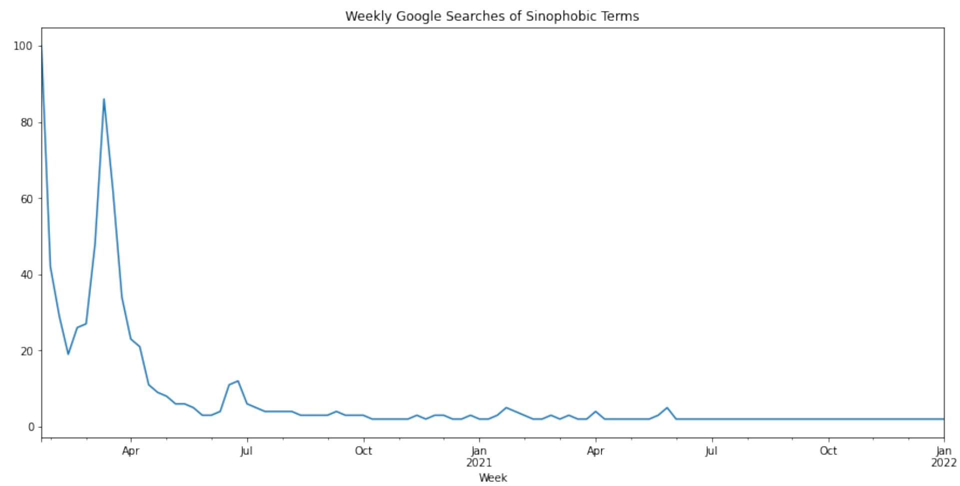

Figure 12: Weekly Google Search Trends for Sinophobic Terms

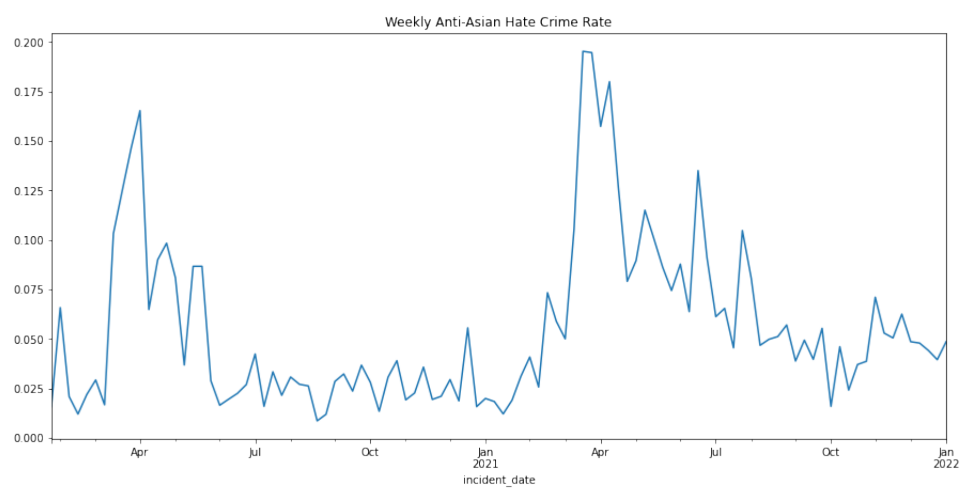

Figure 13: Weekly Anti-Asian Hate Crime Rate

In 2020, there were 357 anti-Asian hate crime incidents among total 9949 hate crime incidents; while in 2021, there were 778 anti-Asian hate crime incidents among total 10889 hate crime incidents. This increase in anti-Asian hate crimes from 2020 to 2021 can be attributed to multiple factors, including a lagged effect of pandemic-related rhetoric that associated COVID-19 with Asian communities, the rise in awareness and reporting mechanisms among hate crime victims, and the gradual restrictions-lifting polices that allowed people to enter public spaces. Noticeably, there was a sudden surge in anti-Asian hate crimes from March to May 2021. This sudden crime surge might be influenced by the Stop Asian Hate movement, catalyzed by the Atlanta shootings on March 16, 2021, where six of the eight victims were women of Asian descent. This tragic event brought widespread attention to the issue, prompting increased reporting and awareness of racially motivated attacks against Asians.

5.3.2 OLS Regression

The OLS regression analysis reveals nuanced relationships between online Sinophobia and real-world hate crimes during COVID-19 (Appendix B). First, as indicated by the temporal trends of weekly YouTube hate amount and intensity in the previous hate analysis (Figure 9), these two time series objects are indeed highly correlated. According to Figure 15 (Appendix B), a significant positive correlation exists between hate intensity and hate amount, evidenced by a coefficient of 1.157. The model's R-squared value of 0.916 indicates that hate intensity accounts for approximately 91.6% of the variance in hate amount, denoting a remarkably high level of linear correlation. These findings imply that extreme Sinophobic content on YouTube, represented by heightened hate intensity, may act as a catalyst, normalizing online Sinophobia and potentially encouraging a broader participation in the posting of such content.

While it is presumed that the number of new confirmed infection cases might instigate Sinophobic sentiments online and offline, OLS regression analyses yield surprising results. As illustrated in Figures 16, 17, and 18 (Appendix B), the regression models with confirmed infection cases as the sole independent variable show minimal explanatory power for YouTube comment hate amount, Sinophobic Google searches, and anti-Asian hate crime rates, demonstrated by very low R-squared values of 0.000, 0.138, and 0.078, respectively. The introduction of cumulative confirmed infection cases as an additional independent variable, however, increases the models' R-squared to 0.042, 0.227, and 0.191, respectively, indicating a significant improvement in explanatory power. Furthermore, when examining the YouTube comment hate amount, the coefficient for weekly confirmed infection cases shifts from positive to negative upon the inclusion of cumulative infection cases, which may indicate the presence of multicollinearity. This suggests that the independent variables may be correlated with each other, challenging the model's ability to isolate their individual effects on the dependent variable. Additionally, the reduced magnitude of the weekly infection cases' coefficient, when accounting for cumulative infection cases, suggests that while there is a correlation with Sinophobic Google searches, the effect decreases as the cumulative infection count rises. This implies that the public's responsiveness to weekly new cases may be dampened in the context of a larger ongoing outbreak, while an initial small amount of infection cases tend to trigger larger effect on people's perception and sentiment.

Lastly, when accounting for weekly confirmed new cases and cumulative infection cases, both YouTube comment hate amount and Sinophobic Google searches are positively correlated with anti-Asian hate crime rate, with R-squared 0.215 and 0.197 respectively (Figure 19, Appendix B). This indicates that roughly 21.5% and 19.7% of the variation in anti-Asian hate crimes can be explained by these online factors. This suggests that an increase of online Sinophobic sentiment may contribute to or reflect an increase in real-world anti-Asian hate crime incidents, pointing to an important relationship between online discourse and offline behaviors.

5.3.3 VAR Model and Granger Causality

	Hate Amount_x	Hate Intensity_x	Google Searches_x	Anti-Asian Crimes_x	Confirmed Infectious Cases_x
Hate Amount_y	1.0000	0.1111	0.1329	0.2985	0.4081
Hate Intensity_y	0.3696	1.0000	0.1075	0.2229	0.5769
Google Searches_y	0.2899	0.2316	1.0000	**0.0198	0.8181
Anti-Asian Crimes_y	0.5840	0.6790	**0.0386	1.0000	0.2957
Confirmed Infectious Cases_y	0.0571	0.1796	0.7898	0.4562	1.0000

Table 4: P-values table at lag=2 week

	Hate Amount_x	Hate Intensity_x	Google Searches_x	Anti-Asian Crimes_x	Confirmed Infectious Cases_x
Hate Amount_y	1.0000	*0.0959	0.1968	0.3409	*0.0687
Hate Intensity_y	0.2404	1.0000	0.1024	0.2774	0.5048
Google Searches_y	0.4023	0.2807	1.0000	0.1458	0.1808
Anti-Asian Crimes_y	0.4554	0.6534	**0.0000	1.0000	**0.0409
Confirmed Infectious Cases_y	**0.0253	*0.0651	0.8490	0.4856	1.0000

Table 5: P-values table at lag=6 week

	Hate Amount_x	Hate Intensity_x	Google Searches_x	Anti-Asian Crimes_x	Confirmed Infectious Cases_x
Hate Amount_y	1.0000	0.1111	**0.0004	0.2985	**0.0259
Hate Intensity_y	0.1223	1.0000	**0.0007	0.2229	0.2832
Google Searches_y	**0.0002	**0.0003	1.0000	**0.0041	*0.0862
Anti-Asian Crimes_y	0.1679	0.2620	**0.0000	1.0000	**0.0000
Confirmed Infectious Cases_y	**0.0266	*0.0677	0.7898	0.4562	1.0000

Table 6: P-values table at lag=12 week

After applying the Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests, it is determined that the time series for Google searches on Sinophobic terms was non-stationary. To address this non-stationary issue, I difference the Google search series to achieve stationarity, exclude its NA values, and align the other time series data to match this modified time frame. An optimal lag of 12 weeks is further identified for the VAR model using the AIC and BIC lag selection criteria. At a 2-week lag, Table 4 presents a significant p-value of 0.0386 with Google search time series as independent variable and anti-Asian crime rate as dependent variable, indicating that weekly Google searches for Sinophobic terms Granger cause weekly anti-Asian hate crime rate. Notably, the Granger causality is bidirectional, suggesting that changes in the rate of anti-Asian hate crimes also forecast subsequent searches for Sinophobic terms. At lag of 6 weeks (Table 5), this bidirectional Granger causality reappears between Google searches for Sinophobic terms and anti-Asian hate crime rate. Furthermore, using a significant level of 0.1, the coronavirus infection cases Granger cause both the YouTube comment hate amount and the anti-Asian hate crime rate.

At a lag of 12 weeks, the Granger causality relationships are more pronounced: the hate amount in YouTube comments Granger cause Sinophobic Google searches with a p-value of 0.0002, and similarly, the hate intensity in YouTube comments also Granger causes Google searches with a p-value of 0.0003, which aligns with the high correlation observed between hate amount and hate intensity; both of these Granger causality relationships are bidirectional. Similar to the 2-week lag, the rate of anti-Asian hate crimes is shown to Granger cause Sinophobic searches with a p-value of 0.0041. Furthermore, confirmed coronavirus infection cases demonstrate strong predictive power at lag 12, Granger causing the YouTube hate amount amount, Sinophobic searches, and the rate of anti-Asian hate crimes with p-values of 0.0259, 0.0862, 0.0000, respectively. However, this result should be analyzed with caution because a lag value of 12 weeks is relative large, which may introduce the concern of over-fitting the model.

6 Discussion

6.1 Main Takeaway

The content analysis of discussions under COVID-19 news reports indicates that while China and the Chinese community were significant topics, they were among a variety of subjects that captured public interest. These ranged from skepticism towards national policies (masking, quarantine orders, and vaccination) to critiques and support of government leaders (Trump, Biden, and Obama) for their pandemic response. Additionally, a large number of discussions feature conspiracy theories unrelated to China, such as the purported use of 5G networks to spread the coronavirus, Bill Gates' alleged plan to implant microchips through vaccination, and the U.S. government bio-engineered the coronavirus. Other topics also include criticism of the media for spreading misinformation, the origin of the coronavirus, and commentary on the 2020 presidential election.

Focusing on comments containing keywords related to China or the Chinese, a pattern of criticism and anger emerges. The dominant theme is blame for the Chinese government and the Communist Party of China (CCP) for the virus's origin and spread, coupled with conspiracy theories suggesting the virus was either leaked from a Chinese lab or deliberately engineered as a bioweapon. YouTube news videos, particularly from the FOX channel, often feature content related to controversial and unverified conspiracy theories. Titles such as "Exclusive: Former CDC director believes COVID-19 came from Wuhan lab," "Bongino says 'evidence was everywhere' for COVID lab leak theory," and "COVID cover-up efforts by Chinese Communist Party are 'staggering'" exemplify the channel's focus on these narratives. Such provocative content appears to spark a large wave of Sinophobic commentary, with comments posted under these videos stating "it wasn't a lab leak, it was all by design," "Nobody will ever believe this happened naturally," and "You are blind to the evil CCP who made the bioweapon and killed 1,000,000 people around the world. Are you blind of the deadly virus? Are you blind of the killing?" These detailed criticisms regarding historical events, suggesting a direct link between the video content and the emergence of hostile online sentiments toward China. The frequent references to the CCP highlight a significant association of China's policies with its ruling party, reflecting deep-rooted skepticism towards the CCP on social media. Additionally, mentions of SARS in discussions about China and the pandemic often carry undertones of "yellow peril" racism, further accusing the Chinese community of being responsible for the emergence and dissemination of contagious diseases.

The sentiment analysis reveals an overwhelming angry emotion among comments posted under COVID-19 news report filtered by Chinese keywords, which manifests the frustration against China and the Chinese government for mishandling the emergence of the pandemic and for providing misinformation regarding to the virus. The varying outcomes from different machine learning methods and the low agreement score among human annotators underscore the challenges in assigning definitive sentiment labels to social media posts. Nevertheless, large language models have shown a surprising level of accuracy, surpassing traditional supervised machine learning classifiers in this task. Statistically, the strong predictive power of hate intensity, measured by LLM generated hate score, over hate amount, measured by the count of hateful comments exceeding a certain hate score threshold, reveals a potential trend of extreme Sinophobia being normalized and adopted on social media. Visually, the trend of weekly hate speech score precedes that of weekly hate speech count, indicating a decrease in hate intensity and an increase in hate amount over time. This pattern implies that a smaller volume of highly hateful comments typically surfaces on YouTube before similar, albeit less intense, sentiments are adopted more broadly by the general public.

Time series analysis reveals a consistent positive correlation between online and offline

Sinophobia, with the relationship being moderated by the progression of the pandemic as measured by weekly new confirmed infection cases and cumulative infection cases. This suggests that as the pandemic advances, the rise in real-life anti-Asian hate crimes corresponds with the increase in explicit and implicit online Sinophobic sentiment, evidenced by metrics such as Sinophobic Google searches and the amount and intensity of YouTube comment hate. Moreover, the pandemic's severity and infection rates themselves are significant predictors of Sinophobic behavior both on the internet and in the real world, although this influence is tempered as the cumulative number of cases grows. The analysis reveals that public Sinophobic reactions tend to diminish over time; increases in an already large base of infection cases do not elicit as substantial a response as they do from a smaller base.

The significant bidirectional Granger causality identified between Sinophobic online activity and real-world anti-Asian hate crimes, at both short (2 and 6-week) and longer (12-week) lags, suggests a strong reciprocal influence where increases in one can predict rises in the other. This relationship emphasizes the role of online platforms in reflecting and potentially amplifying discriminatory behavior. Conversely, real-life movements and crimes may intensify online anti-Chinese sentiments. Additionally, the pandemic's progression, represented by coronavirus infection rates, emerges as a key predictor of Sinophobic sentiment, underlining the intersection of public health issues with racial animosity. However, caution is advised in interpreting these patterns over longer periods due to the risk of overfitting, which calls for careful consideration in policy-making and prevention strategies.

6.2 Limitation

This study acknowledges several limitations that should be considered. First, the use of comments from YouTube news videos as an indicator of online Sinophobia may not capture the full landscape of public views, as YouTube may not be the predominant platform for expressing opinions, and the chosen news channels may cater to a niche audience. Additionally, the platform's moderation practices and algorithmic filtering could selectively remove content, thereby distorting the representation of visible comments and limiting the generalizability of findings.

Second, the decision to use comment publish dates rather than video publish dates in time series analysis hinges on the assumption that YouTube users actively seek out content that interests them, engaging in discussions irrespective of the video's original release date. This approach presupposes that users' engagements are driven by specific content searches rather than passive consumption of newly uploaded videos. However, this assumption may underestimate the role of the video itself as the primary catalyst for subsequent discussions. If the initial release of a video significantly prompts viewer reactions and comments, then ignoring video publish dates could lead to a misunderstanding of the dynamics that drive user engagement and discourse on YouTube.

This study also has identified several limitations regarding the hate crime dataset compiled by the FBI. First, underreporting is a common issue in police-recorded hate crime data, as victims may not always recognize or report the offense they experienced. Second, there is a possibility of incorrect categorization of incidents as anti-Asian hate crimes, as the categorization is based on police officers' descriptions. Additionally, using anti-Asian as a criterion may not fully represent anti-Chinese violence, despite potential spillover effects.

6.3 Future Research

Based on the findings of this study, future research can explore several topics. One potential direction for content analysis is to distinguish language used to initiate or reinforce online Sinophobia in comparison to that used to combat it. For instance, the presence of Sinophobic terms in text does not inherently signal endorsement of Sinophobia; such terms might be employed to highlight their negative connotations and address the issue of stigma. Furthermore, investigating the effectiveness of YouTube language regulation and counter-speech campaigns in combating hate speech and fostering social inclusion online would also be valuable. Additionally, analyzing sentiment trends using video publish dates might yield fresh perspectives on how audiences were triggered by news content and articulate their viewpoints instead of actively searching for target news videos to express their opinions. This approach has the potential to reveal not only the dynamics of how sentiment evolves but also potentially uncover mechanisms of video content in shaping people's viewpoints. Lastly, given the issues of under-reporting and mis-classification of the official hate crime report, additional self-report victim survey data might contribute to a more accurate hate incident result.

Data and Code Availability Statement

The datasets analysed during the current study are available in the replication-materials repository hosted here: https://github.com/yuzhouw313/thesis_clean

References

Benesch, S. (2012). Dangerous speech: A proposal to prevent group violence. *Voices That Poison: Dangerous Speech Project*.

Bentley, R. A., & Ormerod, P. (2009). Social versus independent interest in'bird flu'and'swine flu'. *PLoS Currents*, *1*.

Biswas, A., Masud, Z., Mokhberi, M., Kteily-Hawa, R., Goldstein, A., Gillis, J. R., Rayana, S., & Ahmed, S. I. (2022). Covid-stigma: A dataset of anti-asian stigmatizing tweets during covid-19.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bose, E., Hravnak, M., & Sereika, S. M. (2017). Vector autoregressive models and granger causality in time series analysis in nursing research: Dynamic changes among vital signs prior to cardiorespiratory instability events as an example. *Nursing research*, *66*(1), 12–19.

Bragazzi, N. L., Watad, A., Brigo, F., Adawi, M., Amital, H., & Shoenfeld, Y. (2017). Public health awareness of autoimmune diseases after the death of a celebrity. *Clinical rheumatology*, *36*, 1911–1917.

Breazu, P., & Machin, D. (2023). Racism is not just hate speech: Ethnonationalist victimhood in youtube comments about the roma during covid-19. *Language in Society*, *52*(3), 511–531.

Brigo, F., Lochner, P., Tezzon, F., & Nardone, R. (2014). Web search behavior for multiple sclerosis: An infodemiological study. *Multiple sclerosis and related disorders*, *3*(4), 440–443.

Cabral, S. (2021). Covid 'hate crimes' against asian americans on rise. *BBC news*, *21*.

Chan, J., Ghose, A., & Seamans, R. (2016). The internet and racial hate crime. *Mis Quarterly*, *40*(2), 381–404.

Fadel, L. (2021). With racial attacks on the rise, asian americans fear for their safety. *Heard on All Things Considered, National Public Radio. Retrieved October*, *22*, 2021.

Gover, A. R., Harper, S. B., & Langton, L. (2020). Anti-asian hate crime during the covid-19 pandemic: Exploring the reproduction of inequality. *American journal of criminal justice*, *45*(4), 647–667.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Gray, C., & Hansen, K. (2021). Did covid-19 lead to an increase in hate crimes toward chinese people in london? *Journal of Contemporary Criminal Justice*, *37*(4), 569–588.

Hossain, L., Kam, D., Kong, F., Wigand, R., & Bossomaier, T. (2016). Social media in ebola outbreak. *Epidemiology & Infection*, *144*(10), 2136–2143.

Hswen, Y., Xu, X., Hing, A., Hawkins, J. B., Brownstein, J. S., & Gee, G. C. (2021). Association of "# covid19" versus "# chinesevirus" with anti-asian sentiments on twitter: March 9–23, 2020. *American Journal of Public Health*, *111*(5), 956–964.

Huang, J., Krupenkin, M., Rothschild, D., & Lee, J. (2022). From anti-china rhetoric to anti-asian behavior: The social and economic cost of" kung flu".

Kapadia, F. (2022). Violence and the covid-19 pandemic: A public health of consequence, may 2022.

Langlois, S. (2018). How biased is your news source? you probably won't agree with this chart. *Market Watch*.

Leader Maynard, J., & Benesch, S. (2016). Dangerous speech and dangerous ideology: An integrated model for monitoring and prevention. *Genocide Studies and Prevention*, *9*(3).

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Mavragani, A., & Ochoa, G. (2018). The internet and the anti-vaccine movement: Tracking the 2017 eu measles outbreak. *Big Data and Cognitive Computing*, *2*(1), 2.

Pandey, A., Abdullah, K., & Drazner, M. H. (2014). Impact of vice president cheney on public interest in left ventricular assist devices and heart transplantation. *The American Journal of Cardiology*, *113*(9), 1529–1531.

Poletto, C., Boëlle, P.-Y., & Colizza, V. (2016). Risk of mers importation and onward transmission: A systematic review and analysis of cases reported to who. *BMC infectious diseases*, *16*, 1–13.

Reja, M. (2021). Trump's 'chinese virus' tweet helped lead to rise in racist anti-asian twitter content: Study. *ABC News*, *18*.

Röttger, P., Vidgen, B., Nguyen, D., Waseem, Z., Margetts, H., & Pierrehumbert, J. B. (2020). Hatecheck: Functional tests for hate speech detection models. *arXiv preprint arXiv:2012.15606*.

Salcedo, A. (2021). Racist anti-asian hashtags spiked after trump first tweeted 'chinese virus,'study finds. *The Washington Post*, *19*.

Schild, L., Ling, C., Blackburn, J., Stringhini, G., Zhang, Y., & Zannettou, S. (2020). " go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19. *arXiv preprint arXiv:2004.04046*.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics*, *118*, 26–40.

Stocking, G., Van Kessel, P., Barthel, M., Matsa, K. E., & Khuzam, M. (2020). Many americans get news on youtube, where news organizations and independent producers thrive side by side. *Pew Research Center*, *28*.

Tekumalla, R., Baig, Z., Pan, M., Hernandez, L. A. R., Wang, M., & Banda, J. M. (2022). Characterizing anti-asian rhetoric during the covid-19 pandemic: A sentiment analysis case study on twitter. *ICWSM Workshops*.

Than, M. (2021). 9000 anti-asian incidents have been reported since the pandemic began. *URL: https://www. npr. org/2021/08/12/1027236499/anti-asian-hate-crimes-assaults-pandemic-incidents-aapi (accessed 10.09. 2022)*.

Vidgen, B., Thrush, T., Waseem, Z., & Kiela, D. (2020). Learning from the worst: Dynamically generated datasets to improve online hate detection. *arXiv preprint arXiv:2012.15761*.

Wang, P., & Catalano, T. (2023). 'chinese virus': A critical discourse analysis of anti-asian racist discourse during the covid-19 pandemic.

Wang, Y. (2023, April). *The relationship between online Sinophobic sentiment and anti-Chinese hate crimes* (Version 1.0.0). https://github.com/macs30200-s23/replication-materials-yuzhouw313

Wiedlitzka, S., Prati, G., Brown, R., Smith, J., & Walters, M. A. (2023). Hate in word and deed: The temporal association between online and offline islamophobia. *Journal of quantitative criminology*, *39*(1), 75–96.

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, *60*(1), 93–117.

Zhang, D. (2022). Sinophobic epidemics in america: Historical discontinuity in disease-related yellow peril imaginaries of the past and present. In *The covid pandemic: Essays, book reviews, and poems* (pp. 63–80). Springer.

Appendix

A. Visualization of All Time Series

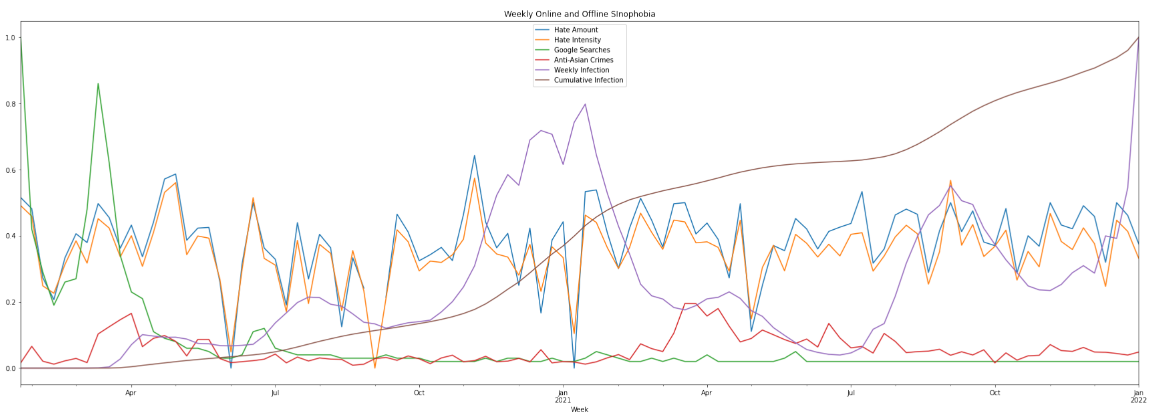

Figure 14: The temporal trend of all Sinophobia indicators with reference to confirmed infection cases and cumulative infection cases

B. OLS Regression Tables

OLS Regression Results

Dep. Variable:	Hate Amount	R-squared:	0.916
Model:	OLS	Adj. R-squared:	0.915
Method:	Least Squares	F-statistic:	1088.
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	1.49e-55
Time:	20:52:27	Log-Likelihood:	200.51
No. Observations:	102	AIC:	-397.0
Df Residuals:	100	BIC:	-391.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0247	0.013	-1.922	0.057	-0.050	0.001
Hate Intensity	1.1570	0.035	32.990	0.000	1.087	1.227

Omnibus:	16.353	Durbin-Watson:	2.112
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27.549
Skew:	-0.676	Prob(JB):	1.04e-06
Kurtosis:	5.157	Cond. No.	11.7

Figure 15: OLS regression table of YouTube comment hate amount and intensity

OLS Regression Results

Dep. Variable:	Hate Amount	R-squared:	0.000
Model:	OLS	Adj. R-squared:	−0.010
Method:	Least Squares	F-statistic:	0.03819
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	0.845
Time:	20:48:47	Log-Likelihood:	74.297
No. Observations:	102	AIC:	−144.6
Df Residuals:	100	BIC:	−139.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.3813	0.018	21.191	0.000	0.346	0.417
Weekly Infection	0.0109	0.056	0.195	0.845	−0.100	0.121

Omnibus:	25.655	Durbin-Watson:	1.618
Prob(Omnibus):	0.000	Jarque-Bera (JB):	39.227
Skew:	−1.139	Prob(JB):	3.03e−09
Kurtosis:	5.011	Cond. No.	5.07

OLS Regression Results

Dep. Variable:	Hate Amount	R-squared:	0.042
Model:	OLS	Adj. R-squared:	0.022
Method:	Least Squares	F-statistic:	2.151
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	0.122
Time:	20:48:49	Log-Likelihood:	76.448
No. Observations:	102	AIC:	−146.9
Df Residuals:	99	BIC:	−139.0
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.3623	0.020	18.154	0.000	0.323	0.402
Weekly Infection	−0.0433	0.061	−0.712	0.478	−0.164	0.077
Cumulative Infection	0.0826	0.040	2.065	0.042	0.003	0.162

Omnibus:	19.956	Durbin-Watson:	1.682
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28.379
Skew:	−0.924	Prob(JB):	6.88e−07
Kurtosis:	4.806	Cond. No.	6.24

Figure 16: OLS regression table of infection cases as independent variable and YouTube hate amount as dependent variable (left), moderated by the cumulative infection cases (right)

OLS Regression Results

Dep. Variable:	Google Searches	R-squared:	0.138
Model:	OLS	Adj. R-squared:	0.129
Method:	Least Squares	F-statistic:	15.98
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	0.000123
Time:	20:52:48	Log-Likelihood:	51.838
No. Observations:	102	AIC:	−99.68
Df Residuals:	100	BIC:	−94.43
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1462	0.022	6.521	0.000	0.102	0.191
Weekly Infection	−0.2777	0.069	−3.998	0.000	−0.416	−0.140

Omnibus:	108.305	Durbin-Watson:	0.351
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1301.691
Skew:	3.648	Prob(JB):	2.20e−283
Kurtosis:	18.907	Cond. No.	5.07

OLS Regression Results

Dep. Variable:	Google Searches	R-squared:	0.227
Model:	OLS	Adj. R-squared:	0.211
Method:	Least Squares	F-statistic:	14.53
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	2.92e−06
Time:	20:53:12	Log-Likelihood:	57.405
No. Observations:	102	AIC:	−108.8
Df Residuals:	99	BIC:	−100.9
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1837	0.024	7.639	0.000	0.136	0.231
Weekly Infection	−0.1708	0.073	−2.331	0.022	−0.316	−0.025
Cumulative Infection	−0.1629	0.048	−3.379	0.001	−0.259	−0.067

Omnibus:	107.269	Durbin-Watson:	0.382
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1330.035
Skew:	3.572	Prob(JB):	1.54e−289
Kurtosis:	19.184	Cond. No.	6.24

Figure 17: OLS regression table of infection cases as independent variable and Sinophobic Google searches as dependent variable (left), moderated by the cumulative infection cases (right)

OLS Regression Results

Dep. Variable:	Anti-Asian Crimes	R-squared:	0.078
Model:	OLS	Adj. R-squared:	0.069
Method:	Least Squares	F-statistic:	8.444
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	0.00451
Time:	20:56:12	Log-Likelihood:	183.42
No. Observations:	102	AIC:	-362.8
Df Residuals:	100	BIC:	-357.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0692	0.006	11.217	0.000	0.057	0.081
Weekly Infection	-0.0556	0.019	-2.906	0.005	-0.094	-0.018

Omnibus:	30.004	Durbin-Watson:	0.503
Prob(Omnibus):	0.000	Jarque-Bera (JB):	47.456
Skew:	1.322	Prob(JB):	4.95e-11
Kurtosis:	5.043	Cond. No.	5.07

OLS Regression Results

Dep. Variable:	Anti-Asian Crimes	R-squared:	0.191
Model:	OLS	Adj. R-squared:	0.175
Method:	Least Squares	F-statistic:	11.72
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	2.70e-05
Time:	20:56:27	Log-Likelihood:	190.12
No. Observations:	102	AIC:	-374.2
Df Residuals:	99	BIC:	-366.4
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0580	0.007	8.855	0.000	0.045	0.071
Weekly Infection	-0.0877	0.020	-4.395	0.000	-0.127	-0.048
Cumulative Infection	0.0489	0.013	3.729	0.000	0.023	0.075

Omnibus:	36.126	Durbin-Watson:	0.580
Prob(Omnibus):	0.000	Jarque-Bera (JB):	63.355
Skew:	1.531	Prob(JB):	1.75e-14
Kurtosis:	5.352	Cond. No.	6.24

Figure 18: OLS regression table of infection cases as independent variable and anti-Asian hate crime rate as dependent variable (left), moderated by the cumulative infection cases (right)

OLS Regression Results

Dep. Variable:	Anti-Asian Crimes	R-squared:	0.215
Model:	OLS	Adj. R-squared:	0.191
Method:	Least Squares	F-statistic:	8.949
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	2.69e-05
Time:	21:00:28	Log-Likelihood:	191.64
No. Observations:	102	AIC:	-375.3
Df Residuals:	98	BIC:	-364.8
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0377	0.013	2.792	0.006	0.011	0.064
Hate Amount	0.0561	0.033	1.717	0.089	-0.009	0.121
Weekly Infection	-0.0853	0.020	-4.304	0.000	-0.125	-0.046
Cumulative Infection	0.0443	0.013	3.338	0.001	0.018	0.071

Omnibus:	36.208	Durbin-Watson:	0.613
Prob(Omnibus):	0.000	Jarque-Bera (JB):	63.047
Skew:	1.543	Prob(JB):	2.04e-14
Kurtosis:	5.304	Cond. No.	11.0

OLS Regression Results

Dep. Variable:	Anti-Asian Crimes	R-squared:	0.197
Model:	OLS	Adj. R-squared:	0.173
Method:	Least Squares	F-statistic:	8.022
Date:	Mon, 22 Apr 2024	Prob (F-statistic):	7.80e-05
Time:	21:00:48	Log-Likelihood:	190.49
No. Observations:	102	AIC:	-373.0
Df Residuals:	98	BIC:	-362.5
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0538	0.008	6.504	0.000	0.037	0.070
Google Searches	0.0229	0.027	0.836	0.405	-0.031	0.077
Weekly Infection	-0.0838	0.021	-4.082	0.000	-0.124	-0.043
Cumulative Infection	0.0527	0.014	3.794	0.000	0.025	0.080

Omnibus:	35.818	Durbin-Watson:	0.588
Prob(Omnibus):	0.000	Jarque-Bera (JB):	63.758
Skew:	1.500	Prob(JB):	1.43e-14
Kurtosis:	5.450	Cond. No.	8.55

Figure 19: OLS regression table of anti-Asian hate crime rate as dependent variable and YouTube comment hate amount as independent variable moderated by infection cases (left); OLS regression table of anti-Asian hate crime rate as dependent variable and Sinophobic Google searches as independent variable moderated by infection cases (right)