Journal of Public Diplomacy Vol. 5, No. 2: 56-80.
DOI: 10.23045/jpd.2025.5.2.003

Research Article

The International Expansion of China's Online Propaganda: A Case Study of the Uyghurs and Xinjiang

Neeley Pate[1] · Inga K. Trauthig[2] · Jo Lukito[3] · Samuel Woolley[4]

[1]University of Rochester, npate@ur.rochester.edu

[2]Florida International University, itrauthi@fiu.edu

[3]The University of Texas at Austin, jlukito@utexas.edu

[4]University of Pittsburgh, scw97@pitt.edu

Received: March 27, 2025 | Accepted: August 15, 2025

Abstract

This study expands understandings of pro-Chinese Communist Party (CCP) international propaganda efforts focusing on the Uyghurs in China's Xinjiang region. Comparative topic modeling and time series analysis of relevant tweets and *China Daily* (operated by the Chinese Communist Party's Central Propaganda Department) headlines yield three findings: 1) Ethnically-charged tweets target a range of foreign audiences, 2) Relevant tweets contain high amounts of negatively-coded language in contrast to both the *China Daily* messaging and past studies of pro-CCP online propaganda campaigns, 3) Creators of the tweets leverage particular advantages of social media in order to spread this aggressive propaganda. With these findings, this paper contributes cross-medium understandings of how pro-CCP messaging has shifted when using social media platforms versus traditional media outlets for the purposes of internationally focused propaganda.

Keywords: China; Xinjiang; Chinese Communist Party (CCP); Propaganda; Computational Propaganda; Ethnic Propaganda

*Corresponding author: npate@ur.rochester.edu

Introduction

Over the last decade, China has become the United States' main geopolitical competitor (Hass, 2021). Cyberspace, including social media platforms used for computational propaganda and public diplomacy, are crucial domains in this competition (Bolsover, 2018; Monaco, Smith, & Studdart, 2020). Computational propaganda employs, "algorithms, automation, and human curation to purposefully manage and distribute misleading information over social media networks" (Woolley & Howard, 2018). In the case of China, previous studies have shown that Chinese computational propaganda mostly consists of effusively supportive messages about the Chinese Communist Party (CCP), primarily targeting the domestic population, but also citizens of bordering states (i.e., Taiwan) with positive messaging (Brady, 2012; King, Pan, & Roberts, 2017; Rawnsley, 2005; Stockmann & Gallagher, 2011). These strategies are done in tandem with more traditional forms of public diplomacy, such as the production of *China Daily*, which creates a morally ambiguous international communication strategy for China (Nip & Sun, 2022).

Against this background, our study focuses on how ethnically focused, international propaganda regarding the Uyghurs is tied to efforts to amplify "Chinese Communist Party narratives related to the treatment of the Uyghur population in Xinjiang" (Twitter, 2021, para. 6). This study expands our understanding by showing that the substance and focus of pro-CCP online manipulation campaigns have evolved: our data reveals that pro-CCP messaging now targets a range of foreign audiences and contains more negativity and aggression. The study hones in on the importance of Twitter[1)] for international public diplomacy, a trend that has been noted previously by researchers such as Jia and Li (2020) and Guo (2021), but expands upon this prior work by exploring how public diplomacy efforts are used to shape foreign publics' opinions about domestic affairs. Further, this study extends burgeoning research on the Uyghur genocide and related propaganda efforts specifically (Gilsinan, 2020; Yusupov, 2022), as well as research on propaganda in non-U.S. contexts (Lukito, 2024; Abhishek, 2021).

Literature Review

Uyghurs and the Communist Chinese State

Tensions between the Uyghur population and the Chinese Communist government are longstanding. In 1945 (before the CCP's rise to power), the Uyghurs attempted to secede from China unsuccessfully. In 1955, the Xinjiang Uyghur Autonomous Region (XUAR) was

1) In October 2022, Elon Musk completed his purchase and acquisition of Twitter, subsequently rebranding it as "X." As our analysis focuses on a case prior to this acquisition, we continue to use the name "Twitter" for the platform.

established, reintegrating the region into Chinese law and marking the start of the CCP's efforts to force the assimilation of non-Han-Chinese people into a "unitary" Chinese nation-state (Finnegan, 2020). Increased tension can be seen in the region's recent history, leading to "an outbreak of Uyghur demonstrations and interethnic unrest in 2009, and sporadic clashes involving Uyghurs and Xinjiang security personnel that spiked between 2013 and 2015" that ultimately resulted in "large scale criminal arrests and intensive security measures in the XUAR, aimed at combating 'terrorism, separatism, and religious extremism'" (Lum & Weber, 2023, para. 5). Since asserting dominance in the region, the CCP has targeted the Uyghur community through various assimilation policies and human rights atrocities (Lim, 2021). For example, since 2017, the XUAR: "placed restrictions upon dress and grooming, traditional Uyghur customs, and adherence to Islamic dietary laws (halal), closed or demolished thousands of mosques, and banned the use of 'Uyghur language' in the region" (Lum & Weber, 2023, para. 6). Furthermore, Chinese authorities have detained over one million Uyghurs in so-called reeducation centers, where captives are compelled to renounce many of their Islamic beliefs, and undergo food deprivation as well as forced labor or torture (Lum & Weber, 2023). In addition, the Chinese government's genocidal actions have gained international attention, prompting the United Nations Office of the High Commissioner for Human Rights to publicly call on China in 2022 to address these human rights violations (Aoláin et al., 2022).

Because of the need to instill a sense of national cohesion, the Uyghurs in Xinjiang present a two-fold challenge to the government and its external presentation. First, the Uyghurs' Turkic ethnicity and religion (Islam) separates them from the majority Chinese population (Han Chinese), who are officially atheist, but often follow religious practices close to Buddhism (Maizland & Albert, 2020). These differences have been interpreted by Chinese authorities as a threat to the Chinese political system, which rests on a vision of communal identity (Campbell, 2021). Second, acts of repression by the Chinese authorities have attracted international, especially Western, attention and led to public denouncements of the Chinese state (Aoláin et al., 2022; Lum & Weber, 2023). This harms China's carefully cultivated soft power efforts internationally (Albert, 2018).

Importantly, treatment of the Uyghur people rapidly deteriorated (as anti-Uyghur sentiment became more radicalized) when they were declared a "terrorist threat" by the Chinese state in the early 2000s (Arablouei & Abdelfatah, 2021). This securitization of Muslims was in line with the Zeitgeist of the "Global War on Terror" (Tazamal, 2018). In 2023, this attitude changed, and the suppression of Muslim populations is no longer justified with terror rationales (Gilbert, 2022). Instead, the U.S. has politicized China's mistreat of the Uyghurs and is cited as the main reason for U.S. diplomats boycotting the Winter Olympics in 2022 (Brant, 2021a; Gilbert, 2022). This ethnic and religious minority in China continues to experience human rights abuses at the hands of the Chinese government and, as such, are the focus of serious international concern (Maizland, 2022; Shichor, 2005).

CCP Propaganda and Soft Power

Propaganda and public diplomacy are considered crucial to Chinese government's soft power—"the ability to affect others to obtain the outcomes one wants through attraction rather than coercion or payment" (Nye, 2008, p. 94)—as soft power is not only essential to the regime survival, but also to China's international success (Chang & Lin, 2014; Mulvenon & Chase, 2008). Protecting China's national interests and its soft power means leveraging well-versed communication capabilities, which are supposed to benefit Chinese messaging internationally but can also strengthen national cohesion domestically (Albert, 2018; Ding, 2009). Previous research reveals coordination of the online activities of the Chinese Ministry of Foreign Affairs, finding that the ministry "uses traditional propaganda-based methods for public diplomacy practices and ignores online interaction with foreign publics" (Huang & Wang, 2021). This has been historically true for government-led communication about ethnicities, which has been strictly managed to achieve the CCP's goal of promoting the image of ethnic harmony in China, with the intention of it functioning as an example for the world (Albert, 2018; Brady, 2012; Hagström & Nordin, 2019).

To protect this soft power, the CCP relies on public diplomacy and propaganda efforts. Often, scholars differentiate between Chinese public diplomacy and propaganda by defining public diplomacy as a spin-off of propaganda, focused on official state figures and outlets, though there is acknowledged overlapping historical roots and features (Gilboa, 2008; Melissen, 2005). For example, recent studies on Chinese communication around the COVID-19 pandemic note the success of Chinese public diplomacy by randomly exposing Indian citizens to real Twitter messages from Chinese diplomats and arguing that "positive public diplomacy" that emphasizes aid and friendship works, while "negative public diplomacy" that criticizes international rivals is ineffective (Mattingly & Sundquist, 2022). Since this study does not focus on specific state officials, but rather a detected information operation combined with a state affiliated news outlet, we rely on the more inclusive terminology of international propaganda, including both traditional propaganda (via state affiliated news outlets) as well as computational propaganda (on social media) (Woolley & Howard, 2018).

In the early 2000s, Chinese propaganda efforts focused on traditional media, with China's soft power being "based on 'strong propaganda methods and strong propaganda capabilities'" (Bolsover & Howard, 2018, p. 4). Only in recent years has China moved to online media to spread propaganda. "Internet Power" became a prominent concept and Chinese online propaganda increased beginning around 2016. This not only highlights how international propaganda relates to soft power, but also, “suggest[s] the Chinese government is paying more attention to foreign social media” in recent years (Bolsover & Howard, 2018, p. 5). Since President Xi Jinping came to office in 2013, Chinese public diplomacy has become more digitally oriented (Guo, 2021) and assertive (Brady, 2015).

Social Media Propaganda and Public Diplomacy

Existing studies examining social media as a tool to spread propaganda cover communication studies but also interdisciplinary scholarship in area studies. They mainly address China’s relationship with national media and social media and related impacts on its public diplomacy, as well as different strains of propaganda – such as ethnic propaganda. Our study advances existing understandings about Chinese propaganda by outlining how Chinese computational and traditional propaganda is evolving online, specifically with regard to the Uyghur minority.

Chaudhari and Pawar (2021) present a holistic analysis of propaganda on social media by studying how other researchers have identified propaganda. In the paper, the authors list types of propaganda identification on social media, including biased news, bots, and political news. They find that the study of propaganda on social media is a thriving field, and emphasize the importance of propaganda detection. Highlighting the history of China’s relationship with social media, deLisle, Goldstein, and Yang (2016) emphasize how social media can be used to shape public opinion (such as with the “fifty-cent party”) or to target those who may form organized opposition against party policies. The study also notes the role of the internet in keeping China from “maintaining a monopoly on organized politics, limiting dissent, and censoring some ideas while privileging others” (deLisle, Goldstein, & Yang, 2016). The “fifty-cent party” is a key example of how China engages with computational propaganda. A study by King, Pan, and Roberts (2017) reported an estimate of 448 million posts per year, and found that the content matter was overwhelmingly positive:

> The 50c party engages in almost no argument of any kind and is instead devoted primarily to cheerleading for the state, symbols of the regime, or the revolutionary history of the Communist Party (p. 497).

Their findings highlighted that most members of the “fifty-cent party” were “composed of government employees contributing part time outside their regular jobs, not, as has been claimed, ordinary citizens” because the individuals were identifiable (p. 497). What is unique about this study is the lack of ability to identify the individuals running the removed accounts.

Focusing on ethnic issues, Brady (2012) emphasizes that any communication related to ethnicity in China must be strictly managed—promoting an image of ethnic harmony internally and externally. One major challenge to this hegemonic narrative are the Uyghurs in western China. While early research assessed that, "the existing internal and external challenge of Uyghur separatism and Islamic radicalism to Chinese rule in Xinjiang is, at best, marginal and, at worst, manageable," (Shichor, 2005, pp. 132-133) the CCP insists on painting the Uyghurs as a terrorist threat (Purbrick, 2017). More recently, however, the Uyghurs have entered the "centre of a propaganda war between the US and China" such that the U.S. may potentially exaggerate the maltreatment of the Uyghurs (framing the issue as a human rights crisis that justifies restrictive economic retaliation), whereas Chinese officials emphasize the alleged imminent threat at hand while simultaneously downplaying any presumed maltreatment of the Uyghurs (Gilbert, 2022, p. 60).

News reporting on propaganda efforts about Uyghurs in Xinjiang on platforms such as TikTok and Facebook have been picked up by *The Wall Street Journal, Rappler*, and *The China Project*. These reports all highlight the overwhelmingly positive nature of the ads about Xinjiang on Facebook and the censorship of Uyghur oppression on TikTok (Isobel, 2020; Niewenhuis, 2021; Purnell, 2021). Existing studies have started to include the Uyghurs in the Chinese state's specific tactics and practices regarding public diplomacy and propaganda. Yusupov (2022) specifically focuses on how the Chinese state aims to "influence and shape the international court of opinion about its measures in Xinjiang", relying on the anthropological framework of human rights employed regarding the Uyghurs as well as other Muslim minorities in the region (p. 1). He draws attention to important processes of Chinese public diplomacy and argues that the Uyghur crisis is met with ambiguity and uncertainty by the international public. Our research enhances understandings of pro-CCP messaging regarding the Uyghurs by adding a comparative analysis when bringing together traditional and computational media as well as by being more holistic, as we did not limit ourselves to one framework or angle (such as human rights).

Methods

We focus our study on content from Twitter — the most important social media platform for China's international propaganda efforts (Potkin, Baptista, & Munroe 2022; Wang & Xu 2022) – and *China Daily*, an English language daily newspaper owned and operated by CCP and arguably the country's most important English-language newspaper (*China Daily*, n.d.; Hartig, 2020). It is imperative to recognize both of these tools as a means to spread ethnic propaganda regarding the Uyghurs, and how the difference in presentation shapes the strategies used with each one. Against this backdrop, we ask:

RQ1: What strategies are utilized to spread computational propaganda regarding messaging about the Uyghurs?

RQ2: How does computational propaganda on social media differ from traditional propaganda with regards to the Uyghurs?

Data collection

Twitter.

In December 2021, Twitter released a dataset of tweets with posts between March 2019 and March 2021 from 2,048 accounts that were removed because of the amplification of "Chinese Communist Party narratives related to the treatment of the Uyghur population in Xinjiang" (Twitter, 2021, para. 6). Twitter, in a prior dataset curation effort, described these accounts as those occurring as part of a state-affiliated "information operation", which produces "coordinated malicious activity" (Roth, 2019, paras. 1 and 23). This dataset specifically focuses on accounts that appear to be affiliated with the Chinese Communist Party, as determined by Twitter. Furthermore, as the content has since been deleted by Twitter, it would not be possible to retrieve this data were it not for Twitter providing it publicly.[2]

In the beginning, analysis on several languages was considered; ultimately, only English tweets were used for topic modeling and time series analysis. For example, preliminary analysis of the dataset found that tweets were written in several languages, with most of the tweets being written in either Chinese, English, or Japanese. Because of the mass amounts of data provided (n = 31,270 for the Twitter data across all languages), substring searches were used to pull data relevant to the scope of the study. First, the team decided to focus on Chinese and English tweets, filtered by "tweet_language" ("en" for English, resulting in n = 17,582 tweets, and "zh" for Chinese, resulting in n = 6,757 tweets) and removing all duplicates from the set. It is important to note that some tweets do contain characters outside of their detected language, and that those characters were removed during the analysis process. For both the English and the Chinese tweets, the substrings of "Xinjiang", "China", "Uyghur", "Uygur" (as sometimes Uyghur was spelled Uygur), and their lowercase counterparts were used to identify tweets of relevance to the study (for the Chinese tweets, the substrings were translated into "新疆", "中国", "维吾尔", and "维吾尔族". It is important to note that "Uyghur" would capture "Uyghurs", since "Uyghur" is a proper substring of "Uyghurs" in both languages. From there, URLs and the keywords used in the search were removed. If

2) The Twitter Transparency effort, which housed the Information Operations datasets, began in 2018 and ended in 2022, following Elon Musk's acquisition of Twitter and subsequent transformation into X. It remains one of the few examples of a social media company providing datasets of deleted state-sponsored propaganda and was a critical resource for international communication scholars. A version of the website can be found on the Internet Archive: https://web.archive.org/web/20220404202253/https://transparency.twitter.com/en/reports/information-operations.html

these were not removed, every topic would have the same top words (because we filtered all data to have one of those words). This resulted in n = 4,344 English tweets. However, this selection of relevant tweets only revealed n = 847 Chinese tweets. The team chose to focus solely on the relevant English tweets for further analysis. While the relevant English tweets represent a rather small subset of the original dataset, the number of tweets utilized is still large enough to offer interpretable results.

China Daily

To understand the relation between the tweets and the news articles written, the team utilized a dataset from LexisNexis containing news article headlines from *China Daily*. *China Daily* is accredited with potential global reach and constitutes one of "the 'Big Four' government-owned Chinese media organizations that are most actively involved in China's global media campaign" (Hartig, 2020, p. 9). This outlet is also described as the "mouthpiece for the Party in its efforts to communicate with the wider world" (Chen, 2004, p. 700). To align this dataset with the one gathered from Twitter, the data were limited to article headlines published from March 2019 to March 2021. The dataset was also restricted to headlines for similar reasons; because the Twitter data had a cap on the number of characters when pulling data, the length of the China Daily headlines is more comparable to the length of the tweets. To match the data pre-processing methods used for the tweets, a substring search with the same words as the English tweets was conducted. One result from preliminary data processing was that this search needed to be slightly more refined. Because the period of analysis was March 2019 - March 2021, searching for "China" pulled in headlines about COVID-19 that were unrelated to the Uyghurs in Xinjiang (and caused the model to have worse performance, based on the number of headlines possible to classify and assign coherence scores). For headlines that specifically contained "COVID", "covid", "Covid", or "virus", another check was made to ensure that any of the keywords except "China" (or "china") was in the headline. This resulted in n = 414 headlines.

Analysis methods

Topic model classification.

Topic modeling was conducted to understand the trends of the datasets. Topic modeling is the process of taking "documents" (in this case, tweets or headlines) and creating topic clusters based on how frequently words appear together. To parse all the data, a list of stop words in the appropriate languages were utilized to remove common words from appearing in the topic model results. Then, a bag-of-words is created from the frequency of words in each document. Using the bag-of-words, BERT topic models were generated for both datasets. The BERT topic model plots the generated clusters on a graph meant to show the distance between perceived clusters. The appropriate number of topics for each model was determined by two factors: examining the number of documents in each topic and the coherence score. Following prior studies (Bender et al., 2023; Fedorov & Datyev, 2022), we use the u_mass coherence score metric. Details of the model setup can be found within the appendix.

The English tweet topic model had 4,344 data points, but the unrestricted model generated over 100 topics, with several overlapping each other and a majority of them having fewer than 100 elements. This model was then reduced to 10 topics, and then 6, where most topics had at least 100 elements (with one topic being those documents unclassified). Similarly, the *China Daily* model with the set parameters generated 4 topics (with one topic being those documents unclassified).

Figure 1 and Figure 2 were created using the built-in figure generation capabilities of BERT. In Figure 1, five unique topic clusters are plotted. The size of each cluster is representative of the number of documents classified into that topic. The distance between the topics is representative of how different each topic is from the other ones (in two dimensions). Figure 2 provides a scatter plot of every document, color coded by the assigned topic of each document. This is still representative of the unique regions of each topic, as well as the determined distance of documents from each other. Both plots show the distinct regions between each of the determined topics.

Table 1. English Tweets per Topic.

Topic Number	Number of Tweets
0	2938
1	1005
2	185
3	181
4	14

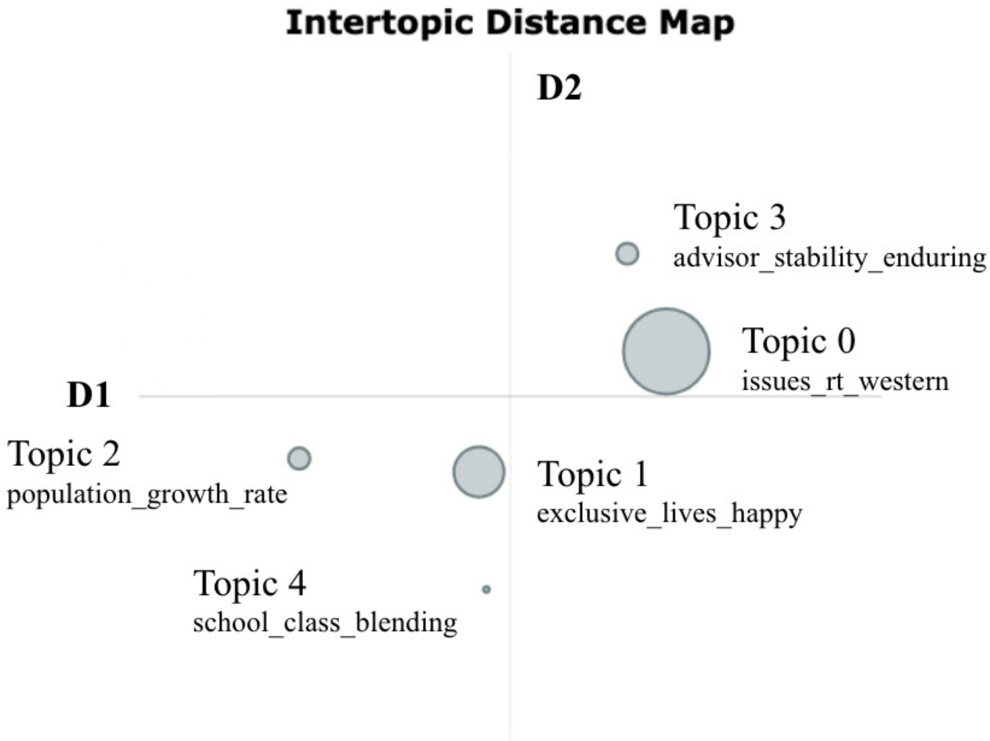

Figure 1. English Tweet Topic Model.

Figure 1 shows the results of the topic modeling for the English tweets, and Table 1 shows the number of tweets per cluster. 111 tweets were not classified. From these topics, we can see that topic 0 and topic 1 had significantly more tweets than the other topics, accounting for around 91% of tweets. The top words for topic 0 consisted of the following: issues, rt, western, rumors, region, terrorism, official, rights, officials, us. The top words for topic 1 consisted of the following: exclusive, lives, happy, peaceful, tianshan, interviews, net, comrades, place, wonderful. Topic 2 included words such as population, growth, rate, and increased. Topic 3 top words included advisor, stability, enduring, and political. Topic 4 included words such as school, class, blending, and exercise. The coherence score for this model was -9.24. Further analysis of these findings is provided in the Results section.

Table 2. *China Daily* Headlines per Topic.

Topic Number	Number of Headlines
0	13
1	209
2	164

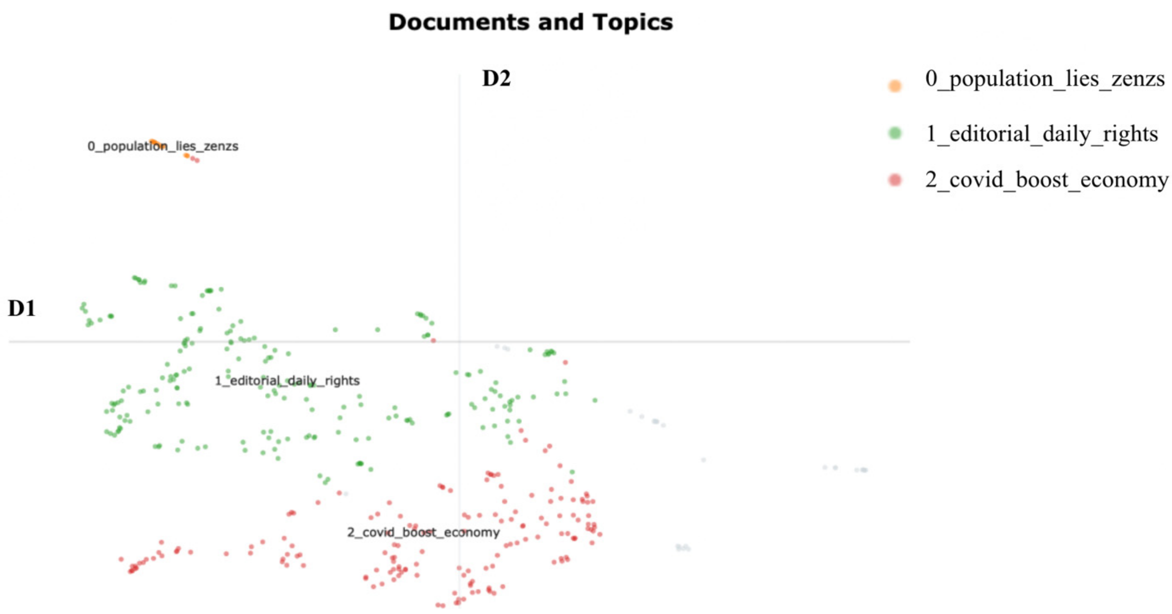

Figure 2. China Daily Topic Model.

Figure 2 shows the results of the topic modeling for the *China Daily* headlines. Because of the limited number of headlines, the BERTopic model was not able to generate a plot like the one for the English tweets, but it was able to plot the documents spatially and colored by each topic. This gives a sense of the proximity of topics to each other, like Figure 1. Table 2 shows the breakdown of the number of headlines per topic. 28 headlines were not classified. From these results, we can see that topics 1 and 2 were the most encompassing, covering over 90% of the headlines. The top words for topic 0 consisted of the following: population, lies, zenzs, adrian, debunked, change, socalled, fabrication, full, report. The top words for topic 1 consisted of the following: editorial, daily, rights, report, human, labor, forced, west, claims, muslims. The top words for topic 2 consisted of the following: covid, boost, economy, southern, stability, villagers, cases, outbreak, prosperity, virus. The coherence score for this model was -16.18. Further analysis of these findings is provided in the Results section.

Time series analysis.

After identifying all relevant tweets and headlines, time series plots were generated for the English tweets and the *China Daily* headlines to indicate temporal spikes in activity. For our results, we utilize the monthly aggregate to detect periods of major activity, and then utilize daily counts to explain such activity within the context of current events at that time. Figure 3a shows the frequency of relevant tweets per month, and Figure 3b shows the frequency of relevant tweets per day. Notably, there is a huge increase in activity beginning in December of 2020, leading to two spikes in January 2021 and March 2021.

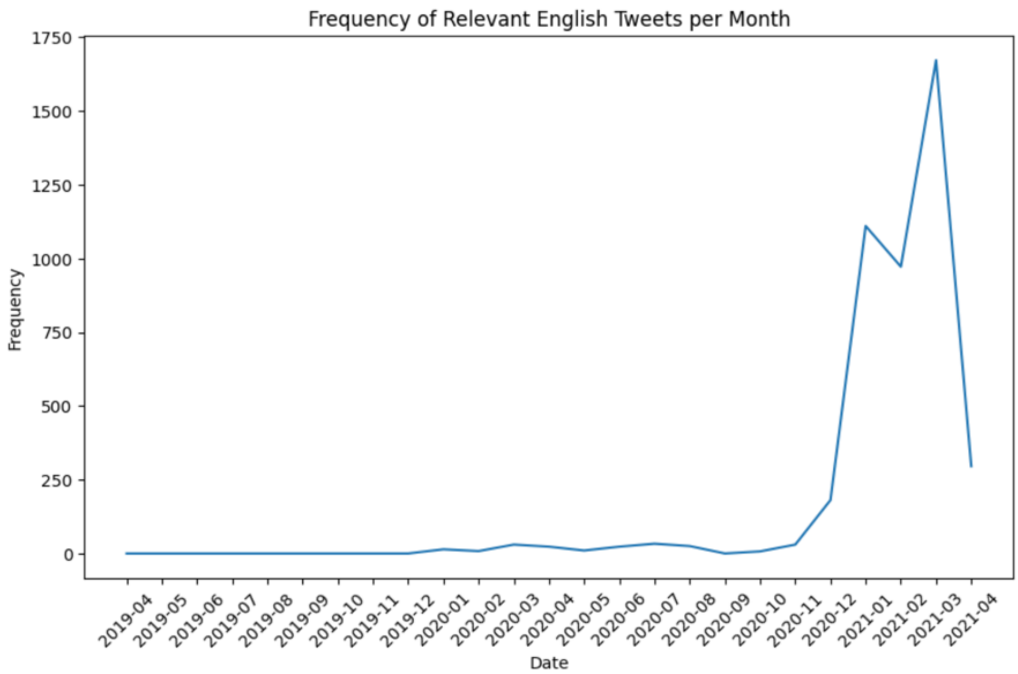

Figure 3a. Monthly Count of English Tweets.

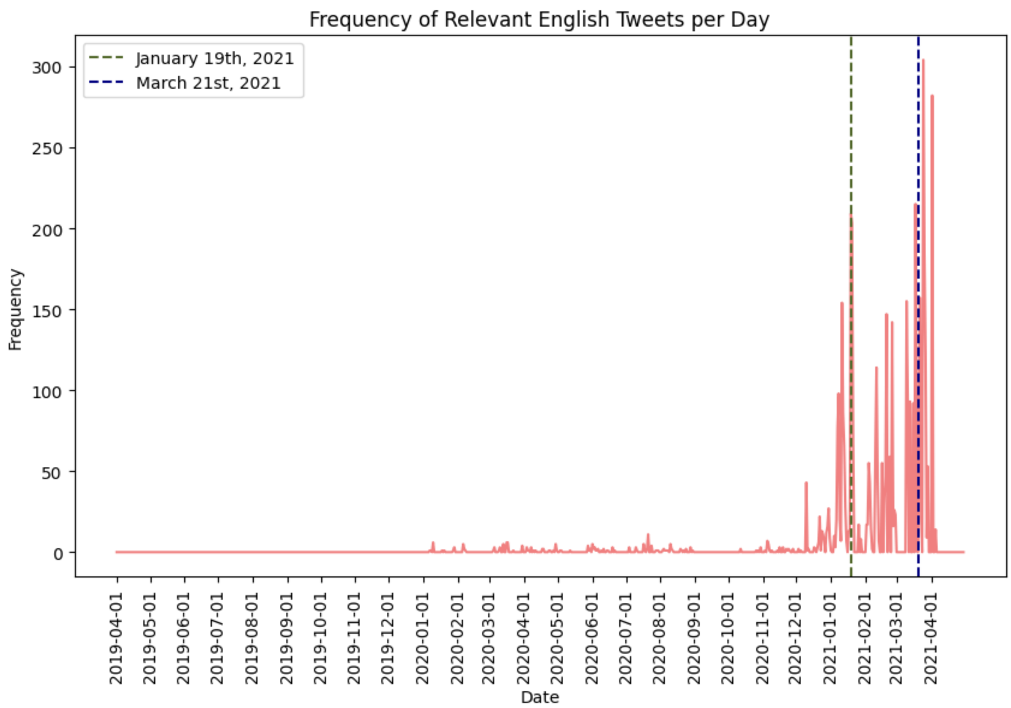

Figure 3b. Daily Count of English Tweets

Figure 4a shows the frequency of *China Daily* headlines per month, and Figure 4b shows the frequency of headlines per day. This graph reveals relative spikes around July 2019 and December 2019. One trend that is notable when comparing the English tweets to the *China Daily* headlines is the significant decrease in content between relative spikes. At maximum, there are about 50 headlines, while there are thousands of tweets at any given spike. Second, the spikes occur at very different times, with the *China Daily* spikes occurring in 2019, the English tweet spikes occurring in 2021. Further analysis was conducted in the Results section to understand global opinion and policies at the time in relation to the Uyghurs in Xinjiang.

Figure 4a. Monthly Count of *China Daily* Headlines.

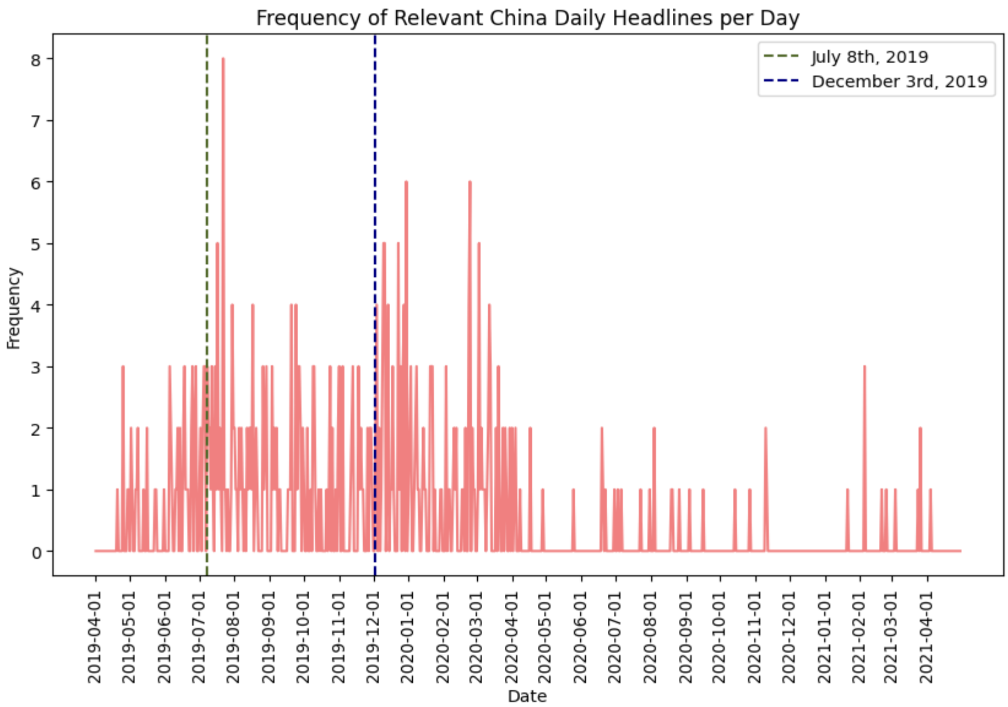

Figure 4b. Daily Count of *China Daily* Headlines

Results

Topic model classification

The resulting topics from the topic models can be explained in two broad themes – spreading positively charged messaging about Xinjiang and China and spreading negatively charged messaging, often by depicting Western nations negatively or by implying that Western narratives about the topic are false. From the topic modeling, the team interprets the topics in the following way:

- English Tweet Model: Topic 0 is negative messaging; topic 1, topic 2, topic 3, and topic 4 are positive messaging.
- *China Daily* Model: Topic 0 and topic 1 are negative messaging; topic 2 is positive messaging.

Of the English tweets classified, 68% fell within topics 0, and 32% fell within topics 1, 2, and 3. Of the *China Daily* headlines classified, 54% of headlines fell within topic 0 and topic 1, and 40% of headlines fell within topic 2.

After categorizing each tweet and headline, the documents were scanned based on the top keywords generated for each topic to identify more nuanced trends within each of the categories.

The English tweets (looking at the first 4 with the most samples) reflected the following trends:

- Topic 0 consisted heavily of tweets focused on refuting “Western lies”.
- Topic 1 focused on trying to draw tourists to the region. Several tweets mention the region being a “wonderful place” and wanting to visit the area. It also emphasizes that those who live in the area have happy lives.
- Topic 2 emphasized the growth of the population in Xinjiang. Several of these tweets were reposting tweets from news sites, such as Global Times News or *China Daily*, which had released articles discussing the population growth.
- Topic 3 had 150 tweets (of 181) repeating the same message: “top political advisor stresses enduring stability in Xinjiang”.

One notable trend from the Twitter data is the number of tweets that are nearly identical —pushing the same message in the same wording (save for a different URL or random characters at the end of the tweet). While the topics fall into general categories such as “positive” and “negative”, it is worth noting the unique ways to spread certain messaging, such as using social media as a platform to push “traditional propaganda” for topic 2 or messaging from political advisors for topic 3. Another notable trend is the emphasis on appealing to tourists to come to Xinjiang. While the positive messaging for this category showcases the region as a great place to live, the additional finding of discussing tourism would further suggest a slight emphasis on international or non-local audiences as the target for this messaging as well.

The *China Daily* headlines (looking at each topic) reflected the following trends:

- Topic 0 focused on discussing the Xinjiang population, emphasizing that it is growing rather than declining. An interesting note for this topic is that it contained a series of 6 different *China Daily* stories specifically addressing the work of Adrian Zenz, an academic researcher studying the Uyghurs’ treatment in Xinjiang. The topic had several headlines that were identical but published at different dates, suggesting that the content of each article was similar to the previously posted version.

- Topic 1 focused on denouncing lies about the region of Xinjiang, with several of them written in the “*China Daily* editorial”. These drew heavily on painting “the west” as a spreader of these lies.

- Topic 2 was surprisingly positive. Even though it included words such as “covid” and “virus”, most headlines were referring to the disease in the region declining. Non-COVID headlines depicted the Xinjiang region as a place with a great economy and stability.

This generalization of the topic material allows us to notice a few trends. First, the positive propaganda is focused heavily on showing Xinjiang as a great place to live (as a place of stability, creativity, an increasing population, and a great economy). While most propaganda that focused on denouncing lies emphasized the role of the West in spreading rumors, topic 0 emphasizes how some of this content was republished on different dates.

These findings suggest a few things. First, propaganda on social media has a higher frequency of negativity than propaganda in traditional media. Second, social media can be used to amplify messaging from traditional propaganda. Third, while the practice of republishing the same content appears as a trend in both traditional propaganda and computational propaganda, these findings demonstrate how different the scales are (6 headlines vs. 150 tweets). Fourth, unique nuances between “positive” propaganda efforts exist between social media and traditional means based on intended audience, while “negative” propaganda efforts utilize similar tactics regardless of audience or medium.

Time series analysis

Figure 3a shows the frequency of relevant tweets per month. From this figure, we see that January and March of 2021 resulted in the two largest spikes in the number of tweets posted. Looking at the daily spikes in Figure 3b, we can contextualize the increase in posts in response to current events. For example, on January 19th, 2021, Secretary of State Mike Pompeo issued a determination about the treatment of the Uyghurs, stating:

> under the direction and control of the Chinese Communist Party (CCP), [China] has committed crimes against humanity against the predominantly Muslim Uyghurs and other members of ethnic and religious minority groups in Xinjian g… [and] has committed genocide against the predominantly Muslim Uyghurs and other ethnic and religious minority groups in Xinjiang (para. 6 and 7).

Similarly, in March of 2021, the European Union “approved sanctions against four Chinese officials involved in running internment camps for hundreds of thousands of Uyghurs in the region of Xinjiang” (Lau & Barigazzi, 2021, para. 1), leading to statements regarding concerns over worker treatment in Xinjiang by companies such as H&M and Nike to resurface, and calls for boycotts of those companies within China in the days following the sanction approval (Brant, 2021b). These two examples of global attention to the Uyghurs’ treatment in combination with the massive spikes in activity online highlight how important Twitter was as a platform to spread computational propaganda that pushed back against global opinion at the time.

Figure 4a shows the frequency of relevant headlines per month. We see two peaks in article publication, in July and December of 2019. We further examine the headline production by day in Figure 4b to contextualize these spikes. On July 8th, 2019, 22 nations at the United Nations Human Rights Council (UNHRC) issued a joint statement to the UNHRC president and U.N. High Commissioner on Human Rights, calling on China to “refrain from the arbitrary detention and restrictions on freedom of movement of Uighurs, and other Muslim and minority communities in Xinjiang” (Lawrence et al., p. 32). Notably, the United States is not one of the 22 signers of this document. However, in December of 2019, the United States government took a stance regarding the Uyghurs, passing a bill (almost unanimously) to “condemn gross human rights violations of ethnic Turkic Muslims in Xinjiang, and calling for an end to arbitrary detention, torture, and harassment of these communities inside and outside China” (S.178 - Uighur Act of 2019, 2019, para. 1). Again, it is clear that the spikes in relevant articles align with current events of the time.

There are several differences when comparing the spikes in tweets and headlines. First, while tweet spikes appear to occur instantaneously, the headline spikes often occur a few days after the incident of significance. This highlights one of the main differences in traditional and computational propaganda: the ability to respond in real-time. It is also interesting to note that the spikes occur during different times, with tweet spikes occurring in 2021 and headline spikes occurring in 2019. This emphasizes the growing importance of social media as a medium to spread propaganda, and the growth of computational propaganda in recent years.

Conclusion

In relation to RQ1, it is apparent that social media offers unique advantages for spreading propaganda. While having access to a foreign audience is an advantage, there are other advantages to Twitter and similar platforms. On Twitter, propagandists can quickly generate a multitude of responses in a less formal setting than via a more traditional news outlet like *China Daily*. Twitter messages can quickly and anonymously counter some narratives directly, as is demonstrated in the roughly 4,500 tweets directly related to our topic of study. On Twitter, users can also flood the online space with noise in efforts to drown out conversation about a given topic while amplifying others. Traditional media organizations, on the other hand, are bound to more structured releases of information. They cannot expeditiously generate the same level of noise. Nor can such outlets quickly and repetitively republish the same content the way social media propagandists can.

Our analysis demonstrates these high-speed messaging efforts via the spikes across the tweets per month and the corresponding magnitude of those spikes. The data reveals that these spikes happen in rapid response to current events involving the Uyghurs in Xinjiang. While traditional media like *China Daily* can certainly be used to respond to global events, they generate significantly less content, with more journalistic restrictions on where content is sourced, and thus must take a more targeted approach in their propaganda efforts. Finally, social media allow for various degrees of anonymity and plausible deniability. Twitter stands out in this regard, given it lacks a “real name” policy and is amenable to complete anonymity. Established publishing sites (such as *China Daily*) are limited in how they frame and push propaganda because they are public-facing and rely on some amount of credibility to maintain appearances. Those spreading computational propaganda, in contrast, can utilize anonymity and other affordances of social media (including massive, real-time interactions during news events) to both better conceal their identities and amplify their communication efforts.

The results of our data analysis answer RQ2, which focuses on understanding the difference between pro-CCP propaganda on social media and traditional methods. We found that pro-CCP messaging has evolved on Twitter to be more internationally-oriented. This is evidenced by the range of languages that the larger corpus of tweets were written in: Chinese, Japanese (a regional power and historical adversary of China), English (for Western audiences and the U.S.), and 34 other languages. This multinational, multilingual campaign strongly suggests that the content is meant for audiences beyond China's borders. The trends found in the results of our topic modeling highlight how the propaganda messaging differs between social and traditional media. Speaking generally, the propaganda on Twitter emphasized negative messaging campaigns critiquing others over positive content about China, demonstrated by the 68% of English tweets containing negative messaging and the 32% of English tweets spreading positive messaging. When examining the literature and comparing the Twitter content to *China Daily* headlines (wherein 40% emphasize positive messaging and 54% negative messaging), it is clear how pro-CCP messaging has evolved on social media platforms from strategically focused upon positive messaging to leveraging negative attacks seemingly aimed at foreign enemies. While positive messaging on both Twitter and *China Daily* depicted Xinjiang as a great place to live, they used other unique methods of persuasion as well; topic 2 focused on appealing to tourists to visit Xinjiang while the *China Daily* headlines focused on the population growth in the region and receding COVID-19 cases. These subtle changes in rhetoric further demonstrate how foreign influence campaigns draw from both computational propaganda and public diplomacy techniques, depending on the topic discussed.

Discussion

In the context of the Uyghurs in Xinjiang, this study particularly focuses upon pro-CCP computational propaganda efforts. We show how these tactics include negative content that vilifies the West to publicly and internationally shape public opinion related to the Uyghurs and their ongoing persecution in China. This work pushes back against the misperception that soft power efforts are slower or more cumbersome (Zaharna, 2007), by highlighting the ability those hoping to shape public opinion on social media have by being able to respond to real time events. We also extend current studies of public diplomacy by analyzing the changing nature of pro-CCP computational propaganda (Nip & Sun, 2022), the relationship between pro-CCP computational propaganda and world opinion, and the utilization of social media to propagate pro-CCP messages. In doing so, this study goes beyond an attempt at distinguishing public diplomacy from propaganda (an argument which has been repeated ad nauseam for decades; see Guth, 2009), to understanding how state governments can leverage multiple strategies to achieve their foreign communication goals, utilizing both traditional and new forms of communication.

Limitations and future work

In the scope of this work, there are several limitations that can be expanded upon in future work. First, both datasets are limited. The tweet dataset was provided by Twitter Transparency, which has been shuttered following Elon Musk's acquisition of Twitter. While these datasets present a unique opportunity to study content deleted from their platforms, their comprehensiveness relies on trust that a social media company would provide a complete dataset. Thus, the timeframe of the analysis is driven in part by the data availability. The timeframe of the data provided coincides with the beginning of the COVID-19 pandemic, and is reflected in the current body of tweets, which warranted data cleaning to capture trends regarding the Uyghurs on Twitter. Future studies could expand their collection using additional API data from different actors or data from other platforms. This data also focused on tweet text and did not include any images or videos that may have been posted. Future work could also include exploring the visual aspects of propaganda during this timeframe. We further only focused on English and Chinese tweets, and did not consider data in other languages such as Japanese or Arabic. Nevertheless, owing to the challenges of otherwise collecting *and* sharing government propaganda data, these datasets warrant greater scholarly consideration. Similarly, the *China Daily* dataset only considered one publisher; other publishers could be included in the headline analysis to provide a more holistic analysis of internal messaging in China.

Other limitations involve the analysis of the datasets. The keywords to scan the tweets for could easily be expanded to include words such as "Muslim" or "muslim" for document inclusion or "outbreak" to reject documents. Regarding the time series plots, including irrelevant tweets alongside the subset of relevant tweets could also yield interesting results. Future work can emphasize the importance of understanding current computational propaganda efforts across time and meda, diplomacy efforts as reactions to the global ecosystem, and the role social media will inevitably continue to play in propagating public diplomacy.

References

Abhishek, A. (2021). Overlooking the political economy in the research on propaganda. *Harvard Kennedy School Misinformation Review.* https://misinforeview.hks.harvard.edu/article/overlooking-the-political-economy-in-the-research-on-propaganda/

Albert, E (2018). China's Big Bet on Soft Power. Council on Foreign Relations. https://www.cfr.org/backgrounder/chinas-big-bet-soft-power

Aoláin, F., Ní, A. E., Mullally, S., Voule, C. N., & Rajagopal, B (2022). Xinjiang report: China must address grave human rights violations and the world must not turn a blind eye, say UN experts. https://www.ohchr.org/en/press-releases/2022/09/xinjiang-report-china-must-address-grave-human-rights-violations-and-world

Arablouei, R., & Abdelfatah, R (2021). “Who The Uyghurs Are And Why China Is Targeting Them”. *NPR*. https://www.npr.org/2021/05/31/1001936433/who-the-uyghurs-are-and-why-china-is-targeting-them

Bender, M., Braun, T., Möller, R., & Gehrke, M (2023). “Unsupervised Estimation of Subjective Content Descriptions.” In 2023 *IEEE 17th International Conference on Semantic Computing (ICSC)*, 266-73. ISSN: 2325-6516. March 20, 2023. https://doi.org/10.1109/ICSC56153.2023.00052

Bolsover, G. (2018). China: An Alternative Method of a Widespread Practice. Vol. 1. *Oxford University Press*. https://doi.org/10.1093/oso/9780190931407.003.0010

Bolsover, G, & Howard, P. (2018). “Chinese computational propaganda: automation, algorithms and the manipulation of information about Chinese politics on Twitter and Weibo”. *Information, Communication & Society,* 22(14), 2063-80. https://doi.org/10.1080/1369118X.2018.1476576

Brady, Anne-Marie. (2012). ““We Are All Part of the Same Family”: China’s Ethnic Propaganda”. *Journal of Current Chinese Affairs,* 41(4), 159-181. https://doi.org/10.1177/186810261204100406

Brady, Anne-Marie. (2015). “China's foreign propaganda machine”. *Journal of Democracy,* 26(4), 51-59. https://doi.org/10.1353/jod.2015.0056

Brant, R. (2021a). “US diplomats to boycott 2022 Beijing Winter Olympics”. *BBC News* https://www.bbc.com/news/world-us-canada-59556613

Brant, Robin (2021b). “Nike, H&M face China fury over Xinjiang cotton ‘concerns’”. *BBC.* https://www.bbc.com/news/world-asia-china-56519411

Campbell, C. (2021). How Beijing Is Redefining What It Means to Be Chinese, from Xinjiang to Inner Mongolia. https://time.com/6078961/china-ccp-anniversary-identity/

Chang, T. K., & Lin, F. (2014). “From propaganda to public diplomacy: Assessing China’s international practice and its image, 1950-2009”. *Public Relations Review,* 40(3), 450-458. https://doi.org/10.1016/j.pubrev.2014.04.008

Chaudhari, D. D., & Pawar, A. V. (2021). Propaganda analysis in social media: A bibliometric review. Information Discovery and Delivery, 49(1), 57-70. https://doi.org/10.1108/IDD-06-2020-0065

Chen, L. (2004). “Evaluation in media texts: A cross-cultural linguistic investigation”. *Language in Society,* 33(5), 673-702. https://doi.org/10.1017/S0047404504045026

China Daily. About Us | About China Daily. Accessed April 19, 2023. https://usa.chinadaily.com.cn/2011-04/13/content_12319449.htm

deLisle, J., Goldstein, A., & Yang, G. (2016). The Internet, Social Media, and a Changing China. *University of Pennsylvania Press*, March. https://doi.org/10.9783/9780812292664

Ding, S. (2009). “Informing the Masses and Heeding Public Opinion: China’s New Internet-Related Policy Initiatives to Deal with Its Governance Crisis.” *Journal of Information Technology & Politics,* 6(1), 31-42. https://doi.org/10.1080/19331680802681857

Fedorov, A. M., & Datyev, I. O. (2022). “The Effect of Additive Regularization for Topic Modeling of Social Media Communities”. *Artificial Intelligence Trends in Systems*, 557-67. https://doi.org/10.1007/978-3-031-09076-9_51

Finnegan, C. (2020). “The Uyghur Minority in China: A Case Study of Cultural Genocide, Minority Rights and the Insufficiency of the International Legal Framework in Preventing State-Imposed Extinction”. *Laws,* 9(1), https://doi.org/10.3390/laws9010001

Gilbert, S. (2022). “Caught in the crossfire: The Uyghurs in America’s propaganda war with China”. *Human Geography,* (1), 60-63. https://doi.org/10.1177/19427786221091007

Gilboa, E. (2008). “Searching for a Theory of Public Diplomacy”. The ANNALS of the *American Academy of Political and Social Science,* 616(1), 55-77. https://doi.org/10.1177/0002716207312142

Gilsinan, K. (2020, May 28). How China Is Planning to Win Back the World. The Atlantic. https://www.theatlantic.com/politics/archive/2020/05/china-disinformation-propaganda-united-states-xi-jinping/612085/

Guo, J. (2021). Crossing the “Great Fire Wall”: A Study with Grounded Theory Examining How China Uses Twitter as a New Battlefield for Public Diplomacy. *Journal of Public Diplomacy,* 1(2), 49-74. https://doi.org/10.23045/jpd.2021.1.2.049

Guth, D. W. (2009). Black, white, and shades of gray: The sixty-year debate over propaganda versus public diplomacy. Journal of Promotion Management, 14(3-4), 309-325. https://doi.org/10.1080/10496490802624083

Hagström, L, & Nordin, A. H. M. (2019). “China’s “Politics of Harmony” and the Quest for Soft Power in International Politics”. *International Studies Review,* 22(3), 507-25. https://doi.org/10.1093/isr/viz023

Hartig, F. (2020). “Rethinking China’s global ‘propaganda’ blitz”. *Global Media and Communication,* 16(1), 3-18. https://doi.org/10.1177/1742766519871694

Hass, R. (2021). The “new normal” in US-China relations: Hardening competition and deep interdependence. https://www.brookings.edu/blog/order-from-chaos/2021/08/12/the-new-normal-in-us-china-relations-hardening-competition-and-deep-interdependence/

Huang, Z. A., & Wang, R. (2021). “Exploring China’s Digitalization of Public Diplomacy on Weibo and Twitter: A Case Study of the U.S. – China Trade War”. International *Journal of Communication,* 15(0), 28. https://ijoc.org/index.php/ijoc/article/view/15105

Isobel, C. (2020). Welcome to TikTok’s sanitized version of Xinjiang. https://www.rappler.com/technology/features/tiktok-sanitized-version-xinjiang-coda-story/

Jia, R., & Li, W. (2020). Public diplomacy networks: China’s public diplomacy communication practices in Twitter during Two Sessions. *Public Relations Review*, 46(1), 1-12. https://doi.org/10.1016/j.pubrev.2019.101818

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*. https://doi.org/10.1017/S0003055417000144

Lau, S., & Barigazzi, J. (2021). "EU imposes sanctions on four Chinese officials". https://www.politico.eu/article/eu-imposes-sanctions-on-four-chinese-officials/

Lawrence, S., Campbell, C., Fefer, R. Leggett, J., Lum, T., Martin, M., & Schwarzenberg, A. (2019). U.S.-China Relations. https://crsreports.congress.gov/product/pdf/R/R45898

Lim, P. H. (2021). China Steps Up Assimilation of Ethnic Minorities by Banning Languages in Schools. https://www.voanews.com/a/china-steps-up-assimilation-of-ethnic-minorities-by-banning-languages-in-schools-/6281558.html

Lukito, J. (2024). Global Misinformation & Disinformation Special Issue Introduction. International Journal of Public Opinion Research, 36(3). https://doi.org/10.1093/ijpor/edae030

Lum, T., & Weber, M. A. (2023). China Primer: Uyghurs. https://crsreports.congress.gov/product/pdf/IF/IF10281

Maizland, L. (2022). China's Repression of Uyghurs in Xinjiang. Technical report. *Council on Foreign Relations*. https://www.cfr.org/backgrounder/china-xinjiang-uyghurs-muslims-repression-genocide-human-rights

Maizland, L, & Albert, E. (2020). The State of Religion in China. *Council on Foreign Relations*. https://www.cfr.org/backgrounder/religion-china

Mattingly, D. C., & Sundquist, J. (2022). "When does public diplomacy work? Evidence from China's "wolf warrior" diplomats". *Political Science Research and Methods*, 1-9. https://doi.org/10.1017/psrm.2022.41

Melissen, J. (2005). "The New Public Diplomacy: Between Theory and Practice". *Studies in Diplomacy and International Relations*, 3-27. https://doi.org/10.1057/9780230554931_1

Monaco, N., Smith, M., & Studdart, A. (2020). Detecting Digital Fingerprints: Tracing Chinese Disinformation in Taiwan. Technical report. https://graphika.com/reports/de-tecting-digital-fingerprints-tracing-chinese-disinformation-in-taiwan

Mulvenon, J. C., & Chase, M. S. (2008). "Breaching the Great Firewall: External Challenges to China's Internet Controls". *Journal of E-Government,* 2(4). https://doi.org/10.1300/J399v02n04_05

Niewenhuis, L. (2021). How will Facebook deal with Beijing's propaganda on Xinjiang?. https://thechinaproject.com/2021/04/02/how-will-facebook-deal-with-beijings-propaganda-on-xinjiang/

Nip, J. Y. M., & Sun, C. (2022). Public Diplomacy, Propaganda, or What? China's Communication Practices in the South China Sea Dispute on Twitter. *Journal of Public Diplomacy*, 2(1), 43-68. https://doi.org/10.23045/JPD.2022.2.1.4

Nye, J. S. (2008). "Public Diplomacy and Soft Power". The ANNALS of the *American Academy of Political and Social Science,* 616(1), 94-109. https://doi.org/10.1177/0002716207311699

Pompeo, M. R. (2021). Determination of the Secretary of State on Atrocities in Xinjiang (Press Statement). https://2017-2021.state.gov/determination-of-the-secretary-of-state-on-atrocities-in-xinjiang/

Potkin, F., Baptista, E., & Munroe, T. (2022). "How China became big business for Twitter" Reuters. https://www.reuters.com/technology/block-blue-ticks-how-china-became-big-business-twitter-2022-09-13/

Purbrick, M. (2017). "Maintaining a Unitary State: Counter-Terrorism, Separatism, and Extremism in Xinjiang and China." *Asian Affairs,* 48(2), 236-56. https://doi.org/10.1080/03068374.2017.1313595

Purnell, Newley. (2021). Facebook Staff Fret Over China's Ads Portraying Happy Muslims in Xinjiang. *Wall Street Journal*. https://web.archive.org/web/20210516193405/https://www.wsj.com/articles/facebook-staff-fret-over-chinas-ads-portraying-happy-muslims-in-xinjiang-11617366096

Rawnsley, G. D. (2005). "Old wine in new bottles: China—Taiwan computer-based 'information warfare' and propaganda". *International Affairs,* 81(5), 1061-1078 https://doi.org/10.1111/j.1468-2346.2005.00502.x

Roth, Y. (2019). "Information operations on Twitter: principles, process, and disclosure". https://blog.x.com/en_us/topics/company/2019/information-ops-on-twitter

S.178 - Uighur Act of 2019. (2019). https://docs.house.gov/billsthisweek/20191202/BILLS-116s178-SUS.pdf

Shichor, Y. (2005). "Blow Up: Internal and External Challenges of Uyghur Separatism and Islamic Radicalism to Chinese Rule in Xinjiang". *Asian Affairs: An American Review,* 32(2), 119-36. https://doi.org/10.3200/AAFS.32.2.119-136

Stockmann, D, & Gallagher, M. E. (2011). "Remote Control: How the Media Sustain Authoritarian Rule in China". *Comparative Political Studies,* 44(4), 436-467. https://doi.org/10.1177/0010414010394773

Tazamal, M. (2018). How anti-Muslim rhetoric drives the imperialist "Global War on Terror" https://bridge.georgetown.edu/research/how-anti-muslim-rhetoric-drives-the-imperialist-global-war-on-terror/

Twitter. (2021). Disclosing state-linked information operations we've removed. https://blog.x.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed

Wang, R., & Xu, W. W. (2022). "Hashtag framing and stakeholder targeting: An affordance perspective on China's digital public diplomacy campaign during COVID-19." *Journal of Information Technology & Politics,* 0(0), 1-19. https://doi.org/10.1080/19331681.2022.2096742

Woolley, S. C., & Howard, P. N. (2018). Computational propaganda: Political parties, politicians, and political manipulation on social media. *Oxford University Press*. https://doi.org/10.1093/oso/9780190931407.001.0001

Yusupov, R. 2022. "Crimes against sovereignty": Foreign diplomacy, state propaganda, and the Uyghur crisis in Xinjiang." *HAU: Journal of Ethnographic Theory,* 12(2), 382-391. https://doi.org/10.1086/720565

Zaharna, R.S. (2007). The soft power differential: Network communication and mass communication in public diplomacy. The Hague journal of diplomacy, 2(3), 213-228. https://doi.org/10.1163/187119007X240505

▌Neeley Pate

Neeley Pate is a third year computer science PhD student at the University of Rochester. Her current research revolves around belief rigidity, misinformation, and technical interventions impacting message interpretation and reception. Her domain is broadly categorized as human-computer interaction, with an emphasis on the political and psychological domains. Previously, she attended The University of Texas at Austin, where she received her Bachelor's of Science in electrical and computer engineering and conducted research with the Center for Media Engagement.

▌Inga K Trauthig

Inga K Trauthig is research professor at the Jack D. Gordon Institute for Public Policy at Florida International University (FIU).

▌Josephine Lukito

Josephine ("Jo") Lukito (she/her) is an Assistant Professor at the University of Texas at Austin's School of Journalism and Media. She is also the Director of the Media & Democracy Data Cooperative and a Senior Faculty Research Affiliate for the Center for Media Engagement.

▌Samuel C Woolley

Samuel C Woolley is the inaugural William S. Dietrich II Endowed Chair in Disinformation Studies and an Associate Professor in the Department of Communication at the University of Pittsburgh.

Appendix

Specifics of the Topic Modeling

BERT Model Parameters

Parameters were set in the BERTopic models to ensure reproducible results. $n\_neighbors$ is defined as "the number of neighboring sample points used when making the manifold approximation" or in the creation of the topic clusters (Grootendorst, n.d., para. 19). $n\_neighbors$ ultimately determines the number of clusters generated (at maximum). min_dist is utilized to determine, "the minimum distance apart that points are allowed to be in the low dimensional representation" (McInnes, n.d., para. 15). $metric$ is defined as "the method used to compute the distances in high dimensional space" or how the model determines which documents are "closer" or "farther" (Grootendorst, n.d., para. 21). low_memory is set to False for both models because the datasets are small enough (fewer than 1,000,000 data points) to avoid memory issues when classifying. $random\_state$ sets the random seed of the model, which ensures the same result can be produced over different runs. For the English topic model, $n\_neighbors$ was set to 200, min_dist was set to 0.0, $metric$ was set to "cosine", and $random\_state$ was set to 123. For the China Daily topic model, $n\_neighbors$ was set to 12, min_dist was set to 0.0, $metric$ was set to "cosine", and $random\_state$ was set to 123.

Coherence Score Parameters

The coherence score parameters include the following: $topics$, which was populated with the top 10 words in each topic, $corpus$, which was the bag-of-words generated for the model, $dictionary$, which consists of the word-to-ID mapping for all words, and $coherence$, which was set to "u_mass".