Understanding Global News Discourse on Chinese Foreign Policy Initiatives

A Natural Language Processing Approach

Jens Koning [a]

[a] Peace Research Institute Oslo (PRIO)

https://github.com/jens-koning/dsr_media_nlp*

1 Introduction

The current competition between China and the United States (US) takes many forms. Somewhat reminecent of the Cold War, the two countries are engaged in a conflict of narratives, excerting their discursive ability to set agendas and win over the perceptions of policy-makers and citizens around the world. At the forefront of this competition are a collection of state-sanctioned foreign policy initiatives and strategies, most notably China's Belt and Road Initative (BRI) announced in 2013, China's recent push to expand the BRICS bloc to BRICS+, and the US's Build Back Better World Initiative (B3W) proposed in 2021 at the G7 meeting. While all of these initiatives have varying degrees of scope and realism, they are often widely discussed in news media and among scholars, affecting perceptions about US-China competition among global audiences.

In this article, I show how natural language processing (NLP) offers new opportunities for scholars of China to investigate the impact of news discourse on Chinese state-sanctioned initiatives such as the BRI, by using the Digital Silk Road (DSR, a sub-branch of the BRI) as a case study. The DSR presents a particularly interesting case of a foreign policy, as it sits in the between the extremes of a high-level strategic initiative, and a branding effort for Beijing to promote its global vision on technology (Carrozza [unpublished]; Cheng & Zeng 2023; Triolo 2022). Specifically, I aim to illustrate that we can use existing large news databases and topic modelling approaches to desegregate and systematically infer meaningful insights by addressing where articles on the DSR are published, as well as study what topics these articles include.

Observers of the DSR tend to form different camps depending on geographic and political

*DRAFT - please do not circulate.

affiliation. US and European observers tend to focus on the DSR being a bulwark of Chinese digital authoritarianism, and see the DSR in strategic competition with the Western technology for the future of digital connectivity (e.g. Abramowitz & Chertoff 2018; Ben-Avie 2022; Markowska 2022; Strub 2023). On the other hand, Chinese observers and state media tend to portray the cooperative benefits of DSR projects and its ability to boost digitalization and trade (e.g. Chu, Wei & Xiao 2022; CGTN 2023; Gong & Li 2019; Zihan 2022). Lastly, there tends to be third camp of observers, often from countries historically part of the Global South that strategically attempt to extract the most benefits from both China and the US through a mix of cooperative and confrontational elements, using a tactic that has been referred to as *hedging*. The countries part of the Association of Southeast Asian States (ASEAN) have been noted to exemplify such behavior (Kuik 2024). Equally, many observers in *hedging* states tend to have a pragmatic view of the utility of the DSR for their own countries, and highlight both the cooperative advantages as well as the risks involved (e.g. Lohalo 2023; Ngeow 2021; Wang 2024).

By combining news data with contextual knowledge about the DSR, I find observational evidence supporting the impression that global news discourse on the DSR include a plurality of news narratives which both highlight strategic competition between China and the US, as well as the the potential cooperative gains from the DSR. Regression models give further evidence that different media 'blocs' have consistently different narratives on the DSR the last three years, and I argue that it reflects overall political attitudes towards the foreign policy initiative. Notably, the models reveal that Western media overwhelmingly highlight competitive topics when discussing the DSR, Chinese media features cooperative topics, while the ASEAN countries

This article is structured into four sections, where I first outline a theory on the importance of international news narratives in international politics, and second move on to outline my research design and methodology - topic modelling with BERTopic. Third, I present the DSR as a typical case of a foreign policy imitative, and forth present the results of my topic and regression models. Lastly, I conclude with a discussion on what the recent media discussions on the DSR could tell us about the projects future prospects.

2 Theory: News Media, Politics and Influence

It is widely acknowledged that language and text is a important medium to shape international politics (Gimmer & Stewart 2013). One way we can understand how foreign policy initiatives are received by domestic and foreign audiences is by reading news content. News content reveal insights on how topics in international politics vary in different contexts as a product of the narrative presented. During the Cold War, news coverage about US and Soviet Union (USSR) would vary between actors depending on context and bloc-affiliation. Indian news tended to convey more positive news narratives about foreign policy initiatives presented

by the USSR, reflecting the government's non-aligned position, while new media in treaty allies of the US would often default to convey more sceptical narratives (Mastny 2010; Stein 1967). The current media dynamic between China and the US is more convoluted than that of the US and the USSR during the Cold War (in part due to the rise of the internet and social media), yet the underlying assumption that news narratives matter in a competition for global influence is pervasive. The variation in these narratives across newspapers globally could reflect different domestic beliefs about the intent of Chinese or American foreign policy, hence shaping the preferences of policy-makers having to align themselves in a world characterized by two dominant powers.

A growing body of social science literature is computationally documenting how news and speech narratives can affect decision-making on economic- and foreign affairs (Ash, Gauthier & Widmer 2023; Bénabou, Armin & Jean 2018; Eliaz & Spiegler 2020; Shiller 2020). In this arm of the literature, domestic news media can act as a constraint on policy-makers by heightening their *audience costs* for certain actions. Existing work shows that politicians are sensitive to both traditional and social media across both democratic and autocratic regimes (Baum & Potter 2019; Li & Chen 2021; Weeks 2008). As such, there is much to be gained by creating international news attention for your foreign policy initiative (given 'positive' news coverage), as it can potentially influence politicians in other countries to align their preferences with your own.

However, receiving global news attention is no easy feat. It requires the ability of states to create foreign policy initiatives that captures the attention of foreign audiences. By doing so, the state in question exhorts a form of discursive ability (or 'power'). Theoretically, we can assume that states with a large economy and powerful diplomatic and military apparatus are more likely to have an out-sized impact on news narratives in international politics.[1] Vanuatu's ideas on global governance will likely get less attention than that of Japan in global news media. In this regard, China and the United States compete in a league of their own. With access to a global network of embassies, their global new channel networks (i.e. CGTN and Radio Free Europe/Asia), economic ties to almost every country in the world, as well controlling armed forces with global reach - China and the US have the necessary infrastructure to enable themselves to put forth new foreign policy ideas that will receive media attention around the world.

The last decade China has attempted to put forth its own agenda and narratives about how the world should be governed, creating new large foreign policy initiatives like the BRI and the DSR. As noted by Brown (2020), China has "[a] need to proactively spell out what its rise means to the rest of the world" (Brown 2020, p. 323). In Chinese state media, these initiatives have been tied to themes such as economic development, collaboration and community-building with developing countries and presented as an alternative to Western-led institutions institutions and initiatives (e.g. Lin 2022). US and European think-thanks, rather highlight the security

[1] These states are often referred to as 'great powers' in the international relations literature.

implications of joining the DSR, the surveillance potential of using Chinese technology, and the threat of expanding Chinese digital authoritarianism (e.g. Abramowitz & Chertoff 2018).

> **H1:** *In systems characterized by bipolar competition, news narratives on one large power's foreign policy initiative will display topics that highlight competition, but also gains for joining the initiative.*

Lastly, there are a many countries who attempt to position themselves in between China and the United States, effectively hedging their bets and attempting to extract the most value out of each bloc. As touched upon briefly in the introduction, the ASEAN countries fit this category, hedging between the two dominant powers (Kuik 2016). While not unison, we can expect there to be a higher plurality of media voices in the region voice either that caution against the DSR, or to look at the potential commercial and developmental benefits of the DSR and/or Chinese technology. From this I deduce the following hypotheses:

> **H2:** *In systems characterized by bipolar competition, news narratives in hedging countries will highlight a plurality of topics.*

3 Research Design

Automated text analysis or NLP still represents a new branch of the literature in the study of Chinese domestic and foreign policy, with a few notable previous publications (King, Pan & Roberts 2013; Mochtak & Turksanyi 2021; Pan et al. 2020; Thunø & Nielbo 2023; Turksanyi, Liškutin & Mochtak 2022). So far automated analysis of large news data is relatively new only been used in the study of conflict (Mueller & Rauh 2022), but the method is increasingly used in industry and communication research (e.g. Guo and Vargo 2017). I recognize that a similar approach can be used to study the discursive impact of China's foreign policy, as large news data can convey information about political themes and narratives. There are several advantages of automated text analysis methods over manual methods. NLP excels in terms of speed, efficiently handling large data sets and delivering consistent and replicable results. Its multilingual capabilities make it adaptable to a diverse set of and languages, while its pattern recognition abilities can uncover insights that humans cannot. However, while NLP is efficient and replicable, a balanced approach will still require manual analysis and verification to ensure a nuanced and context-specific understanding of the data. Automated text analysis fundamentally serves a qualitative task: to extract meaning from collections of text (Grimmer, Roberts & Stweart 2022, p. 26).

In the following subsections, I outline the research design for the article. In brief, I derive news data on the DSR from NewsCatcherAPI, wrangle the data using the Python open-source programming language, manually verify the collected data, and subsequently analyze the data using a new topic model approach (BERTopic) to extract meaningful themes from the data.

3.1 News Data

The data used in this article is derived from using NewsCatcherAPI (`https://www.newscatcherapi.com/`), a technology company specializing a internet news crawling. They offer a database of global news containing more than 900 million full-text articles from the last 3 years from over 100 different countries. In this regard, their database is unique in its geographical and linguistic reach. Using Python, I call the REST API of NewsCatcher and query for articles containing quotes of "Digital Silk Road" and "China" and/or "DSR", however, I query a variety of different iterations, including and excluding some geographic areas. These are filtered using ISO 3166-1 alpha-2 country-codes. The key search term used for my data set is shown in appendix A.

The data returned from the API is structured, containing a myriad of variables including *title*, *summary* (the complete article text scraped from the internet), a *country* variable denominating the focus area of the articles (meaning the country the articles focuses on), a *link* containing the URL of the scraped article, a *clean_url* showing the root of the webpage (e.g. *thediplomat.com*). These variables have been identified by NewsCacherAPI's own data scraping methodology (CatcherAPI, 2023). The most important variables are diplayed in the Table 1 below, while a full list can be found in the appendix.

Table 1: Example of Data Structure

	title	date	clean_url	summary	country
1	China's Digital Silk Road...	2020-12-17	thediplomat.com	As part of China...	IDN
2	Bumps on the Digital Silk...	2021-04-30	straitstimes.com	Opinion - Chinese tecnology...	IDN
3	Digital Silk Road to further...	2020-10-02	nst.com.my	KUALA LUMPUR: The Digital...	MY
4	Xi calls for more fruitful...	2021-08-31	english.cctv.com	Chinese President Xi Jinp...	MY
5	China and ASEAN embrace...	2023-08-27	thejakartapost.com	The digital economy will...	IDN
6	Can the G7's B3W Initiative...	2021-12-23	thediplomat.com	The U.S. and China might...	US
7	[...]	[...]	[...]	[...]	[...]

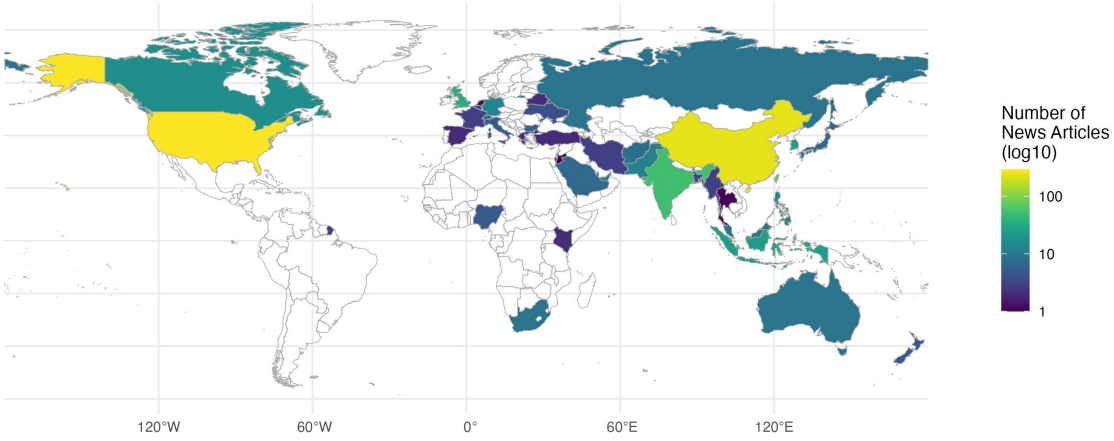


Figure 1: News Articles by 'Country' Variable (n = 975)

The quality of the data, meaning the output relevance to the DSR varies depending on input parameters. To make sure that duplicates are not included in the final data set, I make sure all duplicates are removed. Furthermore, I manually verify all the queried data and remove irrelevant articles before filtering parts of data based on preliminary results for. The reason for their removal is stated in the comments of the Python code. Replication of the search strategy can easily be done by sequentially running the scripts nested in the GitHub repository for this article.

In total my final dataset contains 975 news articles from the period from September 1st 2020 to September 31st 2023. One obvious trend we can infer by looking at the projected data in Figure 1 is that most news articles appear to talk about the DSR in the context of the two dominant powers, either China (n = 237) or the US (n = 293). Other large or populated countries are also in focus of global news media, i.e. Indonesia, Germany and India being noted to have many news articles.

3.2 Topic Modelling with BERTopic

BERTopic is a well-established deep learning framework that addresses the task of extracting coherent topics from data (Grootendorst 2022). The BERT transformer model outperforms previous topic modelling algorithms such as LDA, NMF and Top2Vec (Egger & Yu 2023), and was originally trained on a large corpus of news media, making it a particularly suitable method

for the purposes of this paper.

Technically, BERTopic works by embedding all of the documents within a data set, a process wherein each document is transformed into dense vector (numerical) representations. Using these numbers, it reduces the dimensionality of the text corpus, identifies clusters, and then uses a context-specific c-TF-IDF to create labels for each cluster based on words unique to the cluster.[2] In essence, the key advantage of BERTopic over older topic modelling approaches is that it considers the context in which words appear within a document. For example, in a longer news segment, certain words might be highly relevant and important in one section (e.g. BRI or Huawei), but not as important in another section. c-TF-IDF helps with identifying such context-specific words importance by assigning weights. The weight W of a word t in class/context c is given by:

$$W_{t,c} = tf_{t,c} \cdot \log\left(1 + \frac{A}{tf_t}\right) \tag{3.1}$$

where $tf_{t,c}$ is the term frequency of word t in class c, and A represents the average number of words per class divided by the term frequency tf_t of the word across all classes.

I use the BERTopic model in a semi-supervised way, first allowing the algorithm to find as many topics as it sees fit in the corups, then I run a iterations of the original model where I restict the minimum number of topics ($min\_topic\_size = n$). However, the corpus is not very sensitive to the minimum number of topics, although restricting it to a relatively high number such as $n = 20$ makes the model falter as it begins to integrate "stop words" in the topic outputs.[3] I found that the default model approach where I allow BERT to find the appropriate number of topics, yields the best and most consistent results across model specifications. I then extract the 15 most frequent topics, which make up my final dataset. While BERTopic uses stop words as a means to better understand the text context, Grootendorst (2022) recommends experimenting with additional treatments that can subdue their effect on the raw BERT model. One such method is CountVectorizer, which put finds a middle ground between removing stopwords and utilizing them for context. The model used in my final analysis used includes CountVectorizer (see appendix A).

4 Case Study: The 'Digital Silk Road'

The Digital Silk Road (*shuzi sichou zhilu*) or DSR is an initiative spearheaded by Beijing in 2015 as a component of the BRI. The DSR encompasses a wide array of investments in cutting-edge technologies. These investments cover various sectors, including telecommunications networks (such as roll-out of 5G networks), surveillance technology, cloud computing, e-commerce, and

[2] The Contextual Term Frequency-Inverse Document Frequency (c-TF-IDF) consists of a term frequency (TF) counter, that essentially counts words, while the invese document frequency (IDF) measues the importance of the term relative to the entire corpus of documents (the entire data set). c-TF-IDF extend this by considering the context (c) of sections within documents, allowing for more nuanced and context-aware word weighting.

[3] Stop words are words in a stop that are filtered out (i.e. stopped) before or after processing of text because they are insignificant, these are words such as "at, be, an" etc. (Rajaraman and Ullman 2011, p. 1-17).

Smart City programs, among others (NDRC, 2015). In 2022, 19 countries had signed MoUs on strengthening cooperation in building the DSR and 23 countries had signed e-commerce agreements (Lin, 2022), while over 170 countries have received or planned Chinese digital connectivity-related infrastructure and investments (IISS China Connects, 2022). The DSR is rooted in a mixture of interests: China's overall global economic ambitions, the overcapacity of Chinese technology companies eager to expand their global market presence, as well as a strong demand by recipient developing countries in need of advanced technologies and expertise.

The DSR has evolved significantly since its inception. As highlighted by Gordon and Nouwens (2022, pp. 15–16), the DSR's initial focus on digital infrastructure has expanded to include collaborations in emerging areas such as e-health and artificial intelligence (AI). This evolution reflects the dynamic nature of the initiative, shaped not only by the Chinese central state but also by non-state Chinese actors, as well as the interests of recipient countries.

Within the DSR's broader context, China has intensified its efforts to shape global internet and tech governance. Initiatives like the Global Data Security Initiative (*quanqiu shuju anquan changyi*) (State Council, 2020b) and a strong emphasis on AI development demonstrate China's commitment to technological advancement as a means to enhance its military and economic power on the global stage. The New Generation Artificial Intelligence Development Plan (*xin yidai rengong zhineng fazhan guihua*) of 2017, for instance, underscores China's aspirations to key technologies in the DSR for military and national defense purposes (State Council, 2017).

Despite its significance, the DSR remains somewhat enigmatic, lacking explicit details, and not all technology-focused projects undertaken by Chinese companies are explicitly labeled as DSR projects. This ambiguity raises questions about the primary drivers behind the initiative, whether they are commercial or strategic in nature, as there is no universally accepted understanding of its scope (Gordon & Nouwens, 2022). However, this flexibility and adaptability enables the DSR to align with different agendas and objectives, depending on the specific context (Carrozza [unpublished]).

As such, the DSR can be placed in a theoretical middle-ground of foreign policy initiatives, which sits in between being a high-level incentiveced strategy to promote the country's technological prowess, but also used a 'slogan' meant to broaden the appeal of the "Beijing model" [...] among global audiences. The DSR can also be said to be partially a product of domestic commercial interests, making it comparable to foreign policy initiatives of other great powers such as the United States (Kim and Milner 2021). In this regard, I classify the DSR as what Gerring (2017) calls an exemplifying *typical case* which may have external validity in other cases where a dominant power seek to gain influence through major foreign policy initiatives.

From a topic modelling perspective, I expect to encounter topics touching on various different elements relevant to the DSR, whether it be 'digital infrastructure', '5G' or 'standard setting.'

It is also likely that news media would cover events such as the signing of Memorandums of Understandings (MoUs) with participant countries, or locations where promotion of the DSR would take place. Moreover, as mentioned previously, many US or European publications may highlight the more strategic topics, or the DSRs challenge to reflecting the potential security implications for host-countries in joining the DSR, or some topics trying to counteract the DSR. Lastly we can also expect to see some 'positive' topics highlighting cooperation or development, in line with Chinese state media narratives.

5 Results

My BERTtopic model returns 15 relevant topics for based on 975 articles containing references to the DSR. The model output is interesting as the topic clusters highlight both competative as well as cooperative narratives. In additon, the topics shed light on important events in the duration of the dataset (September 2020 to September 2023).

Figure 2: Most Common Topics (n = 975)

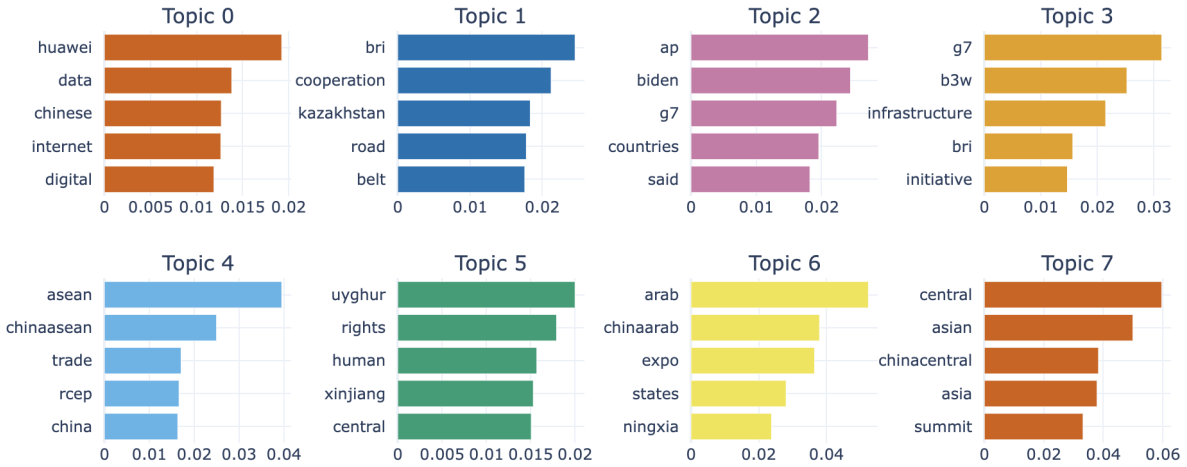


Figure 2 shows the eight most common topics in the news corpus, as well as the five most common keywords representing each topic.[4] The first topic (0) delves into the intricate landscape of China's technological advancements, particularly emphasizing the digital sector. Major tech conglomerates such as Huawei, serve as the linchpin in this discourse, representing China's burgeoning influence and burgeoning presence in the digital realm. The incorporation of terms such as "data," "Chinese internet" suggests a narrative that not only scrutinizes China's internal digital progression but also examines its strategic ambition to position itself as a global digital powerhouse.

The second thematic area (1) pivots towards China's diplomatic and economic initiatives,

[4] The x-axis represents the topic representation score for each keyword in the model. It quantifies the relevance of the keywords to the particular topic, with higher values indicating a stronger association between the keyword and the topic.

specifically through the lens of the BRI. This topic unravels the layers of China's collaborative endeavors, particularly with nations in Central Asia, exemplified by the mention of "Kazakhstan." The use of "cooperation" in conjunction with the BRI underscores China's commitment to forging partnerships that transcend borders, fostering regional integration and connectivity. This narrative is further enriched by the discussions on the BRI and DSR by global peers at the G7 meeting in 2021, as highlighted in topic (2), where keywords such as "Biden said" and "G7". The fixation on the US' administrations actions during the G7 further underscores how Chinese foreign policy is often contrasted by the US' actions.

The third topic (3) highlights that the DSR (and the BRI) is often discussed in contrast to the the G7 group and the B3W-initative, consistently including terms like "G7," "b3w," "bri," and "initiative". Concurrently, the exploration of China's relations within the ASEAN bloc in topic (4) highlights the technology partnerships and agreements, as evidenced by references to "trade" and "RCEP" trade agreement. This narrative is juxtaposed with the sensitive domain of human rights in topic (5), with a particular focus on the surveillance technology employed in Xinjiang, bringing to the fore discussions on human rights violations. In addition, topic (6) provides insights into China's burgeoning relations with Arab nations, illuminated through terms such as "expo" and "Ningxia," signifying the multifaceted nature of China's global partnerships. Lastly, topic (7) underlines the significance of summits and diplomatic engagements between China and Central Asian countries. Collectively, these topics areas provide an overview of the key news narratives on the DSR. The additional topics found in the main corpus can be seen in appendix A.

5.1 Validity of the Model

In order to test whether the model is picking up on real-world events, I probe for the validity by cross referencing the key topics with events noted in Austrian the yearbooks summarizing key events in Chinese foreign policy in 2021 and 2022.

Reading of news on the DSR in the time period 2020 to 2023 does underscore the relevance of the model as it picks up on key events in the time period, such as the G7 meeting in 2021 where the B3W is presented as an alternative to China's BRI (White House, 2021), or the 2023 Ningxia expo where Arab states and China discussed broadened collaboration on digital infrastructure (Wang & Hu, 2023). The focus on Huawei is also relevant for the time period given European countries discussion on whether to ban Huawei products in 2020 following the US' ban in 2019. In addition to these events, the model is able to discern distinct political key-words linked political attitudes about the DSR, such as 'cooperation' and 'trade', as well as more competitive topics such as 'Chinese internet' and 'human rights' and the 'B3W' as an alternative to the BRI - all of which reflects real world media chatter on the DSR. Given the plurality of topics on display, I argue that this offers support for my first hypothesis (**H1**), arguing that in a bipolar system, news narratives on one large power's foreign policy initiative will display topics that highlight competition, but also gains for joining the initiative.

5.2 News Location and Topics

Having established that the news data displays topics highlighting both competition as well as cooperation and development, I probe for whether the *location* of the news source matters in terms of the probable narrative by a running a set of binomial regression models. To do this, I manually categorize the 241 news web-pages based on the cleaned URLs variable into three different regional blocs: (1) Media outlets in the US as well as the US' key allies (Europe, Australia, South Korea, and Japan), (2) Chinese state media, and (3) media in the ASEAN countries as well. I then code three new categorical predictors for each bloc (where 1 = the news media belongs to the bloc, and 0 = news media is not in bloc). These new predictors are then regressed against the topics highlighting competition between China and the US (*'Competitive Topics'*), and the more cooperative topic, emphasising the benefits of digital trade and cooperation with China through the DSR and BRI (*'Cooperative Topics'*).[5]

Table 2: The Predicted Effect of Media Location on 'Competitive' and 'Cooperative' Topics

	Dependent variable:					
	'Competitive Topics'		'Cooperative Topics'		'Competitive Topics'	'Cooperative Topics'
	(1)	(2)	(3)	(4)	(5)	(6)
US & allies media	1.191***		−1.789***			
	(0.134)		(0.155)			
China state media		−1.683***		1.980***		
		(0.161)		(0.158)		
ASEAN media					−0.042	0.179
					(0.265)	(0.276)
Constant	−0.468***	0.585***	0.043	−1.381***	0.140**	−0.751***
	(0.095)	(0.078)	(0.092)	(0.094)	(0.066)	(0.071)
Observations	975	975	975	975	975	975
Log Likelihood	−469.957	−462.411	−400.142	−395.874	−475.146	−416.052
Akaike Inf. Crit.	943.914	928.823	804.284	795.748	954.291	836.104

Binomial GLM *p<0.1; **p<0.05; ***p<0.01

In models 1 and 3, US & Allies media exhibits a significant positive log-odds on the 'Competitive Topics' ($\beta = 1.191$, $p < 0.01$), indicating that an increase in media narratives from the US and its allies are associated with an 76.7% probability of showing more competitive news narratives. On the other hand, model 2 shows that Chinese state media narratives exhibit a negative log-odds ($\beta = -1.683$, $p < 0.01$), indicating that an increase in media narratives only have a 15.6% probability of highlighting more competitive narratives.

Model 3 and 4 further underscore the bifurcation of narratives, with model 3 showing US & Allies media exhibiting a significant negative log-odds on 'Cooperative Topics' ($\beta = -1.789$, $p < 0.01$; 14.3% probability), and model 4 unsurprisingly showing that Chinese state media

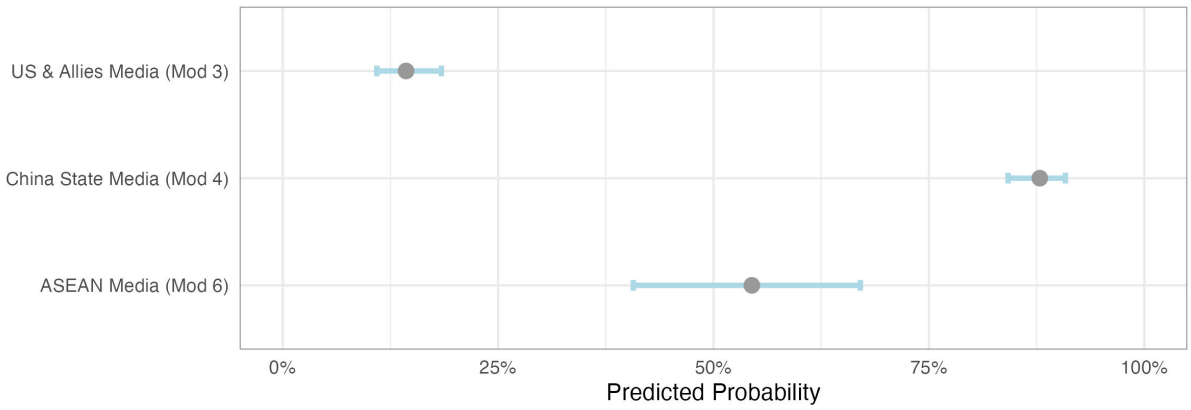
[5] The selection topics included in these categories is explained in appendix A.

exhibits a positive log-odds on the 'BRI digital cooperation' topic ($\beta = 1.980$, $p < 0.01$), being associated with an 87.9% probability of showing more cooperative news narratives.

Interestingly, model 5 and 6 targeting the effect of ASEAN media do not exhibit a significant log-odds on either the 'Cooperative Topics' or the 'Competitive Topics', indicating that the media narratives displayed in ASEAN media emphasizes a more 'neutral' narrative. The constant terms across all models are significant, but cannot be interpreted substantively as the predictors are dichotomous. The models exhibit satisfactory fit, as indicated by the low variation in log-likelihood values and Akaike Information Criteria.

Figure 3 effectively summarizes the main take-away from these models: Western media does not emphasise cooperative narratives in coverage of the DSR, ASEAN media is 'neutral', while Chinese state media has a high predicted probability of featuring cooperative narratives.

Figure 3: Predicted Probability of Displaying 'Cooperative Topics' by Media Location



In sum, the regression analysis provides further observational evidence of a world featuring different 'media blocs' that highlight different narratives on Chinese foreign policy. Notably, the pattern provides evidence to support my second hypothesis (**H2**), as it suggest that media in 'hedging' countries, such as the ASEAN group, displays a higher plurality of media narratives, and do not conform to competitive or cooperative narratives.

5.3 Limitations

The results presented above come with some limitations. The article only evaluates new media aggregated from NewsCatcherAPI, looking explicitly at anglophone new media. This is a natural constraint of data availability as other aggregators do not offer full-text articles, which would severely limit the ability of the topic model to pick up relevant narratives.

Moreover, the geographical distribution of news articles by location is skewed towards both Western and Chinese news media (see appendix A), with the ASEAN category having the lowest number of articles (see Appendix A). This naturally weakens the robustness of the results, however as noted, the argument in this articles is showing general associations, not

causality. In this regard, future research addressing news discourse on Chinese foreign policy can attempt a multilingual approach and translate topics from a pluarality of languages into comparable categories.

6 Conclusion and Discussion

In this article I have demonstrated that using a topic model can be a analytically productive way to understand news narratives on Chinese foreign policy by using the DSR as a case study. My analysis supports the impression of fragmented media narratives of China's Digital Silk road, and highlights that media 'bloc'-affiliation can serve as predictor of attitudes in foreign policy. Given the relatively recent time-frame of the news data (2020 to 2023), it is notable that the mode media coverage of the DSR focuses on Huawei and other important events such as the G7 meeting in 2021. Overall, the media narratives support the notion that there is a 'conflict of narratives' between the US and China regarding the intentions of the DSR, but also that media in 'hedging' countries in the ASEAN bloc display a higher plurality of narratives.

Given the influence domestic news media has on policy-makers, the implications of these findings suggests that projects related to the DSR will continue to be perceived in the West as a competitive and strategic effort by China to pursue increased global influence. On the contrary, Chinese state media will continue to push narratives of cooperation and mutual benefit arising from the DSR both within China and across its global news networks (i.e. CGTN), and is likely to gain headwinds in DSR host-countries with authoritarian governments and limited freedom of expression.

As hedging remains a dominant strategy for the ASEAN countries facing US-China competition (Kuik 2023), it is not surprising that news in the region reflects the prevailing beliefs of policy-makers about the DSR: Regional projects affiliated with the DSR will remain a pragmatic option to American and European technology as long as the risks attached can be controlled. This raises interesting implications for countries in the Global South

Lastly, the article adds further quantitative evidence to existing assumptions on media attitudes and coverage of Chinese foreign policy, which can aid scholars in understanding the trajectory of the competition between China and the United States. The article encourages other scholars to use a similar approach when studying media discourse on China it can offer generalizable perceptions of into

References

[1] Abramowitz, M. and Chertoff, M. (2018). Opinion | The global threat of China's digital authoritarianism. *Washington Post*.

[2] Ash, E., Gauthier, G., and Widmer, P. (2023). Relatio: Text Semantics Capture Political and Economic Narratives. *Political Analysis*, pages 1–18.

[3] Baum, M. A. and Potter, P. B. K. (2019). Media, Public Opinion, and Foreign Policy in the Age of Social Media. *The Journal of Politics*, 81(2):747–756.

[4] Ben-Avie, J. (2022). Don't let digital authoritarians lead the way in connecting the world - Atlantic Council.

[5] Brown, K. (2020). Chinese Storytelling in the Xi Jinping Era. *The Hague Journal of Diplomacy*, 16(2-3):323–333.

[6] Bénabou, R., Falk, A., and Tirole, J. (2018). Narratives, Imperatives, and Moral Reasoning. *National Bureau of Economic Research*.

[7] Carrozza, I. (2023). Connectivity with Chinese Characteristics: China's Normative Power along the Digital Silk Road. Unpublished draft.

[8] CGTN (2023). Southeast Asia: The Silk Road to a digital future - CGTN.

[9] Cheng, J. and Zeng, J. (2023). "Digital Silk Road" as a Slogan Instead of a Grand Strategy. *Journal of Contemporary China*, 0(0):1–16.

[10] Chu, Y., Wei, M., and Xiao, J. (2022). Digital trade spurs growth in Belt and Road partner countries-Xinhua.

[11] Egger, R. and Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 7.

[12] Eliaz, K. and Spiegler, R. (2020). A Model of Competing Narratives. *American Economic Review*, 110(12):3786–3816.

[13] Gerring, J. (2017). Qualitative Methods. *Annual Review of Political Science*, 20(1):15–36.

[14] Gordon, D. and Meia, N. (2022). *The Digital Silk Road: China's Technological Rise and the Geopolitics of Cyberspace*. Taylor & Francis.

[15] Grimmer, J., Roberts, M. E., and Stewart, B. M. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

[16] Grimmer, J. and Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297.

[17] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794 [cs].

[18] Guo, L. and Vargo, C. J. (2017). Global Intermedia Agenda Setting: A Big Data Analysis of International News Flow. *Journal of Communication*, 67(4):499–520.

[19] Hasmath, R. (2023). Convergence and Divergence in Policy Topics Among Think Tanks in China.

[20] IISS (2022). IISS China Connects Project. International Institute for Strategic Studies. URL: https://chinaconnects.iiss.org/methodology.aspx.

[21] Kim, I. S. and Milner, H. V. (2021). Multinational Corporations and their Influence Through Lobbying on Foreign Policy. In *Global Goliaths: Multinational Corporations in a Changing Global Economy*, pages 497–536. The Brookings Institution.

[22] King, G., Pan, J., and Roberts, M. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review*, 107(2

(May)):1–18.

[23] Kuik, C.-C. (2016). How Do Weaker States Hedge? Unpacking ASEAN states' alignment behavior towards China. *Journal of Contemporary China*, 25(100):500–514.

[24] Kuik, C.-C. (2023). Shades of grey: riskification and hedging in the Indo-Pacific. *The Pacific Review*, 36(6):1181–1214. Publisher: Routledge _eprint: https://doi.org/10.1080/09512748.2022.2110608.

[25] Li, X. and Chen, D. (2021). Public opinion, international reputation, and audience costs in an authoritarian regime. *Conflict Management and Peace Science*, 38(5):543–560.

[26] Lin, Z. (2022). China-proposed Digital Silk Road creates new opportunities for global development. *People's Daily Online*.

[Lohalo] Lohalo, N. The Digital Silk Road: A Path to Africa-China Win-Win Cooperation.

[28] Makowska, M. (2022). China's Digital Authoritarianism vs. EU Technological Sovereignty: The Impact on Central and Eastern Europe.

[29] Mastny, V. (2010). The Soviet Union's Partnership with India. *Journal of Cold War Studies*, 12(3):50–90.

[30] Mochtak, M. and Turcsanyi, R. Q. (2021). Studying Chinese Foreign Policy Narratives: Introducing the Ministry of Foreign Affairs Press Conferences Corpus. *Journal of Chinese Political Science*, 26(4):743–761.

[31] Mueller, H. and Rauh, C. (2022). Using past violence and current news to predict changes in violence. *International Interactions*, 48(4):579–596.

[32] Ngeow, C. B. (2021). The Digital Silk Road: From South-China to Southeast Asia.

[33] Pan, D., Yang, J., Zhou, G., and Kong, F. (2020). The influence of COVID-19 on agricultural economy and emergency mitigation measures in China: A text mining analysis. *PLOS ONE*, 15(10):e0241167. Publisher: Public Library of Science.

[34] Rajaraman, A. and Ullman, J. D., editors (2011). *Data Mining*. Cambridge University Press, Cambridge.

[35] Sen, G. and Bingqin, L. (2019). The Digital Silk Road and the Sustainable Development Goals. *IDS Bulletin*, 50(4). Number: 4.

[36] Shiller, R. J. (2020). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

[37] Stein, A. (1967). India and the USSR: The Post-Nehru Period. *Asian Survey*, 7(3):165–175.

[Strub] Strub, D. Confronting the Rise of Digital Authoritarianism.

[39] Thunø, M. and Nielbo, K. L. (2023). The Initial Digitalization of Chinese Diplomacy (2019–2021): Establishing Global Communication Networks on Twitter. *Journal of Contemporary China*, 0(0):1–23.

[40] Triolo, P. (2022). The Digital Silk Road and the evolving role of Chinese technology companies. In *The Digital Silk Road*. Routledge. Num Pages: 23.

[41] Turcsanyi, R. Q., Liškutin, K., and Mochtak, M. (2023). Diffusion of Influence? Detecting China's Footprint in Foreign Policies of Other Countries. *Chinese Political Science Review*,

8(3):461–486.

[42] Wang and Hu (2023). China seeks strategic energy partnership with Arab States | investinchina.chinadaily.com.cn.

[43] Wang, Z. (2024). China's Digital Silk Road (DSR) in Southeast Asia: Progress and Challenges.

[44] Weeks, J. L. (2008). Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization*, 62(1):35–64.

[45] WhiteHouse (2021). FACT SHEET: President Biden and G7 Leaders Launch Build Back Better World (B3W) Partnership.

[46] Zihan, L. (2022). China-proposed Digital Silk Road creates new opportunities for global development - People's Daily Online. *People's Daily Online*.

A Supplementary Material

List of the top 15 topics found by the model. The column 'CustomName' denotes my own interpretation of the topic, while 'Name' denotes the machine-given name of the topic. For the regression analysis, I create two main categories: For the **'Cooperative Topics' category** I include 'BRI Digital Cooperation', 'China-ASEAN Digital Trade and RCEP', 'Ningxia Expo and China-Arab Cooperation' and 'Central Asia'. For the **'Competitive Topics' category** I include 'Chinese Digital Alternative', 'B3W as Alternative to BRI', 'Technology and HR abuses in Xinjiang', 'BRI dept trap' and 'Strategic Competition' topics. All the topics are shown in table 3.

Table 3: Top 15 most Salient Topics (n = 975). BERTopic model.

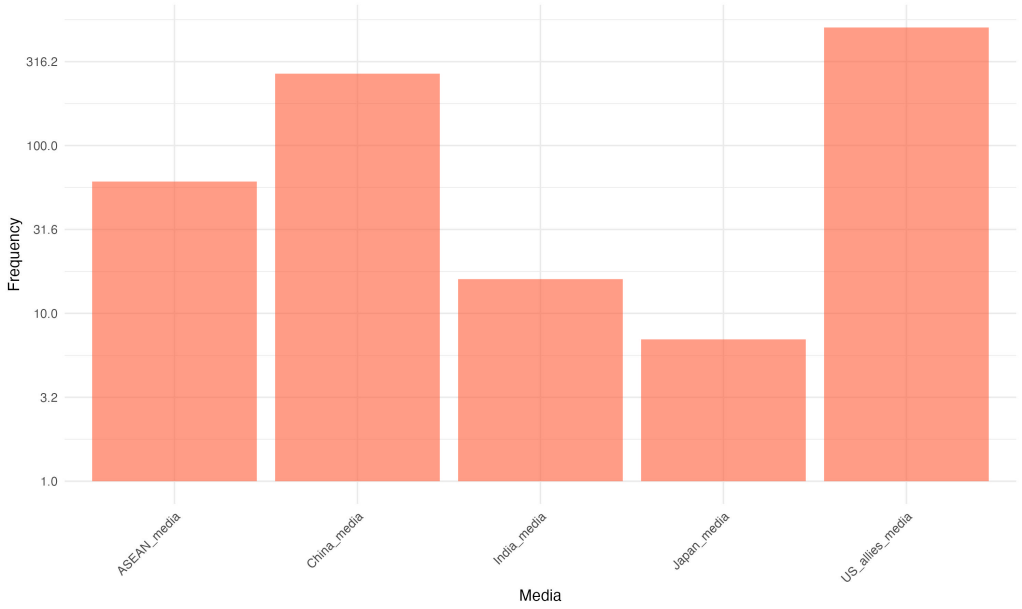
	Topic	Count	Name	CustomName	Representation
1	0	186	0_huawei_data_chinese_internet	Chinese Digital Alternative	['huawei', 'data', 'chinese', 'internet', 'digital', '5g', 'technology', 'companies', 'china', 'cable']
2	1	149	1_bri_cooperation_kazakhstan_road	BRI Digital Cooperation	['bri', 'cooperation', 'kazakhstan', 'road', 'belt', 'countries', 'china', 'initiative', 'xi', 'development']
3	2	124	2_ap_biden_g7_countries	B3W as Alternative to BRI	['ap', 'biden', 'g7', 'countries', 'said', 'developing', 'president', 'funds', 'investment', 'project']
4	3	84	3_g7_b3w_infrastructure_bri	B3W as Alternative to BRI	['g7', 'b3w', 'infrastructure', 'bri', 'initiative', 'plan', 'projects', 'chinas', 'china', 'labour']
5	4	51	4_asean_chinaasean_trade_rcep	China-ASEAN digital trade & RCEP	['asean', 'chinaasean', 'trade', 'rcep', 'china', 'region', 'economic', 'regional', 'asiapacific', 'cooperation']
6	5	51	5_uyghur_rights_human_xinjiang	Technology and HR Abuses in Xinjiang	['uyghur', 'rights', 'human', 'xinjiang', 'central', 'asia', 'asian', 'chinacentral', 'chinese', 'ethnic']
7	6	42	6_arab_chinaarab_expo_states	Ningxia Expo and China-Arab Cooperation	['arab', 'chinaarab', 'expo', 'states', 'ningxia', 'yinchuan', 'cooperation', 'trade', 'shanxi', 'crossborder']
8	7	41	7_central_asian_chinacentral_asia	Central Asia	['central', 'asian', 'chinacentral', 'asia', 'summit', 'countries', 'cooperation', 'xian', 'xi', 'china']
9	8	40	8_big_data_teaching_terminology	Technology Education	['big', 'data', 'teaching', 'terminology', 'education', 'digital', 'summit', 'fuzhou', 'language', 'characters']
10	9	38	9_chinese_bri_debt_financial	BRI Dept Trap	['chinese', 'bri', 'debt', 'financial', 'china', 'loans', 'countries', 'projects', 'road', 'belt']
11	10	38	10_russia_china_united_states	Strategic Competition	['russia', 'china', 'united', 'states', 'military', 'chinas', 'war', 'chinese', 'defense', 'nato']
12	11	35	11_ciie_openingup_trade_import	CIIE Expo and Chinese Tech Export	['ciie', 'openingup', 'trade', 'import', 'opportunities', 'opening', 'xi', 'world', 'china', 'share']
13	12	34	12_china_chinese_party_xi	Xi Jinping and Chinese Communist Party	['china', 'chinese', 'party', 'xi', 'chinas', 'beijing', 'economy', 'economic', 'growth', 'xis']
14	13	32	13_saudi_gulf_arabia_gcc	GCC and China-Gulf Cooperation	['saudi', 'gulf', 'arabia', 'gcc', 'china', 'oil', 'chinese', 'middle', 'trade', 'east']
15	14	30	14_iot_managed_service_industry	IOT and Digital Services	['iot', 'managed', 'service', 'industry', 'wuxi', 'services', 'management', 'db', 'case', 'study']

I use the BERTopic algorithm with standard settings + CountVectorized. Subsequently, I picked the top 15 output topics.

Table 4: The BERTopic sequence and algorithm used at each step

Step	**Algorithm**
Sentence embedding	all-MiniLM-L6-v2
Dimensionality reduction	UMAP
Cluster	HDBSCAN
Tokenize	CountVectorizer
Topic representation	Cluster-specific TF-IDF

Figure 4: Media Counts by Country/Region



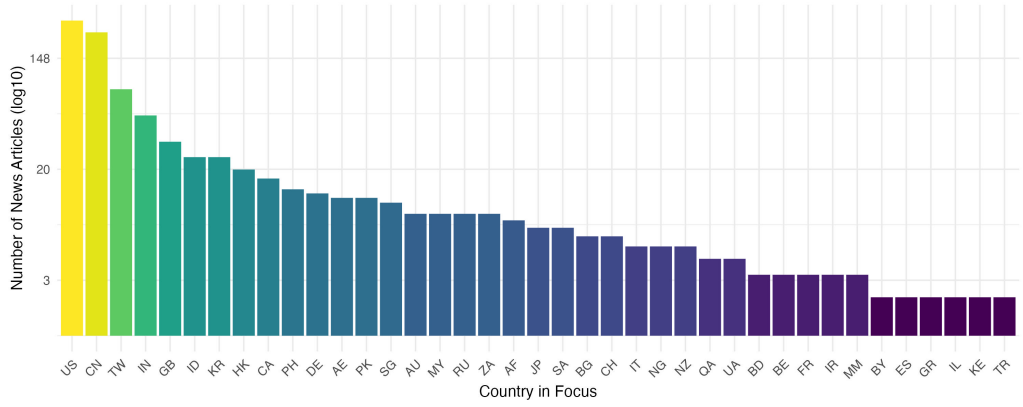


Figure 5: News Articles by 'Country' Variable Distribution (n = 975)

Listing 1: Search Query in NewsCatcherAPI

```
params = [
    {
    'q': 'Digital Silk Road AND DSR',
    'lang': 'en',
    'to_rank': 10000,
    'page_size': 100,
    'from': '3 years ago',
    'page': 1
    },
    {
    'q': 'Digital Silk Road AND China',
    'lang': 'en',
    'to_rank': 10000,
    'page_size': 100,
    'from': '3 years ago',
    'page': 1
    },
    {
    'q': 'Digital Silk Road AND DSR',
    'lang': 'en',
    'countries': 'MY,PH,ID,VN,TH,LA,KH,MM,SG,BN,TL',
    'to_rank': 10000,
    'page_size': 100,
    'sort_by': 'rank',
    'from': '3 years ago',
    'page': 1
    },
    {
    'q': 'Digital Silk Road AND China',
    'lang': 'en',
    'countries': 'MY,PH,ID,VN,TH,LA,KH,MM,SG,BN,TL',
    'to_rank': 10000,
    'page_size': 100,
    'sort_by': 'rank',
    'from': '3 years ago',
    'page': 1
    }
]
```