



# Agenda-Setting in Cross-National Coverage of COVID-19: An Analysis of Elite Newspapers in US and China with Topic Modeling

Kan Wu 1\*

 0000-0002-4411-1410

<sup>1</sup> Zhejiang University of Finance and Economics Dongfang College, CHINA

\* Corresponding author: [wukanq@163.com](mailto:wukanq@163.com)

**Citation:** Wu, K. (2021). Agenda-Setting in Cross-National Coverage of COVID-19: An Analysis of Elite Newspapers in US and China with Topic Modeling. *Online Journal of Communication and Media Technologies*, 11(4), e202116. <https://doi.org/10.30935/ojcm/11083>

## ARTICLE INFO

Received: 9 Jan 2021

Accepted: 16 Jun 2021

## ABSTRACT

This study examines agenda-setting in US-China elite newspapers coverage of COVID-19 through topic modeling. It attempts to contribute to studies of media agenda first by demonstrating the relevance of text-mining in agenda-setting research and second by comparing how elite newspapers from different countries choose topics as part of agenda-setting when they report a single event. Topic-modeling the news corpora collected between 15 January 2020 and 15 June 2020 from the four US-China elite newspapers, the study finds that “domestic economy” and “international relations” are the two dominant topics that help shape the agenda in the Chinese newspapers, whereas “family & friends” and “daily life” are the topics playing the same role in the US newspapers. The study argues that such differences may associate with ideological gaps between the two countries in terms of “concepts of development”, “media bias” and “views of individualism”.

**Keywords:** agenda-setting, COVID-19, topic modeling, elite newspapers, China, US

## INTRODUCTION

The present research investigates how elite newspapers in US and China choose their topics as part of agenda-setting when reporting COVID-19 with topic modeling, a text-mining technique for automatic retrieval of latent topics in a corpus (Murakami, 2017, p. 243). It attempts to contribute to studies of media agenda first by demonstrating the relevance of text-mining in agenda-setting research, a field in mass communication that has traditionally been dominated by qualitatively exploration and second by comparing how elite newspapers from different countries choose topics as part of agenda-setting when they cover COVID-19, a recent pandemic that has arrested much media attention worldwide.

Since the outbreak of COVID-19 in China back to January 2020, the pandemic has swept the world within months, leaving the total number of confirmed cases reaching 11,669,259 and death 539,906 at the time of writing (WHO, 2020). Apart from its observed threats to life and health, the pandemic has created far-reaching impacts on the politics, economy, diplomacy, daily life, etc. of those infected countries around the globe. These impacts, together with the pandemic itself, are exploited by world media and have become a recurring theme in their pandemic reports. This is especially true to the media in US and China, two superpowers with the former carrying the greatest number of current cases and the latter being the place of initial outbreak. However, due to different sociopolitical and ideological landscapes, the ways elite newspapers in US and China set their agenda when reporting COVID-19 could be different. To make the situation more complex, US-China conflicts in trade and politics over the past years seem to fuel the media bias in the two countries as they cover the pandemic in each other's territories. This is likely to intensify such difference in the agenda-setting, making it scholarly noteworthy. Finding out such difference is then the motivation of this work and where topic modeling as a methodological choice cuts in.

Topic modeling is dedicated to automatically retrieving the “aboutness” of a given corpus. As an alternative to keywords analysis, topic modeling is unsupervised and needs neither a reference corpus nor pre-identified semantic categories in the identification of corpus “aboutness”, thus improving analytical rigor. Methodologically, topic modeling defines “topic” as “a recurring pattern of co-occurring words” in a corpus (Brett, 2012, p. 12). The working mechanism behind this lies in identifying these recurring patterns based on either sampling-based or variational algorithms (e.g., LDA, LSA, NMF) (Blei, 2012, p. 81), in which all the words in a given corpus are treated as “a bag of words”, taking no account of their sequences and lexico-grammatical connections in the same corpus. It is these methodological features and edges of topic modeling that make it a proper analytical tool for agenda-setting research in news media. Yet, there are only a very limited number of agenda-setting research (e.g., Korenčić et al., 2015; Pinto et al., 2019) takes this empirical ground. I believe the potential of topic modeling is still largely untapped in agenda-setting research against the background of digital humanity. Therefore, this work employs topic modeling to explore how elite newspapers in US and China choose possibly different topics as part of agenda-setting when reporting COVID-19. This is another motivation of the study.

The sections below will begin by positioning the present work within studies of agenda-setting, which is followed by a brief overview of the development elite newspapers in US and China. Then, I will discuss the relevance of topic modeling in agenda-setting research, before explaining the research data and procedure. Finally, I will present the modeling results, which is accompanied by a brief discussion of some possible ideological factors that may partly shape the modeling results.

## **AGENDA AND AGENDA-SETTING IN NEWSPAPERS**

---

Media agenda is a fuzzy concept that defies a single definition. Of its various scholarly definitions, some conceptualize it as “distribution of topic’s coverage” (Pinto et al., 2019, p. 614), others as “the set of issue that get media coverage” (Korenčić et al., 2015, p. 1). In this work, however, we define it as a set of statistically salient topics in newspapers. Since, McCombs and Shaw (1972) raised the agenda-setting theory, research on agenda-setting has explored issues like the differences between media agenda and public agenda (e.g., Tan & Weaver, 2007; Wanta et al., 2004), the effect of media agenda on public opinions (e.g., Protesse & McCombs, 2016; Zhang et al., 2012), media bias in agenda-setting (e.g. Elejalde et al., 2018; Lazaridou et al., 2016), media agenda diversities (e.g., Guo, 2019; Humprecht and Esser, 2018), etc. These studies demonstrate not only the role of agenda as a bridge between mass media and the public, but also how media differ in agenda-setting when reporting a single event --- a multifactorial issue and the concern of this work. To show the contribution of the present work, I will position it within studies of agenda-setting by reviewing related works with the special focus on the research methodology and perspective.

Manual content analysis has methodologically dominated previous work on agenda-setting, while automated analysis using text-mining is a more recent fashion. This is especially true in the case of news media. Before the rise of digital humanity, researchers (e.g., Elder, 1997; Matthes & Kohring, 2008) would manually code the news content to investigate aspects of agenda-setting. The results of such manual coding are likely to depend on coders’ understanding of news content, even though the coding process in those studies often involves different coders for the same content. One reason for this preference of manual coding might be a lack of proper digital tools to carry out the analysis. In more recent years (the 2010s), however, the development of text-mining tools has made automated analyses possible. For instance, Kim et al. (2014) measure the agenda diversity between the newspaper and the public through topic modeling. Korenčić et al. (2015) employ a semi-supervised topic model to explore the newspaper agenda in a US news corpus through the automatic tagging of news articles. Pinto et al. (2019) compare newspaper agenda with public interest by automatic detecting latent topics in an Argentina news corpus. Compared with those manually content analyses, automated analyses using text-mining tools have improved the rigor of data processing, making large quantities of news data easily accessible to researchers.

For the coverage of a single event, investigations of agenda-setting are mostly confined to domestic news sources, while those based on cross-national sources are few. This is because the news corpora selected in a majority of previous studies involve only local/domestic events (e.g., political events). For example, Larcinese et al. (2011) study the agenda of economic news between 1996 and 2005 in major US newspapers. Bowe et

al. (2013) investigate the effects of second level agenda-setting through an exploration of the connections between US newspaper coverage of Islam and public views of Islam in US. Guo (2019) investigates media agenda diversities in the Chinese news coverage of an annual CCP political event. These studies have offered multisided insights into various aspects of agenda-setting. Nevertheless, the domestic nature of the news sources used in those studies could mean that agenda-setting in the cross-national coverage of a single event is often underexplored, except for in a very limited number of studies (e.g., Cui & Wu, 2017). Furthermore, these studies seem to concentrate mostly on the overall topical patterns in agenda-setting, while special investigations of individual topics are relatively few. The present work argues that such exploration and investigations would be meaningful when it comes to elite newspapers, because they help reveal possible ideologies of different agenda-setters across nations, illustrating thematic ingredients of agenda-building topics when those newspapers report a single event. Therefore, this study will examine these topics as part of agenda-setting in cross-national coverage of COVID-19, with elite newspapers in US and China as examples.

## ELITE NEWSPAPERS IN US AND CHINA

---

Elite newspaper in this article is defined as news publications that can influence the agenda of other news outlets. With better journalistic resources, elite newspapers often produce higher-quality information with balanced views and delivers it at the national level (Carpenter, 2007, p. 761-763). Furthermore, elite newspapers sometimes play the role of intermedia agenda setters for other non-elite news outlets, particularly in the coverage of international events and issues (Izadi, 2007, p. 148). To have a more rounded understanding of the elite newspapers and their agenda-setting roles in the coverage of COVID-19 in US and China, a brief overview of their historical development in the two countries may offer some insights.

Development of newspapers in the US is largely a “narrowcasting” process, where their target readership has shifted from mass audiences to a small number of elite/ interested readers with the development of society and advancement of technology (Singer, 2013, p. 21). The 19<sup>th</sup> Century has witnessed the narrowcast of newspaper in the US from “populist press” that associates itself with wider audiences to “penny press” newspaper that chooses *not* to conform to the taste of readers in communities by retaining its editorial independence and “partisan press” that are often politically affiliated (Schudson, 1998, p. 123). In the 20<sup>th</sup> Century, increasing literacy rates and newspaper professionalism have further accelerated this “narrowcast” trend, where two major schools of newspapers, the “populist” and “elite”, were well established to appeal to both the working class who needs daily tips for survival and the wealthy who needs commercial information for business. Stepping into the 21<sup>st</sup> Century, the advent of social networking began to gradually drain up the demand of traditional print newspapers, forcing some of the latter to reposition themselves as elite publications. This “narrowcast” of the US newspapers over the past centuries have helped shape national elite newspapers that represent the views and interests of social elites and often “have an intermedia agenda setting effect on the news agendas of local newspapers” (Golan, 2006, p. 326).

Newspapers in China follow a different path of development. In the early days of PRC, the Chinese journalist system operated under strict monitor and censor, where the government controlled the news media and the official media were the exclusive source of news outlet. However, the Chinese economic reform in 1978 and proliferation of the internet industry in the 21<sup>st</sup> Century have added “a layer of liberating force” to this Chinese journalist system (Guo, 2019, p. 2461). Today, those state-owned official newspapers are no longer the exclusive news outlet for people. They have to compete with the burgeoning commercial newspapers for the reader market, despite they are the traditional agenda setters. Commercial newspapers, mostly privately owned, come nowadays as an alternative voice for daily news consumption. They are allowed to operate with some journalist freedom under governmental monitor and censorship. Recent studies (Guo, 2019; Tang et al., 2012) show that editorial decisions of commercial newspapers reflect not only governmental agenda, but audience-based profit considerations as well. Consequently, some commercial newspapers have become an important competing force for their official counterparts, becoming the true elite through market competition. This development trajectory reveals the overall journalistic landscape in China, which started from a strictly controlled one that takes official newspapers as the exclusive agenda setters to a more recent semi-controlled one where official and elite commercial newspapers compete for market and set the agenda for each other.

**Table 1.** The sources and sizes of the media reports

Source	Number of reports	Size (tokens)
<i>The New York Times</i>	5,711	43,347,625
<i>The Washington Post</i>	4,893	32,869,639
<i>China Daily</i>	4,492	17,502,341
<i>Global Times</i>	2,004	8,400,716
Total	17,100	102,120,321

While the elite newspapers in US come as the result of “narrowcast”, their Chinese counterparts that of “competition”. In some sense, they are both “natural-selected” in their rise to elite. In most cases, elite newspapers are the public opinion leaders and the agenda setters in the coverage of important social events. Yet, the question remains as how those US-China elite newspapers choose topics as part of agenda-setting when reporting a global event like COVID-19. To answer this question, the present work takes topic modeling as the methodological solution.

## METHOD, DATA AND PROCEDURE

### Topic Modeling

Since topic modeling came to methodological prominence against the rise of digital humanities, it has been widely used in sociology (e.g. DiMaggio et al., 2013; Sterling et al., 2019), literary research (e.g., Jokers et al., 2013; Murakami, 2017), art and culture studies (Allen et al., 2017; Garcia-Zorita et al., 2018), finance & economics (e.g., Härdle et al., 2017; Wehrheim, 2019 ) and critical discourse analysis (e.g. Kimo Stine et al., 2020; Törnberg et al., 2016), etc. By uncovering topical structures of research samples/objects, these studies have explored some latent functions (e.g., representations, images, opinions, concerns) of these structures, offering fresh insights into the data under investigation and proving the worth of topic modeling in humanities research. Hence, the present work chooses topic modeling as the methodological underpin in its quest of agenda-setting.

### Data and Procedure

The data used in this research are news reports in English about COVID-19 collected between 15 January 2020 and 15 June 2020 from *The New York Times*, *The Washington Post*, *China Daily* and *Global Times* --- four elite newspapers in US and China. All the reports are collected from Factiva<sup>1</sup>, where the keywords used to search these reports in the news archives are “COVID-19”, “coronavirus” and “virus”. After careful proofreading, we remove reports irrelevant to COVID-19 and obtain a total number of 102,120,321 reports. Details regarding the collected reports are reproduced in **Table 1**. Before introducing the analysis methods, it is worth stressing that the selection of the source and language of these reports is a twofold consideration. For one thing, given the recognized reputation of the four newspapers, the quality of these reports could be well guaranteed; for another, the choice of English instead of both English and Chinese as the language included in the reports could avoid possible inconveniences stemming from cross-linguistic comparisons.

The analyses are performed on Python 3.7, where the study resorts to single words (1-gram) in the modeling process, according the “bag of words” principle (Blei, 2012, p. 81). The primary packages used are “nltk” and “Gensim” and specific steps are taken as follows: 1) preprocessing the corpus of reports with the purpose of cleansing unnecessary data noise, removing stop words, tokenizing and lemmatizing all tokens; 2) generating document-term matrix (DTM) from the preprocessed corpus to get it statistically prepared for the modeling; 3) topic modeling the corpus based on LDA, because it is so far one of the most widely used topic models (Murakami et al., 2017, p. 276); 4) evaluating the topical structure of the topic model by calculating the “Cv” coherence score to dynamically adjust the topic numbers and modeling passes for more interpretable modeling results. The reason to choose the “Cv” out of many possible ways of coherence calculation is because that it shares a higher correlation with available human topic ranking data (Syed & Spruit, 2017, p. 168); 5) visualizing the modeling results to give them better readability and comparability.

<sup>1</sup> Factiva is an international business information and news database owned by Dow Jones.

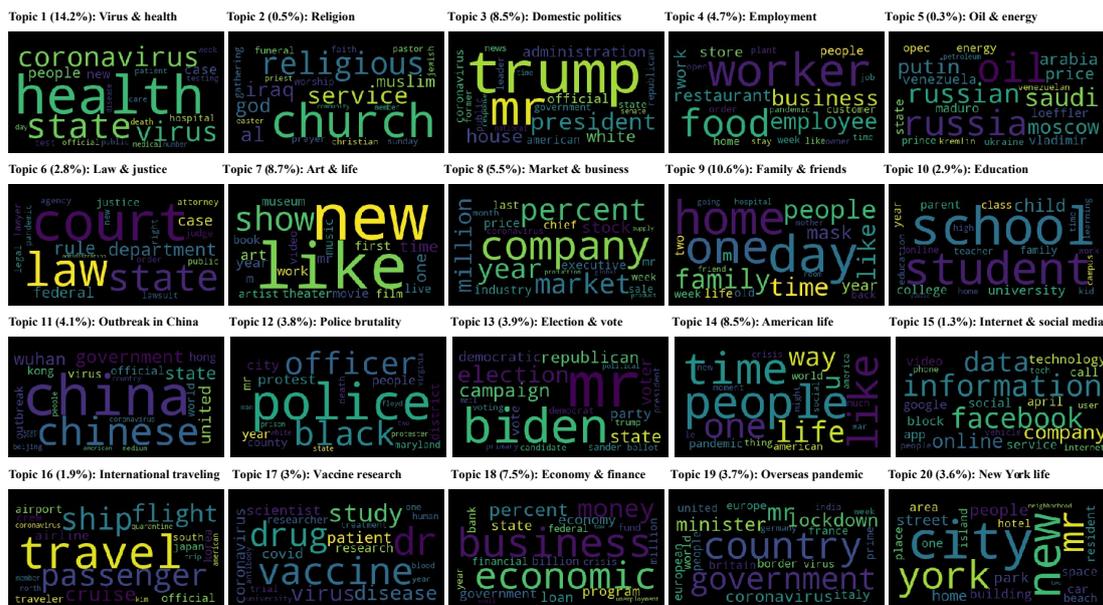


Figure 1. Topics with top 20 keywords in the US coverage of COVID-19

To have a multifaceted view of how these elite newspapers in US and China choose topics as part of agenda-setting, I will topic-model the corpus<sup>2</sup> and compare the results at both macro and micro levels. At the macro level, topic modeling is performed on all the reports from US and China, respectively. At the micro level, the modeling is separately conducted according to different time span and sources to find out how the event's development and different news agencies may affect the newspapers' choices of topics. In the cross-temporal modeling, I will segment the reports by the two countries into ten spans at the interval of 14-15 days to trace how the newspapers' topics may develop over time and compare the differences in such development between the two countries. In the cross-source modeling, I will sort all the reports according to their sources to compare to what extent does the topics by each news agency differ when reporting COVID-19. Finally, before proceeding to the results, there's a need to introduce the working definitions of two key concepts in this article: "topic strength" and "dominant topic", in which the former, represented by the number of tokens in the corpus, describes the statistical salience of a topic in a news corpus; and the latter refers to a topic with the greatest of such salience within a modelled topical structure.

## RESULTS

### The Overall Topical Patterns

Repeated modeling with topic numbers ranging from 3 to 26 on the corpus suggests that the optimized topic numbers in the US and Chinese reports are both 20, with the "Cv" coherence scores being 0.54 and 0.51, respectively. According to findings in some previous research (e.g., Nanda et al., 2021; Röder et al., 2015; Syed & Spruit, 2017), a "Cv" coherence score equals 0.50 or above is more likely to efficiently reflect the topical structure of the corpus under investigation. Based on their attached keywords, each topic is given a label that could properly summarize its topical theme. Details (label, strength, keywords, etc.) of these topics in the reports from US and China are separately presented in Figures 1 and 2. Topic strength in Figure 1 suggests that topics related to "international relations/cooperation" and "economy & finance" dominate the Chinese pandemic reports, whereas topics relevant to "family & life" and "health" dominate the US reports. This may partly point to the agenda-setting patterns between the US-China newspapers as they report COVID-19. Delving into the topics, the following diversities come into view.

<sup>2</sup> The coding process in the present research is available on figshare.com with the DOI: 10.6084/m9.figshare.14446965.



**Figure 2.** Topics with top 20 keywords in the Chinese coverage of COVID-19

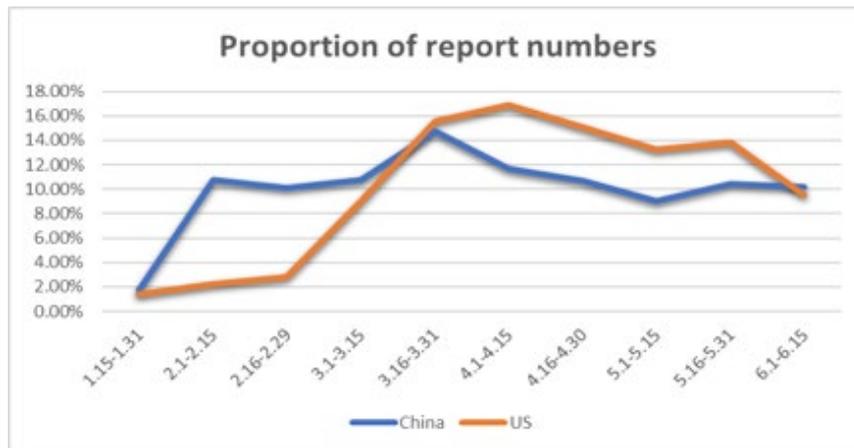
First, the US and Chinese newspapers seem to take different perspectives to report the impacts of COVID-19 on economy/finance and health. In economy/finance-related topics, keywords such as “growth”, “government”, “measure” are found in the Chinese reports, while in the US reports, keywords like “loan”, “unemployment”, “crisis” are found. This may suggest that the two Chinese newspapers take a more positive view on the impact, since they may believe that under the guidance of government policies, China can restore economic development and promote financial revitalization. By contrast, the two US newspapers hold a less optimistic view, in which they seem to worry about how this health crisis may bring economic burdens to people’s life. In medical/health related topics, keywords like “team”, “nurse”, “volunteer” are found in the Chinese reports, while these words are not found in the US reports. This could possibly mean that the Chinese reports appreciate the effort made by the medical staff and volunteers in the fighting the pandemic, whereas this appreciation is less frequently found in the US reports. Taken together, these varied topical perspectives could contribute to the ways the newspapers in the two countries set their agenda for the pandemic coverage.

Second, the newspapers in US and China appear to cover different areas of daily life amid COVID-19. The Chinese reports focus particularly on how COVID-19 affect education and traveling, as are the cases with keywords such as “student”, “online”, “education”, etc. in the topic of “Online education” as well as “control”, “prevention” and “lockdown”, etc. in “Infection report” and “Infection control”. The US reports, on the other hand, seem to care more about “family & friends” and other life-related topics, with keywords like “art”, “friend”, “theater”, “life”, “family” and “home”, etc. under these topics. I argue that these differences may reflect the lifestyles of people in the two countries, in which the Chinese may care more about daily commuting/traveling and education, while the Americans more about their entertainment and interpersonal relationships amid the pandemic.

Finally, the newspapers in US and China seem to divide in their attention when it comes to political topics. This is also likely to divide the ways the newspapers in the two countries set their agenda. Keywords like “Trump”, “international”, “cooperation”, etc. under the “International relation” in [Table 2](#) might indicate the Chinese newspapers crave for international cooperation to fight the coronavirus, while their US counterparts seem to care more about how the virus could affect the domestic political campaign and election with such keywords as “election”, “ballot”, “voter”, etc. in [Table 3](#) under “Domestic politics”. Such differences may partly mirror the political landscapes of the two countries: while China may attempt to improve its geopolitical ties when fighting against the epidemic, US is focusing on the ways COVID-19 may bring impacts on its presidential election and the public opinions.

**Table 2.** Reports with dominant Topics in the Chinese media across time

Period	Label	Terms associated with the topic
1.15-1.31	The Wuhan outbreak	Wuhan, disease, virus, spread, pneumonia, Sars, outbreak, measure
2.1-2.15	International response	international, threat, Tedros, emergency, response, pandemic, virus, racist
2.16-2.29	Medical staff & patient	patient, hospital, medical, doctor, treatment, team, worker, nurse,
3.1-3.15	World outbreak	country, Italy, health, world, death, spread, reported, confirmed
3.16-3.31	American xenophobia	racist, racism, blame, hatred, ideological, xenophobia, attacking, populism
4.1-4.15	Imported cases control	health, infection, imported, confirmed, Russia, death, quarantine, control
4.16-4.30	US politics & global cooperation	virus, Trump, international, political, politician, health, cooperation, election,
5.1-5.15	US politics & politician	Trump, administration, politician, house, white, blame, Pompeo, election
5.16-5.31	International cooperation	world, international, cooperation, global, pandemic, virus, support, state
6.1-6.15	Economic recovery	market, business, company, sale, growth, economy, trade, consumption



**Figure 3.** Proportions for the numbers of reports from US-China media across time

**Table 3.** Reports with dominant Topics in the US media across time

Period	The dominant topic	Terms associated with the topic
1.15-1.31	The Wuhan outbreak	Wuhan, virus, hospital, new, case, symptom, patient, mask
2.1-2.15	Chinese government	Beijing, official, outbreak, controllable, observe, noncitizen, preventable, restriction
2.16-2.29	Li Wenliang & Weibo	Li, Wenliang, Weibo, Wuhan, Chinese, death, censor, police
3.1-3.15	Domestic outbreak	Health, case, virus, patient, hospital, state, test, symptom
3.16-3.31	Domestic politics	Trump, president, state, official, administration, Cuomo, government, governor
4.1-4.15	Test & control	Testing, death, infection, test, vaccine, antibody, infected, lockdown
4.16-4.30	Family, friends, home life	family, home, care, mother, feel, life, friend, together
5.1-5.15	Personal finance	loan, unemployment, economic, paycheck, aid, payroll, rate, money
5.16-5.31	New York life	New, patient, city, work, York, life, disease, restaurant
6.1-6.15	Police brutality	police, protest, protester, Floyd, black, people, death, killing

### The Cross-temporal Topical Patterns

To find out how the newspapers in US and China choose topics as part of agenda-setting when reporting COVID-19 across time, I will segment the US and Chinese reports with the method described in Section *Method, Data and Procedure*, before describing and topic-modeling the reports from the two countries in different periods. The results are summarized in **Figure 3** and **Tables 2-3**.

In the outbreak of COVID-19, the newspapers of both countries take less journalistic heed of the pandemic. The Chinese reports of COVID-19 between 15 and 31 January accounts for 1.74% of all its pandemic reports. Likewise, before the large-scale outbreak, the two US newspapers reported the pandemic sporadically between 15 January and 29 February, where the proportions for the numbers of reports during these periods are invariably below 3%. This may indicate that the outbreak of COVID-19 has not got enough media attention in both countries, as it has been covered-up by the Chinese authorities and brought very few cases in the US. After the outbreak, however, the numbers of COVID-19 reports rise rapidly from 1 February in China and 1 March in the US. At the time of writing, the pandemic is under development in both countries, yet a slight drop in numbers can be detected in China from 1 April, as the situation is steadily improving. To explore how

these newspapers set their agenda for the pandemic reports over time, I will focus on the dominant topics in these reports from each period, as they are more likely to show the agenda-setting patterns.

Topic modeling results reveal that the dominant topics in the Chinese reports change more noticeable over the months than those in the US reports. This may partly suggest the two Chinese newspapers adjust their agenda more frequently than their US counterparts. From initial concerns about the Wuhan outbreak and its international responses to such wider concerns as the US outbreak, infection control, international cooperation and recovery of national economy, the shifts of dominant topics in the Chinese reports could mean that the pandemic development has imposed a far-reaching influence on the Chinese society and its geopolitics, and is thus likely to play a greater role in the agenda-setting. Dominant topics in the US reports, by contrast, start with the Chinese outbreak and comments on the Chinese government before turning into full concerns of the domestic pandemic and its multi-dimensional impacts on daily life. Such differences seem to hint that the Chinese care more about the sociopolitical consequences of the epidemic, whereas the Americans more about the varied pandemic impacts upon daily life.

Differences in the cross-temporal topical patterns are also manifested in the keywords of those dominant topics. In addition to place names and general descriptions of the virus, most keywords under the dominant topics in the Chinese reports (**Table 2**) are connected with infection report and control (e.g., “confirmed”, “control”, “quarantine”), international cooperation (e.g., “global”, “cooperation”, “support”), economic recovery (e.g., “growth”, “economy”, “consumption”) and above all US politics and racism (e.g., Trump, politician, racist, hatred). Nevertheless, keywords under the dominant topic in the US reports within the same periods as shown in **Table 3** are mostly associated with domestic pandemic and politics (e.g., “symptom”, “test”, “president”, “governor”) and a variety of daily life issues (e.g., “family”, “loan”, “restaurant”, “police”). These diversities in the keywords may once again suggest that the two elite newspapers in China have wider international concerns than and thus differ from their US counterparts in the pandemic reports. Furthermore, such differences may also imply that apart from epidemic control, the Chinese authorities do cherish their years of efforts in developing the economy and value international cooperation in combating the epidemic with the hope to enhance their global image.

### The Cross-source Topical Patterns

The idiosyncratic topics chosen by the four elite newspapers as part of agenda-setting over the pandemic reports are also noteworthy, since they may reveal the stance taken by those news agencies. The following parts will explore these topics in terms of their numbers, labels and keywords.

**Figures 4-7** capture these topical details, in which repeated modeling shows there are 24 topics of best coherence (Cv coherence score =0.52) in the reports by *China Daily*, a figure far more than the 16 topics (Cv coherence score =0.46) in the reports by *The Global Times*. In terms of the dominant topic, “international cooperation” prevails in the former, whereas “US politics” is the constant theme in the latter. This result is well within the expectation, because at the time of writing, the Sino-US relation is faced with bulk of unprecedented challenges in trade, technology and politics. Meanwhile, the Chinese authorities are seeking for international cooperation in fighting the COVID-19 amid its worldwide outbreak. As is argued previously, I interpret this as the “wannabe” role the Chinese government plays in the global political arena, so as to enhance its international image and influence. Therefore, *The Global Times* as the mouthpiece of the Chinese Communist Party and *China Daily* as a government-controlled English newspaper naturally reflect such official concerns of China. For the two US newspapers, repeated modeling reveals that the topic number of best coherence in *The New York Times* is 25 with a coherence score of 0.54, while that in *The Washington Post* is 23 with the score being 0.55. On the other hand, the dominant topic in both newspapers appear to be “family & friends”. I understand this result as the American view of interpersonal relationships, which may partly explain this unanimous concern amid the pandemic.



Figure 4. Topics with top 20 keywords in the reports of COVID-19 by *The New York Times*



Figure 5. Topics with top 20 keywords in the reports of COVID-19 by *The Washington Post*

Aside from the dominant topics, modeling results show that the four newspapers take different perspectives towards the pandemic impact on domestic economy. For the two Chinese newspapers, topics related to market, economy or finance are not only shared but outstanding (top 3 regarding topic strength), where their keywords often include ones (e.g., “technology”, “development”, “growth”, “support”, “measure”) that depict the bright side of domestic economy, even though they connect to different economic aspects in



Figure 6. Topics with top 20 keywords in the reports of COVID-19 by *China Daily*

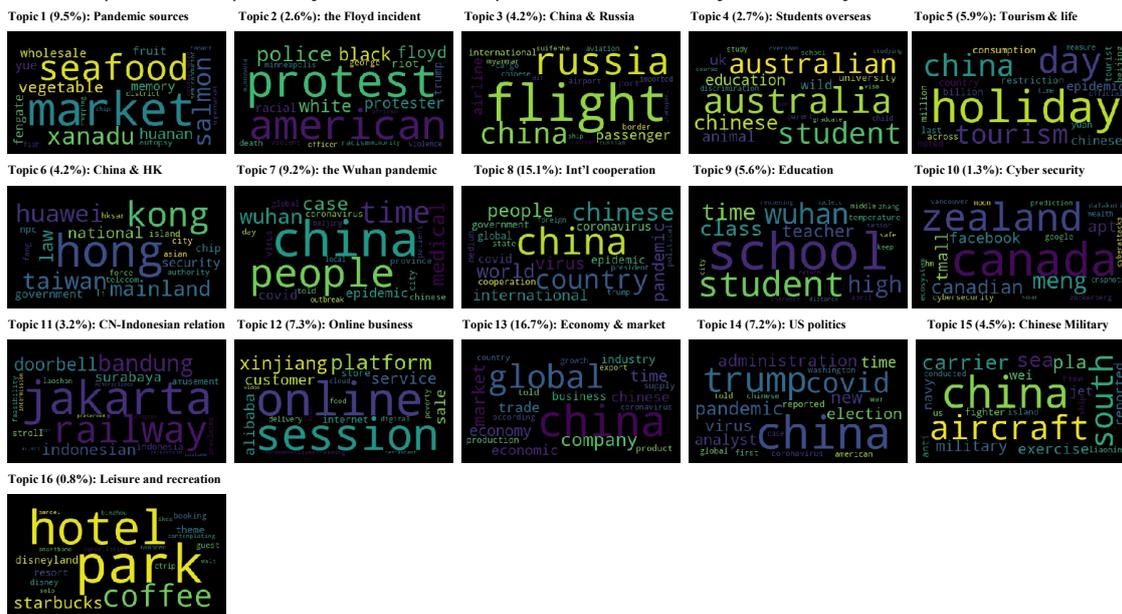


Figure 7. Topics with top 20 keywords in the reports of COVID-19 by *The Global Times*

*China Daily* and *The Global Times*. For the two US newspapers, topics like “stock & market” and “business” are not only less prominent in the corpus of reports, but sometimes imbued with pessimistic tones, as keywords such as “loan”, “unemployment”, “cut”, “crisis”, etc. are prevalent. Such differences could signify varied agenda-setting patterns, implying the contrasted media confidence in domestic economy over the pandemic: the Chinese newspapers and authorities appear to hold that the pandemic will not undermine the Chinese economy in the long run; while the US newspapers might be less optimistic in the same regard, focusing more on actual “money matters” related to individuals and enterprises.

Topics focusing on different aspects of daily life also contribute to the varied ways in which the newspapers in US and China set their agenda. For example, “life & religion”, represented by top 20 keywords like

"Christian", "church", "God", "service", etc., receives due attention by both US newspapers, but not by the two Chinese newspapers. Instead, the Chinese newspapers are more interested in tourism-related topics, which are instantiated by keywords like "visitor", "holiday", "travel", "scenic", "EU", etc. Possible reasons for such difference could be ideological and economic. As PRC is an atheism state, where religious worships are respected but not encouraged, elite media at the national level are therefore less likely to report religious issues. Yet, the US is totally different in this regard, as the Christian faith is something prevalent in the country. Furthermore, the soaring of the Chinese economy over the past decades has witnessed the rise of middle Chinese class, who has the money and passion for overseas tourism. Likewise, life relevant topics like "communities" and "care & nursing", built up by such keywords as "care", "facility", "nursing", "park", etc., seem to be the concern of both US newspapers. This is not saying that the Chinese newspapers do not care about these topics, but that keywords which could build similar topics may scatter across different reports.

Finally, the modeling results show that the four newspapers have topics of their own concern. As a result, the agenda-setting patterns in these newspapers could be individualized as they cover COVID-19. For instance, *China Daily* is particularly interested in "Wildlife & food", in which top 20 keywords like "animal", "food", "meat", "ban", etc. are frequently found in its reports. This may partly indicate the newspaper is more concerned about the illegality and possible threats of wildlife meat consumption, which may help spread the virus. I interpret this concern as the newspaper's appeal for a healthy and safe dieting culture. *The Global Times*, on the other hand, solely focuses on the topic of "Chinese military", which comprises top 20 keywords like "carrier", "PLA", "navy", "fighter", etc. As *The Global Times* is a mouthpiece, I understand this military concern as the determination of the Chinese government to defend the Chinese global interests during this pandemic crisis. For the US newspapers, *the New York Times* is especially interested in the topic of "migrants", despite its topic strength is relatively minor (0.4%). It is represented by top 20 keywords such as "immigration", "migrant", "detention", "prison", "abortion", "asylum", etc. These words appear to form a negative image of the migrants to US at the time of COVID-19. As explored by previous studies (e.g., Saeed, 2007; Van Dijk, 1989), media racism is a commonplace. Hence, I tend to view this "migrant" concern of *the New York Times* as media bias. *The Washington Post*, however, brings the "crime" issue onto the table. Even though the topic "Crime" carries a relatively minor topic strength (1.2%), keywords such as "stolen", "arrested", "theft", "police", etc. are fairly eye-catching. Since the US is a "melting pot", people from different sociocultural backgrounds may take varied responses, even criminal ones, to the ways this pandemic crisis disrupts their lives. This might also partly explain why *the Post*, as a national US newspaper, has a unique concern for the "Crime".

## DISCUSSION

Because newspaper agenda are understood in this research as a set of statistically salient topics, dominant topics which carry high topical strength are believed to partly represent patterns of agenda-setting. Topic modeling results from the overall, cross-temporal and cross-source patterns in the previous three sections show that the two Chinese newspapers are keen on topics like "economy & finance" and international relations", with a positive tone towards the pandemic impacts on economy and politics. By contrast, the two US newspapers focus mostly on such topics as "family & friends" and "life", with a negative tone towards the pandemic impacts. To understand such differences, the following part will interpret them from ideological perspectives by focusing on the ideological gaps between the two countries in terms of "economic development", "media bias" and "cultural values".

First, the study argues that the Chinese economic achievement and its developing trajectory over the past decades could be one possible reason why the two Chinese elite newspapers are obsessed with topics related to economy and diplomacy. The forty-year reform and open-up has brought the country an economic miracle that tightens the trade connection between China and the rest of world. Today, the Chinese economy has become increasingly dependent on international trade and overseas market (Wang, 2020, p. 1-2). Reverting to the news corpus, concordance search of "economy" and "finance" in the Chinese reports returns many L5:R5 collocates with negative prosody such as "crisis", "blow", "risks", "depression", etc., suggesting a media concern of the negative impacts of COVID-19 upon the Chinese economy. Such concern by the two Chinese newspapers may also connect to a governmental determination to keep the Chinese economic fruit gained over the past decades. According to IMF, China is expected to slow down its economic growth by 0.4% at the

initial outbreak of COVID-19 (Fernandes, 2020, p. 3) --- a situation which may become something on the mind of the Chinese government. Furthermore, the present economic achievements in the country are also the results of strong international cooperation (Zhang, 2020). Hence, we may understand the two Chinese newspapers' concern of international relations amid the pandemic as another facet of media concern of the Chinese economy.

Second, the study holds that "individualism" under the American cultural values may partly explain why the two elite US newspapers focus especially on life-related issues like "family & friends". For a long time, the concept of individualism has well rooted in the American cultural values, imposing a huge influence on the contemporary US society (Triandis, 1995). One study by Bianchi (2016) shows that interpersonal intimacy is inversely related to the economic situation under this individualistic culture: a robust economic will highlight individualism and weaken interpersonal relationships; while a recessed economic will increase the level of interpersonal interdependence. Similarly, a recent study by Ahmed et al. (2020, p. 332) reveals that despite some families suffered from conflicts, instability, and even divorce due to strict lockdown or social distancing, other families established stronger ties and bonds using the positive spillovers effect of this pandemic. Identified topics in the two US newspapers illustrate that COVID-19 has brought a negative impact on the US economy, which according to Bianchi's study could lead to increased interpersonal relationship and interdependence. This is well attested in the US reports, where concordance research based on the two words "family" and "friends" show that most of their L5:R5 collocates are "love", "care", "support", "sustain", etc. Therefore, it seems natural that "family & friends" becomes a more important part in daily life during the pandemic and thus one of the primary topics shared by the two US newspapers.

Finally, entrenched media bias could also be one reason why the US-China newspapers differ in their choices of topics. As China is playing an increasingly significant role in the world economy (Siddiqua, 2020, p. 1271), some US media have used a series of journalistic discourses such as "equality", "rules" and "fairness" to restrain China's international competition for the sake of safeguarding the present US edge in global economy, trade and politics (Guo, 2020, p. 106). With such ideology, some elite US newspapers aim to compress the room of China's international development by unilaterally disseminating negative information about China over the pandemic outbreak. For instance, the findings by Abbas (2020) illustrates that the stance taken by the New York Times towards the pandemic in China has shifted from "China is unable to stop the spread of the outbreak" at the outbreak to "although China clamps down on coronavirus but it is moving and spreading faster" and "China's efforts to stop the outbreak might not be real or effective" during the pandemic development --- a negative and relatively biased tone. In addition, this is also partly demonstrated in the US reports through concordance search: with "China" and "Chinese" as the nodes, many L5:R5 collocates turn out to be words with negative prosody like "animosity", "vulnerable", "crippled", "mistakes", etc. This may point to the media bias, which weakens the efforts made by the Chinese government to fight the pandemic. Furthermore, this also presents the possibility that some elite US newspapers are using media discourse to limit a favorable Chinese image in the global community.

## CONCLUSION

---

As a text-mining technique that detects latent topics in a large amount of language data, topic modeling could be a useful tool for agenda-setting research. By topic-modeling 17,100 reports of COVID-19 from four elite newspapers in US and China between 15 January and 15 June, the present work finds that those newspapers have different topics of concern when setting their agenda, despite economy, politics, daily life, etc. are their common interest. Judging from the dominant topics, the two Chinese newspapers focus more on topics related to domestic economy and international relations, whereas their US counterparts more on "family & friends" as well as "daily life". From the ideological perspective, I argue that these diversities are likely to associate with such ideological gaps between the two countries as "concept of development", "media bias" and "view towards individualism". Nevertheless, it is worth pointing out that the interpretation of modeling results in this article is largely confined to the ideological plane, even though this consideration is out of the present research needs. For further work along this line, broader perspectives (e.g., sociocultural, geopolitical) could be taken to explain the modeling results and understand possible reasons behind the division between these elite newspapers in US and China when reporting events like COVID-19.

**Funding:** Author received no financial support for the research and/or authorship of this article.

**Declaration of interest:** Author declares no competing interest.

**Data availability:** Data generated or analysed during this study are available from the author on request.

## REFERENCES

- Abbas, A. H. (2020). Politicizing the pandemic: A schemata analysis of COVID-19 news in two selected newspapers. *International Journal for the Semiotics of Law-Revue Internationale de Sémiotique Juridique*, 1-20. <https://doi.org/10.1007/s11196-020-09745-2>
- Allen, C., Luo, H., Murdock, J., Pu, J., Wang, X., Zhai, Y., & Zhao, K. (2017). Topic modeling the hàn diǎn ancient classics. *Journal of Cultural Analytics*, 1(2). <https://doi.org/10.22148/16.016>
- Ahmed, D., Buheji, M., & Fardan, S. M. (2020). Re-emphasising the future family role in "care economy" as a result of COVID-19 pandemic spillovers. *American Journal of Economics*, 10(6), 332-338. <https://doi.org/10.5923/j.economics.20201006.03>
- Bianchi, E.C. (2016). American individualism rises and falls with the economy: cross-temporal evidence that individualism declines when the economy falters. *Journal of Personality and Social Psychology*, 111(4), 567-584. <https://doi.org/10.1037/pspp0000114>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. <https://doi.org/10.1145/2107736.2107741>
- Bowe, B. J., Fahmy, S., & Wanta, W. (2013). Missing religion: second level agenda setting and Islam in American newspapers. *International Communication Gazette*, 75(7), 636-652. <https://doi.org/10.1177/1748048513482544>
- Brett, M.R. (2012). Topic modeling: a basic introduction, *Journal of Digital Humanities*, 2(1), 12-16.
- Carpenter, S. (2007). US elite and non-elite newspapers' portrayal of the Iraq War: A comparison of frames and source use. *Journalism & Mass Communication Quarterly*, 84(4), 761-776. <https://doi.org/10.1177/107769900708400407>
- Cui, D., & Wu, F. (2017). Inter-media agenda setting in global news production: Examining agenda attributes in newspaper coverage of the MH370 incident in the US, China, and Hong Kong. *Asian Journal of Communication*, 27(6), 582-600. <https://doi.org/10.1080/01292986.2017.1305426>
- Davie, W. R., & Maher, T. M. (2006). Review and criticism: research pioneer tribute —Maxwell McCombs: Agenda-setting explorer. *Journal of Broadcasting & Electronic Media*, 50(2), 358-364. [https://doi.org/10.1207/s15506878jobem5002\\_13](https://doi.org/10.1207/s15506878jobem5002_13)
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6), 570-606. <https://doi.org/10.1016/j.poetic.2013.08.004>
- Elejalde, E., Ferres, L., & Herder, E. (2018). On the nature of real and perceived bias in the mainstream media. *Plos One*, 13(3). <https://doi.org/10.1371/journal.pone.0193765>
- Elder, R. W. C. C. D. (1997). *Participation in American politics: The dynamics of agenda-building*. Allyn and Bacon, Inc.
- Fernandes, N. (2020). Economic effects of coronavirus outbreak (COVID-19) on the world economy. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3557504>
- Golan, G. (2006). Inter-media agenda setting and global news coverage: Assessing the influence of the New York Times on three network television evening news programs. *Journalism Studies*, 7(2), 323-333. <https://doi.org/10.1080/14616700500533643>
- Guo, L. (2019). Media agenda diversity and intermedia agenda setting in a controlled media environment: A computational analysis of china's online news, *Journalism Studies*, 20(16), 2460-2477. <https://doi.org/10.1080/1461670x.2019.1601029>
- Guo, L. (2020). China's "fake news" problem: Exploring the spread of online rumors in the government-controlled news media. *Digital Journalism*, 8(8), 992-1010. <https://doi.org/10.1080/21670811.2020.1766986>
- Härdle, W. K., Chen, C. Y. H., & Overbeck, L. (Eds.). (2017). *Applied Quantitative Finance*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-69179-2>

- Humprecht, E., & Esser, F. (2018). Diversity in online news: On the importance of ownership types and media system types. *Journalism Studies*, 19(12), 1825-1847. <https://doi.org/10.1080/1461670x.2017.1308229>
- Izadi, F., & Saghave-Biria, H. (2007). A discourse analysis of elite American newspaper editorials: The case of Iran's nuclear program. *Journal of Communication Inquiry*, 31(2), 140-165. <https://doi.org/10.1177/0196859906298073>
- Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics*, 41(6), 750-769. <https://doi.org/10.1016/j.poetic.2013.08.005>
- Kim, Y., Kim, S., Jaimes, A., & Oh, A. (2014, April). A computational analysis of agenda setting. In *Proceedings of the 23rd International Conference on World Wide Web*, 323-324. <https://doi.org/10.1145/2567948.2577379>
- Kimo Stine, Z., & Agarwal, N. (2020). Comparative discourse analysis using topic models: Contrasting perspectives on China from Reddit. In *International Conference on Social Media and Society*, 73-84. <https://doi.org/10.1145/3400806.3400816>
- Korenčić, D., Grbeša-Zenzerovi, M., & Šnajder, J. (2015). *Topics and their salience in the 2015 parliamentary election in Croatia: A topic model based analysis of the media agenda*.
- Larcinese, V., Puglisi, R., & Snyder Jr, J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers. *Journal of Public Economics*, 95(9-10), 1178-1189. <https://doi.org/10.3386/w13378>
- Lazaridou, K., & Krestel, R. (2016). Identifying political bias in news articles. *Bulletin of the IEEE Technical Committee on Digital Libraries*, 12.
- Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, 58(2), 258-279. <https://doi.org/10.1111/j.1460-2466.2008.00384.x>
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176-187. <https://doi.org/10.1086/267990>
- Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). "What is this corpus about?": Using topic modelling to explore a specialised corpus. *Corpora*, 12(2), 243-277. <https://doi.org/10.3366/cor.2017.0118>
- Nanda, G., Douglas, K. A., Waller, D. R., Merzdorf, H. E., & Goldwasser, D. (2021). Analyzing large collections of open-ended feedback from MOOC learners using LDA topic modeling and qualitative analysis. *IEEE Transactions on Learning Technologies*. <https://doi.org/10.1109/tlt.2021.3064798>
- Pinto, S., Albanese, F., Dorso, C. O., & Balenzuela, P. (2019). Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and its Applications*, 524, 614-624. <https://doi.org/10.1016/j.physa.2019.04.108>
- Protest, D., & McCombs, M. E. (Eds.). (2016). *Agenda setting: Readings on media, public opinion, and policymaking*. Routledge.
- Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of The Eighth ACM International Conference on Web Search and Data Mining* (pp. 399-408). <https://doi.org/10.1145/2684822.2685324>
- Saeed, A. (2007). Media, racism, and Islamophobia: The representation of Islam and Muslims in the media. *Sociology Compass*, 1(2), 443-462. <https://doi.org/10.1111/j.1751-9020.2007.00039.x>
- Schudson, M. (1998). The public journalism movement and its problems. *The Politics of News, The News of Politics*, 132-149.
- Singer, J. (2013). The ethical implications of an elite press. *Journal of Mass Media Ethics*, 28(3), 203-216. <https://doi.org/10.1080/08900523.2013.802163>
- Sterling, J., Jost, J. T., & Hardin, C. D. (2019). Liberal and conservative representations of the good society: A (social) structural topic modeling approach. *Sage Open*, 9(2), 2158244019846211. <https://doi.org/10.1177/2158244019846211>
- Syed, S., & Spruit, M. (2017, October). Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International Conference on Data Science and Advanced Analytics* (pp. 165-174). IEEE. <https://doi.org/10.1109/dsaa.2017.61>
- Tan, Y., & Weaver, D. H. (2007). Agenda-setting effects among the media, the public, and congress, 1946-2004. *Journalism & Mass Communication Quarterly*, 84(4), 729-744. <https://doi.org/10.1177/107769900708400405>

- Tang, L., & Sampson, H. (2012). The interaction between mass media and the internet in non-democratic states: The case of China. *Media, Culture & Society*, 34(4), 457-471. <https://doi.org/10.1177/0163443711436358>
- Törnberg, A., & Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media*, 13, 132-142. <https://doi.org/10.1016/j.dcm.2016.04.003>
- Triandis, H. C. (1995). *Individualism and Collectivism*. Westview Press.
- Van Dijk, T. A. (1989). Mediating racism: The role of the media in the reproduction of racism. *Language, Power and Ideology*, 199-226.
- Wanta, W., Golan, G., & Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication Quarterly*, 81(2), 364-377. <https://doi.org/10.1177/107769900408100209>
- Wang, F. (2020). Foreign dependence and strategic autonomy of China's diplomacy. *China Foreign Affairs Review*, 37(02), 1-22+156-157.
- Wehrheim, L. (2019). Economic history goes digital: Topic modeling the Journal of Economic History. *Cliometrica*, 13(1), 83-125. <https://doi.org/10.1007/s11698-018-0171-7>
- WHO. (2020). <https://COVID19.who.int/>
- Zhang, C., & Meadows III, C. W. (2012). International coverage, foreign policy, and national image: Exploring the complexities of media coverage, public opinion, and presidential agenda. *International Journal of Communication*, 6, 20.
- Zhang, J. (2020). Seeing through the "anti-China" public opinion trap of Western politicians. *Global Times*, 2020-04-30(015).

