THE UNIVERSITY OF CHICAGO

When a Chinese Balloon Is Shot Down, Does It Echo?

By

Dylan T. Clairmont

June 2023

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor and Preceptor: Marshall Kramer

Abstract

This thesis employs a computational linguistic methodology to investigate the framing of echo chambers on Twitter, using similarity and difference over time as a proxy. The analysis centers on the 2023 Chinese Balloon Incident, which involved the U.S. military shooting down a Chinese balloon, and examines coalescence or divergence in discourse using string similarity comparisons, sentiment scores, topic modeling, word embedding, and top terms. The thesis aims to problematize existing methodologies and instead focuses on everyday language use to better observe political and social agency in technologized spaces. Through this approach, the study contributes to a deeper understanding of the impact of social media on democratic discourse and the broader political landscape.

Contents

I. Introduction

On February 4, 2023, an object of unknown origin was shot down by the U.S. Air Force off the coast of South Carolina. While the Chinese Government claims that the object was a weather balloon that had drifted off course, the U.S. and Canadian militaries publicly contend that it was used for surveillance purposes. Following the incident, social media was ablaze with tweets and reactions, as is typical for breaking news events in the twenty-first century.

Twitter, in particular, has emerged as a platform of choice for political and economic discourse, with the reinstatement of Jack Dorsey as CEO and Donald Trump's unprecedented use of the platform during his presidential campaign in 2015 propelling it to new heights of prominence. While Trump leveraged Twitter to advance his political agenda, Elon Musk, Twitter's current CEO, has utilized the platform to promote his economic interests. Musk's tweet linking to a Reddit board hyping the stock of GameStop, and the subsequent surge in GameStop's stock value, serves as an example of this (Bursztynsky 2021).

Despite China's domestic ban on Twitter, Musk's economic influence extends into China, where he has a complex relationship that he is keen to maintain. Given the vested interests of powerful figures like Trump and Musk to control certain narratives around China within the U.S., and the contentious nature of the recent balloon incident, it is worthwhile to investigate how Twitter specifically facilitated discourse around this event. Does politics primarily occur at the level of top-down messaging, or is there an iterative process at work, whereby users gradually reshape

discourse and influence ideology? To what extent do actors have influence over the shape of ideological discourse in a virtual space like Twitter?

This paper endeavors to investigate the extent of existence of echo chambers on Twitter by analyzing tweets related to the 2023 Chinese Balloon Incident. Using computational linguistic methods, I examine whether there is a convergence or divergence of language in these tweets, and explore the broader implications of echo chambers for online politics and democracy. Rather than adopting a traditional focus on pre-ordained ideology and network analysis (Nikolov et al. 2015; Hong and Kim 2016; Quattrociocchi et al. 2016; Törnberg 2018; Wang and Qian 2021), this study emphasizes the potential of natural language processing models to uncover patterns of discourse over time. By taking a literal interpretation of echoing as the repetition and distortion of particular words or phrases over a certain period of time, this research offers a novel perspective on echo chambers.

The paper provides background information on several key topics that are fundamental to understanding the research. First, a detailed overview is presented of the 2023 Chinese balloon incident, followed by an examination of Twitter, echo chambers, citizen journalism, and tracking events over time. The present study employs a diverse range of computational linguistic methods to serve as a proxy for analyzing the evolution of textual content over time, with a focus on both syntactic and semantic transformations, including string comparison techniques, diachronic word embedding, top k terms, sentiment analysis, and dynamic topic modeling. The findings of each of these methods are presented and analyzed. Ultimately, this paper contributes to a better understanding of how narratives surrounding breaking news take shape on social media and

offers new insights into the dynamics of online political discourse. In addition, it points to the emergence of a new techno public sphere which has provided everyday users with a platform to engage in discourse around news and reclaim some agency in shaping ideology.

II. Background

A. The 2023 Chinese Balloon Incident

Fig. 1: The approximate route of the balloon

From January 28 to February 4, 2023, a large Chinese airship drifted across U.S. airspace, making its way from Alaska through the continental U.S. (Kutllovci 2023), as shown in Fig. 1. The balloon's trajectory indicated that it could fly over a number of sensitive military sites, raising concerns that it was for surveillance purposes. The incident heightened tensions between the U.S. and China, which have been exacerbated in recent years due to a variety of political and

economic issues, including the COVID-19 pandemic, the trade war, territorial disputes, and human rights concerns in Hong Kong and Xinjiang (Bera 2023).

The role of China in the discourse and policies of the Trump administration was highly contentious, with the former president making China a focal point of his campaign and presidency. The stance taken by the Trump administration contributed significantly to the increased tension and negative rhetoric that characterized the relationship between the United States and China during that time. The extreme nature of this rhetoric was exemplified by a guest on Fox News' Tucker Carlson show who made an incendiary comment, stating that the United States needs a military composed of Type A men who seek to sit on thrones made of Chinese skulls (Media Matters Staff 2021). Such language is highly inflammatory and underscores the virulent nature of the discourse that has characterized the U.S.-China relationship in recent years. Despite some efforts at diplomatic engagement, the relationship between the two countries remains highly strained, and the inflammatory rhetoric of the past continues to reverberate through the current political climate.

The U.S. Air Force eventually shot down the Chinese balloon off the coast of South Carolina on orders from President Joe Biden, which further complicated the already-tense relationship between the two countries and raised concerns about the potential for future incidents to escalate into conflict. Despite the seemingly trivial term "balloon," the object's considerable size, with a substructure roughly 90 feet in length, as shown in Fig. 2, prompted U.S. officials to refrain from shooting it down over land and instead wait until it moved off the coast to avoid collateral damage (Bera 2023). During the week following the incident, several additional high-altitude

objects were observed in the airspace over the United States and subsequently downed by American military forces, although later assessments revealed these objects to be unrelated to China (Kutllovci 2023).

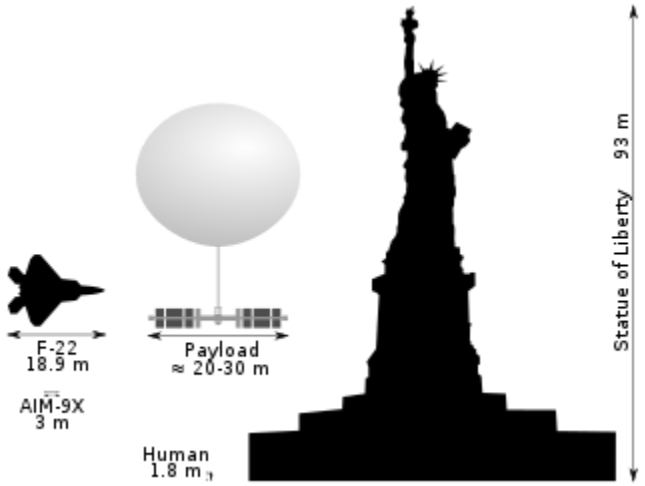

Fig. 2: Estimated size of balloon comparison

The recent Chinese balloon incident has brought to light the potential hazards and consequences of surveillance technologies, which can be employed to monitor and track individuals or groups without their consent or knowledge (Zuboff 2020, 50). In recent years, there has been a growing concern over the emergence of Surveillance Capitalism, a term coined by the scholar Shoshana Zuboff to describe a novel economic system in which data is extracted from individuals and exploited for profit (2020, 7). This incident has led to widespread speculation and discussions about Chinese surveillance practices and their potential to gather intelligence on other countries or individuals. Zuboff's work underscores the power dynamics that are at play in these situations, and the potential for individuals to be exploited for profit or political gain through the collection and use of personal data (Zuboff 2020, 268). In the context of the Chinese balloon incident, the use of surveillance technology by the Chinese government highlights the potential for abuse of power and the need for greater transparency and accountability in surveillance practices.

The Chinese government has maintained that the balloon was a meteorological device that had gone off course, while the American and Canadian militaries maintain that it was for surveillance purposes and illegal by international law (Byers 2023). The incident occurred at a challenging time for China, which is facing a variety of complex political issues and has expressed a desire to annex Taiwan, a move that President Biden has made clear he will not allow (Bera 2023). The event has also drawn comparisons to similar episodes of espionage via balloon that have taken place in Hawaii and Guam in the past (Barnes 2023).

Overall, the incident has highlighted the desire of key actors like Elon Musk to tightly control the narrative in such international incidents, particularly in the age of social media, and in a context where President Biden has taken a rather hands-off approach to his presidency. While the Chinese government has a sprawling propaganda apparatus to control discussion, the United States faces a different challenge in the form of the decentralized and often chaotic nature of social media. As we will explore in the next section, the Chinese airship incident provides a case study in how social media can both amplify and distort the narrative surrounding an international incident, and how this can emerge from a variety of actors.

B. Twitter

The Chinese Balloon Incident presents an opportunity to apply the lens of Surveillance Capitalism to Twitter. Twitter, a microblogging and online social networking service, was launched in 2006 with a prompt asking "What's happening?" (Marwick and boyd 2011). This prompt results in a wide range of tweets, from mundane thoughts on daily life to breaking news.

Co-founder Jack Dorsey chose the name 'Twitter' based on the Oxford English Dictionary definition of the word as "a short inconsequential burst of information, chirps from birds" (Sarno 2011). The length of tweets has been a defining feature of the platform, with an initial maximum of 140 characters that was later doubled to 280 characters in November 2017 (Rosen 2017). However, recent rumors suggest that Elon Musk, who acquired the site in 2022, may eliminate the character limit altogether, potentially altering the site's fundamental structure.

Jack Dorsey, a co-founder of Twitter in 2006, and a brief CEO in 2007, had a significant influence on the platform's development when he was reinstated as CEO in 2015. Dorsey's original objective was to enable communication and connect individuals globally, a goal that has been mostly achieved. While acknowledging that certain negative elements have emerged on the platform, this thesis aims to demonstrate that Dorsey's vision has brought forth positive outcomes as well. In essence, Dorsey's actions have resulted in an unexpected social experiment that has brought about new publics through the use of technologies that promote open and expansive dialogue. Thus, it can be posited that Dorsey's objectives have been largely successful, despite some unintended consequences.

Although Jack Dorsey founded Twitter with the goal of facilitating communication and bringing people together, Elon Musk's vision for the platform differs significantly. In 2022, Musk acquired Twitter in an unconventional and dramatic manner, and since then has been accused of grossly mishandling the site (Clayton and Hoskins 2022). Moreover, Musk's alleged violation of German hate speech laws could result in a fine of up to $33 billion, an amount greater than the company's current worth (Palli 2023). Musk overplayed his hand in acquiring Twitter, and has since lost a

considerable amount of money, with the site now worth less than half of what he bought it for (Palli 2023). As previously noted, Musk has a history of making disruptive claims that challenge economic order, and Twitter has been a platform where many recent crypto trends, such as the proliferation of NFT profile pictures, have gained traction. Additionally, Musk's economic interests in China due to Tesla may motivate him to control the narrative on his platform. Musk's complex character looms over this event and the discourse on his site.

As mentioned in the introduction, an additional prominent figure in Twitter's history is former President Donald Trump. Trump's tweets gained notoriety and attracted a wide range of reactions, from amusement and insight to dismay. People became engrossed with Trump's tweets, and political journalists were drawn to Twitter during his administration since Trump's tweets frequently dominated the daily news cycle. Although it remains to be determined how Twitter will fare under Musk's stewardship, it is evident that Trump's use of the platform had a significant impact on the media landscape (Robertson 2021).

The collection of data from Twitter poses a significant challenge, as the platform is vast, public, and difficult to constrain, let alone identify the subjects of study at any given time (Marwick, 2013; Alshaabi et al., 2021). Thus, the concept of an "echo chamber" on Twitter is brought into question: Can such a chamber exist on a platform that is inherently open? This question will be revisited later. Much of the research on Twitter to date has focused on analyzing hashtags, which offer a popular and straightforward way of mapping discourse. Hashtags create communities around them, providing a cohesive space for producing and exchanging ideas (Bruns and Burgess, 2012; Maireder and Schlogl, 2014; Bonilla and Rosa, 2015; Jackson and Foucault

Welles, 2015; Williams et al., 2015; Storniauolo and Thomas, 2017). However, hashtags are only used in a small fraction of tweets, ranging from 5% to 11% (Suh, Hong, Piroli, & Chi, 2010, as cited in Marwick, 2013), making them an unrepresentative method for mapping discourse. Rather than limiting discourse to clunky, unrepresentative hashtags, this study considers all tweets related to the incident, regardless of whether it fits in neatly or wishes to fit in at all. Through this approach, a contested space of ideology emerges, where various camps, which may or may not interact with one another, represent diverse talking points that fall under the ideological and discursive umbrella of the unmediated internet environment.

The incident discussed in this study goes beyond the realm of online social media discourse. The involvement of key actors with past prejudices and the political and economic implications elevate it to a matter of national security, with implications for national sovereignty. While these actors may perceive themselves as owning Twitter and controlling discourse on it, this perspective raises a critique of top-down ideology dissemination. In this context, it is useful to consider the concept of the echo chamber, which will be explored in detail below.

C. Echo Chambers

The term ‘echo chamber’ is a well-established concept with a rich history dating back to 1842 when it was first coined in its literal sense to denote a room with sound-reflecting walls. In contemporary usage, however, it has acquired several figurative meanings. Cass R. Sunstein is widely recognized for introducing the modern definition of this term in his book, *Echo Chambers*, published in 2001. According to Sunstein, an echo chamber refers to a virtual space where individuals tend to engage solely with like-minded individuals, thereby limiting exposure

to diverse perspectives. Echo chambers have been described in academic literature as spaces that reinforce pre-existing beliefs, leading to the intensification of opinions and expressions (Du and Gregory 2016). The concept of echo chambers has gained immense currency in academic research and serves as a conceptual anchor for investigating a variety of phenomena (Du and Gregory 2016; Hong and Kim 2016; Michailidou 2017; Bouvier 2020; Geiss et al. 2021).

The diverse range of contexts in which echo chambers have been examined has resulted in a range of perspectives regarding their validity. Some scholars, such as Du and Gregory (2016) and Hong and Kim (2016), have found evidence to support the idea that Twitter fosters echo chambers, while others, including Shore et al. (2016), Vaccari et al. (2016), Dubois and Blank (2018), and Geiss et al. (2021), have found this to be impossible or at least improbable. The lack of a standardized definition for the term echo chamber has resulted in varying methodologies for its study, providing researchers with the freedom to define and measure the phenomenon in their own ways. This has allowed for a more quantitative approach to be taken in the present study.

Network analysis is a popular method for studying the existence of echo chambers in social media platforms (Nikolov et al. 2015; Hong and Kim 2016; Quattrociocchi et al. 2016; Törnberg 2018; Wang and Qian 2021). In particular, researchers often use social media connection networks to measure closeness, assign various ideologies to users, and then measure the purity of networks. In addition to these network analysis methods, there are other approaches researchers have used to investigate the phenomenon of echo chambers in social media. These methods include analyzing the content shared on social media, as well as examining the activity and engagement levels of users.

In addition to these network analysis methods, there are also qualitative approaches to studying echo chambers. For example, Bouvier (2020) conducted in-depth interviews with #MeToo movement participants and found that participants experienced a sense of community and support within the movement but also acknowledged that the movement could be an echo chamber, limiting exposure to alternative viewpoints. Similarly, Michailidou (2017) conducted qualitative content analysis of Greek referendum tweets and found that while echo chambers existed, they were not as extreme as some previous studies had suggested.

It is important to note that these various methods have their own limitations and biases, and it is challenging to fully capture the complex and dynamic nature of echo chambers. The challenges of studying echo chambers on social media platforms are compounded by the difficulty of identifying users' ideological preferences, which are often multifaceted and constantly evolving in the online space. While some users may openly self-identify with a particular political or social ideology, others may be less explicit or actively attempt to conceal their views. Moreover, the same user may express different opinions on different topics or in different contexts, making it difficult to assign them to a single ideological category.

Even when users do openly express their ideological preferences, it can be challenging to determine the degree to which they hold those beliefs or how influential they are within their networks. Additionally, users may engage in strategic behavior, such as following or retweeting accounts that represent opposing views, in order to appear more moderate or to gain attention. All of these factors make it challenging to accurately capture and measure users' ideological

preferences on social media platforms, and raise questions about the validity of the methods used to study echo chambers. Despite these challenges, continued research on echo chambers is important as they can have significant implications for public discourse and democracy. It is essential to recognize, however, that the semi-popular construction of ideology is found in the everyday and is inherently iterative and transformative. I aim to demonstrate this by using computational linguistic methods to identify linguistic qualities that reveal whether the online platform promotes creative divergence or whether there is a gradual coming together and echoing of opinions. Ultimately, my goal is to challenge studies that imbue alarming sincerity to online interactions in order to provide a social framework for connections that may be far more casual.

Echo chambers, as commonly understood, suggest the amplification of preexisting ideas (Geiss et al. 2021). It is worth exploring whether a figure with top-down influence like Trump can propagate an idea to such an extent that it is echoed by the masses, resulting in ideological reproduction. However, this phenomenon has yet to be rigorously investigated. To address this gap, a new approach to studying echo chambers is proposed, drawing upon the literal definition of echoing and reverberation. Echoes are not exact repetitions; they involve slight distortions. Specifically, computational linguistic methods are employed to identify instances of linguistic coalescence or divergence. This linguistic analysis provides a unique perspective on the evolution of language and opinion in online political discussions, exploring whether the online platform promotes creative divergence or a gradual coming together and echoing of opinions. This approach moves beyond traditional analyses focused on social network structure and static ideological orientation, instead examining the linguistic characteristics of online political discourse.

At the core of the concept of ‘echo chambers’ lies the closely associated idea of homophily, which suggests that people tend to form connections with those who are similar to them (McPherson et al. 2001). McPherson et al. (2001) discussed the concept of homophily in social networks in their seminal work “Birds of a Feather,” which was published before the advent of Twitter. This raises the question of what sets the internet apart from offline networks, and why we should not expect similar homogeneity in personal networks online, given that people's networks offline tend to be homogeneous with respect to various sociodemographic, behavioral, and intrapersonal characteristics (Granovetter 1973; McPherson and Smith-Lovin 1987; McPherson et al. 2001; Schmitt et al. 2008; Shalizi and Thomas 2011; Susarla et al. 2012). Homophily is known to constrain people's social circles, and this has significant implications for the information they receive, the attitudes they form, and the interactions they experience. Thus, the same fears that underlie the concept of the modern-day, technologized ‘echo chamber’ are embodied in homophily (McPherson et al. 2001).

Sunstein (2001) was one of the first to suggest that social media platforms encourage a lack of exposure to contrary perspectives, which is synonymous with echo chambers. Indeed, social media has been criticized for exacerbating political polarization and encouraging echo chambers (Du and Gregory 2016), with some researchers accusing social media of promoting social extremism (Barbera et al. 2015). There is a widespread belief that social media users are particularly vulnerable to echo chambers since they are able to selectively choose the information to which they are exposed, which can result in self-isolation from viewpoints with which they disagree, leading to a constant positive feedback loop (Vaccari et al. 2016).

Consequently, users are often segregated into self-selecting bubbles that can create an information and knowledge gap (Dubois and Blank 2018). Despite the prevalence of research on echo chambers and polarization, there is still a lack of accurate data to support the claim that these phenomena are taking place on social media beyond superficial measures of homophily, which has existed well before the advent of social media.

This study seeks to explore the possibility that the term "echo chamber" may, in fact, be an instance of its own namesake, wherein a particular bias is repeatedly echoed and perpetuated. However, the bias being examined is not one that necessarily manifests exclusively within virtual spaces, but rather may reflect a broader societal fear and misunderstanding of the potential impact and implications of these spaces. By interrogating this phenomenon, this research aims to provide a nuanced understanding of the role and influence of virtual spaces in shaping contemporary discourse and media consumption.

Despite being often characterized as a threat to democracy (Sunstein 2001; Pariser 2011; Tufekci 2017), the effects of echo chambers are rarely studied in depth. While the establishment of echo chambers is typically based on metrics of homophily, little attention is given to what actually occurs within these echo chambers on a day-to-day basis. This raises the question of whether it is even appropriate to call them echo chambers if there is no evidence of echoing taking place. Furthermore, if we find that language, thought, or opinion are not being stifled, but rather creatively blossoming among like-minded individuals, is it still accurate to label the phenomenon as an echo chamber? The current understanding of echo chambers is limited in this regard and

there is a need for further research to better understand the dynamics and effects of echo chambers in online spaces.

D. Citizen Journalism

Despite the lack of conventional association, it is imperative to incorporate the concept of citizen journalism in conjunction with echo chambers. While the latter embodies many apprehensions concerning the internet, the former provides a sanguine trajectory for leveraging online communication to fortify democracy. If echo chambers have the proclivity to constrict exposure to heterogeneous perspectives and foster societal polarization, citizen journalism has the potential to widen the range of voices and perspectives that are expressed, which can culminate in a more variegated and all-encompassing public discourse. Consequently, examining the interplay between citizen journalism and echo chambers is crucial in comprehending the internet's potential to impede or amplify democratic participation and dialogue.

The internet's ability to transcend nation-state borders poses a challenge to the concept of citizen journalism. The notion of citizenship itself is in question when users are citizens of the internet rather than any particular nation-state. Is this distinction a helpful one? The emergence of this "new citizenship" sheds light on the new public sphere that has arisen through technologized mediated environments. This novel public sphere exists beyond traditional publics, which are often founded on nationalist beliefs and rhetoric (Anderson 1983). Instead, it represents a global citizenship, wherein discourse and ideology are free to flow unencumbered, though we know that censorship laws, copyright, and algorithms complicate this utopian narrative.

Atton (2009) characterizes citizen journalism as a burgeoning phenomenon of user-generated content or homemade media, whereby ordinary citizens produce journalism from their homes. Brigadir (2016) contends that this process entails a transformation of news consumers into producers, whereby citizens become actively involved in the observation, selection, filtering, distribution, and interpretation of events in unprecedented ways. According to Hermida (2010), many scholars have linked the rise of citizen journalism to social media platforms like Twitter. Hermida (2010), Bruns (2011), and Zubiaga (2019) have explored this connection in their work. Hermida (2010) introduces the concept of 'ambient journalism' as a related phenomenon to echo chambers, underscoring how the continuous consumption of news, or 'ambient news,' has evolved into 'ambient journalism' through the increased capacity for individuals to partake in the news production process.

Social media has enabled individuals to frame conversations, set agendas, and disseminate information, thus creating a more level playing field (Lin et al. 2014). As Murthy (2013) suggests, Twitter has led to a significant 'demotic turn,' providing ordinary people with the ability to break news, produce media content, and express their opinions publicly. This newfound autonomy has been harnessed for social causes and activism, as evidenced by its prominent role in the Arab Spring of 2011, #MeToo, the Black Lives Matter movement, and more (Lindgren 2013; Tufekci 2017).

The interplay between citizen journalism and echo chambers raises questions about agency in online discourse. Scholars such as Baker and McEnergy (2015) and Ulbricht (2020) argue that Twitter has the potential to democratize discourse by enabling ordinary users to engage with and

contribute to online discussions. Cha et al. (2012) note that grassroots users can trigger the initial discussion and dissemination of news stories. However, Wihbey (2014) questions the internet's transformative potential for democratizing discourse, pointing to the reassertion of legacy media patterns. Despite the internet's supposed freedom from real-world hierarchies, research has identified the emergence of "opinion leaders" on social media (Bergstrom and Belfrage 2018) and the creation of online social hierarchies through "celebrity practice" (Marwick and boyd 2011). These studies illustrate that political figures and corporations can wield disproportionate influence on Twitter and speak to, rather than with, their followers.

The issue of communication flow in news dissemination is a recurring theme in the literature. Katz (1957) outlines the conventional flow of ideas from radio and print media to opinion leaders, who then disseminate them to less active segments of the population. However, social media has disrupted this traditional model, enabling individuals to create and share content without the need for institutional gatekeepers. Despite this, Bergstrom and Belfrage (2018) contend that established news media still dominate the news content circulated on social media platforms. Given Twitter's algorithmic complexity, the potential for citizen journalism to thrive and promote original, divergent thought remains a question that this thesis seeks to address.

Furthermore, the issue of the visibility of divergent thought on social media platforms such as Twitter cannot be ignored. While there may be a plethora of content available, it is possible that only a select few accounts have a significant impact on others' news consumption (Bergstrom and Belfrage 2018). Consequently, the relevance and significance of free-flowing content on Twitter may be limited if it is not being seen or shared widely. Despite this, prior studies have

tended to focus on user influence and reach (Cha et al. 2010), rather than language and opinion production, which warrants further investigation. In other terms, there has been an overreliance on figuring out methods of transmission that has led to ideology being treated as static, rather than a metapragmatically re/shaped series of statements and claims (Barthes 1972).

It is important to recognize that not all tweets are created equal, and hence, it is necessary to track the development of language and discourse even if it does not have a high level of reach or status. By integrating computational linguistic methods into the conversation, this study aims to contribute to the discourse on the democratization of media ecosystems (Wihbey 2014). The question of whether social media can lead to a more radically democratized media landscape remains open, and this research seeks to provide insights into this topic through an examination of language and opinion production on Twitter.

As a result, this thesis seeks to address the question of whether a new public sphere has emerged, distinct from the public sphere envisioned by Habermas in *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (1962). This new public sphere is characterized by its own centers, focus points, and accessory points, and raises the possibility of reemergent techno-liberalisms. Through the agency of the digital citizen, this emerging public sphere enables discourse and ideological developments to evolve in new and potentially transformative ways. By examining the distinctive features of this new public sphere, this thesis contributes to a deeper understanding of the ways in which technology is reshaping the contours of public life (Fraser 2017). This thesis considers the role of the digital citizen as an active participant in this emerging public sphere, reclaiming agency over evolving discourse and

enabling new forms of political engagement (Bimber et al. 2012). By exploring the contours of this new public sphere, this thesis seeks to illuminate the evolving relationship between technology, politics, and society, offering insights into the complex ways in which digital media are transforming the nature of public life in the 21st century.

E. Tracking Events Over Time

Over the past few decades, there has been a growing body of literature dedicated to digitally tracking events over time. Starting with Allan et al. (1998), this literature has continued to grow and develop up to the present day. While many of these studies focus on tracking news events, there has been a lack of attention paid to the political and social contexts that surround them. Some studies approach event tracking from a purely computational perspective, while others use more specialized approaches informed by journalism or discourse studies (Bouvier 2015; Chen et al. 2020; KhosraviNik 2020; Peng et al. 2021).

In the context of online event tracking over time, Brigadir (2016) and Lin et al. (2014) have approached the topic in a manner similar to the one I have conceptualized. However, their methods and analyses differ from mine, and their studies focus on different aspects of event tracking. Brigadir (2016) looks at qualitatively tracking events through social media platforms such as Twitter, while Lin et al. (2014) examine the ability of social media to level the playing field and empower ordinary individuals to contribute to mass events.

In contrast, my approach seeks to integrate computational linguistic methods into the event tracking process, with a focus on the language and opinions expressed during the events. By

analyzing the language used in social media conversations surrounding events over time, my aim is to gain a deeper understanding of the ways in which public discourse is shaped and how social relations and politics are implicated in this process. This approach has the potential to shed new light on the complex interplay between language, power, and social change in contemporary society.

III. Methods

This study adopted a data collection methodology utilizing the web scraper tool, snscraper in Python, with the aim of extracting all tweets that referenced the term "Chinese balloon" to investigate the discourse surrounding this event on Twitter. The snscraper tool is a robust social media scraper capable of bypassing API limitations and effectively collecting hundreds of thousands of social media contents. The scraping process resulted in a total of 428,789 tweets, and subsequently, a subset of 413,249 tweets was selected for analysis. The subset was chosen to cover the period from when the event broke to the public and the two weeks that ensued, which is a critical time frame for examining the discourse and reactions related to the event.

The utilization of web scraper tools like snscraper in research is increasingly popular, allowing researchers to gather significant amounts of data from social media platforms, such as Twitter, without relying on the platform's API. However, ethical concerns relating to data privacy and ownership, as well as potential data misuse, emerge from this practice. Therefore, it is vital for researchers to consider the ethical implications of using such tools and abide by ethical guidelines when collecting and utilizing data from social media platforms. Adhering to these ethical considerations is paramount for ensuring that research methods are responsible and

respectful of the individuals and communities represented in the data. Ethics is discussed further in Section V.

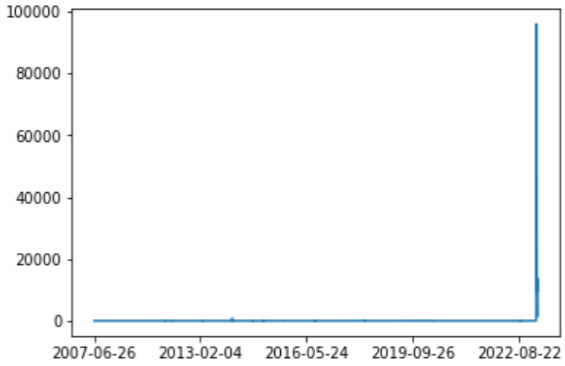

Fig. 3: Num. of tweets that include ‘chinese balloon’ across all time

As demonstrated in Fig. 3, it is evident that the overwhelming majority of tweets containing the term "Chinese balloon" from 2007 to 2023 were posted during the month of February 2023, following the occurrence of the Chinese Balloon Incident. Collecting data from only the month of February 2023 as opposed to all-time data had little impact on the computational results as 99% of the tweets were posted during this period. Notably, on February 1, there were only a handful of tweets containing the term, but this number skyrocketed to over 3,000 on February 2 and to nearly 70,000 by February 3, underscoring the rapidity and magnitude with which this event gained public attention. While it would have been desirable to analyze data beyond the two weeks immediately following the incident, Fig. 4 illustrates that tweet volume rapidly declined in the aftermath of the initial surge and continued to decrease over time.

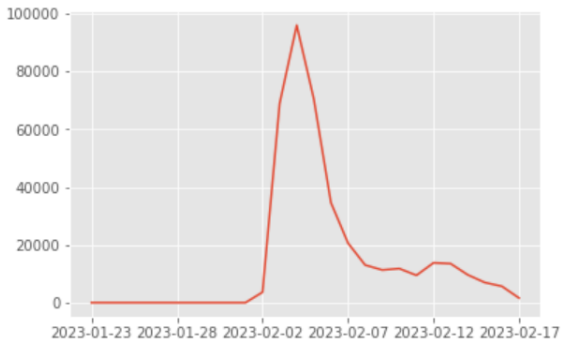

Fig. 4: Num. of tweets that include 'chinese balloon' from ~ mid Jan. to mid Feb. 2023

After acquiring the data, I grouped the tweets by the day on which they were posted. Since the discourse began on February 2, I used that as the starting point and categorized the tweets by day from February 2 to February 16. I then preprocessed the tweets by removing URLs, stop words (common, unnecessary terms), punctuation, and converting all text to lowercase to account for noise in the data. I also tokenized and lemmatized the data. Tokenization involves splitting sentences into lists of words, while lemmatization involves reducing each word to its simplest form to standardize language and account for variations in syntax. For example, "leaves" would become "leaf", "is" would become "be", and "better" would become "good."

To analyze the tweets, a combination of natural language processing techniques were employed. Similarity metrics were calculated to test for similarities between tweets tweeted on each day, and these metrics were compared across days to detect changes in similarity over time. The presence, meaning, and connections of words were also analyzed, providing an interpretive methodology to complement the quantitative approach. The combination of qualitative content analysis and computational methodologies enabled a more comprehensive understanding of the discourse surrounding the Chinese Balloon Incident. Although the social science literature on ideological circulation has long employed qualitative arguments (Baudrillard 1968; Foucault

1969), the availability of quantitative data is now an essential tool for providing a more precise understanding of the scale and scope of these processes.

The techniques employed in this study include: (a) string comparison methods such as Levenshtein and Jaro-Winkler distance measures, (b) diachronic word embedding analysis, (c) identification of top terms, (d) sentiment analysis, and (e) dynamic topic modeling. In the forthcoming section, an introduction to each of the methods will be provided, followed by a detailed exposition of their respective interpretative frameworks in Section IV.

A. String Comparisons

In this study, string comparisons were employed using two distinct methods, namely Jaro-Winkler and Levenshtein. These techniques involve a character-by-character or word-by-word comparison of textual data, allowing for the calculation of similarity scores between tweets. Specifically, the tweets that were tweeted on the same day were compared with one another, and an average similarity score was obtained for each day. This approach enabled an assessment of the degree to which tweets on each day were similar to one another in terms of their syntactic makeup. By examining how these similarity scores evolved over time, it was possible to determine whether there was evidence of increasing homogeneity in the discourse, which may support allegations of an echo chamber effect.

B. Diachronic Word Embedding

Diachronic word embedding is a method developed by Hamilton et al. (2016) that encodes unique words in a high-dimensional semantically-represented vector space. This allows for

comparison of words based on their co-occurrence with other semantically-similar terms. By encoding words as vectors, we can examine the similarity or difference between certain terms using metrics such as cosine distance. This method provides both a quantitative perspective, with scores indicating how similar the terms are, which can then be averaged and compared across dates to determine the prevalence of certain terms on particular days. Additionally, it provides a qualitative lens through which to analyze the associated meaning of these terms.

C. Top K Terms

To ascertain the most frequently occurring terms, I implemented a methodology in which the top k terms were calculated for each day, where k represents the number of terms. Specifically, I selected ten as k for the purpose of my analysis. The percentage of the total text occupied by the top k terms was then determined, and these scores were compared over time. This methodology not only allowed for a qualitative examination of popular terms and their meanings over time, but also provided a quantitative measure of "cohesion" and the level of frequency in the discourse on a daily basis.

D. Sentiment Analysis

Sentiment analysis is a computational linguistic method that quantifies the subjective emotions and attitudes expressed within a given text. The process involves generating a score that reflects the text's positivity or negativity, with a neutral sentiment score of 0, peak positivity represented by a score of 1, and peak negativity represented by a score of -1. In my analysis, I compared the average daily sentiment scores over time and examined the standard deviation of each day's data to determine the amount of variance present. To perform sentiment analysis on the tweets, I

utilized the Valence Aware Dictionary and sEntiment Reasoner (VADER), a rule-based model designed for social media text that is pre-trained and can be applied to unlabeled data. Given the subjective nature of sentiment analysis and the inherent challenge of distinguishing between sincere and ironic tweets, this study acknowledges the potential for errors associated with the use of sentiment analysis as a standalone tool. Therefore, in order to mitigate these limitations, sentiment analysis is employed as one component among several in this investigation.

E. Dynamic Topic Modeling

The final methodology that I employ to analyze my data is the continuous time dynamic topic model (cDTM), which was first introduced by Wang et al. (2012). Topic modeling is a powerful approach for uncovering underlying themes that give structure to a body of text. The most commonly used technique for topic modeling is Latent Dirichlet Allocation (LDA), which involves stochastic updates of topic-document and topic-word distributions to minimize within-topic variance and maximize between-topic variance. The cDTM, building upon LDA, is a model that analyzes similarities and differences in corpora over time, giving a sense of how topic discussion is evolving during a given period. It is an extension of the dynamic topic model (DTM), which is a family of probabilistic time series models developed to analyze the time evolution of topics in large document collections, such as tweets (Blei and Lafferty, 2006). The cDTM offers both quantitative and qualitative insights into the contents of a large document collection, making it an ideal model for my mixed-method analysis. While de Melo and Figueiredo (2021) present a similar model for using topic modeling to track COVID-19 discourse, they use a different analytical framework.

Ultimately, a variety of methods are employed to test the robustness of using similarity metrics like these and to identify patterns that emerge with all of these textual analytical methods. The analysis exclusively focuses on textual Twitter data to apply computational linguistic models. However, it is important to note that certain limitations arise from omitting other media forms such as gifs, images, and emojis, which are integral to communication on Twitter. These limitations should be taken into account when interpreting the findings.

IV. Findings

A. String Comparisons

a. Levenshtein

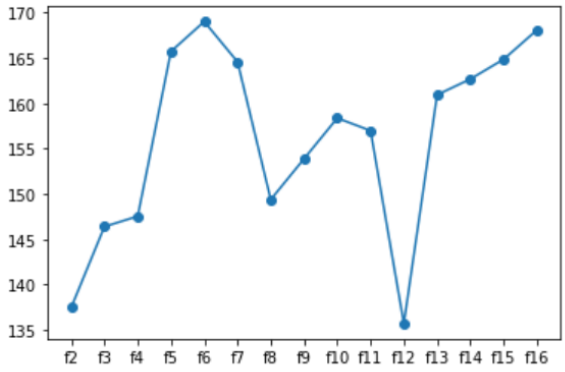

Fig. 5: Levenshtein distance scores over time

According to the Levenshtein distance metric, a higher score indicates greater dissimilarity between texts. Therefore, a visual inspection of Fig. 5 reveals that the average similarity score on February 2 ranged between 135 and 140, while on February 16, it was between 165 and 170. This indicates that, on average, the tweets became increasingly different over time. Notably, there is a spike in the maximum similarity score around February 5, 6, and 7, which coincides with the period when the corpus received the most tweets. Conversely, an anomaly appears on February 12. Explanations for why that may be the case include: news articles or tweets that used

similar wording or phrasing about the Chinese balloon incident were published on February 12, causing a temporary convergence in language; a prominent figure or organization tweeted or commented on the incident using specific language that was picked up and repeated by others, leading to a temporary increase in similarity of tweets; a Twitter algorithm change or bug on February 12 caused tweets to be categorized or recommended using certain keywords or phrases related to the Chinese balloon incident, resulting in an increase in similarity of language across tweets; and, other unrelated events on February 12 could have influenced the language used in tweets about the Chinese balloon incident. Unfortunately, subsequent research did not expose a clear candidate for this anomaly, perhaps suggesting an algorithmic explanation.

b. Jaro-Winkler

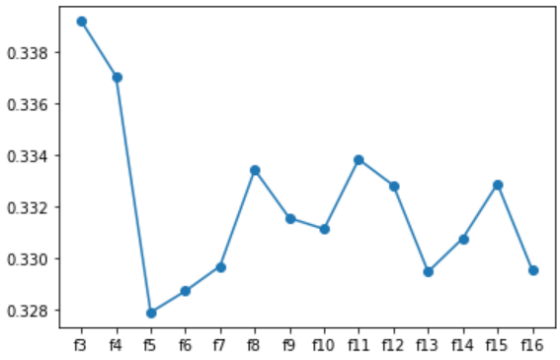

Fig. 6: Jaro-Winkler distance scores over time

The Jaro-Winkler similarity metric differs from Levenshtein in that the lower the score, the greater the difference between the strings, as the score ranges between 0 and 1 and can be interpreted as a percentage of the similarity between the strings. Thus, although Fig. 6 appears to convey a different story than Fig. 5, when imagined in reverse, they share a similar trend, although it is less evident than in the Levenshtein visualization. Notably, we observe a

comparable peak in similarity scores on February 5, 6, and 7, but unlike the Levenshtein visualization, there are no anomalies present in this graph.

B. Diachronic Word Embedding

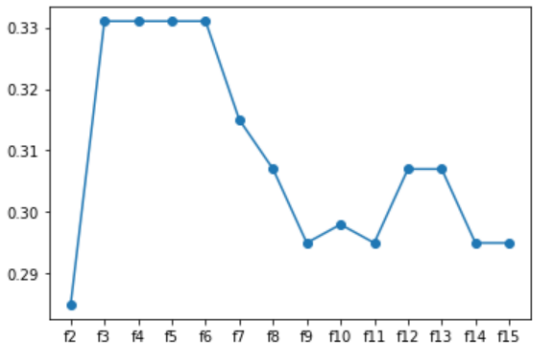

Fig. 7: Average vector scores for top 5 most similar terms to ‘balloon’

In the diachronic word embedding analysis, I focused on examining the word vectors for the term 'balloon' in the corpus. Notably, this approach proved to be more successful in capturing the dynamics of discourse surrounding the Chinese balloon incident than searching for the term 'Chinese'. As depicted in Fig. 7, the results of this quantitative analysis reveal a significant shift in the usage of the word 'balloon' over time. The graph portrays a decreasing trend in the average daily vector scores for the terms most similar to ‘balloon’, which can be interpreted as a decreasing similarity between the term 'balloon' and its related terms. In other words, the lower the score, the greater the variance or dissimilarity, whereas the higher the score, the less variance or more similarity. Notably, the results reveal that the discourse around the event initially exhibited a high degree of variance, as evidenced by the lowest point in Figure 7, but subsequently peaked in homogeneity, resembling an echo chamber. This was followed by a downward trend in similarity, suggesting a diversification in the discourse surrounding the event.

Fig. 8, on the other hand, provides a more qualitative perspective, showcasing how the context of the word 'balloon' evolves and changes over time. In the early days of the incident (February 2-4), the top associated words were generally related to transportation (traffic), energy (carbon, hydrogen), and technical issues (reedtimmeraccu - a meteorologist Twitter account). This may suggest that people were focused on the logistics and technical aspects of the incident, such as how the balloon was powered and how it affected transportation. On February 5-6, the top words shifted to topics like social media (vid), foreign policy (Quincyinst), and politics (Dinesh D'Souza). This may suggest that people were starting to discuss the broader implications of the incident, such as its impact on relations between China and the US.

On February 7-8, the top words shifted again to more general topics like memory (remember), politics (Bullschiff, Zelenski), and pets (dog), perhaps suggesting that public attention on the incident was starting to wane. On February 9-10, the top words included profanity (shit), and topics like defense (Raytheon) and writing (written). This may suggest that people were starting to express frustration with the incident and the response to it. On February 12, humor was introduced (lmfao), suggesting that people were starting to joke about the incident, or that it had become less of a major concern for many. On February 13-15, the top words shifted again to more general topics like security (secure), animals (bird), and family (mother). This may suggest that the incident was no longer a major topic of conversation, or that people were trying to move on from it. Overall, these changes in associated words may reflect the evolving attitudes and perceptions of the public towards the Chinese balloon incident, and may be useful for understanding how events like this are discussed and processed in real-time.

Date	1st Most Similar	2nd MS	3rd MS	4th MS	5th MS
Feb. 2	traffic	buse	keep	north	realli
Feb. 3	depart	carbon	reedtimmeraccu	christian	hydrogen
Feb. 4	appar	consider	squat	tremend	empir
Feb. 5	intern	vid	quincyinst	spycraft	solid
Feb. 6	dineshdsouza	encourag	angie_anson	audienc	dishonest
Feb. 7	rememb	bullschiff	neighbor	crow	went
Feb. 8	ask	maintain	lie	zelenski	dog
Feb. 9	shit	last	buy	green	proudelephantus
Feb. 10	determin	want	raytheon	written	sent
Feb. 11	friend	made	georg	involv	collin
Feb. 12	earth	alright	north	lmfao	phoenix
Feb. 13	issu	classic	secur	qeclova	estim
Feb. 14	lead	day	kid	blame	corner
Feb. 15	agenc	across	crew	bird	mother

Fig. 8: The five most similar terms to ‘balloon’ over the two weeks

C. Top K Terms

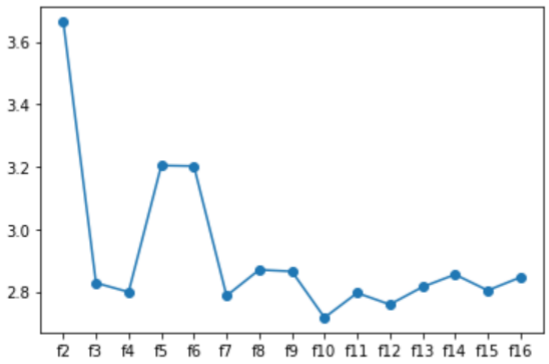

Fig. 9: Top 10 terms as percent of total text over time

The visualization presented in Fig. 9 allows us to interpret that a lower score indicates a lesser amount of text space occupied by the top terms in the entire corpus, leaving room for other terms to surface. The visual depicts a discernible negative trend, indicating that the top terms occupied a significant portion of the corpus on February 2, implying that individuals were initially reacting

to the event without many distinct opinions or thoughts. Moreover, it is crucial to note that while the analysis commenced on February 2, the shooting incident occurred on February 4, and several other significant developments ensued, which may have resulted in an increase in content to discuss in the subsequent days. Consequently, it is plausible that the top k terms could only cover a limited amount of space in the corpus.

Looking at the daily top 10 most common terms from Feb. 2 to Feb. 16 about the Chinese balloon incident (Fig. 10), we can observe a few general trends. There seems to be a shift in focus from the initial sighting of the balloons on Feb. 2 towards a focus on political figures and military action in the following days. The term ‘biden’ appears more frequently starting on Feb. 3 and ‘trump’ also begins to appear on Feb. 5. Additionally, as expected, the terms ‘shoot’ and ‘shot’ become more common in the days following the initial sighting. This suggests that the incident was becoming increasingly politicized and viewed through a lens of potential military action.

Throughout the time period, the term ‘China’ also remains a prominent topic of discussion, suggesting a broader interest in China’s actions and intentions beyond just the specific incident with the balloons. Overall, the top 10 most common terms suggest a complex web of topics and themes surrounding the Chinese balloon incident, including military surveillance, political implications, and broader concerns about China’s actions.

Date	1st Top Term	2nd TT	3rd TT	4th TT	5th TT	6th TT	7th TT	8th TT	9th TT	10th TT
Feb. 2	balloon	chinese	suspected	pentagon	surveillance	spotted	tracking	northern	montana	states
Feb. 3	chinese	balloon	shoot	biden	china	montana	pentagon	balloons	would	suspected
Feb. 4	chinese	balloon	biden	balloons	shoot	shot	china	america	would	military
Feb. 5	chinese	balloon	balloons	trump	biden	shot	three	china	administration	shoot
Feb. 6	chinese	balloon	balloons	trump	biden	administration	china	shot	would	three
Feb. 7	chinese	balloon	balloons	trump	biden	china	shot	administration	military	president
Feb. 8	chinese	balloon	balloons	biden	trump	china	surveillance	shot	military	like
Feb. 9	chinese	balloon	balloons	says	china	biden	surveillance	trump	military	signals
Feb. 10	chinese	balloon	balloons	biden	shot	china	trump	says	alaska	object
Feb. 11	chinese	balloon	balloons	shot	biden	china	object	alaska	entities	like
Feb. 12	chinese	balloon	balloons	shot	biden	china	like	object	shoot	canada
Feb. 13	chinese	balloon	balloons	china	shot	biden	airspace	objects	like	military
Feb. 14	chinese	balloon	balloons	biden	shot	china	military	says	sensors	trump
Feb. 15	chinese	balloon	balloons	china	biden	shot	objects	trump	airspace	military
Feb. 16	chinese	balloon	balloons	biden	china	shot	objects	president	military	weather

Fig. 10: The top ten terms daily

D. Sentiment Analysis

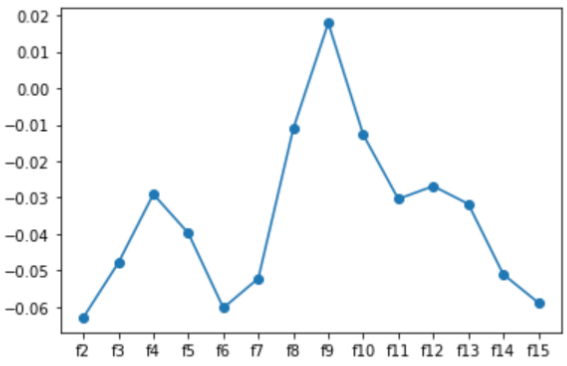

Fig. 11: Average sentiment score over time

Here, I present the results of analyzing the average sentiment score over time. The sentiment score is a measure of the overall positivity or negativity of the tweets, with higher scores indicating more positive sentiment. In Fig. 11, we can observe a peak in positivity on February 9. However, it is important to note that throughout the two-week period, the sentiment scores remain around neutral (0). Thus, it appears that there is not much meaningful information to be gained from this analysis compared to the other methods employed.

E. Dynamic Topic Modeling

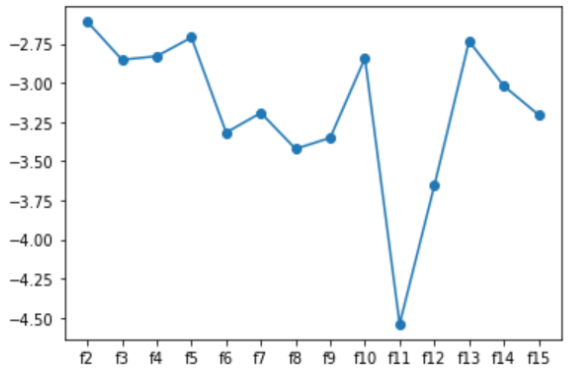

Fig. 12: Topic coherence scores over time

Fig. 12 charts topic coherence scores over time. The interpretation of topic coherence scores is that the more negative the score, the less coherent the topics are, and therefore the more differentiated and less resembling of an echo chamber for each topic. In contrast to the anomaly noted on February 12 in the Levenshtein graph, we observe an anomaly on February 11. These differences suggest that perhaps these findings should not be viewed as synonymous and placed under the same umbrella of simple similarity measurements.

Fig. 13 is an example of the pyLDAvis visualization that accompanied the topic modeling analysis for each day. It includes an intratopic distance map, or how far in similarity/difference each topic is from each other. By observing these differences, we can also get a sense of how varied the discourse is for each day. The visualization also provides a qualitative glimpse into this issue, showing on the right the most common terms for the event that you are currently hovering over, and also how common those terms are in the overall corpus.

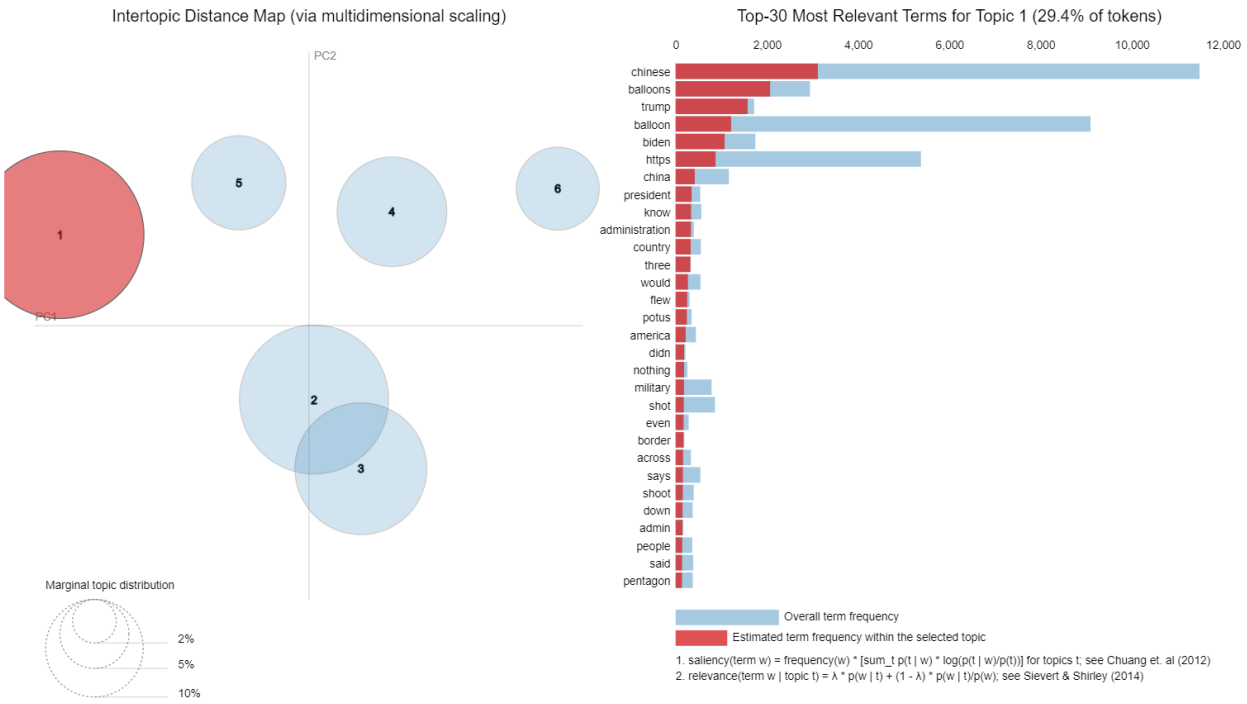

Fig. 13: pyLDAvis visualization for topic modeling for Feb. 8

V. Discussion

This study focuses on examining the dynamics of online discourse pertaining to a particular event. It addresses the issue of echo chambers, which are often characterized as online spaces where discussion becomes stagnant and unproductive. To investigate the possibility of echo chambers, we utilize natural language processing models to measure similarity and dissimilarity in discourse over time. The results reveal that there is no evidence of a general convergence of discourse on Twitter. While the volume of discussion may decrease over time, there is no observable trend towards greater syntactic or semantic similarity. This finding suggests that online discourse can remain diverse and dynamic even in the face of a common event. The study

sheds light on the importance of examining the speed and amplification of discourse online and its potential impact on shaping public opinion and understanding.

This research presents a novel perspective on the concept of echo chambers and raises questions about the validity of existing methods to capture them. However, several limitations can be identified that could be addressed in future work to further strengthen the findings. One of the primary limitations is the challenge of determining whether the observed patterns of similarity and dissimilarity represent meaningful coalescence or simply reflect the inherent variability of discourse on Twitter. Additionally, the focus on the internet as a whole may obscure important differences in sub-communities and the ways in which their language use may or may not coalesce over time. This limitation raises questions about the extent to which the current study captures the phenomenon of echo chambers, which is itself a complex and multifaceted concept. Overall, while the present study provides valuable insights into the dynamics of online discourse, future research could address these limitations to further refine our understanding of echo chambers and their impact on public discourse.

It is essential to consider the nonhuman technological actants that engage with the actors in the analysis, as highlighted by Lewis and Westlund (2014). Additionally, Woolley and Guilbeault (2017) have emphasized the significant role that bots play in Twitter discourses. Therefore, while not explicitly coding for bots in the data, it is important to acknowledge their potential impact on the results and their ability to exacerbate the effects of echo chambers. As such, the role of bots should be considered when interpreting the findings and identifying potential limitations of the study.

It is important to acknowledge the ethical implications of this research, as it is situated in the realm of digital communication and social media. The work draws upon the seminal work of boyd and Crawford (2012), who highlighted the ethical challenges surrounding big data research, including issues of informed consent, privacy, and data protection. In addition, Salganik (2019) has argued that ethical considerations should be an integral part of the research design process, as it is critical to ensure that researchers are aware of the potential risks and benefits of their work.

As such, this research adheres to strict ethical guidelines in order to protect the privacy and confidentiality of the individuals involved in the Twitter corpus. The data has been anonymized and no personally identifiable information has been collected. In addition, this research recognizes the limitations of relying on social media data, as it is a highly curated and mediated environment that may not reflect the broader population. As such, it is critical to consider the representativeness of the data and to be cautious in drawing generalizations from it. Overall, this research is committed to responsible and ethical data practices, and aims to contribute to the ongoing conversations around ethical considerations in digital research.

The present study emphasizes the significance of diverse and dynamic online discourse. The findings suggest that even in the presence of a common event, online discourse can remain heterogeneous and dynamic, which challenges the popular notion of echo chambers. This outcome carries critical implications for democratic societies, where diverse viewpoints and perspectives are vital for informed decision-making and a flourishing public sphere. Furthermore, this research highlights the pivotal role that social media platforms play in shaping

public discourse. These platforms can influence the structure of discourse and potentially amplify certain viewpoints, which could have significant implications for public opinion and political decision-making.

Utilizing natural language processing models to measure similarity and dissimilarity in discourse over time could advance the field of computational linguistics and contribute to the development of more accurate and effective methods for analyzing public discourse. By providing a new perspective on the concept of echo chambers and highlighting the limitations of existing methods to capture them, this research adds to the ongoing efforts to promote healthy and diverse public discourse. The findings also raise important questions about the role of social media platforms in shaping public discourse and the ethical considerations associated with big data research, including issues of informed consent, privacy, and data protection.

The remarks made by elite figures such as Trump and Musk regarding China have been noted as xenophobic (Rogers et al. 2020), suggesting that their economic engagements with China have had a significant impact on their discourse. Despite this, it is important to note that discourse on Twitter was not tightly controlled and did not coalesce around a single talking point. This raises questions about the power of elites to manipulate discourse on the platform, as it appears that they are not the only ones with such influence. In fact, the evolution of Chinese balloon discourse did not necessarily align with the interests of Trump or Musk, despite their economic engagements with China. These findings highlight the need for further investigation into the power dynamics of Twitter and the potential for grassroots movements to challenge elite discourse.

VI. Conclusion

In conclusion, this paper has presented a new methodology for studying echo chambers through the analysis of social media text data using natural language processing techniques. By investigating the 2023 Chinese Balloon Incident on Twitter, the study finds no evidence of coalescence or echoing of language, either quantitatively or qualitatively. While the results are inconclusive regarding the presence of meaningful divergence over time, the study provides a framework for future research that seeks to investigate echo chambers using natural language processing methods. Moving forward, this approach allows us to move away from opaque ideology and focus on everyday language practices as a means of studying the actual effects of echo chambers on online discourse. Through this research, we can track discourse around events over time and gain insights into the political makeup of online conversations. Overall, this study contributes to our understanding of political discourse on social media and the limitations, or lack thereof, that it places on it.

Acknowledgements

I wish to express my sincere gratitude to the individuals who have provided invaluable support and guidance throughout my journey in completing this thesis. I am indebted to my loving and supportive mom, whose unwavering encouragement has been a source of strength and inspiration. I also extend my heartfelt thanks to my peers, Mads Conigliaro-Nguyen, Sean Barill and Tanner Krause, for their invaluable feedback and willingness to join me in this adventure. I am grateful to Sabrina Nardin and Jon Clindaniel for imparting their extensive computational knowledge and for their generous support in guiding me through various aspects of this project. Lastly, I would like to extend my deepest appreciation to my advisor and preceptor, Marshall Kramer, for his guidance, patience, and encouragement throughout my research.

Works Cited

Allan, James, Ron Papka, and Victor Lavrenko. “On-Line New Event Detection and Tracking.” *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '98*, 1998. https://doi.org/10.1145/290941.290954.

Alshaabi, Thayer, Jane L. Adams, Michael V. Arnold, Joshua R. Minot, David R. Dewhurst, Andrew J. Reagan, Christopher M. Danforth, and Peter Sheridan Dodds. “Storywrangler: A Massive Exploratorium for Sociolinguistic, Cultural, Socioeconomic, and Political Timelines Using Twitter.” *Science Advances* 7, no. 29 (2021). https://doi.org/10.1126/sciadv.abe6534.

Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso, 1983.

Atton, Chris. “Why Alternative Journalism Matters.” *Journalism* 10, no. 3 (2009): 283–85. https://doi.org/10.1177/1464884909102582.

Baker, Paul, and Tony McEnery. “Who Benefits When Discourse Gets Democratised? Analysing a Twitter Corpus around the British Benefits Street Debate.” *Corpora and Discourse Studies*, 2015, 244–65. https://doi.org/10.1057/9781137431738_12.

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. “Tweeting from Left to Right: Is Online Political Communication More than an Echo Chamber?” *Psychological Science* 26, no. 10 (2015): 1531–42. https://doi.org/10.1177/0956797615594620.

Barnes, Julian E., Edward Wong, Helene Cooper, and Chris Buckley. "China's Balloons Spy on Countries Around the World, U.S. Officials Say." New York Times, February 9, 2023, A12(L). Gale Academic OneFile (accessed April 26, 2023). https://link.gale.com/apps/doc/A736301474/AONE?u=anon~af626ad4&sid=googleScholar&xid=dec31dd4.

Barthes, Roland. *Mythologies*. Translated by Annette Lavers. New York, NY: The Noonday Press, 1972.

Baudrillard, Jean. *The System of Objects*. London: Verso, 2005.

Bergström, Annika, and Maria Jervelycke Belfrage. “News in Social Media.” *Digital Journalism* 6, no. 5 (2018): 583–98. https://doi.org/10.1080/21670811.2018.1423625.

Bimber, Bruce, Andrew Flanagin, and Cynthia Stohl. *Collective action in organizations: Interaction and engagement in an era of technological change.* Cambridge University Press, 2012.

Blei, David M., and John D. Lafferty. “Dynamic Topic Models.” *Proceedings of the 23rd*

international conference on Machine learning - ICML '06, 2006. https://doi.org/10.1145/1143844.1143859.

Bonilla, Yarimar, and Jonathan Rosa. "#Ferguson: Digital Protest, Hashtag Ethnography, and the Racial Politics of Social Media in the United States." *American Ethnologist* 42, no. 1 (2015): 4–17. https://doi.org/10.1111/amet.12112.

Bouvier, Gwen. "From 'Echo Chambers' to 'Chaos Chambers': Discursive Coherence and Contradiction in the #Metoo Twitter Feed." *Critical Discourse Studies* 19, no. 2 (2020): 179–95. https://doi.org/10.1080/17405904.2020.1822898.

Bouvier, Gwen. "What Is a Discourse Approach to Twitter, Facebook, YouTube and Other Social Media: Connecting with Other Academic Fields?" *Journal of Multicultural Discourses* 10, no. 2 (2015): 149–62. https://doi.org/10.1080/17447143.2015.1042381.

boyd, danah, and Kate Crawford. "Critical Questions for Big Data." *Information, Communication & Society* 15, no. 5 (2012): 662–79. https://doi.org/10.1080/1369118x.2012.678878.

Brigadir, Igor. *From Detection to Discourse: Tracking Events and Communities in Breaking News*. Dublin: University College Dublin, 2016.

Bruns, Axel, and Jean Burgess. "Researching News Discussion on Twitter." *Journalism Studies* 13, no. 5-6 (2012): 801–14. https://doi.org/10.1080/1461670x.2012.664428.

Bruns, Axel. "News Produsage in a pro-Am Mediasphere: Why Citizen Journalism Matters." *News Online*, 2011, 132–47. https://doi.org/10.1007/978-1-137-08100-1_9.

Bursztynsky, Jessica. "GameStop Jumps after Hours as Elon Musk Tweets out Reddit Board That's Hyping Stock." CNBC. CNBC, January 27, 2021. https://www.cnbc.com/2021/01/26/gamestop-jumps-as-elon-musk-tweets-out-reddit-board-thats-hyping-stock.html.

Byers, Michael. "Cooler heads prevailed in the Chinese spy balloon crisis." Globe & Mail [Toronto, Canada], February 7, 2023, A13. Gale Academic OneFile (accessed April 26, 2023). https://link.gale.com/apps/doc/A736106148/AONE?u=anon~9c7589bb&sid=googleScholar&xid=ae927332.

Cha, Meeyoung, F. Benevenuto, H. Haddadi, and K. Gummadi. "The World of Connections and Information Flow in Twitter." *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42, no. 4 (2012): 991–98. https://doi.org/10.1109/tsmca.2012.2183359.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi. "Measuring User Influence in Twitter: The Million Follower Fallacy." *Proceedings of the International AAAI Conference on Web and Social Media* 4, no. 1 (2010): 10–17. https://doi.org/10.1609/icwsm.v4i1.14033.

Chen, Emily, Kristina Lerman, and Emilio Ferrara. "Tracking Social Media Discourse about the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set." *JMIR Public Health and Surveillance* 6, no. 2 (2020). https://doi.org/10.2196/19273.

Clayton, James, and Peter Hoskins. “Elon Musk Takes Control of Twitter in $44bn Deal.” BBC News. BBC, October 28, 2022. https://www.bbc.com/news/technology-63402338.

Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson. “Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data.” *Journal of Communication* 64, no. 2 (2014): 317–32. https://doi.org/10.1111/jcom.12084.

Cossard, Alessandro, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. “Falling into the Echo Chamber: The Italian Vaccination Debate on Twitter.” *Proceedings of the International AAAI Conference on Web and Social Media* 14 (2020): 130–40. https://doi.org/10.1609/icwsm.v14i1.7285.

de Melo, Tiago, and Carlos M Figueiredo. “Comparing News Articles and Tweets about COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach.” *JMIR Public Health and Surveillance* 7, no. 2 (2021). https://doi.org/10.2196/24585.

Du, Siying, and Steve Gregory. “The Echo Chamber Effect in Twitter: Does Community Polarization Increase?” *Complex Networks & Their Applications V*, 2016, 373–78. https://doi.org/10.1007/978-3-319-50901-3_30.

Dubois, Elizabeth, and Grant Blank. “The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media.” *Information, Communication & Society* 21, no. 5 (2018): 729–45. https://doi.org/10.1080/1369118x.2018.1428656.

Esteve Del Valle, Marc, and Rosa Borge Bravo. “Echo Chambers in Parliamentary Twitter Networks: The Catalan Case.” *International Journal of Communication* 12 (2018).

Fraser, Nancy. "The theory of the public sphere: The structural transformation of the public sphere (1962)." In *The Habermas Handbook*, pp. 245-255. Columbia University Press, 2017.

Foucault, Michel. *The Archaeology of Knowledge, and the Discourse of Language*. New York: Pantheon Books, 1982.

Geiß, Stefan, Melanie Magin, Pascal Jürgens, and Birgit Stark. “Loopholes in the Echo Chambers: How the Echo Chamber Metaphor Oversimplifies the Effects of Information Gateways on Opinion Expression.” *Digital Journalism* 9, no. 5 (2021): 660–86. https://doi.org/10.1080/21670811.2021.1873811.

Granovetter, Mark S. "The Strength of Weak Ties." American Journal of Sociology 78, no. 6 (1973): 1360-1380.

Habermas, Jürgen. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Translated by Thomas Burger and Frederick Lawrence, MIT Press, 1991.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016. https://doi.org/10.18653/v1/p16-1141.

Hermida, Alfred. “From TV to Twitter: How Ambient News Became Ambient Journalism.” *M/C*

Journal 13, no. 2 (2010). https://doi.org/10.5204/mcj.220.

Hong, Sounman, and Sun Hyoung Kim. "Political Polarization on Twitter: Implications for the Use of Social Media in Digital Governments." *Government Information Quarterly* 33, no. 4 (2016): 777–82. https://doi.org/10.1016/j.giq.2016.04.007.

Jackson, Sarah J., and Brooke Foucault Welles. "Hijacking #MyNYPD: Social Media Dissent and Networked Counterpublics." *Journal of Communication* 65, no. 6 (2015): 932–52. https://doi.org/10.1111/jcom.12185.

Katz, Elihu. "The Two-Step Flow of Communication: An up-to-Date Report on an Hypothesis." Public Opinion Quarterly 21, no. 1, Anniversary Issue Devoted to Twenty Years of Public Opinion Research (1957): 61. https://doi.org/10.1086/266687.

KhosraviNik, Majid. "Digital Meaning-Making across Content and Practice in Social Media Critical Discourse Studies." *Critical Discourse Studies* 19, no. 2 (2020): 119–23. https://doi.org/10.1080/17405904.2020.1835683.

Kutllovci, Liza. "That's No Moon . . . It's a Balloon!" European View 22, no. 1 (March 30, 2023): 140–42. https://doi.org/10.1177/17816858231164477.

Lewis, Seth C., and Oscar Westlund. "Actors, Actants, Audiences, and Activities in Cross-Media News Work." *Digital Journalism* 3, no. 1 (2014): 19–37. https://doi.org/10.1080/21670811.2014.927986.

Lindgren, Simon. "The Potential and Limitations of Twitter Activism: Mapping the 2011 Libyan Uprising." *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 11, no. 1 (2013): 207–20. https://doi.org/10.31269/triplec.v11i1.475.

Lin, Yu-Ru, Brian Keegan, Drew Margolin, and David Lazer. "Rising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events." *PLoS ONE* 9, no. 5 (2014). https://doi.org/10.1371/journal.pone.0094093.

Maireder, Axel, and Stephan Schlögl. "24 Hours of an #Outcry: The Networked Publics of a Socio-Political Debate." *European Journal of Communication* 29, no. 6 (2014): 687–702. https://doi.org/10.1177/0267323114545710.

Marsili, Neri. "Retweeting: Its Linguistic and Epistemic Value." *Synthese* 198, no. 11 (2020): 10457–83. https://doi.org/10.1007/s11229-020-02731-y.

Marwick, Alice, and danah boyd. "To See and Be Seen: Celebrity Practice on Twitter." *Convergence: The International Journal of Research into New Media Technologies* 17, no. 2 (2011): 139–58. https://doi.org/10.1177/1354856510394539.

Marwick, Alice. "Ethnographic and Qualitative Research on Twitter." *Twitter and Society*, 2013.

McPherson, J. Miller, and Lynn Smith-Lovin. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." American Sociological Review 52, no. 3 (1987): 370–79. https://doi.org/10.2307/2095356.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27, no. 1 (2001): 415–44. https://doi.org/10.1146/annurev.soc.27.1.415.

Media Matters Staff. “Fox Guest: ‘We Need a Military Full of Type A Men Who Want to Sit on a Throne of Chinese Skulls.’” Media Matters for America, December 19, 2021. https://www.mediamatters.org/fox-news/fox-guest-we-need-military-full-type-men-who-want-sit-throne-chinese-skulls.

Merriam-Webster.com Dictionary, s.v. “echo chamber,” accessed January 29, 2023, https://www.merriam-webster.com/dictionary/echo%20chamber.

Michailidou, Asimina. “Twitter, Public Engagement and the Eurocrisis: More than an Echo Chamber?” Social Media and European Politics, 2017, 241–66. https://doi.org/10.1057/978-1-137-59890-5_11.

Murthy, Dhiraj. “Towards a Sociological Understanding of Social Media: Theorizing Twitter.” *Sociology* 46, no. 6 (2012): 1059–73. https://doi.org/10.1177/0038038511422553.

Nikolov, Dimitar, Diego F.M. Oliveira, Alessandro Flammini, and Filippo Menczer. “Measuring Online Social Bubbles.” PeerJ Computer Science 1 (December 2, 2015). https://doi.org/10.7717/peerj-cs.38.

Palli, Praneeth. “Elon Musk's Twitter Slapped with a Whopping Fine over Illegal Content; It's More than the Platform's Net-Worth.” Mashable India, April 11, 2023. https://in.mashable.com/tech/50568/elon-musks-twitter-slapped-with-a-whopping-fine-over-illegal-content-its-more-than-the-platforms-net.

Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin Books, 2011.

Peng, Hao, Jianxin Li, Yangqiu Song, Renyu Yang, Rajiv Ranjan, Philip S. Yu, and Lifang He. “Streaming Social Event Detection and Evolution Discovery in Heterogeneous Information Networks.” *ACM Transactions on Knowledge Discovery from Data* 15, no. 5 (2021): 1–33. https://doi.org/10.1145/3447585.

Quattrociocchi, Walter, Antonio Scala, and Cass R. Sunstein. “Echo Chambers on Facebook.” SSRN Electronic Journal, June 13, 2016. https://doi.org/10.2139/ssrn.2795110.

Robertson, Derek. “How @Realdonaldtrump Changed Politics - and America.” POLITICO, January 9, 2021. https://www.politico.com/news/magazine/2021/01/09/trump-twitter-ban-suspended-analysis-456817.

Rogers, Katie, Lara Jakes, and Ana Swanson. “Trump Defends Using 'Chinese Virus' Label, Ignoring Growing Criticism.” The New York Times. The New York Times, March 18, 2020. https://www.nytimes.com/2020/03/18/us/politics/china-virus.html.

Rosen, Aliza. “Tweeting Made Easier.” Twitter. Twitter, 2017. https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier.

Salganik, Matthew J. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press, 2019.

Sarno, David. “Twitter Came to Life Five Years Ago This Week; Creator Jack Dorsey Remembers.” *LA Times*, March 13, 2011.

Schmitt, David P., Anu Realo, Martin Voracek, and Jüri Allik. "Why can't a man be more like a

woman? Sex differences in Big Five personality traits across 55 cultures." Journal of personality and social psychology 94, no. 1 (2008): 168.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. “Homophily and Contagion Are Generically Confounded in Observational Social Network Studies.” Sociological Methods & Research 40, no. 2 (May 17, 2011): 211–39. https://doi.org/10.1177/0049124111404820.

Shore, Jesse Conan, Jiye Baek, and Chrysanthos Dellarocas. “Network Structure and Patterns of Information Diversity on Twitter,” 2016. https://doi.org/10.31235/osf.io/32urj.

Stornaiuolo, Amy, and Ebony Elizabeth Thomas. “Disrupting Educational Inequalities through Youth Digital Activism.” *Review of Research in Education* 41, no. 1 (2017): 337–57. https://doi.org/10.3102/0091732x16687973.

Sunstein, Cass R. *Echo Chambers: Bush v. Gore, Impeachment, and Beyond*. Princeton, NJ: Princeton University Press, 2001.

Susarla, Anjana, Jeong-Ha Oh, and Yong Tan. “Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube.” Information Systems Research 23, no. 1 (April 8, 2012): 23–41. https://doi.org/10.1287/isre.1100.0339.

Törnberg, Petter. “Echo Chambers and Viral Misinformation: Modeling Fake News as Complex Contagion.” PLOS ONE 13, no. 9 (September 20, 2018). https://doi.org/10.1371/journal.pone.0203958.

Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press , 2017.

Ulbricht, Lena. “Scraping the Demos. Digitalization, Web Scraping and the Democratic Project.” *Democratization* 27, no. 3 (2020): 426–42. https://doi.org/10.1080/13510347.2020.1714595.

Vaccari, Cristian, Augusto Valeriani, Pablo Barberá, John T. Jost, Jonathan Nagler, and Joshua A. Tucker. “Of Echo Chambers and Contrarian Clubs: Exposure to Political Disagreement among German and Italian Users of Twitter.” *Social Media + Society* 2, no. 3 (2016): 205630511666422. https://doi.org/10.1177/2056305116664221.

Wang, Chong, David Blei, and David Heckerman . “Continuous-Time Dynamic Models.” *Time Series Modeling of Neuroscience Data*, 2012, 136–67. https://doi.org/10.1201/b11527-9.

Wang, Dandan, and Yuxing Qian. “Echo Chamber Effect in Rumor Rebuttal Discussions about COVID-19 in China: Social Media Content and Network Analysis Study.” Journal of Medical Internet Research 23, no. 3 (March 25, 2021). https://doi.org/10.2196/27009.

Wang, Vivian, and Joy Dong. "China Uses Censorship and Memes to Downplay Balloon Dispute." New York Times, February 8, 2023, A4(L). Gale Academic OneFile (accessed April 26, 2023). https://link.gale.com/apps/doc/A736187390/AONE?u=anon~fb434040&sid=googleScholar&xid=f81a2ba9.

Wihbey, John. “The Challenges of Democratizing News and Information: Examining Data on Social Media, Viral Patterns and Digital Influence.” *SSRN Electronic Journal*, 2014. https://doi.org/10.2139/ssrn.2466058.

Williams, Hywel T.P., James R. McMurray, Tim Kurz, and F. Hugo Lambert. “Network Analysis Reveals Open Forums and Echo Chambers in Social Media Discussions of Climate Change.” *Global Environmental Change* 32 (2015): 126–38. https://doi.org/10.1016/j.gloenvcha.2015.03.006.

Woolley, S, and D Guilbeault. 2017. “Computational Propaganda in the United States of America: Manufacturing Consensus Online.” Edited by S Woolley and P Howard. Computational Propaganda Worldwide. Computational Propaganda Project.

Zubiaga, Arkaitz. “Mining Social Media for Newsgathering: A Review.” *Online Social Networks and Media* 13 (2019): 100049. https://doi.org/10.1016/j.osnem.2019.100049.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, NY: PublicAffairs, 2020.