CHINESE IT COMPANIES UNDER U.S.-CHINA TRADE WAR: A COMPUTATIONAL POLITICAL COMMUNICATION PERSPECTIVE

Yekai Xu* and Mingqi Xie†,**
*Graduate School of Interdisciplinary Information Studies, University of Tokyo, Japan
**Schwarzman College, Tsinghua University, China

ABSTRACT

Computational political communication, based on big data analytics of social media texts, provides a paradigm for understanding the public’s view of and engagement with political events worldwide. This study reviews previous efforts by social and data scientists and offers a demo to show the potential of computational political communication. To characterize online political communication dynamics surrounding U.S.-China tensions and gain a better understanding of the U.S.-China power struggle, a vast amount of user-generated Twitter data is compiled from March 2020 to March 2021 globally. Chinese IT giants (Huawei, Tencent, and ByteDance) and major English-speaking countries (the United States, United Kingdom, Canada, Australia, New Zealand, India, and Pakistan) are chosen as keywords for filtering the tweets gathered. Sentiment analysis of the tweets is carried out automatically. It is found that the popularities of debates regarding certain nations and companies are uneven and might be triggered by events. Furthermore, rather than being segregated, the discourses of all of these companies are intertwined. It is expected that future studies can apply more fine-grained, categorized, and automated sentiment and topic analysis to show a panorama of online public opinion.

KEYWORDS

Big Data, Social Media, Computational Political Communication, U.S.-China Disputes, Tech Giants, Twitter

---

1 The two authors have made equal contributions to this paper.
1. INTRODUCTION

Big data analytics and social media are shedding new light on the interdisciplinary study of both political science and communication studies. For the first time in history, researchers are faced with profuse politics-related information and comments generated by ordinary citizens on social media, empowering them to explore the dynamics of public opinion on critical issues.

U.S.-China bilateral relation has become one of the most influential agendas in contemporary global politics. Over the past few years, starting from increasing mutual tariffs to the escalating conflicts embodied in almost every aspect of the two countries’ interaction including international governance and national security, U.S.-China disputes have drastically altered the ecosystem for global trade, investment, and supply chains. Most specifically, within Trump Administration, from the blacklisting of semi-conductor suppliers for Huawei to the ban of TikTok and WeChat, tech giants with a Chinese background are to a great extent posed as epitomes of the competitive power game scenario (Lu et al. 2019; Rosset 2019).

Despite its friendly and goodwill showbiz in the first two years, the Trump administration has then embedded a pivot of America's policy towards China. Published in December 2017, "National Security Strategy of the United States of America" initiates Washington's novel definition of China as a "competitor", "rival", "adversary", and "revisionist power", which served as a curtain-raiser for the upheaval until this day (The White House 2017). With rising nationalism domestically and intensifying pressure externally, Chinese decision-makers, on the other hand, have also been adopting a more "assertive" and "uncompromising" approach in both Aussenpolitik and Innenpolitik which inevitably accounts for de facto "lex talionis" deadlock (Clarke 2020).

The post-COVID-19 pandemic era has witnessed an exacerbated "decoupling" between the two major powers (Bernes et al. 2020). A vis-à-vis confrontation and retaliation has been projected not only the sphere of trade but more structurally impacting the network of "global tech" constructed via cooperative efforts within the last two decades. As a milestone of the ongoing industrial revolution, a new phase of globalization, and the forefront of free trade, international tech giants have their roots in crossing-border information exchange, and simultaneously contribute to deepening the connection of the "global village" (Barevičiūte 2010; Fatma & Bharti 2019; Wyne 2020).

Huawei, with its headquarters located in Shenzhen, is a leading telecommunication hardware manufacturer and flagship smart device company in China. In 2020, this 5G titan has ultimately met with the suppression from the U.S. featuring the forced outage of its semi-conductor suppliers including TSMC, Samsung, and SK Hynix (Guo et al. 2019). China-based software companies have also been dramatically impacted in 2020 by the "Clean Network" action plan of the U.S. government. TikTok, a ByteDance-made social media app for short videos, and WeChat, another app for instant messaging, audio, and video calls owned by Tencent, are subsequently included in the blacklist by Washington. Not until these latecomer high-tech innovators enjoyed their unprecedented international boom, they are confronted with the strike that may end their game in the U.S. or even all U.S. allies (Secretary of State 2020).

This study looks into Huawei, ByteDance, and Tencent for their representativeness of hardware and software providers impacted by U.S.-China disputes, and for their huge user scale and global influence. Moreover, WeChat has its major income from the Chinese Mainland while TikTok is an international market-oriented app. This can also serve as a comparison. Internet tech giants are considered to be the watershed of China's rise as a global competitor in
technology, innovation, and tertiary industry whereas at the same time an epicenter of China's "challenge" and "threat" to the U.S.

In this digital era, while state behaviors, state-company relations, and geopolitical power plays can be more precisely and directly depicted from the macroscopical policy-making level, the public's perception and interaction also constitute a fundamental sector to the decision-making process, possibly more profoundly than ever. Twitter, among all major global social media platforms, has its characteristics as an internationally connected, weak ties-based "public sphere", facilitating this research to depict a more comprehensive view of U.S.-China disputes (Castells 2008; Habermas et al. 1974; Williams 2017). Therefore, this study chooses Twitter data for delineating online political communication dynamics.

In this work, names of seven major English-speaking countries, namely, the United States, the United Kingdom, Canada, Australia, New Zealand, India, and Pakistan, are selected as keywords for filtering among all the Twitter data collected over twelve months across the globe. The U.S., U.K., Canada, Australia, and New Zealand are also known as the "Five Eyes Alliance", which is an intelligence-based strategic ally group with its origin as "UKUSA" from the Second World War (Albers et al. 2016). This study also notices the strong geopolitical proximity between India, Pakistan, and China. These two South Asian countries, acknowledged as the biggest English-speaking countries adjacent to China and together populated over 1.5 billion, cast a huge impact on China's overseas market and foreign strategies (Sun 2020). Names of the three Chinese tech giants, Huawei, Tencent, and ByteDance, are also used to target relevant tweets.

How does online public opinion perceive and respond to the disputes between America and China? What is the picture of social media discourse on specific tech giants? To our knowledge, this is the first study dedicated to answering these questions. In the second section, we will provide a concise review of the transition of political communication studies and point out its promising future for both social and data scientists. Then, we go on to introduce preliminary findings of our Twitter data analysis. Discussions and conclusions are offered at the end. This study hopes that the publication of the dataset, and the insights the dataset provides after applying certain computational methods, can inspire and facilitate more social and data science researchers to create meaningful scholarly works.

2. EXISTING WORKS: ACHIEVEMENTS, DEFICIENCIES, AND EXPECTATIONS

This section summarizes existing works in understanding the dynamics of public opinion, pointing out its shift from traditional media to social media and the impetus behind it. We show that large-scale social media data ensures a promising future for the interdisciplinary study of political communication as well as social simulating and modeling.
2.1 Traditional Media and Surveys in the Beginning: Efforts by Social Scientists

Media framing of certain countries, issues, or important figures has long been one of the most popular topics among political communication researchers. The manually-coded methods, e.g., text analysis, discourse analysis, and content analysis, are widely adopted in the studies of images and agendas constructed by different media sources, while a majority of the research by social scientists still focus on traditional media, namely newspapers, magazines, TVs, etc.

Elites have long been considered as major stakeholders of policies. Elites’ opinions are attended in ways that the policy makers themselves are interviewed, surveyed, or being experimented on. And with the development of mass media and communication studies as a discipline in last century, elites’ power was also attended via the scrutinization of public media. Research about agenda-setting in policies was an attempt to embed communication theories into political science research. The policy-making process scrutinized are about both domestic and international issues, on a local or national level. Journalists and the press, with the seemingly outdated “Fourth Estate” they presented, were thought to be an important player in policy since they were empowered by the society as an agency to set agendas. "Agendas” that they set, can trigger debate within the society that may alter the government’s decisions. The first phase of agenda setting by the press is to “mention” certain issues that can inform people, attracting their attention on something that may have been ignored otherwise. For instance, when a local factory is to be built, even the government has been bragging about the potential GDP growth and jobs it can bring, the media’s report about environmental concerns related to the factory may possibly alter the public’s attention towards the negative impacts rather than merely positive ones. With an evolving public opinion on the environmental concerns, the local government may face strong pressure to further their plan to build the factory. This also happened in cases with civil societies, NGOs, and industry organizations, for example, while the promotion and proliferation of such discussions mainly come from the press. The next phase of agenda setting may be leading the public sentiments towards certain issues. This, works with methods such as framing, may more directly changes people's views. Therefore, the political communication research effort was mostly made on the traditional media, especially the mainstream press with large audience group. (e.g., Cook et al. 1983; Bartels 1996; Wood & Peake 1998; Liu et al. 2010)

In recent years, China has won more and more attention to its international image presentation. In the early 2000s, a discussion among the intellectuals in China about soft power has led to a national policy-level long-term strategy. (Zhu et al. 2020) China's effort in building a stronger "soft power" and thus better international image, has attracted research endeavors. Golan & Lukito (2015) describes the rise of China via looking into American newspapers' "opinion articles" on China. Golan and Lukito reckon that opinion journalism, including "editorials” and "co-eds” of major newspapers, embodies the most clearly how a country's elite community thinks about specific issues. This research mainly utilizes the inductive qualitative analysis framework, looking into the Wall Street Journal (WSJ) editorials and the op-eds on China's rise. The text analysis focuses on key statements, positive or negative attitudes, and the use of quotes. It presents that "economic partnership,” "internal dispute,” "geopolitical threat,” and "economic threat” are the main categories of WSJ's narrative framework. Golan and Lukito carried out the analysis based on a belief that the elites of society are responsible and influential to foreign policymaking. Similar frameworks are universally adapted in social scientists' research on media framing and international image-related topics.
With the development of social media, its own power towards domestic and foreign policy has gained increasing attention. It has led to an era where elites are no longer the only major stakeholder, whereas the varying opinion from the general public can function as crucial factor in today’s policy making. While public opinion's response to certain countries and issues is becoming more well-attended in academia these days, classic methodologies including surveys are applied in numerous social science research. Yang (2020) carried out a study on how China's image affects China's product selling in the United States. Focusing on public opinion as a critical factor in this research, participants were recruited in a medium city of Ohio in both ways: printed and online questionnaires (on Facebook). However, this approach to depicting public opinion about China's country-of-origin image still faces questions that whether it is representative or comprehensive enough and if there exist more efficient ways to do the research, since the political and value landscape can differ from region to region, and from low income groups to high income groups in the United States.

The Verb In Context System (VICS) is also an early attempt to understand the mentality of people, especially political leaders (Schafer & Walker 2006). Based on the 1969 study of the "operational code" (George 1969), the VICS analyzes people's political opinion and beliefs on power, predictability, the role of chance, etc., through their use of verbs found in public accessible speeches and policy text. An exhaustive dictionary was established to provide reference to the orientation of each verb, e.g., friendly or hostile, optimistic or pessimistic. The VICS is widely used for studying political figures (Cuhadar et al. 2017; Dyson 2007; Renshon 2008; Renshon 2009; Walker 2011), where methods including "Leadership Traits Analysis (LTA)" are adopted to depict the process of political decision making of leaders—for example, analyzing leaders' personality influences during disputes and even armed conflicts (Dyson 2006). However, the methods' generalization to the public remains somehow stagnant, at least partly due to a lack of available texts written by the general public.

### 2.2 The Advent of Social Media: Promises for Data Scientists

Entering the era of social media, the openness and user-generating nature of cyberspace empower the general public to express and construct online discourse. According to a Pew Research Center report, more than 40% of American adults accessed information for the 2016 presidential election via social media, which provides a simple example of this ubiquitous phenomenon nowadays. The rapid development of computer science enables data scientists to directly study the public's opinion for the first time. Opposite to traditional researchers' focus on limited information sources and small target population, data scientists have been working to discover more latent and complex information from broader social media data.

Sentiment analysis is a frequently employed technique for studying social media data in text modality. The basic goals of sentiment analysis are emotion recognition and polarity detection (Cambria 2016; Poria et al. 2017). Many researchers used this method to explore country images, evaluate international relations, and predict electoral results. Chen et al. (2020) and Xu et al. (2020) are both event-based country image studies with Twitter data, observing online public opinion during the 70th anniversary of the People's Republic of China and the COVID-19 pandemic, respectively. Their data was retrieved through Twitter Streaming API, and sentiments towards China (positive, negative, neutral) were analyzed with machine learning algorithms trained on manually labeled data. Their features include: 1) Xu et al. collected and compared English and Chinese data, while Chen et al. focused on English discourse. It was found that a significant opposition existed between the online public opinions towards China of
the two languages. 2) Chen et al. provided fine-grained sentiment analysis by dividing online public opinion towards China into seven categories: Politics, Economy, Foreign affairs, Culture, Epidemic situation, Anti-epidemic measures, and Racism. They revealed that the gradual increase in negative politics-, foreign affairs-, and racism-related tweets and the decrease in non-negative epidemic situation-, anti-epidemic measures-related tweets resulted in the overall sentiments' transition from non-negative to negative towards China. 3) Chen et al. displayed the different patterns in the attitudes of Congress members, media, and social bots, showing that social bots were more likely to spread negative sentiments towards China, while media were usually non-negative. For U.S. congress members, the Republicans were more negative than the Democrats. 4) Xu et al. explored how positive and negative tweets were distributed among different countries and found that states enjoying better diplomatic relations with China generally had a positive view towards China. 5) Xu et al. obtained word vectors for the top 100 frequently and uniquely used words for both English and Chinese, positive and negative tweets through the word2vec technique. Preferred topics of distinct languages and sentiments were analyzed, e.g., positive Chinese tweets primarily focused on celebration activities while negative Chinese tweets tended to talk about broader issues like Hong Kong.

Chambers et al. (2015) modeled relations between states using sentiments revealed in tweets with country names. Seventeen months of Twitter data were collected, and the aggregated sentiments for nation pairs were calculated with a support vector machine. The results indicated an alignment between human polls and social media sentiments and confirmed the relationship between real world political events and sentiment peaks on Twitter, verifying the validity of applying social media data to infer international relations.

Predicting election results with social media data is also a focus for researchers. Related works include (D’Andrea et al. 2019; Jungherr et al. 2017; Lopez et al. 2017; Tsirakis et al. 2017). Other papers addressing online public opinion towards political events include Adams-Cohen (2020), Leong & Ho (2021), McGregor (2019), etc.

2.3 The Era of Interdisciplinary Collaboration: Computational Political Communication

Despite the considerable endeavor and contributions, the above-mentioned works made to the emerging field of computational political communication, their shortcomings are also note-worthy. Firstly, their implications for social challenges are vague. With their vision limited to describing general pictures, the advanced computational techniques are not fully utilized to answer more meaningful questions and bring about possible solutions. Moreover, a lack of real-time, or 'nowcast,' analysis, which has the potential to detect major events at an early stage and provide governments and the society with necessary notifications, also stands out as a significant deficiency for existing studies.

From the global communication perspective, it can be anticipated that a more comprehensive and real-time computational research on social media will become increasingly significant for academia and policymakers. Computational political communication is undoubtedly a rising field for interdisciplinary collaboration, with social scientists' intrinsic dedication to find questions and create meanings and data scientists' capability to initiate more sophisticated quantitative research. Data scientists should be encouraged to engage in more of this interdisciplinary area, honing and experimenting with their methodologies and theories (Margolin 2019; van Atteveldt & Peng 2018). So, below presents a demo analysis with Twitter data.
3. DATA COLLECTION

The data analyzed in this study is collected through Twitter Streaming API, which allows users to retrieve tweets with designated hashtags in real time. Since the dataset is expected to support research in a broader context and is not specific to this work, the hashtags for retrieving data are designed to include the English and Chinese names of all major countries in the world (5 permanent members in the UN security council, G20 countries, OECD countries) and countries enjoying close contact with China (member states of Shanghai Cooperation Organization and the Association of Southeast Asian Nations, and Democratic People's Republic of Korea). All tweets containing these country name hashtags were collected. The collection started on March 4, 2020 and will continue for several years. This dataset is open to all researchers. This study uses the data from March 4, 2020, to March 14, 2021, including a total of 148,725,018 tweets.

Then, we filter the collected data with the following criteria: a tweet should simultaneously include names or synonyms of at least one of the seven English-speaking countries and the names or synonyms of at least one of the three Chinese tech giants. The names and synonyms are shown in Table 1. Please note that the names and synonyms are case-insensitive since all the tweets and names would be changed into the lower case before filtering. Also, we consider only English tweets in this research. At last, a total of 149,339 tweets were selected for further analysis.

<table>
<thead>
<tr>
<th>Type</th>
<th>Names and synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Countries</td>
<td>'the us', 'UnitedStates', 'United States', 'the states', 'America', 'uk', 'UnitedKingdom', 'United Kingdom', 'Britain', 'Canada', 'Australia', 'australia', 'NewZealand', 'New Zealand', 'India', 'Pakistan'</td>
</tr>
<tr>
<td>Companies</td>
<td>'huawei', 'hua wei', 'bytedance', 'byte dance', 'zijietiaodong', 'zi jie tiao dong', 'tiktok', 'tik tok', 'douyin', 'dou yin', 'tencent', 'tengxun', 'teng xun', 'wechat', 'weixin', 'wei xin'</td>
</tr>
</tbody>
</table>

4. DATA ANALYSIS

This section gives examples about what researchers can find from social media data. For further theoretical building, these phenomena discovered can constitute critical empirical support.

4.1 Quantitative Characteristics and Linguistic Preferences

This part reveals the quantitative characteristics and linguistic preferences of tweets linked to different countries and companies.
Figure 1. Number of tweets discussing the seven countries and the three companies, from March 4, 2020 to March 14, 2021 (From top to down: Australia, Canada, India, New Zealand, Pakistan, the United Kingdom, the United States; from left to right: ByteDance, Huawei, Tencent)

Please download all the original figures at https://drive.google.com/file/d/1T_fHpKVz1Z79Fo2I-4QPqya3DIVzaZU2/view?usp=sharing.
Figure 1 displays the number of tweets about the seven chosen countries and the three selected companies from March 4, 2020, to March 14, 2021. It can be found that the ups and downs in popularity of country-company topics are significant: rarely mentioned on most days and intensively discussed in certain periods, possibly after an important event. Some events may induce massive discussions related to the same company in multiple countries, e.g., following India’s ban on TikTok on June 29, 2020, a peak of related tweets appear in Australia, India, the U.K., and the U.S. Also, public attention towards the three companies is uneven. Huawei receives the most mentions when people are simultaneously talking about Australia, Canada, New Zealand, and the U.K., while ByteDance enjoys an overwhelming popularity in India, Pakistan, and the U.S. Tencent, the company that was neither involved in political disputes nor deeply participated in the international market, was comparatively less talked about on Twitter. Similarly, countries attracted diverse levels of attention, with India the most and New Zealand the least. The U.S., Pakistan, and the U.K. were also frequently mentioned countries.

Figure 2 reveals the word vectors of frequently used words (appeared more than 500 times in our collected corpus) in tweets related to different countries and companies. Word vectors are how each word are represented in the semantic space, where words with more similar meanings are usually closer to each other. Therefore, by observing the number and size of clusters in word vectors figures, and the words contained in each cluster, researchers can extract different topics covered in the corpus.

Limited by space, this paper only shows the word vectors of Australia, Pakistan, U.K., and U.S.-related tweets. The word vectors were calculated through the word2vec technique. It can be observed that, for any given country, words from tweets related to different companies generally form only one cluster, entangling with each other rather than distributing separately.
It indicates that the discourse of all these companies is inter-related: no company forms an individual cluster and when people talked about these Chinese tech giants, they were generally talking about one issue rather than three. These results raise the possibility that there exists a single motivation for Twitter users to discuss the three Chinese companies and people were not treating them differently. The colors also provide an intuitive image of which company is more widely discussed, e.g., the prominent green in the U.K. picture tells that Huawei was more frequently talked about.

4.2 Sentiment of the Tweets

This part uses an unsupervised sentiment analysis method to show the attitudes of Twitter users when they mention the countries and companies of interest. Sentiments of text often carry with it how the author view the issue, and exploring the sentiments of the collected tweets can reveal how Twitter users saw the relationship between the respective countries and companies.

Sentiment analysis algorithms have evolved from the primitive dictionary-based methods to machine learning techniques, which can offer better performance in accuracy and implement more complex tasks such as aspect-level sentiment analysis. Machine learning techniques can be generally categorized into supervised and unsupervised approaches. The former requires a labeled dataset to train the algorithm while the latter aims at extracting the inherent features of the samples and complete the tasks without external help. Thus the unsupervised approach offers a relatively cheap and less time-consuming solution in compromise of some accuracy.
The algorithm in this study is an adapted version of the fuzzy rule-based unsupervised sentiment analysis technique developed by Vashishtha and Susan for analyzing social media posts (Vashishtha & Susan 2019). Every tweet was classified as positive, negative, or neutral with the Mamdani system and nine fuzzy rules.

Figures 3 and 4 are the fluctuations of the number of positive and negative tweets mentioning selected country-company pairs. It can be observed that most tweets are neutral, but negative tweets also significantly outnumber positive ones. In the India and ByteDance case, most of the negative tweets were sent around May 23 and June 29, while positive tweets were distributed more evenly. In the U.S. and Tencent case, a similar situation can be observed. These indicate that negative sentiments perhaps were more easily triggered by specific scenario.

Figures 5 and 6 show the number of positive and negative tweets in our data set about every country and company. India and ByteDance are the ones that attract the most attention on Twitter when discussing issues about countries and Chinese tech giants. And the rankings of the number of positive and negative tweets containing country names and company names are generally the same: more positive tweets usually translate into proportionately more negative tweets. Overall, the number of negative tweets was 5 to 7 times of the number of positive tweets, varying according to different countries and different companies.

5. CONCLUSIONS AND DISCUSSIONS

In the past few years, the bilateral relation between U.S.-China has been witnessing a "freefall," posing concern to global governance and international order. With its crossing-border and user-generated nature, social media provides a promising field for computational political communication research, enabling us to understand the mechanism of online public opinion's perception of and interaction with global politics.

Figure 5. Number of positive (left) and negative (right) tweets about Chinese tech giants when mentioning the seven English-speaking countries

Figure 6. Number of positive (left) and negative (right) tweets about the seven English-speaking countries when mentioning Chinese tech giants
This study highlights the prospect computational political communication has for people's understanding of political events in the digital era. After summarizing researchers' efforts spanning traditional media and survey to large-scale social media data, we introduce a new Twitter dataset consisted of country-related corpus. The dataset can support future studies on various international politics and communication issues. Furthermore, we provide an example for the use of such data by revealing its quantitative features and sentiment characteristics. It is found that tweets about the given countries and companies were generally induced by certain events and most of them were sent in the following days of the events. Also, Twitter users tended to treat topics related to the three companies as one coherent issue, and no separation was observed in the topic clusters of the Twitter discourse. As to sentiment analysis, we discover that negative attitudes overwhelmed positive ones when Twitter users mentioned the seven English speaking countries and the three Chinese tech giants, showing an urgent need for the companies to communicate more for the public's understanding in a geopolitical environment confronted with more uncertainty.

Future research will include 1) From social scientists' perspective: generating more political communication questions that can a) be solved with large-scale social media data and computational methods, and b) facilitate high-quality global and social governance that optimizes the living experience of all; 2) From data scientists' perspective: under the principle of respecting privacy, developing more fine-grained sentiment analysis algorithms for digital media contents to discover the political leanings of Internet users and the mechanisms for online political communication. Realizing real-time analysis is another promising task that data scientists can work towards. As an example, it would be scholarly interesting and politically insightful to establish longitudinal and real-time social media datasets to investigate the dynamics of nationalism and populism in the cyberspace which have been influencing offline politics as well in recent years.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to professor NI Shiguang at Shenzhen International Graduate School, Tsinghua University, and Professor David PAN at Schwarzman College, Tsinghua University, for their valuable suggestions and support.

REFERENCES

CHINESE IT COMPANIES UNDER U.S.-CHINA TRADE WAR: A COMPUTATIONAL POLITICAL COMMUNICATION PERSPECTIVE


