PhD Thesis

Addressing Bias and Oversimplification in Measurements of Political Polarization

by

Yijing Chen

Submitted to
Central European University
Department of Network and Data Science

In partial fulfillment of the requirements for the degree of Doctor of Philosophy in Network Science

Supervisor: Prof. Elisa Omodei
Co-supervisor: Prof. Gerardo Iñiguez

Vienna, Austria
2025

For Shingle

Declaration of Authorship

I, the undersigned, Yijing Chen, candidate for the PhD degree in Network Science declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, July 1, 2025

Yijing Chen e-signature on July 1, 2025

Signature

Copyright Notice

[1] Icon by Font Awesome

Abstract

Advances in digital technologies have fundamentally reshaped how individuals participate in political communication. Today, the production, circulation, and (mis)interpretation of political information unfold within a sociotechnical system that is increasingly interpersonally networked, algorithmically curated, and infused with AI-generated content. Thus, challenges such as political polarization are now embedded in a more complex media environment, which calls for renewed examinations of classic theories, the continuous adaptation of methods to new datasets, and critical reflections on how key concepts like political polarization are defined and measured.

Motivated by this growing complexity, my dissertation seeks to operationalize–and critically reflect on the current operationalizations of–classic theories (i.e., intermedia agenda setting) and social constructs (i.e., political polarization) in today's information ecosystem. Specifically, it examines the stability of a classic theory–intermedia agenda setting (IAS)–in a fragmented media environment, and addresses three key limitations of existing polarization research: (1) the conflation of attitudinal and behavioral measures, (2) the reliance on linear, uni-dimensional metrics in modeling political beliefs, and (3) the measurement bias introduced by focusing exclusively on online engagement data that is easily retrievable via public APIs.

The dissertation consists of one conceptual chapter and three empirical chapters. Chapter 1 establishes the theoretical framework, delineating the key developments in political communication in the contemporary media environment and articulating how polarization is conceptualized and operationalized from attitudinal and behavioral perspectives. Chapter 2 assesses the stability of IAS theory by analyzing agenda alignment across different types of news media during the 2016 and 2020 U.S. presidential elections, revealing the increasing divergence and the shifting IAS roles of different media. Chapters 3 and 4 turn to political polarization and focus on the attitudinal and behavioral perspectives, respectively. Chapter 3 introduces a novel application of Response-Item Networks (ResIN) to measure polarization using attitude data, modeling belief systems as statistical networks of interconnected beliefs. This approach captures ideological polarization as a structural transformation in belief systems and provides nuanced, multi-level metrics. Chapter 4 reflects on the current behavioral measurements of polarization that rely on political engagement documented by public platform APIs (i.e., visible engagement). Using a combined dataset of survey responses and user-donated YouTube traces from Hungary, the analysis reveals how a sole focus on visible engagement can distort our understanding of the broader ideological landscape.

Overall, this dissertation bridges network science, computational social science, and polit-

ical communication through conceptual clarification, methodological innovation, and empirical reflection. The main contributions include: (1) clarifying the conceptual distinction between attitudinal and behavioral polarization, enabling more consistent interpretation across studies; (2) expanding belief network analysis (BNA) through the use of ResIN, offering new tools for visualizing ideological divides and quantifying belief system polarization with greater granularity; (3) introducing a novel framework to address measurement bias in behavioral polarization studies by integrating digital traces and survey responses; and (4) advocating for a more globalized approach to polarization research through analyzing data from the U.S., Hungary, and other European countries.

List of publications

Included as dissertation chapters (first-author papers):

[1] **Chen, Y.**, Liu, Y., Singh, L., & Budak, C. (2024). Intermedia Agenda Setting during the 2016 and 2020 US Presidential Elections. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 18, pp. 254-275).

[2] **Chen, Y.**, Speer, A., de Bruin, B., Carpentras, D., & Warncke, P. (2025). A ”Broken Egg” of U.S. Political Beliefs: Using Response-Item Network (ResIN) to Measure Ideological Polarization. OSF preprint https://doi.org/10.31219/osf.io/autkb. [under the 2nd round of review]

[3] **Chen, Y.**, Kmetty, Z., Iñiguez, G., & Omodei, E. (2025). The Public that Engages Invisibly: What Visible Engagement Fails to Capture in Online Political Communication. Communication Methods and Measures, 1-19.

Included as supporting works (collaborations):

[4] Warncke, P., **Chen, Y.**, Speer, A, de Bruin, B., Lüderse, A, Carpentras, D, (2024) ResIN: A New Method to Analyze Social Belief Systems. [Accepted for publication in the Springer book Computational Social Science of Social Cohesion and Polarization] OSF preprint: https://doi.org/10.31219/osf.io/p3x2n_v1.

[5] Baqir, A., **Chen, Y.**, Diaz-Diaz, F., Kiyak, S., Louf, T., Morini, V., Pansanella, V., Torricelli, M. and Galeazzi, A. (2025). Unveiling the drivers of active participation in social media discourse. Scientific Reports, 15(1), 4906.

Acknowledgment

It's been a fun ride of four years with the help of many.

First, my supervisor team with diverse expertise: thanks Elisa Omodei for the guidance throughout the program, for fostering a research group that provided regular feedback and companionship, and for all the loud and quiet support in many stages of my work. Thanks Gerardo Iñiguez for always being inspiring and open for many potential directions to explore and for granting me the autonomy I needed. While their names are not listed as official supervisors, I'm also grateful for the support I received from Júlia Koltai and Zoltan Kmetty, for their hard work in the data donation project, a critical stepping stone that shapes my dissertation topic. I extend my appreciation to all students, postdocs, staff, and faculty members at DNDS, especially my lovely cohort peers (Onkar, Sina, Julia S, and Jan), D305 office mates (Elsa, Gabriela, Adriana, Leo DG, Onur, and Bojan), and research group members (Elisa, Clara, Julia S, Timur, and Andrea). Thanks Onkar and Luka for putting together the step-by-step tax declaration guidance that made my tax report process much less painful. Thank you all for the salad lunch plans, the office ball games, and some of my earliest bouldering training, although I remain quite unambitious in climbing.

I certainly wouldn't have such a list of output without my collaborators: Ceren Budak, without whom I wouldn't have developed my methodological interest in network science, and thus wouldn't have applied for this PhD program. Many thanks to Ceren, Lisa and Yaguang for their valuable contribution in the intermedia agenda setting project. A sincere shoutout to my ResIN team I met at Bremen BIGSSS in 2023: Philip, Dino, Anne, and Bart. Thanks so much for proposing this fascinating idea of ResIN and helping me shape it further into the broken-egg paper, which becomes an essential chapter of this dissertation. Thanks also to another great team of mine I met at the complexity72 workshop in Mallorca–Gale, Maddalena, Valentina, Virginia, Thomas, Sercan, Fernando and Anees–for the interesting project and for the great time we had in the summer heat. I also had the privilege to visit the Center for Information Networks and Democracy, University of Pennsylvania, and I thank Sandra González-Bailón, Yphtach Lelkes, and the CIND members for generously sharing their time, knowledge and perspective.

There are, of course, many others forming an essential part of my support network. Thanks to trainers and class members I met at the Boxing Sisters Vienna and Penn Grad Boxing for all the boxing sessions that maintained my sanity. Thank you, Jordan, Hillary, Lara, Jill and Greg for the sweetest dog pictures and holiday greetings from Michigan. Special thanks to my online friends for their quiet presence, as well as their small but steady interactions with my digital fragments. I'm also deeply grateful to friends I first met offline but now stay in touch with online: Jiaying, for the most random conversations; Jinjin, for the enduring trust;

Yichun, for being such an inspiring young lady. Thanks Lu for cheer-leading along the way and staying courageous despite the setbacks in work. Thanks Yukun for your enthusiasm in launching the global social network scientist initiative. Lastly, thank you, Daniel, for being incredibly kind, understanding, and patient with my constant anxiety.

I was lucky to be born in a southeastern town in China at a time when access to the Internet was just beginning to transform the lives of many. One of my earliest memories of the Internet traces back to a sleepless night: I walked out of my room and saw my mom sitting in front of the PC in the living room corner, perhaps talking to someone or browsing something, her face lit up by the blue glow of the screen. I only realized much later that, in many moments of such, my mom and I shared the urge to look outward, to seek distant voices, and to reach beyond our immediate lives for a bit more freedom and choice. This dissertation is therefore dedicated to my mom, using her online name, "Shingle", which means a small rounded pebble by the sea.

May we all find a solid shore of selfhood as we feel and greet the palpable waves in an age of unrest.

Contents

List of Figures

List of Tables

Chapter 1

Contemporary media environment and political polarization

1.1 Introduction

Deliberation should not be confined to constitutional conventions, Supreme Court Opinions, or their theoretical analogues. It should extend throughout the political process–to what we call the land of middle democracy. The forums of deliberation in middle democracy embrace virtually any setting in which citizens come together on a regular basis to reach collective decisions about political issues–governmental as well as nongovernmental institutions. (12)
– Gutmann and Thompson, Democracy and Disagreement (1996)

Deliberative democracy relies on efficient political communication that meaningfully engages the public in discussions of social issues [1, 2, 3, 4]. While such communication once occurred primarily through real-time, in-person dialogues such as citizen forums, town halls and conventions, it is now often mediated by online platforms that connect distant users through synchronous and asynchronous threads of discussions, constituting today's virtual territory of what Gutmann and Thompson termed as "the land of middle democracy" [2]. As the usage of the Internet and smartphones has risen globally over the past two decades [5, 6], political communication has become deeply embedded in a sociotechnical system that is interpersonally networked, algorithmically curated, and increasingly populated by artificial intelligence (AI)-generated content. Debates over military aid to Ukraine amid its conflict with Russia may unfold within the friend circle we connect with. Election campaigns, op-eds, and messages from elected officials may surface in algorithmically recommended "for you" feeds, readily available for real-time engagement through reposting, liking or commenting. Political events are being reported, circulated and digested in public discourses via text, image, audio and video–human or AI-generated–across various mobile applications and websites, all of which jointly build a multimedia environment of growing complexity. For political communication researchers, such complexity introduces new challenges: long-standing concerns–such as hate speech [7], mis-/dis-information [8], and political polarization [9]–are now entangled with new sociotechnical architectures that demand updated analytical approaches.

Motivated by this agenda, my dissertation is broadly situated in the interdisciplinary area of political communication research (with a topic focus on political polarization) and computational social science (with a methodological focus on network science). Following a background project investigating the intermedia agenda setting (IAS) dynamics of political news among mass media, the main body of my PhD research focuses on a specific challenge in modern democracies–political polarization. Building on existing studies of political polarization across multiple disciplines, my work highlights the bias and oversimplification of its current measurements using new datasets (i.e., combined survey and digital traces in Hungary) and fresh models (i.e., response-item networks).

My dissertation aims to address three key limitations in existing studies of political polarization. First, there is a lack of distinction between the attitudinal and behavioral conceptualizations of political polarization. While scholars in political science and sociology strive to delineate a detailed taxonomy for various types of polarization from the attitude perspective [e.g., 10, 11], many computational social science works take on a rather coarse approach in defining and measuring polarization, umbrellaing various forms of behavioral clustering as a polarized state [11]. Second, when measuring polarization based on attitudinal data, survey-based studies often rely on unidimensional and linear measures, overlooking the multidimensional nature of ideology as a complex system of beliefs and failing to capture the non-linear, non-monotonic inter-dependencies among political beliefs. Although belief network analysis (BNA) offers a more nuanced perspective, its application to polarization research remains limited, with existing models not yet adequately capturing the spatial dividing lines in an ideological space. Third, when studying polarization through online behaviors, many analyses focus exclusively on engagement with retrievable data points through public application programming interface (API) (e.g., posting, commenting), while neglecting engagement that is not well documented and less available to researchers (e.g., viewing, searching). Thus, our observations of online political engagement are often restricted to a small subset of highly engaged users within their political engagement that is visible to researchers. However, online political participation takes numerous forms, and a sole focus on visible engagement may introduce measurement bias for political polarization.

Through addressing these limitations, my dissertation makes the following key contributions. First, through distinguishing between the attitudinal and behavioral perspectives in measuring polarization, the theoretical chapter clarifies the often conflated concepts of political polarization, which has been operationalized through a variety of measures with different datasets. Such a distinction is not only conceptually valuable–allowing researchers to pinpoint what their context-specific measures actually capture–but also practically indispensable for interpreting, organizing, and synthesizing existing findings, as well as identifying potential gaps or discrepancies across studies for future investigations.

Second, my work on response-item networks (ResIN) expands the currently scarce applications of belief network analysis (BNA) in polarization studies, which enriches the understanding of political polarization through a complex-system view with network-based measures. By applying ResIN to the analysis of political polarization, I demonstrate its capacity in visualizing the ideological divide and quantifying both the system-level and attitude-level shifts during the process of ideological polarization. In doing so, I argue that this multidimensional and nonlinear approach offers unique insights that are not easily captured by traditional measures.

Third, my investigation on measurement bias offers a critical reflection on existing measures of polarization that rely solely on digital traces available via public platform APIs. By comparing the outcomes of various polarization measurements across different types of engagement data, this study introduces a novel framework for addressing measurement biases that arise from the use of limited behavioral data, a task that was previously challenging when relying solely on digital traces retrieved via public APIs. This work underscores the benefits of utilizing combined datasets of digital traces and surveys, and incorporating diverse engagement metrics when measuring polarization. Through this work, we motivate future researchers to more concisely scope their findings within the lens of observation (e.g., polarization in terms of what form of engagement?) and to explore the nuanced variation of this phenomenon across a broader range of engagement data.

Fourth, through analyzing a user-donated dataset from Hungary and publicly available survey data from both the U.S. and European countries, my work advocates for a more globalized vision of polarization studies, an endeavor that deserves more research attention in a field still dominated by a U.S.-centric focus [12]. While polarization is often studied in the context of a binary partisan system, with two opposing poles representing the dominant ideological divide, this phenomenon also exists in settings shaped by different cleavages (such as the anti-/pro-government divide in Chapter 4), or in multiparty political systems that structure distinct belief systems (as demonstrated in my supporting work [13]). Therefore, understanding how polarization manifests and evolves in these varied contexts not only broadens the current lens that is largely confined to a single ideological divide, but also deepens our understanding of the fundamental mechanism of polarization across different political contexts.

Overall, my work aims to operationalize–while also reflecting on the current operationalizations of–classic theories (e.g., intermedia agenda setting) and social constructs (e.g., political polarization) in the context of today's information environment where data generation and collection are deeply embedded in complex sociotechnical processes. Drawing on di-

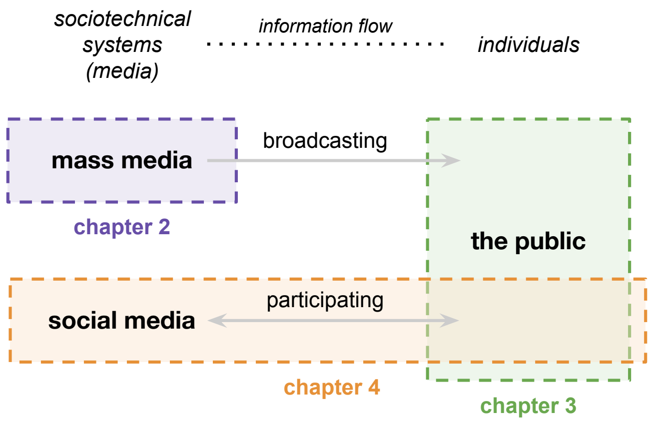

Figure 1.1: A summary graph for the main entities of political communication involved in this thesis, and the focal entity for each chapter.

verse datasets, including news headlines, survey responses, and digital traces, I employ a broad range of analytical approaches that seek to balance the trade-off between achieving computational efficiency and preserving contextual nuances. Through this methodological pluralism, I demonstrate how political communication research–and polarization research in particular–can benefit from the complementary strengths and perspectives of richer methods and more diverse data sources. In particular, my methodological emphasis on network science further bridges the field of complex system with the study of polarization. The ResIN study, for instance, showcases the benefit of modeling belief systems as interconnected networks, specifically that such a network-based approach can deliver fresh observations of belief system dynamics during ideological polarization and enable novel, multi-level quantitative metrics.

1.1.1 Chapter summary

The remainder of this dissertation is structured as follows. Following this brief introduction, Chapter 1 establishes the theoretical foundation of the entire dissertation. I begin with broadly describing political communication in today's media environment and the challenges that come along, motivating an ongoing investigation of political polarization in this complex ecosystem. Then, I summarize how polarization as a social construct has been conceptualized and operationalized across various research disciplines. Specifically, I categorize these measures into attitudinal and behavioral categories, articulate the specific aspects being measured in each of these perspectives, and highlight limitations in these existing measurements.

From Chapter 2 to Chapter 4, I present three empirical projects addressing existing gaps in the studies of political news agenda dynamics and polarization measurements.

To start with, Chapter 2 serves as a background work [14] outlining the information landscape shaped by mass media in a high-choice media environment of growing fragmentation and polarization [4]. Based on a case study of 2016 and 2020 U.S. presidential elections, this chapter looks into the intermedia agenda setting (IAS) dynamics–the process of different news media influencing the agenda of each other–among mass media domains, identifying the main agenda setters in candidate-related news reports and capturing the shifting roles of different media types over the four-year course from 2016 to 2020.

Next, Chapters 3 and 4 showcase the projects that address the limitation in measurements based on attitudinal and behavioral data, respectively. Chapter 3 focuses on the attitudinal perspective and addresses the limitations of unidimensional and linear methods commonly adopted in existing polarization measurements based on attitudinal data. Polarization can happen through sorting [15] across multiple political beliefs, even without a sharp division on a single issue dimension. Here, I emphasize the multidimensional nature of political belief systems, as suggested by Converse [16], and apply a novel belief network analysis (BNA) approach–response-item network (ResIN)–to model the process of ideological sorting. Using American National Election Studies (ANES), I visualize the sorting process as "an egg being broken"–"the transformation of an initial, amorphous, or egg-shaped belief system into a single, polarized, left-right ideological dimension." [17] Using ResIN, I also propose network-based metrics to evaluate the overall polarization level and the structural role of individual beliefs in a polarized system.

In Chapter 4, I turn to the behavioral perspective and address the bias in polarization measurement due to limited behavioral datasets retrieved from social media platforms, i.e., the observed online political behaviors may capture a skewed representation of the broader ideological landscape. What researchers usually observe on social media–politically active users on specific platforms and their political engagement captured by digital traces available in public APIs (i.e., visible engagement)–may not comprehensively cover the underlying ideological composition of a broader population, or depict a complete picture of political engagement of various kinds. I back up this claim through a work that analyzes a combined dataset of survey responses and user-donated YouTube data. In this work, I show how relying only on visible engagement (e.g., commenting) and excluding invisible engagement (e.g., viewing, subscribing), would lead to a biased outcome that paints a more segregated landscape of political communication on YouTube [18].

Finally in Chapter 5, I present the conclusions and discuss several promising avenues for future work related to the two chapters on polarization.

1.2 (Mediated) political communication in the contemporary media environment

Political communication–the process through which individuals exchange information related to politics through some form of interaction–serves as a cornerstone for deliberative democracy [1, 2, 3, 4]. In a rapidly evolving media environment where political communication is no longer confined to face-to-face dialogues and increasingly mediated by a variety of media [19], analyzing (mediated) political communication–and addressing the challenges within–entails understanding both the individuals who participate and the channels that afford such communication. Historically, mass media (e.g., newspapers, television, and radio) have been dominating the public sphere as agenda setters of political discourse [20]. In the recent two decades, such domination has been challenged by the rise of social media platforms [21], which provide virtual infrastructures for online political communication and enable more decentralized, interactive and participatory forms of political engagement [22]. In this section, I outline key developments in political communication in recent times, with special attention to how these developments may relate to political polarization–a challenge that constitutes the focus of later chapters.

To start with, it is helpful to clarify how political communication is afforded in distinct ways by mass media and social media. Let us recall, for instance, the spring of 2022, when Russia launched its military attack on Ukraine. One might have become aware of this political event by skimming news headlines on Guardian's homepage, listening to discussions of the war timeline in a podcast hosted by a columnist they follow on Twitter/X, watching a weekly news digest recommended by YouTube's algorithm, or viewing Instagram stories shared by a friend expressing their strong opinions about the conflict and international aid. In this example, mass media (e.g., news websites, podcasts) and social media (e.g., Twitter/X, YouTube, Instagram) coexist and intertwine in one's day-to-day experiences of political communication; yet, they operate through different channels and afford different types of political engagement. Through mass media, political information typically flows in a one-way direction from content producers to content consumers, uniformly broadcast to a broad audience who mostly consume passively with no instant feedback loop. Social media, in contrast, provides infrastructures for two-way information flows–and thereby blurs the divide–between content producers and content consumers, facilitating not only a global dissemination of news but also localized circulations of political information within users' personal networks.

As contemporaries living in a "hybrid media environment" [23] where individuals have no shortage of information sources or opportunities to connect and interact with distant others, researchers hold mixed views on its impact on deliberative democracy, with both cautious optimism and deepening concerns. On the positive side, some scholars have highlighted the agency of audiences in navigating diverse information streams [24], emphasizing the potential of media–especially social media–to promote a pluralistic democratic ideal and foster participatory politics among the broader public [e.g., 25, 26]. Yet the path to a healthy deliberative democracy appears fraught with numerous challenges for political communication, with political polarization standing out as a key concern [27, 28].

In the context of mass media, concerns about political polarization often co-appear with the observation of media fragmentation and audience segmentation, which is largely driven by partisan selective exposure in a high-choice media environment [e.g., 29, 30, 31, 32, 33, 34]. Selective exposure–formally defined as "any systematic bias in audience composition" [35]–occurs when individuals selectively seek out information sources that align with their previous beliefs or attitudes [34]; partisan selective exposure is thus a special form of such bias along the partisan dimension, that individuals selectively consume content that aligns with their partisan identities [34]. As the media content market expands, individuals now have greater freedom to opt out of political information entirely or to consume only content congruent with their ideological preferences [36, 37]. On the supply side, in order to stay competitive, media outlets are motivated to cater to more niche audiences with clearer preferences, contributing to the rise of partisan media [38, 4] and the erosion of "a common place to meet and debate contrasting views" [31]. According to agenda setting theory, media can exert their influence to the audience through shaping their agenda: determining what issues are important (i.e., first-level agenda setting), what attributes of those issues are important (i.e., second-level agenda setting), and how issues and attributes are associated (i.e., third-level agenda setting) [39, 40]. In an ideologically segmented media landscape, individuals with different partisan leanings are thus likely to get exposure to distinct partisan content, resulting in the separate formation of issue agendas [41, 42], and undergoing a reinforcing trajectory of political predispositions [43]. Such divergence may erode trust in mainstream news and limit shared understanding, deliberation, and reflection on political events–a tendency that can exacerbate political polarization [34, 44].

Given this theoretical linkage between mass media fragmentation and political polarization, my first project (presented in Chapter 2) offers a background view of political communication via mass media. Based on intermedia agenda setting (IAS) theory–which posits that media outlets can influence the agenda of each other [45]–this study analyzes media coverage of presidential candidates during the 2016 and 2020 U.S. elections to evaluate the degree of agenda fragmentation between high- and low-credibility media, as well as between left- and right-leaning media. The results uncover patterns of overall high alignment yet growing divergence of news agenda between different types of media, and reveal the shifting IAS roles of these media types with an alarming signal of the declining IAS power of high-credibility media. Additionally, we notice interesting partisan asymmetries in agenda alignment, that the coverage of Trump is better aligned across different media types, compared to the coverage of his Democratic opponents.

The IAS patterns discussed in Chapter 2 can be viewed as the starting point of a broader chain of dynamics that unfold across the political belief systems and online behavioral patterns examined in Chapters 3 and 4. Growing divergence between partisan media may contribute to an increasingly segmented audience landscape [31], where individuals with different ideological leanings cultivate more distinct media repertoires [46], which, in turn, incentivizes partisan media to produce ideologically targeted content in order to maintain or expand their market share [46, 47]. Through the mechanism of agenda setting, the division in media choice may result in the division in political issue, attributes, and their associations that become salient in consumers' minds [48]; or, in framing theory terms, the "frames in thought" [49]. For belief system structuration, as discussed in detail in Chapter 3, this division has important implications: when different groups of individuals repeatedly encounter distinct clusters of issues framed in particular ways, the cognitive linkages between these issues may solidify along partisan or ideological lines, reinforcing internally coherent yet mutually divergent worldviews. Moreover, as discussed soon in the next paragraph, such a separation in beliefs may inform and motivate distinct patterns of online political behaviors. Thus, the initial divergence in media agendas does not merely signal fragmentation on the supply side of political information; if may also serve as a condition under which polarization can deepen towards the individuals at the demand end, through intertwined processes of belief formation and behavioral reinforcement.

Alongside ongoing investigations about mass media and political polarization, debates on social media's role in the process of political polarization also persist among scholars [e.g., 28, 50, 11, 51]. A common line of argument starts with the suspicion that social media facilitates the formation of "echo chambers" and "filter bubbles", where users encounter ideologically congruent opinions and interact with like-minded others [52]. With reduced exposure to diverse perspectives, individuals may become more entrenched in their pre-existing beliefs [35, 53, 54, 55], which increases the inter-group distance and ultimately, heightens the degree of political polarization. This theory rests on two causal assumptions: first, social media use would decrease ideological diversity in users' information environments; and second, ideologically homogeneous exposure would increase political polarization. Both assumptions remain contested, with mixed empirical findings [50]. For the first assumption, while many studies have documented ideological segregation on multiple social media platforms such as Twitter/X, Facebook, and Reddit [e.g., 56, 57, 58, 59], others argue that such segregation exists primarily around political topics [55], that social media actually facilitates greater exposure to ideologically diverse content and connections than other communication settings [60, 61, 62, 63], and that inter-group interactions are more frequent than conventionally believed [64, 50]. As for the second assumption regarding the relationship between polarization and the ideological homogeneity of consumed information, some studies have indeed found evidence supporting the moderating effect of both counter-attitudinal news exposure [65] and cross-cutting user interactions [66, 67]. Others, however, contend with evidence of a backfire effect, suggesting that encountering opposing views would further polarize users and deepen the ideological gap across groups [68, 69].

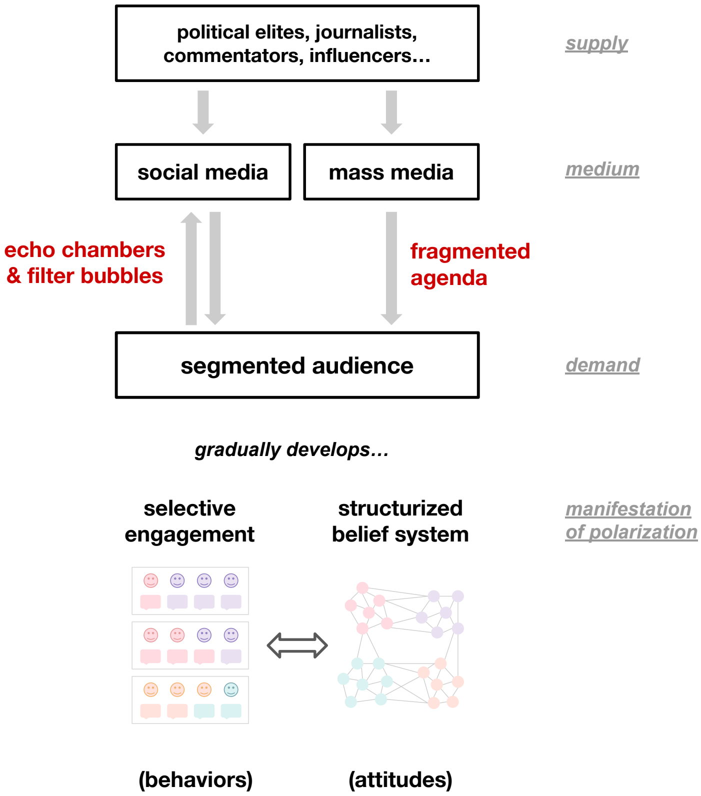

Figure 1.2: A summary diagram for the theoretical framework for later empirical chapters.

While this dissertation does not directly address the role of social media in the process of political polarization, it engages critically with how polarization is measured using a variety of datasets–including social media data. In particular, I reflect on how existing measurements may be biased by analyzing certain types of digital traces captured by public platform APIs while leaving out those that are not captured (see Chapter 4). The aim is to contribute to the field of polarization research by offering recommendations on how scholars shall more carefully scope and interpret polarization observations through digital traces, which, I hope, can be helpful to disentangle multiple factors that contribute to aforementioned inconsistencies regarding the relationship between social media usage and political polarization.

In sum, this dissertation traces political communication from mass media agenda fragmentation on the supply side of information, to the challenge of polarization on the demand side, as reflected in attitudes (structured belief systems) and online behaviors (selective engagement). The overall framework is summarized in Figure 1.2.

1.3 Measuring political polarization

Now, I dive deeper into the concept of political polarization and discuss the measurements of this phenomenon in empirical studies.

As a worldwide issue frequently occurring in warnings from media, politicians and scholars [28, 70], political polarization has become a prevalent concern among many democracies [71, 72], a chasm obstructing the pluralist ideal [73], a catalyst for political hatred and violence [74], a hindrance to social cooperation in public health emergencies [75], and more detrimentally, a key driver of "democratic backsliding" [76, 70]. Understanding its trends, underlying mechanisms, and interactions with today's multimedia environment has become a focal point of interdisciplinary research, with researchers from traditional social sciences and computational fields offering diverse measurement frameworks [11]. Traditional social science disciplines, such as political science and sociology, typically rely on survey data and conceptualize political polarization as a specific structure of political attitudes [e.g., 77, 78, 79]. In contrast, computational social scientists have turned to social media data, analyzing polarization through patterns of behavioral segregation in online political activities [e.g., 56, 80, 58].

At a high level, the variety of datasets and measurement approaches has enabled two main perspectives of observation: attitude and behavior. The attitudinal perspective captures individuals' psychological states related to political beliefs or opinions, typically measured through self-reported surveys with pre-defined questions. These questions often assess ideological leanings (e.g., left/right, liberal/conservative), views on key social issues (e.g., guaranteed jobs and income, abortion legalization), and attitudes toward political parties (e.g., feeling thermometers). Based on attitudinal data, political polarization is generally defined as a divergence or consolidation of public ideologies (i.e., ideological polarization), or as a state of hostility between partisans (i.e., affective polarization) [10]. The behavioral perspective[1] looks into a wide range of political activities in digital environments (e.g., social media). Compared to attitudinal data from surveys, the behavioral data retrieved from social media is usually richer yet noisier, and cannot be directly mapped onto well-defined political constructs. Drawing on behaviors such as following, commenting, and (re-)posting, political polarization can manifest through selective exposure to ideologically congruent content [81, 82], homophilic interactions within like-minded users [55, 57], and toxic communication between opposing ideological communities [83, 84, 85].

The attitudinal and behavioral aspects of polarization, albeit closely intertwined and mutually influential [86, 87], need to be conceptually disambiguated and analytically integrated. In other words, polarization in attitudes and behaviors should be clearly defined as separate

[1] Although political engagement can take forms of both online and offline activities, this dissertation focuses primarily on the online facet as it represents an emerging area of interest in polarization research [11, 28]. Given this emphasis, I generally omit the prefix "online" when discussing behavior-based measurements.

constructs; however, as I will show in Chapter 4, combining data and analysis from both perspectives is crucial for understanding their interrelations and more broadly, the role of social media and other sociotechnical systems in shaping this dyad.

In the following sections, I review how political polarization has been defined and measured in attitudinal and behavioral perspectives, highlighting the limitations of each perspective, and motivating my research in later chapters.

1.3.1 Political polarization reflected by attitudes

Marking a year of political turmoil and 74 national elections[2], Merriam-Webster crowned "polarization" as its Word of the Year for 2024[3], defining it as a "division into two sharply distinct opposites" and in particular a "state in which the opinions, beliefs, or interests of a group or society no longer range along a continuum but become concentrated at opposing extremes." The dictionary definition prioritizes how (political) polarization is frequently conceptualized in the mass media[4] and in traditional social science disciplines–that is, political polarization happens through a significant division in the psychological states related to political opinions, beliefs, or interests between two distinct groups [78, 88].

To measure such a division in public mentalities, researchers have explored–also debated the importance of–ideology-based and affect-based methods [89, 90]. Ideology-based methods focus on the salience of division in terms of political ideologies, which, according to Erikson and Tedin's definition, is a "set of beliefs about the proper order of society and how it can be achieved" [91]. Political ideology includes (i) *symbolic* ideology that describes one's abstract self-identification as left or right, liberal or conservative; and (ii) *operational* ideology that corresponds to one's views on specific social issues or policies, such as universal healthcare or legalizing abortion [92]. The salience of division could be assessed by the strength of inter-group divergence (i.e., "ideological divergence"), which captures the extent to which different groups shift away from moderate positions toward opposing ideological extremes; or by the strength of intra-group consistency ("ideological consistency") [10], which reflects the extent to which individuals within a certain group develop ideologically consistent positions across multiple issues. This aligning process, termed "ideological sorting", often happens along the partisan line in the U.S. context (i.e., "partisan sorting"), such that Democrats would adopt predominantly liberal views and Republicans, conserva-

[2]Source: https://www.idea.int/initiatives/the-2024-global-elections-supercycle, URL accessed on March 3, 2025.

[3]Source: https://www.merriam-webster.com/wordplay/word-of-the-year, URL accessed on March 3, 2025.

[4]See a few examples at https://www.washingtonpost.com/science/2024/01/20/polarization-science-evolution-psychology/, https://edition.cnn.com/2021/11/09/politics/political-typology-pew-poll/index.html, https://www.nytimes.com/2024/09/22/style/storycorps-partisan-divide-election.html. URLs accessed on March 3, 2025.

Term	Brief definition
affective polarization	the degree of division in terms of inter-group hostility/bias/distrust
ideological polarization	the degree of division in terms of political ideologies
ideological divergence	the extent to which groups shift away from moderate positions toward ideological extremes
ideological consistency	the extent to which groups align their issue positions consistently with their liberal or conservative ideals
ideological sorting	the process by which individuals' issue positions become more consistent with their liberal/conservative ideals
partisan sorting	the process by which individuals' ideologies become more aligned with their partisan identities
political ideology	a bundle of political beliefs that delineate one's ideal of how a society should function
symbolic ideology	the ideological label that describes one's political identity at an abstract level (e.g., liberal/conservative, left/right)
operational ideology	the concrete policy preferences or issue positions that reflect one's ideological beliefs in practice (e.g., supporting transgender rights)

Table 1.1: A list of relevant terms with brief definitions.

tive views [15]. An increase in ideological consistency indicates a diminishing cross-cutting space along the main ideological cleavage and a tendency towards all-encompassing social conflict [93].

In contrast, affect-based methods emphasize the *emotional* elements in political polarization (i.e., affective polarization) rather than the ideological stance within or between groups, as scholars argue that polarization may stem not solely from ideological disagreement, but from emotional conflicts such as intergroup disdain and hostility [89]. In practice, researchers measure affective polarization by evaluating the degree to which individuals develop positive feelings for in-group members while harboring negative feelings for out-group members [94], keywords one uses to describe the parties [95], and the degree to which one (dis)trusts the parties [96].

Next, to more concisely delineate each type of polarization, I briefly specify the essence of each polarization type while providing examples of widely adopted measures. It is important to note that the polarization is not inherently limited to bipartisan systems or confined to a binary structure. However, it is often studied in the U.S. context, where the polarized social

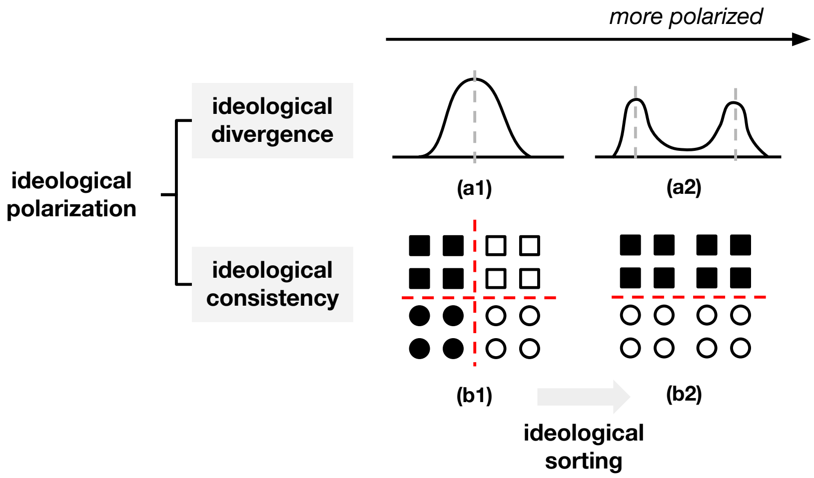

Figure 1.3: Toy models to graphically illustrate (a) ideological divergence and (b) ideological consistency. From left to right, the represented population becomes more polarized in terms of ideological divergence (from a1 to a2), or ideological consistency (from b1 to b2). In panel a, the X-axis can be an ideological spectrum of interest (e.g., left/right, liberal/-conservative), the gray dot lines mark the modes for each distribution. In panel b, each block represents one person, and the color and shape represents different issue positions respectively (e.g., black (white) means support (oppose) transgender rights, square (circle) represents support (oppose) universal healthcare); the red dot lines mark the cleavages that split the population.

landscape is characterized by two prominent groups of opposing partisans. For simplicity, the discussion below considers these measures within a bipolar framework, in which two primary groups hold distinct and opposing ideologies (e.g., liberal and conservative).

Ideological Polarization: Divergence

First, ideological divergence measures the extent to which two opposing groups move further away from each other towards the extremes of a specific ideological spectrum (see Figure 1.3 from a1 to a2). Consider a group of N respondents with self-reported ideological positions, for example, along the liberal/conservative spectrum on a 7-point scale (1 = extremely liberal, 7 = extremely conservative). Let the ideological position of respondent i be denoted as x_i, and the set of all responses as $X = \{x_1, x_2, \ldots, x_N\}$, which forms the empirical ideological distribution $f(x)$.

Intuitively, ideological divergence shall reach its maximum when the distribution is perfectly bimodal, with respondents split 50/50 at two opposing extremes and minimal representation around the moderate center. To quantify this divergence, researchers commonly employ statistical measures such as variance, kurtosis, and the bimodality coefficient (BC) [77, 10, 97], each capturing a unique aspect of the shape and dispersion of the ideological distribution $f(x)$.

- **Variance** measures the average squared deviation from the mean, indicating the statistical spread of ideological positions. A higher variance indicates a more dispersed distribution and thus a greater level of polarization.

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{1.1}$$

  where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ represents the average ideology level for all respondents.

- **Kurtosis** measures the "tailedness" of the distribution, with lower kurtosis suggesting a flatter, more bimodal distribution, corresponding to a more polarized population.

$$Kur(X) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^2} \tag{1.2}$$

- **Bimodality Coefficient (BC)** combines skewness and kurtosis to evaluate the presence of bimodality in the distribution. A higher value of BC means a greater level of polarization.

$$\mathrm{Skew}(X) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{3/2}} \tag{1.3}$$

$$BC(X) = \frac{\mathrm{Skew}(X)^2 + 1}{\mathrm{Kur}(X)} \tag{1.4}$$

Ideological Polarization: Consistency

Unlike ideological divergence, which focuses on how respondents are distributed along a single ideological dimension, ideological consistency examines the alignment across *multiple* ideological dimensions (i.e., constraint), as demonstrated in Figure 1.3 from b1 to b2. At the individual level, constraint measures the degree to which one's position in a certain ideological dimension can reliably predict their position in another dimension. In a low-constraint scenario (b1), such prediction power is so weak that knowing one's position on issue A (e.g., supporting transgender rights) gives no information on their position on issue B (e.g., supporting universal healthcare). On the contrary, in a high-constraint scenario (b2), one's position on issue A can immediately indicate their position on issue B, as these people align their ideological dimensions perfectly. At the population level, the stronger the constraint, the more ideologically distinctive the opposing groups are, the less likely they can find a common ground with partial (if not full) agreement, and eventually the greater the overall level of political polarization we will observe.

For mathematical formulations, let us recall the same example of N respondents who, in addition to reporting their liberal/conservative scale $\mathcal{X}$, also provide their views on K politically relevant issues (e.g., whether the government should implement tax cut or universal healthcare), each measured on a 7-point scale (e.g., 1 = strongly support tax cut, 7 = strongly oppose tax cut). These sets of responses are denoted as $\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, ..., \mathcal{Y}^{(K)}$. With respondents' ideologies captured across multiple dimensions, issue consistency can be assessed either through the internal alignment among operational or symbolic ideologies, or through the alignment between symbolic and operational ideologies.

In previous research, operational ideologies are typically assigned fixed orientations [e.g., 98]–supporting tax cut, for instance, is coded as a liberal stance and opposing tax cut as conservative. These directional codings are taken into account when evaluating cross-dimensional ideological alignment. In practice, scholars have compared the issue consistency scale (i.e., how one consistently chooses liberal or conservative issue positions) between partisan groups [98], calculated identity alignment score (i.e., the absolute distance between party identity and liberal/conservative position) [15], and correlations between multiple ideological dimensions [78, 93, 99] as aggregated measures of ideological consistency.

Specifically, for a given political party A, one can calculate the issue consistency score within that party, denoted as C_A, as:

$$C_A = \frac{1}{N_A} \sum_{i=1}^{N_A} \sum_{j=1}^{K} y_{ij}^{binned} \tag{1.5}$$

where $y_{ij}^{binned} = -1$ (or 1) if y_j is a liberal (or conservative) response, and $y_{ij}^{binned} = 0$ if central. In a polarized population with two parties A and B, the distribution of C_A and C_B is expected to be concentrated on two opposing ends, with one party being consistently liberal and the other consistently conservative.

Meanwhile, the identity alignment score, denoted as D, measures the average distance between partisan identity and ideological position among N respondents:

$$D = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i| \tag{1.6}$$

A lower value of D suggests a higher degree of partisan alignment overall.

Lastly, ideological alignment between any two dimensions can be assessed via the Pearson correlation coefficient r:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1.7}$$

The higher the correlation, the more aligned these ideological dimensions are, and the more ideologically consistent the population is.

Affective Polarization

Affective polarization measures the emotional stance individuals hold towards outgroup members who stand at the opposite end of an ideological spectrum. This aspect of polarization can be reflected by how one feels about the opposing partisan group or how willing they are to interact with outgroup members in a social context (e.g., workplace, friendship, family relationship) [10]. For quantitative measures, the most commonly adopted approach is based on feeling thermometers that assess how warm one feels towards a given political party (e.g., on a 101-point scale, 0 = very cold, 100 = very warm) [89]. The difference between the ingroup and outgroup ratings, reflects the degree of intergroup bias–a greater difference indicates a stronger bias, and thus a higher level of polarization. In practice, one can quantify the level of affective polarization AP as the average feeling thermometer difference (FTD) between the ingroup and outgroup ratings.

$$\mathrm{FTD}_i = \mathrm{FT}_{\mathrm{in},i} - \mathrm{FT}_{\mathrm{out},i} \tag{1.8}$$

$$\mathrm{AP} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{FTD}_i \tag{1.9}$$

where $\mathrm{FT}_{\mathrm{in},i}$ and $\mathrm{FT}_{\mathrm{out},i}$ are the feeling thermometer ratings for the in-party and out-party by individual i.

Alternatively, scholars have developed indicators such as social distance (e.g., whether one would accept the opposing group members as in-laws) [89, 100], as well as assessments of stereotypes about outgroup members [101, 102]. These measures vary more widely in formulation and are therefore not discussed in detail here; for a comprehensive review, see Iyengar et al. [103] and Wagner [104].

1.3.2 Political polarization manifested in online behaviors

Next, we turn to the behavioral perspective–specifically, how political polarization can be studied through human behaviors in political participation. While political participation can occur both online (e.g., sharing political content on social media) and offline (e.g., attending political rallies or protests), online participation is afforded by and intertwined with digital infrastructures of a rapidly evolving sociotechnical system, which urgently calls for ongoing empirical updates to classic theories and the flexible development of large-scale analytical tools [105]. Thus, this section focuses on online political behaviors, particularly on how polarization has been measured through various forms of political participation on social media.

As of February 2025, the number of social media users worldwide has reached 5.2 billion [5], accounting for roughly two-thirds of the global population. Among younger generations, in particular, social media has become a primary channel for accessing news and information (e.g., in the U.K. [106], Europe [107] and the U.S. [108]). Unlike mass media, which typically facilitates a unidirectional flow of information from media to audience, social media enables two-way interactions–both between users and content and among users themselves–fostering what has been termed *participatory politics* [22]. Indeed, the digital infrastructures provided by social media allow for broader civic engagement that goes beyond traditional offline means [109]. Such a transformation from offline, face-to-face activities to online, computer-mediated interactions has generated a vast amount of datasets of online political behaviors, which enables novel observational angles for phenomena that used to be studied only in offline settings, yet at the same time demanding adjustments in measurement frameworks of conventionally defined social constructs into the online context.

To date, there remains a lack of conceptual clarity in how polarization is defined, measured and interpreted in studies based on online behavioral datasets [11, 83, 110]. Meanwhile, the theoretical connection between traditional conceptualizations of polarization in the social sciences (as described in Section 1.3.1) and its operationalization using social media data is insufficiently discussed in many computational studies [11]. These research gaps make it difficult to reconcile findings from earlier survey-based studies with those emerging from contemporary analyses of political polarization on social media, thereby complicating the ongoing debate about the role of social media in exacerbating polarization. In what follows, I aim to articulate specific aspects of polarization manifested through social media activities, to clarify their theoretical relationship with survey-based measures, and to set the stage for a discussion of their limitations in Section 1.3.3.

Depending on the specific data point being analyzed, online political polarization can man-

[5] Source: https://www.statista.com/statistics/617136/digital-population-worldwide/, URL accessed on March 30, 2025

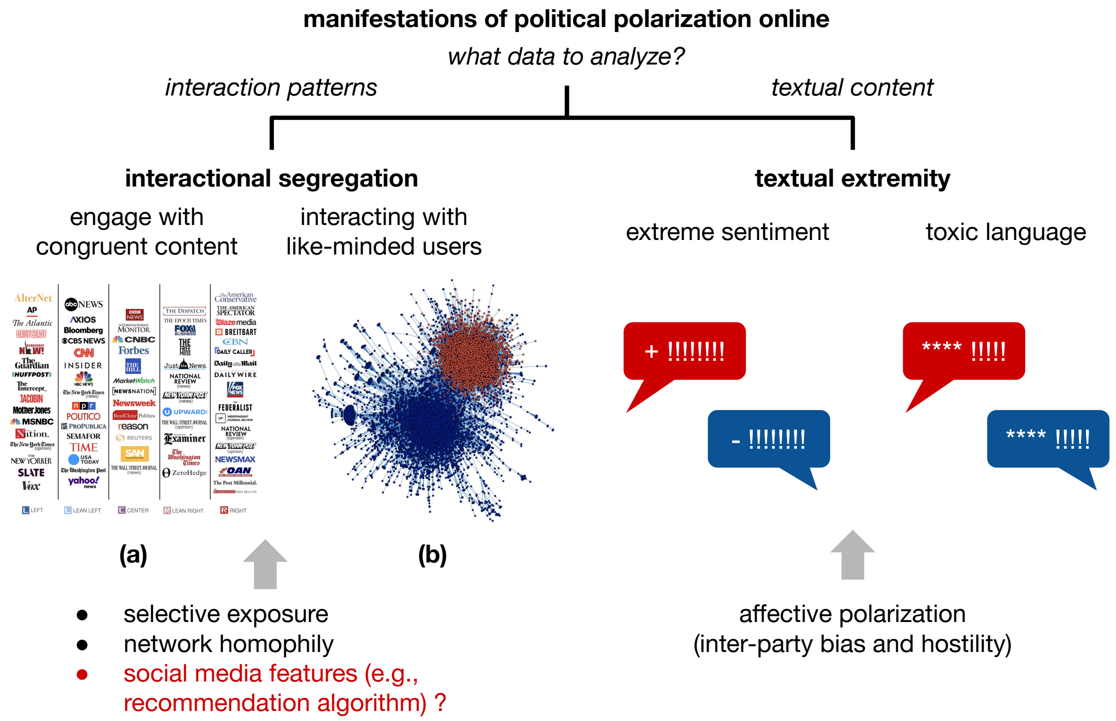

Figure 1.4: A summary graph for two manifestations of political polarization online. Subfigure (a) [111] is a media bias chart that maps out the political leanings of online news outlets in the U.S.; the left (right) side lists media with clearly left-leaning (right-leaning) ideologies, and the center column lists media without a particular political leaning. Subfigure (b) [112] is a directed retweeting network, where nodes represent individual users, and a link from user i to j indicates that i has retweeted from j. Node color reflects the ideological leaning of the corresponding user.

ifest mainly in two ways: interactional segregation (focusing on interaction patterns) and textual extremity (focusing on textual content), as demonstrated by Figure 1.4.

Interactional Segregation

Interactional segregation evaluates the degree to which ideologically homogeneous users interact with congenial content or like-minded others [83]. For instance, let us consider two ideological groups, liberals and conservatives. We would observe a high level of interactional segregation when liberals (conservatives) primarily interact with other liberals (conservatives), subscribe to liberal-(conservative-)leaning media, and share content that aligns with their liberal (conservative) beliefs. In this segregated state, inter-group dialogue is minimal and their media repertoires show little overlap. This segregation pattern can be driven by: (i) selective exposure, i.e., the tendency to seek information that confirms one's preexisting beliefs [35], particularly ideologically aligned content in political contexts [113], and (ii) network homophily, i.e., the tendency to form connections with like-minded others who share similar sociodemographic traits, including political ideology [114]. To operationalize interactional segregation based on behavioral data at scale, scholars have developed a variety of computational methods, including network analysis that illustrates the separate clusters between ideological communities [56, 57], latent variable models that extract users' ideology based on their consumption patterns and further shows how such patterns differ across groups [55, 59], and community embedding training that demonstrates the segregated user base of ideologically opposing communities [58].

Textual Extremity
Textual extremity investigates the language use in online political communication to assess the degree to which it reflects extreme sentiment or radical issue stances. Leveraging natural language processing (NLP) techniques on user-generated texts, researchers extract latent variables such as issue position and sentiment [e.g., 83, 84, 85] to measure the inter-group conflict, as a proxy for (affective) polarization, since such linguistic features could be driven by inter-group hostility and bias. With this measure, a polarized digital sphere would be characterized by toxic and extreme language use in cross-group discussions of social issues. Unlike interactional segregation that views polarization through a "cluster-analytic perspective" [70]–that people would cluster together with in-group members and differ from out-group others, textual extremity underscores the *conflicting* nature of polarization–that people hold negative evaluations of out-group others regardless of their proximity in online interactions [70].

Clarifying these specific polarization measures is critical for us to understand the process of political polarization on social media–specifically, what form of polarization can be impacted, and how social media facilitates the change in the polarization degree. For instance, one commonly held notion is that interactional segregation may be amplified by social media with recommendation algorithms [115], which personalize content and user suggestions based on individual preferences [116]. This algorithmic design can intensify informational silos, creating so-called echo chambers and filter bubbles that further limit users' exposure to counter-attitudinal content [115]. Public discourse has linked such concerns with the increase in social division, as media and political elites adopt a harsher causal language–for instance, describing echo chambers as "straining [and ripping] the social fabric of [Australia]" amid the Israeli–Palestinian conflict [117], "ensure[-ing] we remain divided" around the US presidential election in 2024 [118], and even "destroying democracy" [119]. To further address these concerns, however, researchers need to first articulate whether interactional segregation is actually contributing to a more divided ideological landscape (i.e., ideological polarization), or to greater disdain between opposing partisans (i.e., affective polarization).

To summarize, a widely held view is that social media fosters the formation of echo chambers and filter bubbles, isolating users within ideologically homogeneous communities and limiting their exposure to diverse information sources. As a result, social media use is *perceived* to reinforce–and potentially radicalize–users' preexisting beliefs, thereby exacerbating political polarization. However, this view is not fully supported by empirical research, as discussed by the end of Section 1.2. Indeed, recent reviews have questioned the strength of the evidence supporting the claim that social media inherently exacerbates polarization [28, 11, 50], calling for more nuanced statements rather than sweeping conclusions about social media's impact on political polarization, and more comparative research across understudied countries and underexplored datasets [12].

1.3.3 Limitations of current approaches

Last but not least, I highlight the gaps and limitations in existing measurement frameworks in both attitudinal and behavioral perspectives, as a motivation for my projects to be presented in Chapters 3 and 4.

Unidimensional analyses cannot capture belief system dynamics

In social and political science fields, research on political polarization has yielded the following points of consensus. First, in Western democracies, public opinion has remained largely moderate and stable over decades [10, 120, 121, 122], indicating a low level of ideological divergence overall. However, in the US, polarization is more conspicuously manifested through in partisan sorting–such that partisan identity has grown increasingly salient [93] and consolidated with political beliefs [123], signaling a significant rise in ideological consistency[6]. At the same time, affective polarization in the U.S. has intensified, growing notably faster than in other OECD countries [126], which, according to Mason [15], could be driven by partisan sorting that divides the population into an all-encompassing disagreement along the partisan line.

The relationship among different types of polarization–how they influence one another, or whether one serves as a core driver of the rest–remains ambiguous in the political science literature [89, 90]. Regardless, we can see that one critical shift has taken place in a more complex ideological space beyond the traditional left–right spectrum, prompting scholars to examine the mechanism of polarization via ideological sorting across multiple political dimensions [15, 127], especially how each of these dimensions interrelates one another in a complex system of political beliefs [16]. For instance, one's stance on the Supreme Court's decision to overturn Roe v. Wade is likely to be associated with their stances on transgen-

[6]This trend, however, has not yet been observed in Europe, as suggested by Adams et al. [124] and Adams et al. [125].

der rights or the Black Lives Matter movement. Such associations among multiple beliefs (i.e., belief constraints) constitute a complex system, i.e., belief network, a concept first introduced by Converse [16] and recently operationalized through belief network analysis (BNA) [e.g., 128, 123, 129]. BNA is a powerful tool for modeling and understanding polarization [123], and for dissecting the specific role of certain political issues or beliefs in this process [129, 17]. Despite its potential, the use of BNA in polarization research has been scarce and, as I will argue in Chapter 3, BNA can provide more nuanced insights on system-level and attitude-level shifts in the process of polarization [17].

Reliance on visible online traces ignores invisible political engagement

With the growing availability of diverse data sources, researchers have sought to adapt classic measures of political polarization across different contexts [130]. As online datasets enable researchers to observe facets of polarization absent in survey data, it is crucial to reflect on what types of political engagement are captured by these data–and, just as importantly, what they leave out. Most polarization research drawing on social media focuses on political engagement that are retrievable via platform APIs (e.g., (re-)posting, commenting) [e.g., 59, 131], while overlooking engagement that is generally not well documented and thus less visible to researchers(e.g., viewing videos, searching news). This raises a critical question regarding the bias in polarization measurement: can analyses based solely on visible engagement truly reflect the full picture of political polarization online? For instance, is the level of interactional segregation observed through re-posting behaviors comparable to that observed through viewing behaviors? Are ideological opponents similarly segregated in the content they engage with via commenting, compared to the content they engage with via browsing?

To answer these questions, I discuss the importance of accessing multidimensional behavioral datasets–including visible and invisible engagement–and present an empirical project in Chapter 4, where I utilize digital traces that were once unavailable (or less accessible) through public API queries. Through comparing user-level and content-level variations in political engagement, I argue that leaving out invisible traces may generate significant biases in our observations of the ideological landscape online and mislead our interpretations of the measurement outcome for political polarization.

Chapter 2

Dynamics in the mass media environment: intermedia agenda setting during the 2016 and 2020 U.S. presidential elections

Building on the broader discussion of political communication in a hybrid media environment in Section 1.2, this chapter turns to the *supply* side of political information by analyzing how mass media agendas interact and diverge in the context of U.S. presidential elections. In particular, I focus on intermedia agenda setting (IAS) theory to examine the extent to which media outlets align or differentiate in their coverage of presidential candidates during the 2016 and 2020 U.S. elections. By comparing agenda alignment across media types distinguished by credibility and ideology, this chapter aims to refresh our understanding of whether the pressures of a high-choice media environment are destabilizing traditional IAS relationships, and the implications for the broader media ecosystem and public political discourse.

2.1 Introduction

How do news media select the coverage they present to their audience? Intermedia agenda setting (IAS) theory identifies one important force setting the agenda of a given news producer, that is, other news producers. While this theory is well-established, with a significant amount of early theoretical and empirical support [45], its stability is in question in today's high-choice media environment [132]. Theoretically, one can make a case for either divergence or convergence. News organizations might diverge in their coverage by catering to different audience segments [133]. However, commonalities in journalistic training and the broader social context (e.g., events happening in the real world) can lead to convergence despite economic pressures [134]. Empirical evidence is similarly mixed, with support for both divergence [e.g., 135, 46, 136] and convergence [e.g., 137, 138] of media agendas.

Past work has examined agenda alignment across various media categories such as distribution channels (e.g., TV, newspapers, online news) [138] and ideology [135, 46, 136]. Investigations related to ideology are particularly pertinent and common, given the significant role ideology plays in the U.S. political system and audience preferences. Yet, ideology is no longer the only element that activates the selective news coverage adapted for a segmented audience base. Today's high-choice media environment includes low-credibility news producers that deviate from traditional journalistic standards, at times explicitly providing a "critical meta-discourse on traditional journalism" [139]. This might pose a more fundamental threat to the stability of IAS. It is this threat that motivates our study. We ask: To what extent does the news agenda between low- and high-credibility media diverge? Furthermore, is IAS more significant along the credibility dimension than the ideology dimension? We answer these questions by re-examining IAS across media with varying credibility levels and different partisan leanings.

In this chapter, we present two important case studies, the 2016 and 2020 U.S. presidential elections. We determine the degree to which different media types (low-credibility vs. high-credibility and left-leaning vs. right-leaning) align in terms of the candidate attributes they focus on. We examine two types of attributes: keywords (e.g., how often the word "liar" is associated with Clinton) and topics (e.g., how often the topic "healthcare" is associated with Trump). We focus on the 2016 and 2020 presidential elections for three reasons. First, these case studies are consequential. During nationally pertinent events such as presidential elections, the news agenda can shape the public political discourse, potentially impacting voting behaviors and, ultimately, the election outcome [140]. Second, the similar nature of these case studies helps us determine the degree to which findings from one IAS analysis generalize to other similar contexts. Finally, the four-year course from 2016 to 2020 has witnessed fundamental shifts in the news ecosystem with the growing prominence of and public attention on low-credibility media [141], on top of the longstanding partisan division in the contemporary U.S. media environment [96]. Up-to-date studies are needed to refresh our understanding of the impact brought by these shifts.

Our examination of IAS theory is carried out in two stages. First, we show how media agendas align with one another concurrently in their coverage of each presidential candidate. We measure the degree of alignment by correlating the distributions of the overall attention on various attributes across media of different credibility and ideology. We find that the level of agenda alignment between low- and high-credibility media is comparable to that between left- and right-leaning media. Moreover, we observe (i) a better-aligned coverage for the Republican candidate than the Democratic candidate in general and (ii) an increasing level of divergence from 2016 to 2020. We explain the variation in alignment by highlighting the crucial role controversial candidate attributes play in agenda divergence.

Second, we look into the temporal dynamics of IAS and identify the agenda leader and follower for specific attributes associated with a given candidate. We primarily focus on topic attributes and determine which media type leads the changes in topic coverage. We see that high-credibility media is the dominant agenda setter in general, leading the agenda on more attributes and for longer periods of time than low-credibility media. Meanwhile, we notice the decline in IAS power of high-credibility media for the Democratic candidate, as well as the increased interactions between low- and high-credibility media from 2016 to 2020. Although we observe similar patterns between high-credibility (low-credibility) media and left-leaning (right-leaning) media, there are still subtle differences between these two lines of comparison. For instance, while low-credibility media never takes a persistent agenda leader role, right-leaning media has led a few topics for Trump in 2020.

Finally, although we adopt terms such as "agenda setter" and "Granger causality", it is crucial to bear in mind the constraints of relying solely on temporal correlations to assess causal relationships. Thus, we suggest taking our study as suggestive insights rather than definitive causal assertions.

2.2 Related works

2.2.1 Intermedia agenda setting

Agenda setting theory suggests that news media shape public opinions on issue salience through their coverage–the more media cover a topic, the more important that topic becomes in the public agenda [39]. Alongside the inquiry of agenda flows between media and the public, intermedia agenda setting (IAS) theory looks into the agenda dynamics *among* media and suggests that different news sources can influence one another.

Previous studies have explored the IAS process with these questions: who takes the lead, on what specific issues, and in what time frames? Regarding the agenda leader/follower, researchers have identified the powerful role of elite news media in setting the agenda for others [142, 20], the tendency of junior newspapers to follow the lead of senior ones [143], and more recently, the potential of emerging online media to participate in IAS [144]. The rising prevalence of fake and partisan news media has motivated research efforts to examine IAS through lenses of credibility and ideology.

Most relevant to our study, Vargo et al. [145] found a reciprocal relationship in the network issue agenda between fake news and fact-based news, as well as between fake news and partisan news from 2014 to 2016; Guo and Vargo [141] further pointed out the difference in

IAS dynamics between two presidential candidates in 2016, that (a) compared to attributes associated with Clinton, those associated with Trump were tied closer between fake news and fact-based news, and that (b) partisan media were able to lead the agenda for fake news media on attributes associated with Trump, whereas in Clinton's case, the interaction between partisan and fake news media was much weaker. In terms of the temporality of the IAS process, researchers have distinguished between breaking stories and ongoing debates [146], discussed cases of breaking news being manipulated by false reporting [147], and called for future work to address the nuances in the time scale of the IAS process [144]. Because IAS can happen through linked temporary spikes and correlated ongoing fluctuations, understanding these dynamics requires us to examine the temporal aspect of convergence or divergence with flexible time scales.

Our study contributes to this line of research in the following aspects. First, instead of focusing on a single election, we study two elections to determine the consistency of IAS patterns. Second, the parallel analysis for two media pairings allows us to benchmark the IAS process between low- and high-credibility media, compared to that between left- and right-leaning media. Through this comparative perspective, we are able to reflect upon the significance of agenda divergence, as well as the positioning of agenda leader/follower along the credibility dimension, with respect to a longer-standing media segmentation along the ideology dimension. Third, as we will discuss in the next section, we introduce and validate a dictionary-based topic model that automates text coding and allows for IAS analysis at two different levels of granularity (i.e., aspects and central themes). Combining expert-curated and data-driven attribute schemes, our study outlines an interpretable and well-performing pipeline for computational studies of IAS.

2.2.2 Second-level agenda setting and candidate attributes

Both agenda setting and intermedia agenda setting can be examined at three different levels, each corresponding to distinct units for comparisons of agenda. The first level focuses on broad issues; the second level examines attributes used to describe issues [148, 136]; and the third level investigates the linkages, or co-occurrences, among various issues or attributes [149]. Here, we focus on the second-level agenda setting; that is, we take each presidential candidate as a single issue and ask whether and how the attributes associated with these candidates are aligned and flow between different media types.

In previous studies that also conceptualize political figures as issues, scholars have explored various dimensions of their attributes, the very basic application being the shaping of "candidate image" during political campaigns [e.g., 150, 141]. Among dimensions of attributes compositing such a "candidate image", McCombs et al. [151] specified two fundamental dimensions: the substantive dimension and the affective dimension. The former dimension organizes the candidate image with a set of relevant subtopics (e.g., personality, issue positions); and the latter dimension focuses on sentimental elements (i.e., positive, negative, or neutral) linked to the candidate. As the varying salience of linkages between attributes and a given candidate provides a cognitive frame through which a candidate is portrayed, researchers have connected framing theory with the second-level agenda setting [151, 152, 153]. In our study, we inherit this theoretical connection and model the candidate frame using three main groups of substantive attributes: (i) attributes that discuss general government operations (including election campaigns), (ii) attributes that describe a particular policy-making aspect, and (iii) attributes that mention candidate-related controversies.

In terms of the "granularity" of the frame, scholars have distinguished between two types of attributes–aspects and central themes–when investigating the second and the third level of agenda setting [20]. An aspect is "a micro attribute with a lower level of abstractness" and a central theme is a "macro-level attribute" that "describes a more abstract conceptual category" [154]. With existing studies suggesting a higher level of fragmentation at the aspect level than at the central theme level [155, 45], we consider it necessary to keep incorporating both levels of granularity when examining IAS. Our work measures the degree to which different media types are aligned in terms of both the central themes (e.g., how much do low- and high-credibility media align when associating Trump with various themes?) and the specific aspects of those central themes (e.g., how much do left- and right-leaning media align when associating Biden with various aspects?). We operationalize this dual-level measurement using a dictionary-based approach for text coding. Specifically, we capture aspects by detecting phrases (i.e., keywords) that occur in texts (e.g., "vote", "bank", "tax"), and capture central themes by identifying bundles of phrases (i.e., topics) that correspond to a particular candidate attribute (e.g., the topic "civil rights" includes keywords such as "vote" and "discrimination").

2.3 Dataset and preprocessing

We start our examination by identifying a set of low- and high-credibility online news outlets. We borrow the list of news domains from Bozarth and Budak [156] that combines five sources of domain credibility labels [157, 158, 159, 160, 161]. In our study, a domain belongs to the low-credibility class if it is explicitly marked as "fake" or "low-credibility" by any of the five sources. We then filter out domains that contain satire or mixed-factual content and group the remaining domains into the high-credibility class. We also assign ideology labels (i.e., left-leaning, right-leaning) to these news domains based on the bias rating tags assigned by Media Bias Fact Check [1]. Note that ideology labels only cover approximately 60% of all domains, with an imbalanced distribution across low- and high-credibility categories (the left-to-right ratio is 0.28 for the low-credibility group and 2.67 for the high-

[1] mediabiasfactcheck.com

credibility group). Thus, dimensions of ideology and credibility present two overlapping but not entirely aligned grouping structures.

Next, we collect the headline corpus of the aforementioned news domains. Using Wayback Machine, a webpage scraping API provided by Internet Archive, we retrieve homepage snapshots from 5521 low- and high-credibility news domains[2] during five-month periods of the 2016 and 2020 election seasons (i.e., from July 1 to November 30). From these timestamped snapshots, we extract news headlines that mention the first or the last name of at least one presidential candidate for the corresponding election [3]. To make candidate-wise comparisons in the IAS analysis, we split the headlines into two candidate groups for each year; headlines in a given candidate group capture the media coverage of the corresponding candidate in a certain election.

	media types by credibility		media types by ideology		
year	high-credibility	low-credibility	left-leaning	right-leaning	total # of domains (coverage $\geq$ 50%)
2016	362 (81.7%)	47 (10.6%)	185 (41.8%)	83 (18.7%)	443
2020	504 (62.6%)	222 (27.6%)	249 (30.9%)	233 (28.9%)	805

Table 2.1: Number of domains included in our analysis and group size for each domain category.

Given that the snapshotting frequency varies greatly across domains and across days, we assign a multiplication index to each snapshot, which will be used as a weight when we aggregate topic and keyword counts. The multiplication index equals the inverse of the number of snapshots for a given domain in a given day. This allows us to make sure that each domain will have at most one "average snapshot" per day that describes its overall agenda. We also filter out domains without sufficient snapshot coverage [4] to avoid temporal patterns being distorted by exogenous factors related to the Wayback scraping jobs (e.g., file size, number of parallel ongoing crawls). We report the number of qualified domains and the group sizes for each category in Table 2.1).

2.4 Methods

In this section, we elaborate on specific steps of the dictionary-based topic modeling: how we construct the topic dictionary, how the model computes topic and keyword vectors for

[2]Note that not all domains have snapshots for both years. There are 4540 unique domains for 2016 and 3109 for 2020.

[3]Our data extraction does not include dynamic or nested content.

[4]Domains that do not have at least one snapshot for at least 50% of the days are dropped in downstream analysis.

each input, and how we validate the model output with crowd-sourced labels collected from Amazon Mechanical Turk (MTurk). Then, we explain how we utilize the model output for downstream analysis, first to measure agenda alignment and then to identify agenda leaders and followers.

2.4.1 Dictionary-based topic modeling

Constructing topic dictionaries

For each of the 2016 and 2020 election seasons, we construct a customized dictionary by merging (1) a highly reliable [5] base dictionary [155] that uses the Comparative Agendas Project (CAP) taxonomy [162], and (2) an extended dictionary customized for each election season. The extended dictionaries are curated by political communication experts through consensus labeling (Bode et al. [163] for 2016 and Agiesta [164] for 2020). These dictionaries contain context-specific keywords (e.g., catchphrases, names of political elites) that not only improve keyword coverage in the headline data but also better capture the election-related expressions in both years. The dictionary merging proceeds as follows. We preserve the topic taxonomy established by the CAP codebook, match topic categories from the extended dictionary to the base dictionary, and create a new topic if there exists no reasonable match. After the initial merge, the 2016 dictionary includes 1340 phrases from 26 topics, and the 2020 dictionary contains 1405 phrases from 26 topics. The topics include "government operation" that encompass general administrations and election campaigns, policy-related categories such as "healthcare", "economy" and "international affairs" (included in the CAP taxonomy), as well as scandals-related categories such as "Trump controversies", "Biden controversies" and "Clinton controversies" (added after merging the extended dictionary).[6] The list of topics with their brief descriptions are provided in Table 2.2.

To further enrich the dictionary, we use a semi-supervised topic model, Guided Topic-noise Model (GTM) [165], to identify additional keywords and topics that are salient in the context of presidential elections. GTM utilizes an input dictionary that contains keywords for topics of interest to guide the topic-generation process. It expands the provided lists of keywords using a generalized generative model called Generalized Polya Urn (GPU) [165] to iteratively enhance existing topics and generate new topics containing new keywords and associated weights. Based on previous practices and preliminary runs on our data samples, we set the number of topics to 50; we then inspect all new keyword-topic pairs generated by

[5] Krippendorff alpha was 0.84 across trained coders constructing the dictionary.

[6] Note that both candidates have "DEM candidate controversies" and "REP candidate controversies" in their own attribute lists, since it is possible to mention one candidate within the context of a controversial issue associated with the other candidate (e.g., Trump discusses Clinton's email scandals). Keeping the topic lists consistent across candidates also allows for more convenient comparisons.

GTM and score the degree of relevance for each pair (i.e., 0 for non-relevant, 1 for weakly-relevant, and 2 for highly-relevant), given the possible contexts of a keyword in our dataset. The complete inspection is performed by one author, after two authors reach an acceptable level of inter-rater reliability on their independently-assigned relevance scores for dictionary samples from both years (Krippendorff's alpha = 0.81 for 2016 and 0.7 for 2020). After the inspection, we drop non-relevant keyword-topic pairs and evaluate three strategies of keyword filtering/weighting: (i) only including the highly relevant phrases, (ii) including both the highly and weakly relevant phrases, (iii) including both while giving higher weight to highly relevant phrases. We discuss how we evaluate these strategies in the subsection *Validating Output*. The final model uses only the highly relevant phrases, consisting of 1426 keywords in 2016 and 1453 keywords in 2020.

Identifying topics

The core idea of dictionary-based topic modeling is to detect keywords that occur in a given text, bin those keywords into their corresponding topic categories, and record the count of topics. Given K unique topics, W unique keywords, and text i, we first generate an aspect (keyword) vector $\vec{x}_i$ of length W, where each element $\vec{x}_{i,w}$ equals the raw count of keyword w in text i. Then, we group those identified keywords by topic to obtain a central theme (topic) vector of length K for text i. Finally, the model filters the topic counts and generates a normalized topic vector $\vec{y}_i$ of length K, where each element $\vec{y}_{i,k}$ equals the probability of topic k for text i.

The topic-count filtering controls the number of topic(s) we assign to a single text. We have explored three options: (a) the "primary" option treats this as a single-label classification task, in which each text gets one topic label that is most frequently mentioned and obtains a one-hot $\vec{y}_i$ with a single non-zero element, (b) the "primary + secondary" option assigns the first and the second most frequent topic to a text, with topic weights corresponding to the relevant frequency, and (c) the "all" option includes all topics identified in a text, weighting topics based on relevant frequencies.

Validating output

We finalize and validate our model output by comparing model-human agreement against human-human agreement. Our evaluation rests on the following premise: If the dictionary-based model performs as reliably as a human labeler, the extent to which the model agrees with a random human labeler should be comparable to the extent to which two random human labelers agree with each other. To perform the aforementioned evaluation, we collect human labels through a topic-labeling task on Amazon Mechanical Turk (MTurk), in which we ask MTurk workers to select the primary, secondary and all relevant topics applicable

ID	Topic Fullname	Topic Abbr.	Topic Description (guidance, or a few example sub-categories for each topic)
1	trump controversies	TRUC / REPC	controversial topics related to Trump, such as family or personal scandal, health condition speculations and disputable remarks
2	clinton controversies	CLIC / DEMC	controversial topics related to Hillary, such as family or personal scandal, health condition speculations and disputable remarks
3	biden controversies	BIDC / DEMC	controversial topics related to Biden, such as family or personal scandal, health condition speculations and disputable remarks
5	agriculture	AGRI	agriculture policy, trade & marketing; farmers; fisheries & fishing; animal & crop disease
6	civil rights	CIVR	racial equality; gender equality; voting rights; freedom of speech; gun rights; right to privacy; age discrimination; anti-government activities
7	crime	CRIM	law enforcement agencies; crimes & crime control; police; prisons; court administration; child abuse & family issues
8	culture	CLTR	cultural policy; culture & entertainment
9	defence	DEFC	defence alliance & agreement; military intelligence; nuclear arms; military aid; military procurement; domestic security responses; foreign military operations
10	economy	ECON	banking; small businesses; disaster relief; tax policies; consumer finance; insurance regulation; bankruptcy; corporate management; securities & commodities
11	education	EDUC	education policy; elementary & primary schools; vocational education; higher education, student loans; education of underprivileged students
13	energy	ENRG	energy policy; nuclear; electricity; natural gas & oil; coal; alternative & renewable; conservation & efficiency; research & development
14	environment	ENVR	environmental policy; drinking water; waste disposal; hazardous waste; air pollution; recycling; species & forest; land and water conservation
15	foreign trade	FRTR	trade agreements; exports; private investments; tariff & imports; exchange rates; competitiveness; trade policy
16	government operation	GVOP	general governmental operations; intergovernmental relations; bureaucracy; census & statistics; postal service; procurement & contractors
17	healthcare	HLTH	public health and candidates' health conditions; coronavirus spread & control; healthcare reform; insurance; medical facilities; disease prevention; healthcare research & development
18	housing	HOUS	community development; urban development; rural housing; low-income assistance; housing for veteran, the elderly & the homeless
19	immigration	IMMI	immigration issues & policies; refugees; citizenship
20	international affairs	INTL	international affairs & foreign aid; resources exploitation; developing countries; international finance; human rights issues; terrorism; international organizations
21	labor	LABR	labour, employment & pensions; employee benefits; labor unions; fair labor standards; worker safety; employment training; youth employment
22	religion	RELG	general religious issues; religous groups; church activities; religious freedom
23	social welfare	SOWL	social welfare policy; low-income/elderly/disabled assistance; volunteer associations; child care
24	space, science, technology, & communications	SSTC	issues related to general space, science, technology & communications: mass/social media presence, space programs, telecommunication regulation
25	transportation	TRSP	mass transportation construction; highways, air & railroad travel; maritime transportation; infrastructure

Table 2.2: List of topics in our dictionary. "Candidate controversies" shows up in their corresponding election year (e.g., Biden controversies only show up in 2020).

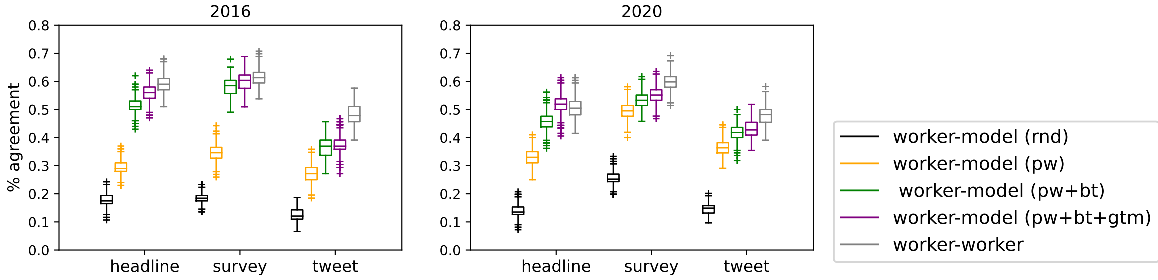

Figure 2.1: Agreement scores for model-human and human-human pairs. The random baseline adopts the topic distribution generated by the best-performing topic model and shuffles the topic label per text. “rnd” refers to the baseline model that randomly assigns a topic label; “pw” refers to the model using the original CAP dictionary; “pw + bt” refers to the version after the initial merge with the extended dictionary; “pw + bt + gtm” refers to the final version that included GTM keywords.

for each of the 10 texts displayed per Human Intelligence Task (HIT). We describe details of the MTurk task in Appendix A.1.

We include detailed evaluation steps in Appendix A.2, and report concrete agreement scores in Appendix Table A.2. We find that the best performance is achieved when using only the strongly relevant keywords and limiting our attention to the primary topic. Using this model version, we show the progression of model-human agreement in Figure 2.1 as we update the topic dictionary. The figure shows that the worker-model agreement scores are comparable to worker-worker agreement scores, especially for the news media and survey data. We also see that the progression is consistent across the two elections, allowing us to make reliable comparisons between the two elections. Although the suboptimal level of overall human-human agreement reflects the inherent difficulty of the labeling task itself, the model does identify meaningful topical cues from the text that make sense to humans in a considerable proportion of texts. In circumstances of conflicts between human-model pairs, roughly 42% of the texts have at least one matching pair of human and model labels, and 57% have at least one human including the primary model topic in their expanded topic list (i.e., primary, secondary or relevant topics). For topics that have low agreement levels between the model and workers and occur rarely in our data[7], we drop them in the downstream analysis.

[7]Topics excluded downstream: “forestry”, “land water management”, “agriculture”, “housing”, “transportation”, “culture”.

2.4.2 Downstream analysis

Measuring agenda alignment

We first assess agenda alignment by comparing the aggregated attention distribution between different media types. The attention distribution can be described at both the keyword level and the topic level. At the keyword level, we compute an aggregated keyword vector $\vec{x}_A$ for media type A by adding up[8] keyword vectors $\vec{x}_i$ for each text i from media type A, and normalizing the output by its sum. Similarly, at the topic level, we obtain an aggregated topic vector $\vec{y}_A$ for media type A, which sums up all topic vectors of texts from media type A and normalizes the output by its sum. An aggregated topic or keyword vector is essentially a probability vector that sums up to 1. With these aggregated vectors, we use Pearson correlation coefficient, a widely-adopted metric in previous agenda-setting studies [e.g., 166, 149] to quantify the degree to which the priorities of candidate attributes align between media type A and B, i.e., $\rho(\vec{x}_A, \vec{x}_B)$ for the alignment at the aspect (keyword) level and $\rho(\vec{y}_A, \vec{y}_B)$ at the central theme (topic) level. The higher the correlation, the better the agendas align.

Since all domain snapshots are timestamped, we can define the time frame of inputs when aggregating topic or keyword vectors and measure the degree of alignment over time (i.e., temporal alignment). For instance, we can measure the daily level of temporal alignment using aggregated topic (keyword) vectors generated from headlines on a given day.

Identifying agenda leader and follower

Agenda alignment reveals how much the priorities of candidate attributes match between two media within concurrent time frames, yet it does not capture the dynamics of agenda flow over time or assess the IAS power of a given media type. Thus, a natural next step is to explore the temporal relationship of media agendas and assess the degree to which a given media type serves as a leading/following actor in the IAS process.

We perform Granger causality tests for daily time series of attribute proportions. Granger causality analysis is a classic approach to evaluate the (intermedia) agenda-setting power using time series [e.g., 167, 168, 169, 145, 141]. It allows us to statistically assess the temporal "causation" between two time series with varying time lags[9]. Let's say we focus on attribute k (e.g., topic "healthcare" for Trump), and extract the time series of its attention proportion in media type A (e.g., low-credibility media) and B (e.g., high-credibility media), $X_{A,k}$ and $X_{B,k}$. If regressing the past of $X_{A,k}$ and $X_{B,k}$ yields a better prediction for $X_{B,k}$ than

[8] As mentioned in Section *Data and Preprocessing*, this step computes a weighted sum of all input vectors. The topic (keyword) vectors for individual texts are weighted by the multiplication index assigned to each snapshot to avoid over-counting (under-counting) domains that have too many (few) snapshots per day.

[9] We apply time lags of 1, 2, 3, 4, 5 days for individual topic time series.

regressing only the past of $X_{B,k}$, we say $X_{A,k}$ "Granger causes" $X_{B,k}$ and in our context, we identify media type A as the agenda leader and B as the agenda follower on attribute k. Before being fed into the Granger causality tests, all time series have been detrended by first-level differencing and have passed augmented Dickey-Fuller (ADF) tests, which indicate that they are stationary.

Finally, we collapse the results yielded by different time lags into four categories: (i) led by media type A if we *only* see significant results in cases of B lagging A; (ii) led by media type B in the reversed situation; (iii) mutual interaction if we see significant results in both directions; and (iv) no relationship if we do not see significant results in either direction. We consider a collapsed result to be significant and robust if (i) the Granger causality test returns $p < 0.05$, so we can reject the null hypothesis that changes of attention on attribute k in media type A fail to Granger cause the changes of attention in B; and (ii) the same result category appears consistently in at least 95% of the bootstrapping runs. Apart from performing the test on the full time series, we also test a shorter time window of 90 days and slide it by daily unit. The daily sliding allows us to distinguish the persisting roles of agenda leaders/followers from flashing patterns boosted by momentary and spurious correlations.

2.5 Results and discussion

We provide two views when comparing agendas across different media types: (1) static alignment, which examines the similarity of aggregate agendas, and (2) temporal dynamics, which characterizes the extent to which a given media type leads the other type in a lagged time frame.

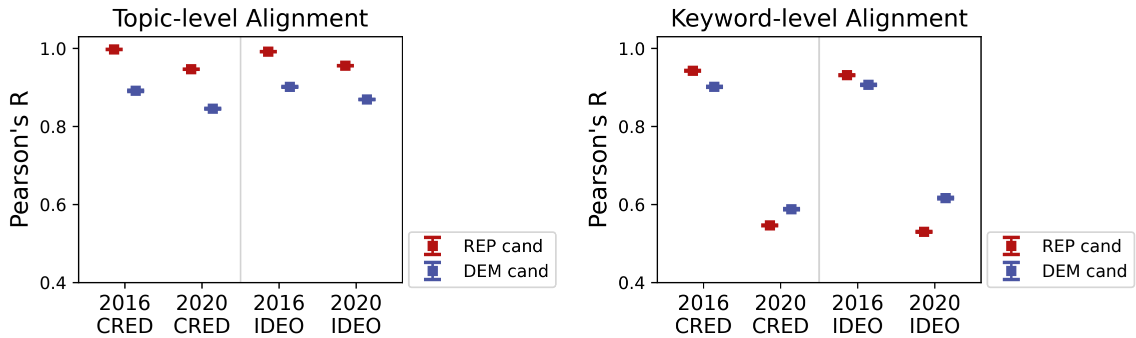

Figure 2.2: Agenda alignment for the 2016 and 2020 election seasons, at the topic level (left) and at the keyword level (right). We keep the top 18 topics that occur frequently and meaningfully in the data, and keep the top 500 keywords with high frequencies. CRED refers to media types by credibility (i.e., low- and high-credibility media), and IDEO refers to media types by ideology (i.e., left- and right-leaning media).

2.5.1 Static alignment: divergence or convergence of media agenda?

We start by presenting the aggregate alignment in Figure 2.2. The first striking finding is the similarity in patterns observed for ideology and credibility. Past work that compares the role ideology and credibility play in news production has found that credibility plays a more significant media fragmentation role [170]. Based on this, we would have expected the media to be more fragmented along the credibility dimension. However, surprisingly, we see that media are no more divided along the credibility line compared to ideology. This could be because ideology is one of many factors that shape broader news production, while it is *the* decisive factor in election campaign coverage. This highlights the enduring role ideology plays in election coverage.

At the topic level, agendas of different media types are still largely aligned in both years ($r > 0.8$), with a slight decrease from 2016 to 2020 (average $\Delta r = -0.041$). In contrast, at the keyword level, the correlations for both candidates have dropped dramatically (average $\Delta r = -0.350$), reaffirming previous findings of a more severe fragmentation at the aspect level than the central theme level [46, 45, 155]. The downtrend in keyword alignment is more pronounced for the Republican candidate (average $\Delta r = -0.399$ per media pairing) than for the Democratic candidate (average $\Delta r = -0.302$ per media pairing), evincing partisan asymmetries in agenda fragmentation at the keyword level.

Candidate-wise, the coverage of the Republican candidate is generally better aligned than that of the Democratic candidate. As shown in Figure 2.2, at the topic level, low- and high-credibility media share highly similar priorities for Trump's attributes in both years ($r = 0.997$ in 2016 and $r = 0.946$ in 2020); so do left- and right-leaning media ($r = 0.991$ in 2016 and $r = 0.956$ in 2020). For his opponent candidate, the correlations between these two media pairings are weaker ($r = 0.891$ in 2016 and 0.845 in 2020 across credibility types; $r = 0.902$ in 2016 and 0.869 in 2020 across ideology types). Interestingly, in 2020, the coverage of Trump achieves a higher level of alignment than that of Biden at the topic level ($\bar{r} = 0.951$ per media pairing for Trump and $\bar{r} = 0.857$ for Biden), but not at the keyword level ($\bar{r} = 0.538$ per media pairing for Trump and $\bar{r} = 0.602$ for Biden). This reveals the level at which media diverge for a given candidate. Different types of media organize similar priorities for Trump-related topics, but the specific keywords used in their discussions are poorly coordinated. Whereas for Biden, although the topic-level agendas are not as well aligned as his Republican counterparts, the keywords used in different media types have relatively greater overlaps.

2.5.2 Candidate controversies: key attributes as divergence drivers

After observing a higher level of agenda divergence for the Democratic candidate (as opposed to the Republican candidate) and for 2020 (as opposed to 2016), we question the source of these differences. On what attributes do media diverge the most? Is the overall pattern of divergence dominated by the divergence on a few attributes or, more or less equally by the divergence on most attributes? Thus, we extend candidate-wise and election-wise comparisons into topic-level and keyword-level breakdowns.

Candidate-wise, we find that higher proportions of attention on "DEM candidate controversies" (topic related to Clinton controversies in 2016 and Biden controversies in 2020) from low-credibility and right-leaning media are the main source of salient agenda divergence for the Democratic candidate. Focusing on blue dots in Figure 2.3, we see that the two largest topics that deviate significantly from the diagonal line are "DEM candidate controversies" and "government operations". Low-credibility and right-leaning media highlight "DEM candidate controversies" more than their counterparts, limiting the attention devoted to "government operations". Such deviations have become more salient in 2020. We again observe similar patterns for the credibility and ideology divide.

The attention on "DEM candidate controversies" not only diverges at the aggregated level but also signals the specific point in time when different media types will diverge. To illustrate this, we examine the temporal dependence between the attention disparity on "DEM candidate controversies" and the temporal fluctuations in overall topic alignment. Specifically, we apply a uni-variate ordinary least squares (OLS) model for topic k, using the time series of temporal alignment as the dependent variable Y, and the time series of the temporal *difference* in the proportional attention on topic k between a given media pairing as the independent variable X. We find that regressing the difference in "DEM candidate controversies" can explain more of the variance (i.e., achieve the highest R-squared values) than any other topic, especially for Biden in 2020 between left- and right-leaning media (R-squared = 0.7565). We report the full results in Appendix Table A.1. Comparing the time series of "DEM candidate controversies" between low- and high-credibility media (see Figure 2.5) with that of temporal topic alignment between low- and high-credibility media (blue lines in Appendix Figure A.3 A1 and A2), we see that agenda divergence is brought forward by the misaligned attention spans or the different spotlighting intensity on "DEM candidate controversies". For Clinton's case in 2016, for instance, we can link some dramatic drops in temporal alignment to the time periods when low-credibility media discussed "Clinton controversies" much longer than high-credibility media after breaking events such as Bill Clinton and Loretta Lynch's meeting in early July and Hillary Clinton fainting in mid-September.

Furthermore, at the keyword level, we notice that the drop in keyword alignment for "REP candidate controversies" contributes greatly to the drop in overall keyword alignment from 2016 to 2020. The patterns observed for ideology are, again, similar and show interesting parallels. As shown in the left two subfigures in Figure 2.4, "REP candidate controversies" is among the noteworthy topics that have a dramatic drop in keyword alignment from 2016 to 2020 and that occur frequently enough to have a sizable impact on the overall alignment.

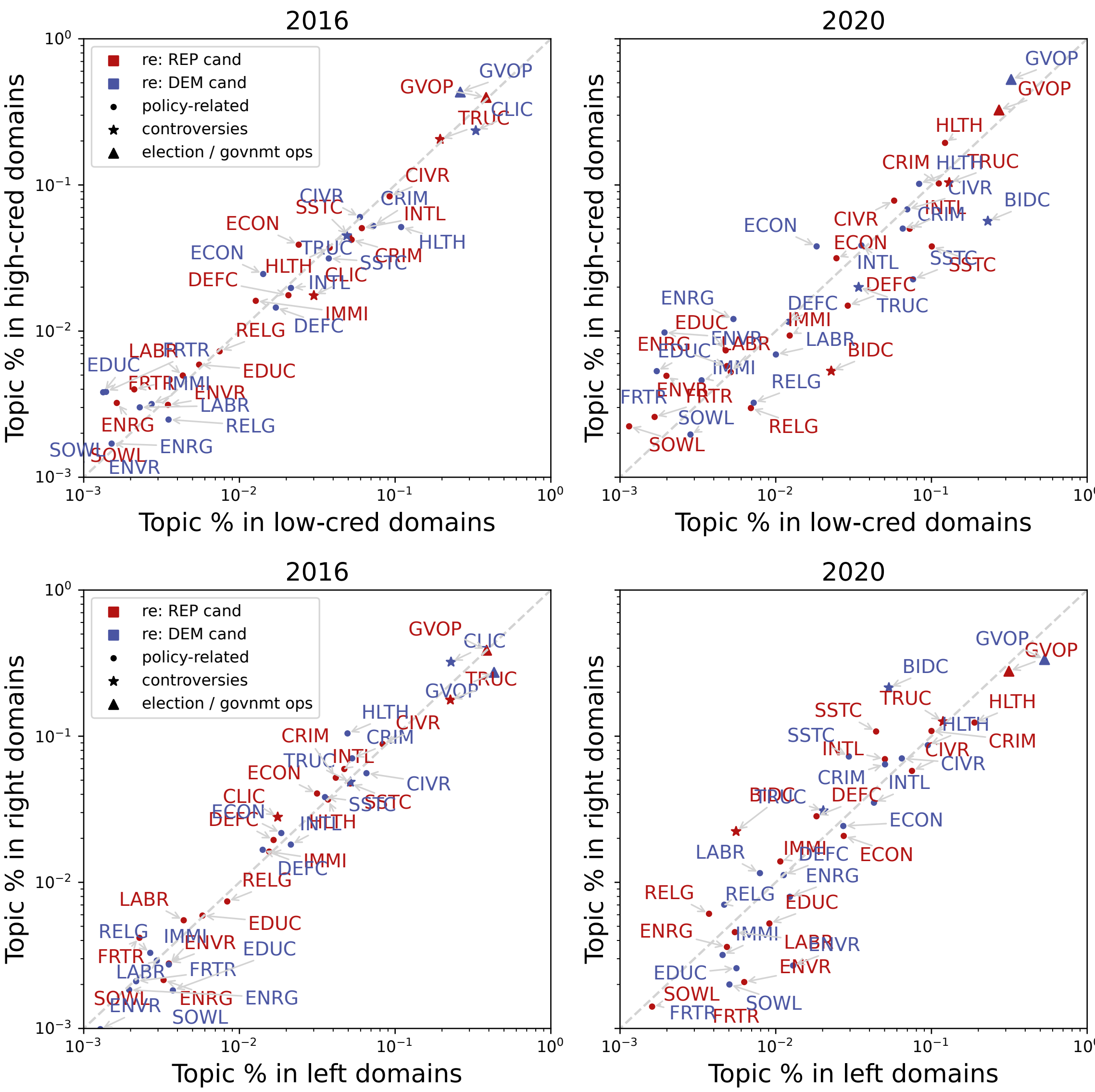

Figure 2.3: Comparisons of topic proportion between different types of domains in 2016 (left) and 2020 (right). The upper two figures compares topic proportions between low- and high-credibility domains, while the bottom two figures compares topic proportions between left- and right-leaning domains. Please refer to Table 2.2 for the list of topics the abbreviations in the figure stand for.

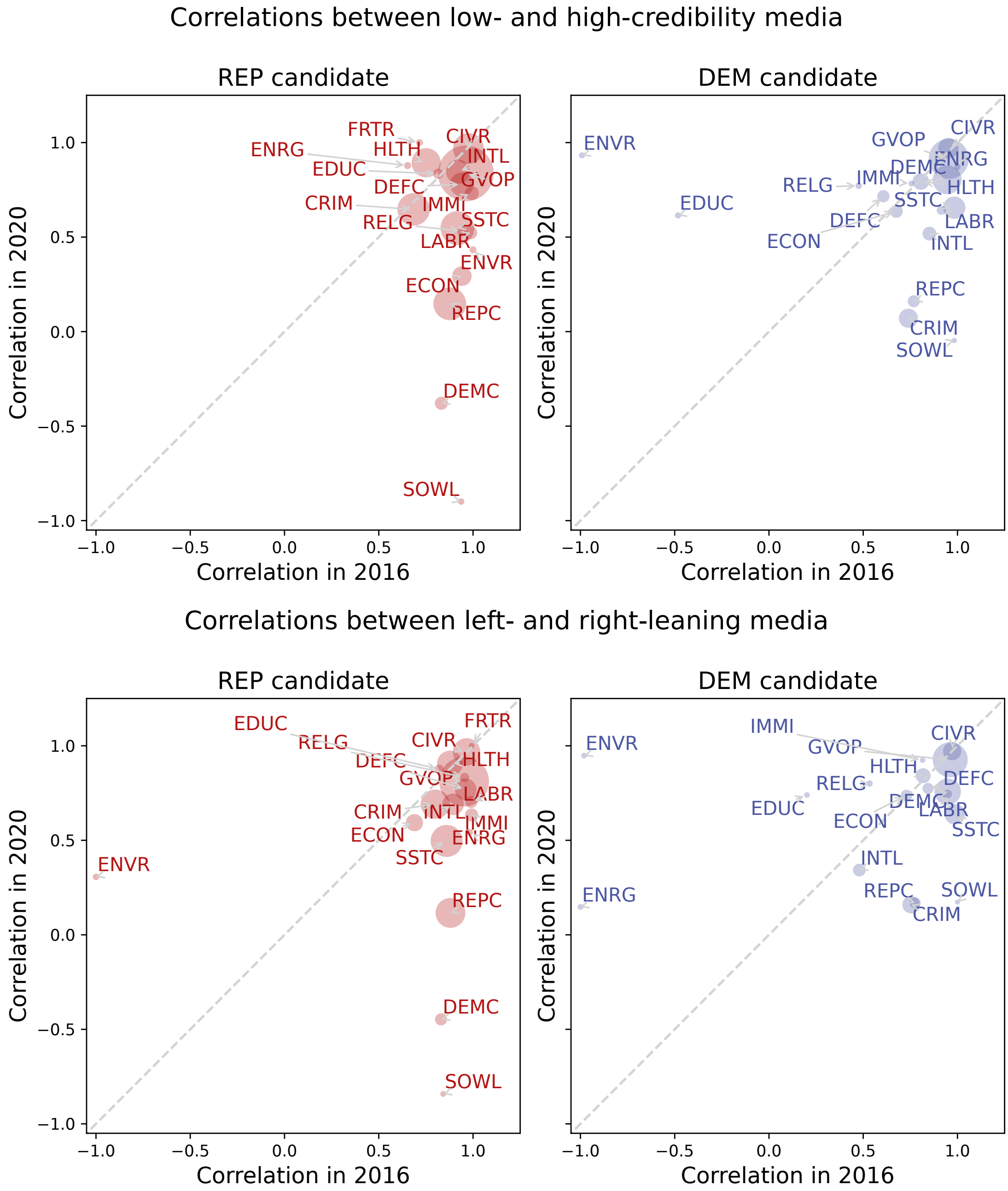

Figure 2.4: By-topic alignment of top 500 keywords across years between different types of domains in 2016 and 2020. The left (right) two figures show the alignment for Republican (Democratic) candidates, and the upper (bottom) two figures show the alignment between low- and high-credibility (between left- and right-leaning) domains. Dot size is a function of the overall frequency of the corresponding topic. Please refer to Table 2.2 for the list of topics the abbreviations in the figure stand for.

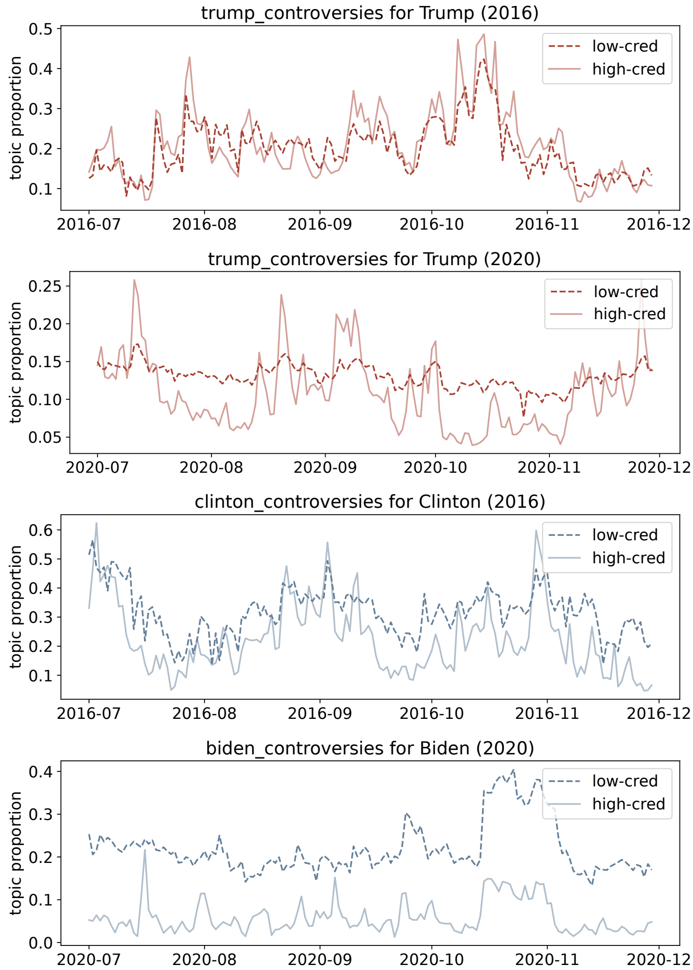

Figure 2.5: Time series of the topic "REP/DEM candidate controversies" from low- and high-credibility media in 2016 (left) and 2020 (right). Blue rectangle boxes in the second row highlight a few periods with drastic drops in overall topic alignment, when low- and high-credibility media have different spotlighting intensity or attention spans on "DEM candidate constroversies".

Looking closer at specific aspects (i.e., keywords) addressed for “REP candidate controversies”, we see that high-credibility media dedicate more attention to Trump’s family members; and that low-credibility media put the spotlight on the deep-state conspiracy, and push stories co-mentioning Trump with figures such as Jeffrey Epstein, Adam Schiff, and Roger Stone. While there is a lack of consensus on central aspects of “REP candidate controversies” across different media types in 2020, the divergence of aspects for “DEM candidate controversies” is much weaker[10] (see the right two subfigures in Figure 2.4). Conditioned on the topic “DEM candidate controversies”, we notice that controversies centered around Hunter Biden are heavily debated on both sides of media with high occurrences of keywords “Hunter Biden”, “laptop” and “Ukraine”.

2.5.3 temporal dynamics: who leads and who follows?

Next, we shift our focus from concurrent correlations within the same time frame to temporal correlations between lagged time frames. We describe the IAS dynamics captured between low- and high-credibility media, and briefly contrast it with the dynamics between the left- and right-leaning media as a reference system.

We assess the IAS power based on (i) the number of attributes one media type leads for the other, as well as (ii) the length of time period during which such IAS power can persist. We summarize this information in the sliding-window plots displayed in Figure 2.6, where the starting points of all 90-day windows with a significant and robust Granger causality result are marked with plus signs ($+$). Each plus sign is followed by 90 dots ($\cdot$) colored the same as the plus sign to visually demonstrate the full length of sliding windows. For example, for Trump 2016 there is only one time window with significant and robust results on the attribute “economy (ECON)”, spanning from early July to early October.

Overall, we see that high-credibility media serve as the dominant actor in IAS, setting the agenda for more candidate attributes than low-credibility media. Out of the top 10 attributes that appear frequently in a given year, high-credibility media lead the agenda of 5.5 attributes for the Republican candidate and 3 attributes for the Democratic candidates on average, with varying window lengths[11]. Meanwhile, we do not see low-credibility media persistently leading the agenda on any attribute in either election season. Despite the encouraging results, we observe that high-credibility media’s IAS power has declined from 2016 to 2020 for the Democratic candidate, with a decrease in terms of the number of attributes it leads

[10] Among the top 500 frequent keywords for Trump-related texts in 2020, we correlate keywords that belong to “Trump controversies”. Pearson’s R = 0.1446 between low- and high-credibility media; Pearson’s R = 0.1139 between left- and right-leaning media. In Biden-related texts, the corresponding Pearson’s Rs for keywords belonging to “Biden controversies” are 0.8095 between low- and high-credibility media, and 0.7597 between left- and right-leaning media.

[11] We count attributes with at least three consecutive windows showing consistent causality results.

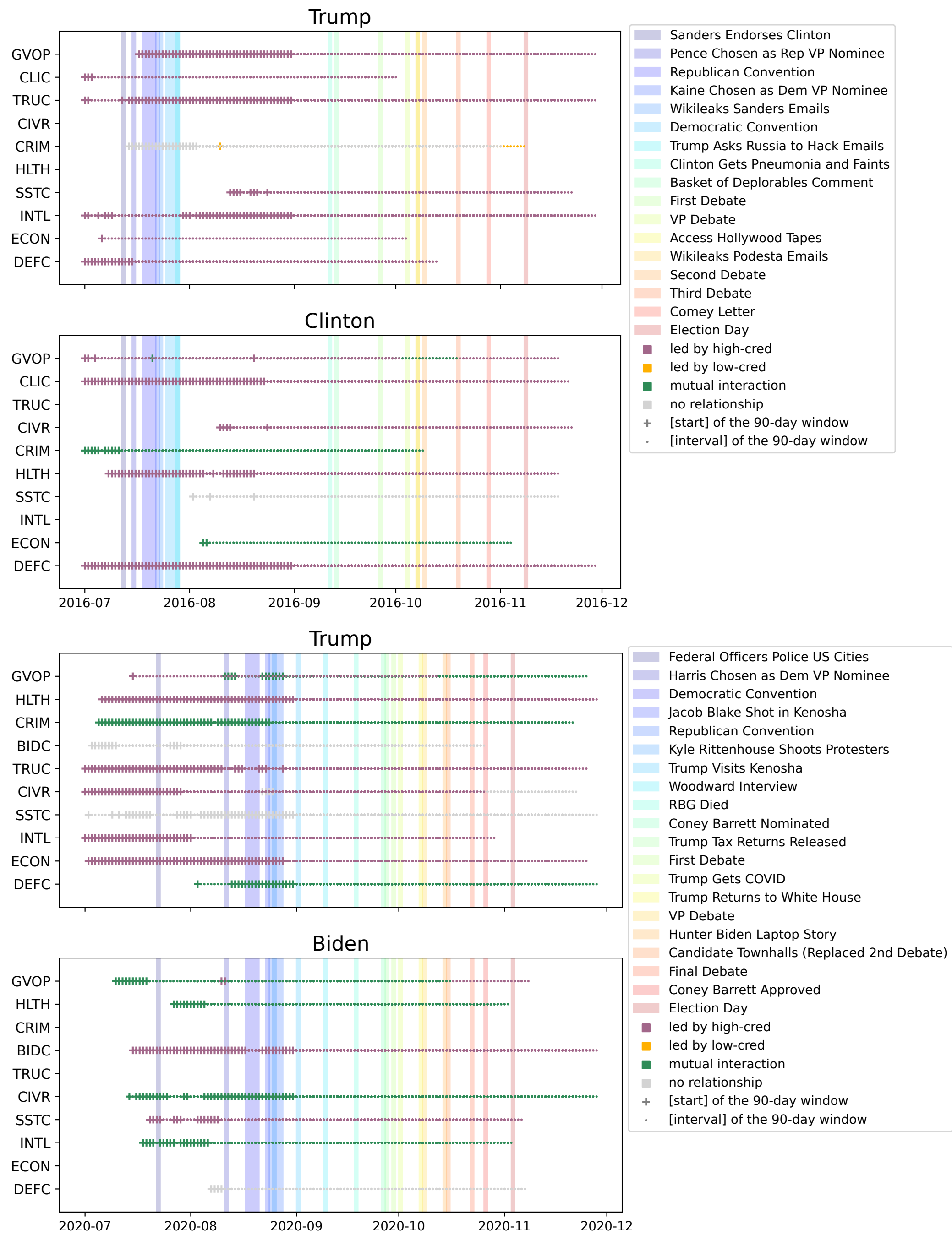

Figure 2.6: Granger causality results between low- and high-credibility media for 2016 (the first and second figures) and 2020 (the third and fourth figures). We test Granger causalities in a sliding window of 90 days and display robust results that appear in more than 95% of the bootstrapping runs. Each plus sign marks the starting point of the 90-day sliding window with a significant result. We include results for the top 10 topics that show frequently in all news headlines for a given election year.

(from 4 to 2), and the total number of windows it leads (from 165 to 60). Furthermore, agendas between high- and low-credibility media appear to be more intertwined, mutually interacting with each other on 3 attributes for Trump and 4 for Biden in 2020, but only 1 for Clinton in 2016.

Notably, while "candidate controversies" acts as a crucial attribute that drives the divergence of the media agenda, discussions of "REP (DEM) candidate controversies" in the coverage of the Republican (Democratic) candidate are always led by high-credibility media. Based on Figure 2.6, the longest consecutive high-credibility-leading windows (i.e., consecutive time windows with a significant and robust result of high-credibility media taking the lead) lasts for 135 days (45 windows) for the Republican candidate and 134 days (44 windows) for the Democratic candidate on average[12]. Moreover, the election-wise comparison re-iterates the diminishing IAS power of high-credibility media specifically on "DEM/REP candidate controversies", as the length of the longest consecutive high-credibility-leading windows shrinks from 141.5 days (51.5 windows) per candidate in 2016 to 127.5 days (37.5 windows) per candidate in 2020. Such shrinking in window length happens more severely for the Democratic candidate (ΔL = -20 days) than the Republican candidate (ΔL = -8 days).

To sum up, high-credibility media is more powerful in IAS compared to low-credibility media, as it leads the agenda for more attributes and consistently for longer periods of time; however, the IAS power of high-credibility media has declined from 2016 to 2020, together with a few more attributes seeing mutually interacting agendas in 2020 (e.g., "crime" for Trump, "government operation" and "healthcare" for Biden). Contrasting these patterns with the IAS dynamics between left- and right-leaning media (see Appendix Figure A.5), we see shared patterns between high-credibility and left-leaning media in terms of their dominant role in IAS in general, as well as their weakening leader advantage from 2016 to 2020, especially for the Democratic candidate. While some level of symmetry does exist between credibility and ideology, we see the value of separately addressing IAS along these two dimensions. Right-leaning media clearly take a more active role than low-credibility media in 2020, persistently setting agenda for a few attributes of the Republican candidate (e.g., leading "civil rights" and "international affairs" for Trump in 2020).

2.6 Ethics statement and broader impact

Given the political context of the case studies, we understand and try to minimize the risk of misinterpreting spurious patterns. We test the significance and robustness of the results by bootstrapping and de-noise the temporal volatility through sliding-window analysis. We also re-iterate the correlational basis of our analysis.

[12]The length of a set of consecutive time windows equals the length of one time window (i.e., 90 days) plus the number of consecutive time windows within the set, as the unit per slide is one day.

Apart from cautiously deriving the implication, we have incorporated the following ethical considerations: (1) once collected and preprocessed, the headline dataset is stored on the server with restricted access; (2) we remove personally identifiable information from the MTurk output; (3) we actively communicate with MTurk workers who raise questions or concerns, and make sure that those who attentively work on the labeling tasks are fairly compensated (even if they fail the screening); and (4) we release the dictionary and the model source code in a GitHub repository[13].

In an era marked by growing concerns of polarization and fake news, our study enhances the understanding of IAS along both the credibility and the ideology dimensions by providing detailed comparisons at multiple levels of granularity. We hope to inspire open dialogues among media entities, policymakers, and the public to address challenges evidenced by the alarming trend in our results–the decline in the IAS power of high-credibility media.

2.7 Conclusions and limitations

In this chapter, we re-examine IAS theory for news headlines related to presidential candidates during the 2016 and 2020 U.S. presidential elections.

Overall, we observe a high level of agenda alignment in candidate coverage between low- and high-credibility media. The agenda convergence indicates that low- and high-credibility media still share a common ground for candidate-related discussions on broad issues; however, the initiator of such assimilation remains unclear. Low-credibility media could be borrowing stories from traditional players with higher credibility levels, due to their limited resources to independently produce impactful news stories in the fierce attention battleground. Alternatively, past work also shows that traditional media can spread misinformation, especially by indexing political elite talking points [171]. High-credibility media might be loosening their journalistic standards in order to attract and retain their audience, generating stories of disputable issues that are easier to be re-packaged into low-credibility clickbaits, as we see in our results the significant proportions of attention devoted to "candidate controversies" rather than policy topics in high-credibility media.

In addition, our study adds to the growing body of literature that highlights partisan asymmetries in the news ecosystem [e.g., 172, 173] by demonstrating the stronger alignment in agendas for the Republican candidate compared to the Democratic candidates. Past work shows that media exhibit their bias largely through negative depictions of the opposing side, as opposed to positive depictions of the preferred side [174, 175]. Here, the diverging agendas for the Democratic candidate provide evidence that such bias may extend to selective coverage of topics.

[13] https://github.com/yijingch/intermedia-agenda-setting

We also observe meaningful shifts of IAS dynamics when comparing results between 2016 and 2020, which underscore two valuable insights. First, agenda divergence between different types of news media has increased noticeably over this period. In line with studies of the same election cycles using social media data [176], this suggests that agenda fragmentation has grown not only among mass media, but also among influencers and users on social platforms as a multimedia phenomenon. This parallel points to a promising direction for future work—extending IAS analysis to social media texts. Second, this chapter reveals a decline in the IAS power of high-credibility media, which may signal that low-credibility media are gaining greater autonomy in setting their own agendas, or even influencing the agendas of traditionally dominant high-credibility media. This finding resonates with previous work documenting the growing visibility and distinctive role of hyperpartisan and low-credibility outlets [e.g., 177, 145, 141]. My results extend this literature by demonstrating that such outlets are not merely competing for audience attention, but may also be destabilizing the traditional IAS hierarchy. This complicates–though does not fully overturn–the classic assumption that elite, high-credibility media primarily lead the agenda [178, 179, 142]. Third, our findings caution against overgeneralizing from single-case studies: the U.S. news ecosystem remains in flux, with IAS power shifting across media types. These results underscore the value of large-scale empirical research to test and refine well-established communication theories using contemporary data.

There are various limitations to this work. First and foremost, although we use terms such as "influence" and "Granger causation" when assessing IAS, the IAS process captured in our study is based on correlational analysis. While we follow the terminology used in scholarship and theorize about setting the agenda, we caution the reader that the associations found here are not sufficient evidence of a causal relationship. Secondly, our dictionary-based model utilizes context-specific topics and keywords related to a certain issue (e.g., presidential candidates), which limits its generalizability. While the dictionaries themselves are not generalizable, the pipeline we introduced for constructing and validating topic dictionaries is. Our modeling approach allows interpretability and cross-year comparisons. Finally, we use a set of existing source lists for determining the credibility of different websites, where sizable disagreements exist across lists constructed by different fact-checkers and scholars [180]. Furthermore, the limited coverage of ideology labels has restricted our scope of analysis when comparing left- and right-leaning media. We encourage future studies to explore more source lists of domain credibility and ideology, and incorporate a better-labeled dataset for such parallel analysis.

Our findings also identify new directions for future work of IAS. First, it is worth following up with more recent datasets to examine if the IAS trends identified in 2016 and 2020 continue in future elections. Second, we notice a few signals for the insufficient explanatory power of the current models in capturing temporal "causation" between left- and right-leaning media. For instance, in Appendix Figure A.5, we see the absence of significant and robust IAS results in most 90-day sliding windows, particularly for the Democratic candidate. This may be a substantive finding: left- and right-wing media fail to set each other's agendas. Or, this may be a result of linear regression models failing to capture the increasingly complicated agenda interactions. Such obscurity invites future explorations of different methodologies to validate or extend our findings. Third, our parallel analysis points out that the divisions along ideology and credibility share some structural features but are not entirely overlapping. Future work could look into the interplay between these two dimensions.

Finally, the findings from this chapter set the stage for a deeper inquiry into the nature of political polarization–a central theme in Chapters 3 and 4. By analyzing IAS patterns between low- and high-credibility outlets as well as between left- and right-leaning media, this chapter has illuminated how the supply side of political information is structured, and how presidential candidate coverage diverges across different types of news producers. These patterns not only shape the information environment to which individuals are exposed but also condition the opportunities and constraints through which citizens develop political attitudes and engage with political content online, both of which are closely linked to the phenomenon of political polarization. This chapter provides an essential context for the following analyses of the *demand* side of political communication, which turn to individuals–their belief systems and their online political engagement–as key sites for addressing the challenge of political polarization.

Chapter 3

Multidimensional belief modeling: Using response-item networks (ResIN) to measure ideological polarization

Following a background study of political news dynamics among mass media, this chapter turns to a specific challenge in political communication: political polarization. As summarized in Section 1.3, polarization can be studied via observations of both attitudes and behaviors. In political and social sciences, scholars typically measure polarization based on attitudinal data collected through surveys and define polarization within the scope of political beliefs. This chapter dives deeper into this attitudinal perspective, elaborating on existing measurement frameworks for political ideologies and belief systems–the fundamental components for understanding polarization in the realm of attitudes, and introducing a novel network-based method–Response-Item Network (ResIN)–to the study of polarization, which addresses the complexity of belief systems and the limitation of current approaches. Through a case study of ANES (2000-2020), I show that ResIN is a valuable tool for visualizing political cleavage lines and measuring polarization at both the system and attitude levels.

3.1 Introduction

Although more than eight decades have passed since Converse famously advocated for analyzing political ideologies as networks of interrelated beliefs [56], scholars have only recently begun to model political attitude systems as dedicated statistical networks [e.g., 128, 181, 123, 129, 182]. By conceptualizing individuals' opinions on political issues as nodes and the associative strength between them as weighted edges, belief network analysis

(BNA) not only provides a methodological framework that well-aligns with Converse's original conception, but also allows researchers to rigorously investigate ideologies as wholistic, system-level phenomena. However, BNA does not adequately capture the spatial dimension of political ideologies, that is, their role in representing abstract dividing lines–such as the familiar left-right or liberal-conservative spectrum–that demarcate political conflict within societies [183, 184, 185, 186, 92, 187, 188]. Since key processes of ideological realignment and polarization are often framed as positional shifts along such dividing lines, BNA offers only limited insights into the structural transformations within belief systems during such dynamics.

We aim to bridge this gap by applying Response Item Networks (ResIN)–a novel class of belief network models operating at the issue response level–to the study of ideological polarization. While previous works on ResIN have either focused on validating its methodology [189], or applied it to static phenomena such as vaccination attitudes [190] and social identity construction [191], our focus here is to explore ResIN's potential to understand political attitude polarization as a dynamic, system-wide process (cf. [192]), without assuming the existence of latent variables. Using six waves of American National Election studies (ANES) from 2000 to 2020, we demonstrate ResIN's capacity in capturing multi-level shifts in U.S. belief systems during ideological polarization, presenting the spatial transition from a largely unstructured, to a fully structured, polarized opinion space.

By visualizing the polarization process among five key political issues in the U.S., we notice that the underlying dynamic resembles the "breaking of an egg", that is, the transformation of an initially amorphous, or egg-shaped belief system into a single, polarized, left-right ideological dimension. Furthermore, we show how ResIN-based metrics can enhance our understanding of this process both at the system and the attitude level. Our proposed system-level measures are able to reveal increasing trends both in belief constraint and polarization, as well as interesting patterns of partisan asymmetries that underlie both processes. At the attitude level, our proposed statistics can identify specific roles different attitudes play in constraining belief systems of different partisan groups as well as which attitudes build bridges between them. We argue that these results and measurements cannot be easily observed in standard BNA, demonstrating that ResIN offers a novel angle to explore ideological polarization with belief networks.

Our contributions to the fields of BNA and ideological polarization are twofold. First, by applying ResIN to study political polarization, our work adds to the relatively scarce literature on conceptualizing polarization in belief networks. Taking this perspective is promising for uncovering psychological mechanisms that drive ideological polarization. Second, our measures at the system and attitude levels evaluate the degree of polarization across various facets in one cohesive framework. Using ResIN, we are able to connect the mental maps of how individuals organize political attitudes, with the collective patterns of how the public structurally constrains different beliefs. A model that encodes this individual-collective nexus, which is a crucial aspect for deconstructing ideological polarization [193], may allow us to synthesize insights of polarization studies at different levels.

The remainder of this chapter is organized as follows. In Section 3.2.1, we review the notion of political belief system since its introduction by Converse [16], articulating its definition and connections with the formation of ideologies in the mass public. We then refresh the conceptualization of ideological constraint and polarization and discuss the strengths and limitations of existing BNA frameworks in Section 3.2.2. Then, in Section 3.3, we describe the dataset for this study, specify the steps of constructing ResIN snapshots, and introduce different measures of ideological constraint and polarization afforded by ResIN. We show the results and demonstrate ResIN's advantages over BNA in Section 3.4, acknowledge the limitations in Section 3.5.

3.2 Related works

3.2.1 Belief systems and ideologies

> Certain postures tend to co-occur and this co-occurrence has obvious roots in the configuration of interest and information that characterize particular niches in the social structure, ..., not simply because both are in the interest of the person holding a particular status but for more abstract and quasi-logical reasons developed from a coherent worldview as well. It is this type of constraint that is closest to the classic meaning of the term "ideology" (8).
>
> – Converse, The nature of belief systems in mass publics (1964)

Political attitudes rarely exist in isolation. For instance, support for increasing government spending on social welfare is usually linked to favoring higher income taxes for wealthy citizens in many Western democracies [194]. Public opinion scholars have long relied on the concept of constraint, i.e., the functional interdependence between two or more issue positions [16], to study the prevalence of connected political beliefs at the individual [195, 196] and the collective level [128, 197, 182]. Viewed through this lens, an ideology, then, is a package of constrained beliefs that have been logically, quasi-logically, or socially developed to form an integrated framework [16, 188].

Building upon this conceptualization, our understanding of mass ideologies further rests on two key premises. First, we view ideology as a latent construct at a level of higher abstraction, which stems from, and might exert influence on multiple issue positions and identities [186, 188]. Yet, ideologies and issue positions are not always aligned [184, 92, 198]; ideological misalignment can have broad and meaningful implications for understanding political cognition and behavior [195, 16, 92, 199]. We therefore refrain from assuming any alignment between ideologies and issue positions a priori, that is, defining ideological orientation based on particular issue positions, and vice versa (e.g., fixating opposition to redistribution as a right-wing attitude or right-wing identity to entail opposing redistribution). Furthermore, we explicitly acknowledge that the way ideologies and issue positions align may vary over time and across different political contexts [200]. We therefore believe it particularly beneficial to let ideological structure emerge organically from observed co-endorsement patterns within political attitude data.

Understanding why variation in ideological alignment exists is key to our second premise: we consider ideologies as socially constructed attitude bundles. This assertion rests on the well-documented observation that the co-occurrence of response patterns is often shaped by an underlying social structure, including socialized partisan leanings [201, 16, 202, 93, 15]. For example, support for social security policies is far from randomly distributed across the population; it is significantly more prevalent among Democrats than among Republicans. The collective identity formed by social and partisan groups often serves as ideological heuristics when individuals try to position themselves on a new issue or update their former beliefs. For instance, when a Democrat positions themself on a new issue, they might first take into account how other Democrats position themselves on this very issue. Consequently, ideological alignment implies more than a sorting process along various ideological dimensions. It should also be understood as a consolidation and polarization process in which political beliefs can become fused with group identities [93, 15].

3.2.2 Ideological polarization and belief network analysis (BNA)

Despite substantial research effort investigating political polarization, debates among social and political scientists persist regarding its overall trends in the mass public and the underlying mechanisms that drive it [121, 78, 203, 204, 188]. One reason such debates persist is that ideological polarization has been conceptualized and measured in various different ways. Two influential conceptualizations of polarization are divergence and alignment [10]. Divergence describes a process of people moving apart from one another towards the extreme ends of an ideological dimension, such as the left-right spectrum [204]. Alignment, by contrast, means that people develop ideologically coherent beliefs (i.e., constraint; [16]) because their partisan identities are increasingly aligned with political issue positions (i.e., partisan sorting [98]). Regarding alignment, a system is considered strongly polarized if people's position on one issue or their partisan identity can be used to reliably predict their position on various other issues. Conversely, a system in which people's position on one issue or their partisan identity is a weak predictor of issue positions would be considered less polarized.

Over the past few decades, researchers have not seen a vast divergence along many political issues in the United States [77, 120, 10]; however they do observe alignment in terms of the increasingly pronounced partisan identities [205, 93] that now have stronger ties with political beliefs [206, 207, 123]. Our study uses a belief network perspective to better understand this process.

To assess the degree of ideological alignment, researchers have explored linear methods [e.g., 15, 208, 93], and more recently belief network analysis (BNA) [e.g., 128, 181, 129]. Following Converse's idea of belief constraint, BNA operationalizes belief systems by representing issues (e.g., abortion restrictions, government spending on welfare, and tax cuts) as nodes while modeling the associations between them as links, assuming that the constraint between two issues can be measured by the degree to which they are aligned in the mass public [128]. Hence, BNA imposes a greater link strength among more highly correlated issue pairs.

Leveraging the network properties of belief system graphs, scholars have attempted to identify the central belief elements [128, 181] to examine how belief centrality may predict the change in the system [129], and more pertinently to our present study, to explain the mechanism of polarization [123]. According to DellaPosta [123], polarization can unfold both in the "heightening alignment across pairs of issues", and in the "broadening alignment across a wider range of issues". In BNA, these processes manifest through a rise in edge weights (i.e., increased correlation among issues), or an expansion in the size of connected components.

While inter-issue correlations can serve as indicators for the strength of dependency between variables, they cannot fully capture non-monotonic relationships where the change of one variable with regard to the other is not simply increasing or decreasing [189]. Moreover, while BNA can capture the issue-wise alignment process, they do not explicitly encode ideologies (i.e., a coherent set of issue positions) as a spatial component of a belief network. Thus, we see an important gap between the theoretical framing of ideologies as the core product of (organized) belief systems [16], and the operational obscurity of how ideologies spatially manifest within a belief network. It is this gap that motivates our application of Response Item Networks (ResIN) to the analysis of ideological polarization within belief networks.

3.3 Data and methods

3.3.1 American National Election Studies (ANES)

The current study utilizes the American National Election Studies (ANES) from 2000 to 2020, a cross-sectional nationally representative survey of the general U.S population. We select six waves conducted in each presidential election year with a four-year interval. Using

ResIN–a spatial belief network model described in further detail below–we create one snapshot for each wave to visualize the shifts in political beliefs throughout this period. In each snapshot, we include a set of politically relevant issues that were consistently assessed in all six waves of ANES (i.e., government service spending, government healthcare insurance, guaranteed jobs and income, aid to blacks, legal abortion), which allows for meaningful temporal comparisons. ANES respondents are asked to self-report their positions (i.e., attitudes) on these issues on either a 7 or 4 point scale. The complete set of issues is listed in Appendix B.1 ANES Included Items.

3.3.2 Creating ResIN snapshots using ANES

While BNA uses individual nodes to represent a single issue (e.g. legalization of abortion), ResIN, instead, treats individual nodes as issue positions (e.g. the response "strongly agree" to legalization of abortion). Furthermore, as a "spatial network", each node in ResIN is located in an N-dimensional space, which can form the basis for a spatial model of ideology as an underlying left-right continuum. In what follows, we describe how we apply the main components of ResIN to the ANES data[1]. Readers should be advised that this procedure follows the exact steps as described in the work by Carpentras et al. [189], which provides a comprehensive description of the method. In this article, we briefly walk through the main steps involved in ResIN-based analysis and focus on how this methodology can be used to explore the polarization dynamics in the belief system.

For a given issue selected from ANES, we first dummy code each issue position as a response variable. For instance, for the issue "guaranteed jobs and income", respondents indicate their attitudes via the response options ranging from 1 (i.e., "strongly agree; government should guarantee job and good standard of living") to 7 (i.e., "strongly disagree; government should let each person get ahead on their own"). For each response x from 1 to 7, we construct a set of binary variables `guar_jobs:x` indicating whether or not a given respondent has chosen response x (1 for yes, 0 for no). These binary variables are represented as nodes in ResIN. A positive[2] association between a given pair of binary variables from two different issues produces a weighted link between the corresponding two nodes. The link weight equals the strength of association, in this case, the phi correlation, which is an equivalent measure of the Pearson's correlation index for binary variables[3] [189]. In-

[1] A full description of ResIN and its properties is beyond the scope of this article but we encourage interested readers to refer to Carpentras et al. [189] and Lüders et al. [191].

[2] Negative edges are excluded from the visualization to avoid visual clutter and because previous research [189] has shown that the positive edges contains already enough information for studying the network structure. Furthermore, many useful visualization algorithms or analysis methods cannot handle negative edges (e.g. force-directed algorithms).

[3] Please note that phi correlation, Pearson correlation and Spearman correlation are identical for binary variables [**?**].

tuitively, the correlation between two responses A and B is a measurement of how often people who select A also select B (and vice versa).

After setting up nodes and weighted links, we implement the force-directed layout algorithm [209], where node positions are determined by an equilibrium state that balances the attractive and repulsive forces among them [189]. The greater the link weights between any two response node pairs, the more powerful the attractive force between them. This allows ResIN to visualize the network structure of political beliefs organically, placing nodes connected by stronger links closer to each other. We also use Principal Component Analysis (PCA) to identify the main dimension among the obtained spatial coordinates among all nodes, which we align with the X-axis by rotating the initial solution. Thus, ResIN enables a spatial organization of attitudes (i.e., nodes) within a 2-dimensional plane, where the distribution of nodes along the (major) X-axis reflects their relative position along the most relevant latent variable. As we will later show in Section 3.4.1, the X-axis happens to align well with partisan identities, a node-level attribute that averages respondents' partisan identities. Carpentras et al. [189] have shown that node positioning in ResIN is practically equivalent to latent variable modeling using Item Response Theory, in that distances between nodes represent corresponding distances in a latent space. Thanks to these unique features, ResIN is capable of identifying asymmetries in attitudinal patterns across the latent space of interest, as is shown in a recent application on vaccination attitudes [190].

To further assist in visualizing how node positions correlate with other covariates–such as party identity–we can use node color to indicate the average covariate value of all respondents who endorse the corresponding attitude in a given year. Here, we choose to color nodes based on partisan leaning, which, at the individual level, is measured by a Likert scale ranging from 1 (i.e., strong Democrat) to 7 (i.e., strong Republican) (see Appendix B.1 Party Identification of Respondent). For instance, the color of the node `guar_jobs:1` in 2020 is determined by the average partisan leaning of all respondents who strongly believe that government should guarantee jobs and a good standard of living, according to their responses in 2020 ANES.

Here, we provide a simple simulation to showcase ResIN's ability to spatially organize nodes in a 2-dimensional space, which is a key feature absent from classic BNA models. Figure 3.1 displays BNA and ResIN using the same synthetic data, where 1000 respondents self-report their positions on 12 issues via 3 possible response levels (i.e., agree, neutral, disagree). When all issues are correlated at similar levels, BNA displays a single connected component with no clear modular structure. In contrast, ResIN can reveal that responses corresponding to the same position are strongly connected (e.g., people who agree on one issue tend to agree on all other issues). Furthermore, we see that the neutral position (position 2) is closer to agree (position 1) than to disagree (position 3), which implies that people who remain neutral on some issues are more likely to agree rather than disagree on other issues. This simulated example shows how ResIN can reveal an asymmetric shape of a belief system that remains hidden in BNA.

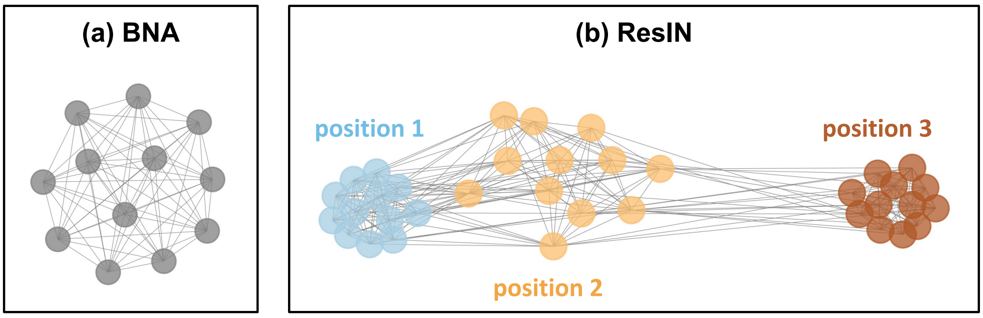

Figure 3.1: Comparison between BNA and ResIN on the same synthetic dataset. While BNA represents issues as nodes and association between issues as links, ResIN represents issue positions (i.e., attitudes) as nodes and association between issue positions as links.

3.3.3 Multi-level constraint and polarization measures afforded by ResIN

Moving beyond revealing simple structural asymmetries within belief networks, we next discuss how ResIN can be used to study attitude structuration and polarization dynamics by offering quantitative, network-based polarization measures. In total, we propose five measures which operate at different levels: (i) two system-level measures that summarize the overall structuration and thus indicate the degree of constraint and polarization for an entire ResIN-network, and (ii) three attitude-level measures that articulate the specific role a given attitude plays in a (polarized) belief system.

System-level: how constrained and polarized is a given attitude space?

Based on Converse's definition of belief system constraint, we expect a highly constrained belief system to produce more interlocked issue positions, that is, more attitude pairs held together via stronger links. Therefore, an intuitive network metric to assess the system-level constraint is link density D, i.e., the proportion of manifest links among all possible links within in a given network. Hence, let $G = (V, E, W)$ denote a ResIN with a set of nodes V, edges E and the corresponding edge weights W, produced by responses to a set of issues K, each with L_k possible response options. Thus, the numerator is the sum of link weights; and the denominator, the number of all possible links, equals the number of links that can connect nodes sourced from different issues. The link density of a given ResIN is therefore:

$$D = \frac{\sum_{e \in E} W_e}{\sum_{m \in K} L_m (\sum_{n \in K, n != m} L_n)}$$

Apart from computing D for the whole graph, we can also compute $D_{G'}$ for a partisan subgraph G' in a single ResIN. For instance, we can compute $D_{G_{rep}}$ for a subgraph G_{rep} consisting of only Republican-leaning nodes, i.e., attitudes that are selected mainly by respondents who lean closer to Republicans.

Next, we introduce linearization as a measure of polarization, i.e., the degree to which a network gets squeezed into a dominant latent ideological dimension [e.g., 210]. In combination with force-directed layout algorithm, node positions in ResIN are determined by a varimax rotation that aligns the main orientation of the network with the X axis (i.e., the dominant latent ideological dimension). In this way, the spatial information of each node meaningfully indicates their position in a latent ideological space – a feature that we will demonstrate empirically in Section 3.4.1. In this space, a highly polarized belief system would flatten-out, or distribute linearly, such that knowing one's position along a single ideological dimension can well predict their positions on a variety of issues (we elaborate on this through simulation in Section 3.3.3). We leverage ResIN's spatial property and quantify the degree to which attitudes towards the five issues collapse into a single, dominant dimension, which can be visually grasped by the degree to which the nodes are being "squeezed" into a flat shape linear line. The linearization score E equals the ratio of X coordinates spread to Y coordinates spread for all nodes in a given ResIN network.

$$E = \frac{X_{max} - X_{min}}{Y_{max} - Y_{min}}$$

To assess the robustness of these system-level metrics, we conducted a subsampling-based sensitivity analysis by randomly re-sampling 80% of the yearly ANES data (without replacement). In each of 200 iterations, we regenerate the ResIN network, yielding a distribution of polarization scores for each year. Here, we report the interquartile ranges (IQRs) to show the degree of variability across subsampled results. We choose subsampling rather than classical bootstrapping, as duplicating responses from individual respondents–necessary in bootstrapping–may distort the network structure in ways that do not reflect realistic data variation. Furthermore, subsampling provides a more conservative estimate of robustness by testing how sensitive the polarization score is to partial data omissions.

Simulation of a "broken egg": the linearization process

Why, one may wonder, does ResIN change its shape in response to variations in belief system polarization? And more precisely, why would an increase in polarization bring forward a "broken-egg" process in ResIN? Remember that proximity between nodes (i.e., issue positions) in ResIN indicates a high degree of attitude co-endorsement. We answer this by observing how ResIN changes in shape as we move from a perfectly polarized system to a fully random one. To control the degree of polarization, we interpolate between a highly polarized model derived from Item Response Theory (IRT) [211] and a random model that encodes no polarization. Details of the simulation are provided in Appendix B.2.

Given the structure of IRT, issue positions with similar mean value in their item characteristic curves are likely to be co-selected by the same people, and thus tend to form clusters in ResIN. Within a structured IRT model (i.e., a polarized belief system), the item characteristic curves are well sorted along the natural order of issue positions, such that respondents with the same θ value are most likely to select the same position, fairly likely to select adjacent positions, but less likely to select more distant positions. For instance, consider three positions (i.e., support, neutral, and against) towards gun control. A structured IRT would predict left-leaning respondents to most likely to choose "support", moderately likely to choose "neutral", and least likely to choose the "against" option. The inverse would be true for right-leaning respondents. Meanwhile, centrist respondents would most likely be neutral towards gun control, and less likely to support or be against gun control. A belief system with multiple issues resembling this very item response structure would naturally lead to a ResIN solution in which the neutral positions are connected to the support and against positions, yet the support and against positions are only weakly connected. A force-directed layout balancing the attractive and repulsive patterns would thus result in a linearized, left-right attitude structure indicative of a highly polarized system.

To sum up, the present simulations help us understand how and why the shape of ResIN-networks is indicative of belief polarization along a latent ideological dimension. A weakly polarized belief systems produces an unstructured cloud-like ResIN, while a strongly polarized belief system produces a linear-shape ResIN that map issue positions closely along a single dimension.

Attitude-level: how (de-)polarizing are certain attitudes?

Next we analyze ResIN at the node level. Our main aim here is to reveal heterogeneity in the structural function of different attitude nodes within ResINs. More particular, we are interested in identifying (a) ideological centroids, i.e., nodes at the center of divided ideological communities and (b) attitudinal bridge(s), i.e., nodes that could help bridge distinct ideological camps.

To accomplish the former, we focus on the node strength centrality as a measure of local importance. Previous research relying on classic BNA models claims that nodes higher in strength are more central to people's belief systems [128, 181, 182]. Leveraging ResIN, we can extend this approach to within-cluster analysis, investigating which issues are relatively more important to Democrats rather than Republicans. Purportedly, these issue positions should also be the most polarizing to the opposite ideological camp.

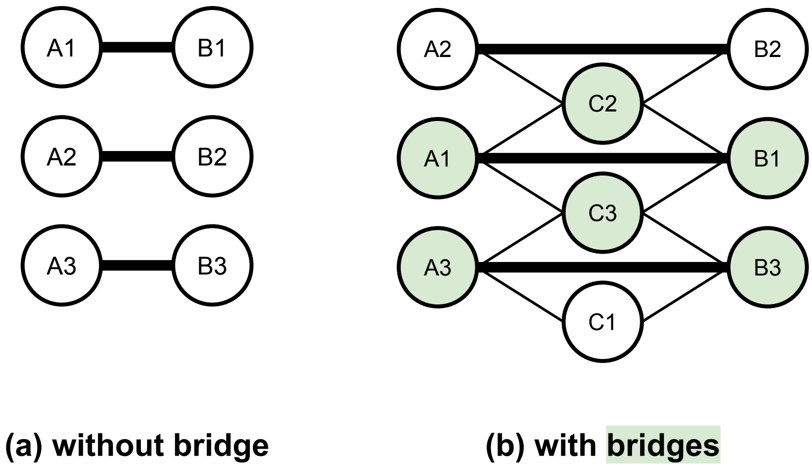

Figure 3.2: Illustration of a toy ResIN model with two issues (a) and three issues (b). Each issue has three positions. The left panel (a) demonstrates the scenario where the population perfectly sorted along issue A and B, and there is no bridge connecting mismatched positions such as A_1 and B_3. The right panel (b) shows a scenario where some positions along issue C serve as bridges connecting previously isolated attitudes; similarly positions along A and B can also serve as bridges for B-C and A-C. The green color highlight those bridge attitudes.

Now, we turn towards attitudinal bridges. To illustrate how we find issue positions with the most de-polarizing potential, let us consider a scenario featuring two completely sorted issues (i.e., A, B; Figure 3.2-a), where knowing one's position on issue A can give full information on their position on issue B (e.g., people who choose position 1 for A, i.e., A_1, would definitely choose position 1 for B, i.e., B_1). Here, two respondents either agree completely on both issues or share no common position at all. However, if a comparable number of respondents who select A_1 and B_1, and those who select A_3 and B_3, can agree on another issue C (e.g., if they have all selected C_3), it would be possible to bridge the debate through finding a common ground on C. In this case, response C_3 would thus function as an attitudinal bridge. In Figure 3.2-b, we highlight six such attitude bridges through which respondents can establish partial agreement on one issue while disagreeing on the other(s). Using ResIN, we can measure such bridging power by computing the node-level betweenness centrality, a measure counting how many times a node lies on the shortest path between any pair of nodes [4] [212]. Nodes that are frequently on the shortest paths connecting two or more ideological communities are more likely to serve as the common ground, thus likely showing the highest

[4]Because ResIN assign link weights based on attitude associations, attitudes with a greater association are thus connected with stronger links and pulled closer, which means node distances would decrease as link weights increase. Hence, we use the inverse of link weight to attribute edges in the calculation of shortest path length.

de-polarization potential in cross-community engagement. Alternatively, we also compute closeness centrality, which is the weighted inverse of the total shortest-path distance from a node to every other reachable node and hence represents a quantitative measure of global connectedness of a given attitude. Conceptually, a high betweenness and closeness centrality should be jointly indicative of node-level bridging power in a belief system.

3.4 Results and discussion

In this section, we first present BNA and ResIN belief networks for the ANES 2000-2020 before discussing the insights we can glean from each. We show that ResIN represents spatial node positions that reflect meaningful ideological differences. BNA solutions, however, do not offer a comparable method of spatial node positioning. Second, we use system-level constraint and polarization measures (i.e., link density and linearization) to summarize the overall trend of polarization over time. Third, our analysis of the ANES 2020 data focuses on more details of our attitude-level metrics, allowing us to discuss the varying roles of particular response nodes in constraining different partisan groups while contributing to system-level (de-)polarization within belief networks.

3.4.1 Meaningful spatial organization of Democratic/Republican Crowd

We start with ResIN's capability to visualize the spatial clustering of co-endorsed issue positions. Figures 3.3 and 3.4 display the BNA and ResIN snapshots of political belief system modeled based on the same six waves of ANES data from 2000 to 2020.

BNA captures belief constraint through bivariate correlations among the issue-level responses. The top pane of 3.3 thus illustrates the overall trend of increasing constraint among the five issues while showing how abortion has evolved from a weakly constrained issue at the periphery of the system, into a well integrated component. In Figure 3.3, we also correlate issue responses with the average partisan leaning and color nodes based on the correlation strength. While BNA-based visualizations could reveal how issue positions are increasingly sorted with partisan leanings over the past two decades, they are unable to capture the increasingly spatial polarization of the attitude space. Because different positions towards a certain issue are collapsed to a single node, BNA is not designed to map beliefs based on their relative positions in an intuitive and theoretically meaningful latent space.

In contrast, ResIN paints a richer and more granular picture of the polarization process. Overall, the evolution of the belief system resembles the breaking of an egg: starting from a very diffuse attitude "blob" in 2000, with mixed endorsements from Republicans and Democrats on many issue positions, the belief system gradually develops a more modular

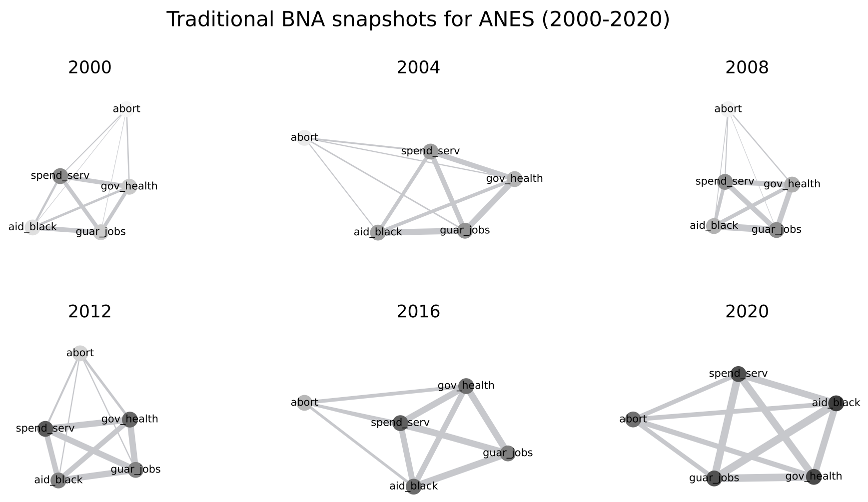

Figure 3.3: Traditional BNA snapshots of ANES data from 2000 to 2020, every four years in the presidential election cycle. Each node represents a single issue (e.g., `spend_serv` represents the issue regarding increase or decrease government's service spending). The edge width is a function of weights, indicating the absolute correlation strength between two corresponding issues. The node color is a function of the absolute correlation between the average partisan leaning and positions toward the given issue. Node positions are determined using the force-directed layout algorithm to ensure a fair comparison between BNA and ResIN.

structure with a clear division between Republican and Democratic attitudes, a trend that has become more pronounced since the 2008-2012 interval. Such crystallization of Democratic and Republican belief clusters may be linked with the rise of a "galvanizing and divisive figure" [213] marked by the historic win of Barack Obama and intensified partisan identities during his administration [214]. Meanwhile, the overall shape of the attitude space becomes gradually more linear as different attitudes towards policy issues increasingly experience alignment with the dominant Republican-Democratic divide. This development largely resembles the simulation results from the least to the most polarized attitude systems discussed in Section 3.3.3.

Moreover, the X coordinate of each node in each ResIN snapshot corresponds reasonably well with the average partisan leaning (see Figure 3.5). This suggests that the X coordinate can be interpreted as a latent variable (as shown in [189]) which in our this corresponds to the latent ideological position of each attitude. We would like to stress that ResIN is able to organically reconstructs the main ideological divide without any prior knowledge about the ideological implications of issue positions in the U.S.

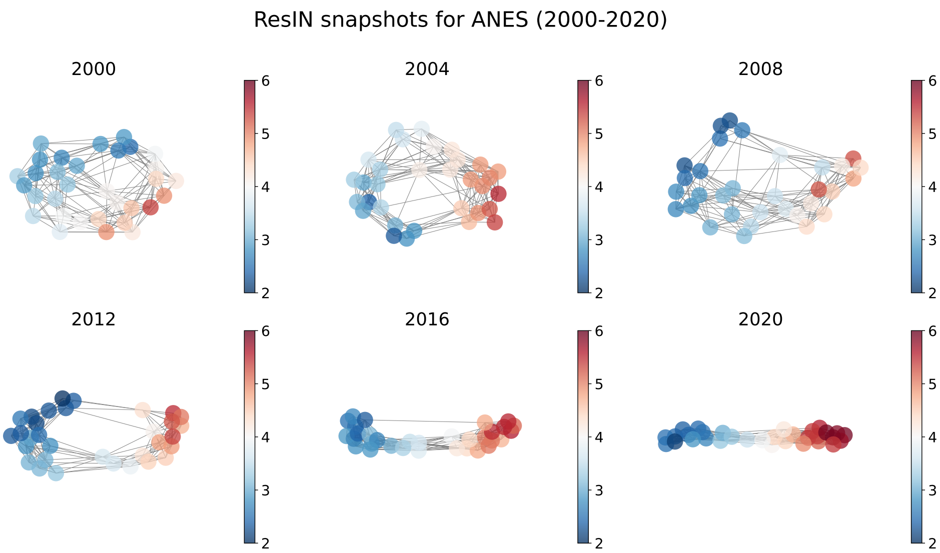

Figure 3.4: ResIN snapshots of ANES data from 2000 to 2020, every four years in the presidential election cycle. Each node represents a specific attitude toward a given issue (e.g., `abort:1.0` represents the attitude "abortion should never be permitted" and `abort:4.0` represents "abortion should never be forbidden"); the link strength between two nodes reveals the extent to which those who choose or did not select these two attitudes overlap. Node color indicates the 7 point scale partisan leaning averaged at the node level; 7 (1) means the attitude is selected only by strong Republicans (strong Democrats). Node positions are determined by the force-directed algorithm that pulls strongly linked nodes closer and a final rotation that aligns the main dimension of the network with X-axis.

In sum, ResIN not only retains the BNA feature of displaying belief constraint, but also captures how the increase in ideological polarization occurs together with the enhance in alignment of issue positions with partisan identities. By moving the analysis to the attitude level, ResIN provides a more detailed picture of evolving belief constraint, polarization, and partisan sorting than issue-level BNA models, while uncovering spatially meaningful organization along latent political ideology.

3.4.2 System-level analysis: ideologically distant, and sorted partisans

We now turn to the quantitative results of system-level constraint and polarization. Using two metrics–link density and linearization–we discuss the overall belief structuration trends among U.S. political beliefs between 2000 and 2020 in this section.

Figure 3.6 shows the results of link density, both for the entire network (left) and for Repub-

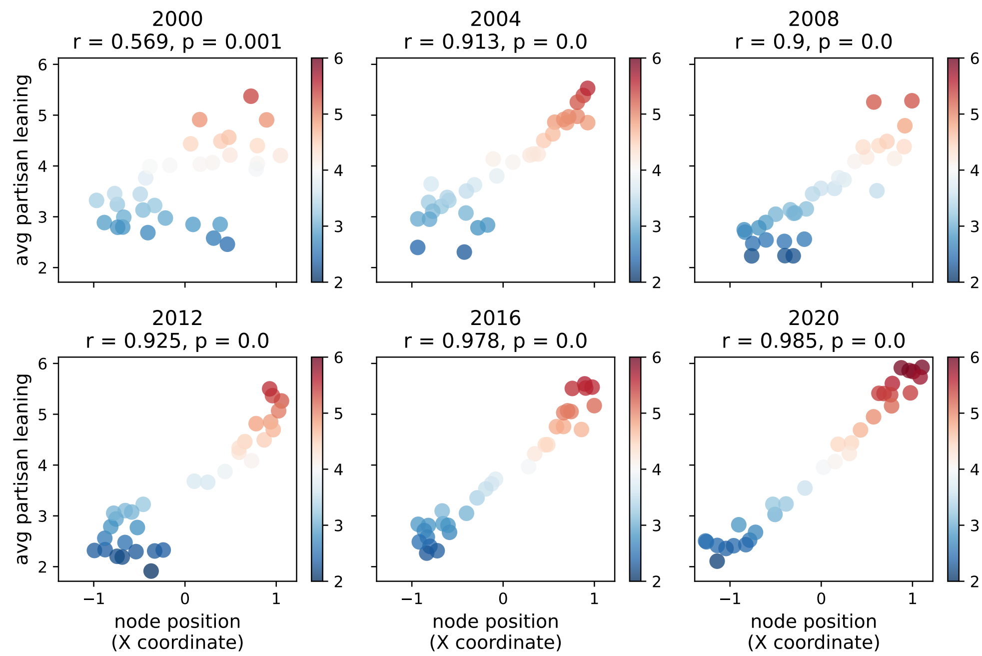

Figure 3.5: The correlation between X coordinates and average partisan leaning at the node level.

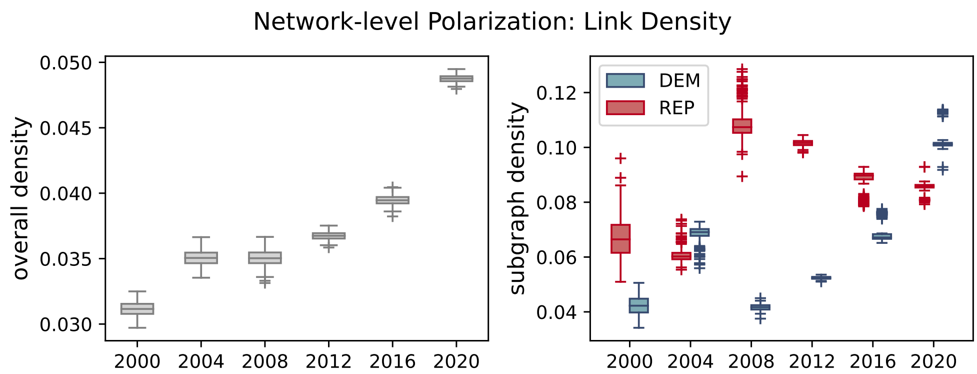

Figure 3.6: System-level polarization measure: link density over years for the entire network (left) and for partisan subgraphs (right). The errorbars show the interquartile ranges (IQRs) of metrics produced by 200 rounds of re-sampling, each taking 80% of the survey responses.

lican and Democratic subgraphs[5] (right). We note that constraint has been steadily rising

[5] As mentioned in Section 3.3.3, the Republican (Democratic) subgraph refers to a subset of ResIN consisting of nodes representing attitudes endorsed primarily by respondents who lean towards Republicans (Democrats).

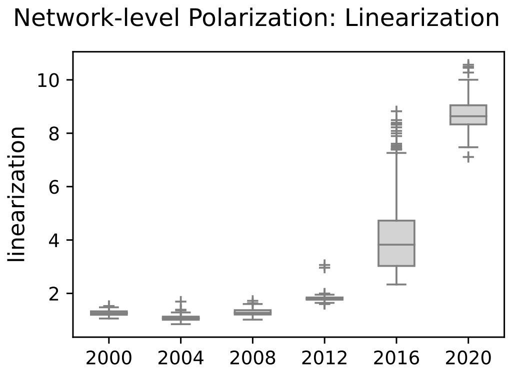

Figure 3.7: System-level polarization measure: linearization over years for the entire network. The errorbars show the interquartile ranges (IQRs) of metrics produced by 200 rounds of re-sampling, each taking 80% of the survey responses.

over the past two decades, as the link density has increased by 51.1% from 0.032 to 0.049. Following a period of relatively mild increases from 2004 to 2016, the public saw its most rapid episode of rising constraint between 2016 and 2020, during the first presidency of Donald Trump. This trend, however, did not unfold symmetrically across partisan groups. As shown in the right subfigure, the Democratic and Republican subgraphs have undergone quite distinct trajectories. Compared to its Republican counterpart, the Democratic subgraph generally showed a lower link density, with the exception of 2020. From 2000 to 2020, the Democratic side reached its first local peak in 2004, bounced back to a low level in 2008 and has been consistently growing since then. Subgraph density among Republicans peaked out early on in 2008, with the density level continuously declining from 2008 to 2020. These differences in belief constraint reflect a partisan asymmetry: the higher link density of the Republican cluster may reflect a focus on "doctrinal purity" and ideological coherence [215], whereas the Democratic cluster initially exhibits more heterogeneous beliefs encompassing interests of different social groups. Convergence on certain social issues among liberal-leaning respondents may explain the observed growth in link density post-2008 [216]. These observations highlight an asymmetry in partisan belief structures, which warrants further investigation.

In terms of linearization, the result delivers a fairly similar message: while the overall belief system has become more polarized from 2000 to 2020, the most dramatic shift occurs between 2016 and 2020 (see Figure 3.7). Beyond attitudes simply forming stronger interlocking structures (which results in a higher link density), our analysis shows that the system is also becoming increasingly well-sorted along a dominant ideological axis. The rising degree of polarization is not driven by some random new ties between attitudes, but by a more systematic sorting that arranges issue positions along the ideological dimension into a consistent, liberal-conservative orientation. This finding is aligned with broader intensifying

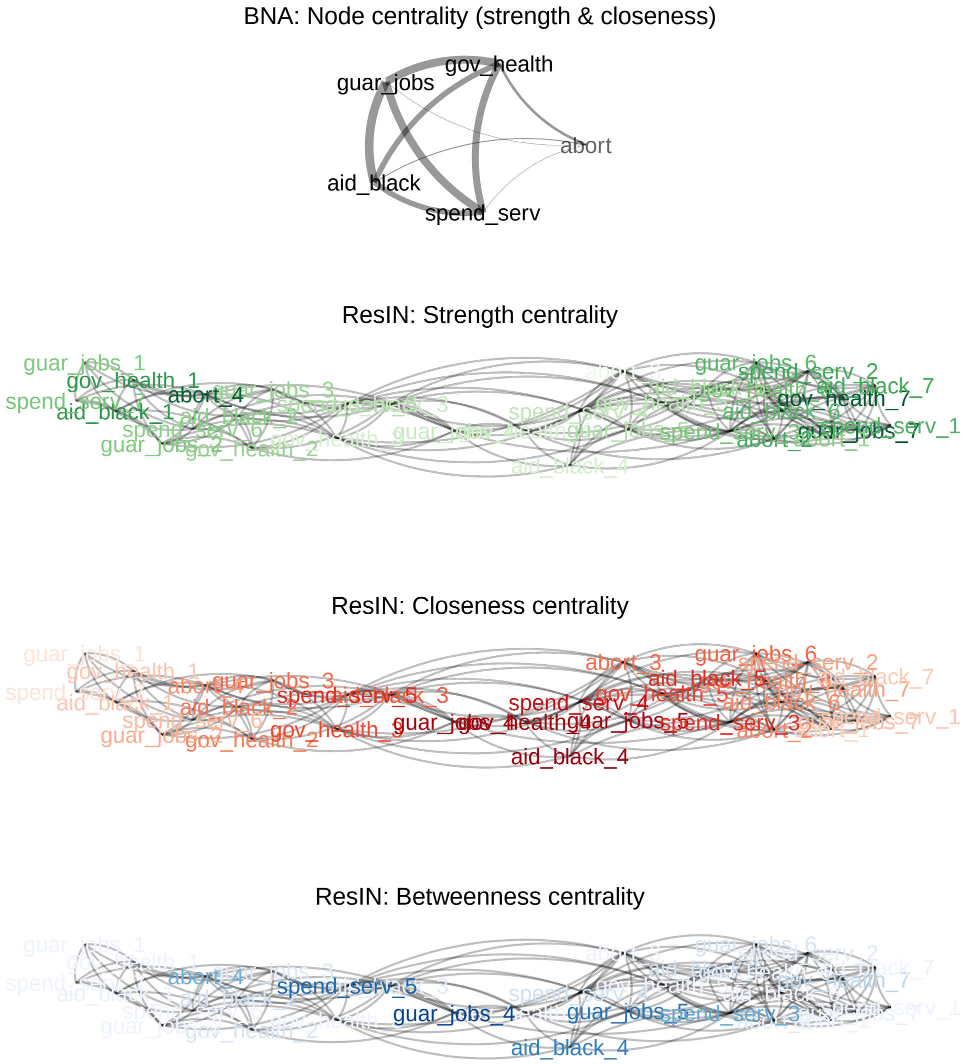

Figure 3.8: Node centrality statistics in BNA and ResIN using the same ANES 2020 data. Color intensity denotes more central nodes. Based on strength and closeness centrality, the most central nodes in the BNA model are government health insurance (0.138), guaranteed jobs (0.137), aid to African Americans (0.132), government service-spending (0.131), followed by legal access to abortion (0.11). Note that strength and closeness centrality are equivalent in BNA (but not in ResIN) as all closest network paths are direct paths.

trends in other aspects of polarization during the 2016-2020 Trump era, which, was indeed marked by heightened affective polarization [217, 218], the consolidation of partisan identities [219], and the alignment of attitudes across economic, cultural, and racial issues into a more unified liberal–conservative structure [220].

3.4.3 Attitude-level analysis: centroids and bridges in a polarized belief system

Delving deeper into the micro-structure of attitude networks, ResIN further enables investigations at the node-level, including details on how much a given attitude is contributing to aggregate polarization dynamics. Mirroring existing work using BNA literature, ResIN can uncover that different political beliefs perform different roles in structuring mass a belief system. A classic yet still prominent argument, for instance, holds that some beliefs are more central and thus exert a greater structural influence over others [16]. A number of recent works have successfully leveraged node-level centrality metrics, such as *strength*, *betweenness*, and *closeness*, to test claims the relative importance of individual nodes to the structural cohesion of belief systems at large [128, 181, 182].

While insightful, *strength*, *betweenness*, and *closeness* arguably capture a very similar phenomenon, that might be labeled general attitude centrality in classic BNA models (see [181]). However, these quantities allow us to discriminate fundamentally different structural roles in ResIN. As opposed to its corresponding interpretation in BNA, high *strength* centrality denotes local importance to a particular belief sub-cluster. A comparison of node strength centrality in ResIN can therefore be indicative of the relative importance of different attitudes to different partisan-ideological communities. As seen in the second panel of Figure 3.8, defending abortion rights is a much more central attitude to Democrats while the most central issues for Republicans deal with government sponsored healthcare and federal job guarantees. These results cannot easily be obtained using classic BNA (see the top panel in Figure 3.8), which would simply locate the guaranteed jobs issue as the most central overall, remaining oblivious to the possibility that different attitudes can be more or less central to different ideological communities.

One peculiar feature inherent in BNA is that strength and closeness centrality are equivalent for networks in which all shortest network paths are direct paths. In this case, betweenness-centrality remains constant at zero across all nodes. Figure 3.8 shows that this is indeed the case if we apply BNA to the ANES 2020 case. In ResIN, however, closeness and betweenness centrality not only indicate different belief system functions, but they are also largely decoupled from metric equivalencies with strength centrality.

Gleaning at the third panel in Figure 3.8, we note that nodes with high betweenness and closeness centrality tend to lie in between the Democrats and Republican clusters. These attitudes are typically moderate in nature and tend to be more frequently endorsed by Independents. According to our model in the bottom panel of Figure 3.8, if one would like to find the most likely issue position providing common ground between Democrats and Republicans, the best bet would be to establish conversations about the general state of public services and about government healthcare. In contrast, one should avoid issues concerning abortion and race.

Then, how stable are the above referenced node-level metrics across time? Do Democrats and Republicans consistently prioritize different issues or was there once more common ground between them? Figure 3.9 summarizes the relative strength centrality estimates within the Democratic and Republican clusters over the past 20 years, highlighting the top, runner-up and bottom two attitude nodes within each cluster. Again, we assign each response node to either partisan camp based on whether more Democrats or more Republicans, on average, endorsed the given item response[6]. Focusing on relative within-cluster strength centrality statistics, we notice that while both partisan groups consistently possess attitudes about government health insurance and job guarantees at the center of their attitude clusters over the past 20 years, there were also marked and evolving variations throughout this process. Whereas abortion attitudes featured among the least important to early 2000's Democrats, their importance in structuring the remainder of their attitude cluster has steadily risen, until becoming the most central attitude in 2020 (that is, by substantial margins). During the same period, aid to African Americans almost entirely lost its once central position within the liberal attitude cluster. Among the Republican community, meanwhile, abortion attitudes never played a pivotal role as an ideological anchor in any of the measured years.

In Figure 3.10, we map the evolution of node closeness and betweenness centrality across time. To avoid visual crowding, we only label top three nodes based on their centrality statistics in each ANES wave. While both quantities generally point to similar nodes as providers of attitudinal bridges between partisan communities within each year, we notice a general trend in falling closeness but growing betweenness centrality estimates as time progresses. This trend appears to be bourne out of the attitude polarization process captured in the ResIN networks depicted in Figure 3.4: as cross-partisan edges disappear and the system linearizes, fewer and fewer attitudes provide effective communicative bridges between partisan camps. This makes the few remaining linking nodes ever more vital, as signified by growing absolute betweeneess centrality statistics of nodes, particularly in the upper right corner in Figure 3.10's lower pane. These nodes, which generally represent issue positions mildly in favor of greater re-distribution and government intervention in the economy, appear most likely to equally appeal to both partisan camps.

[6]Note that ResIN offers other clustering options as well; see Warncke et al. [221].

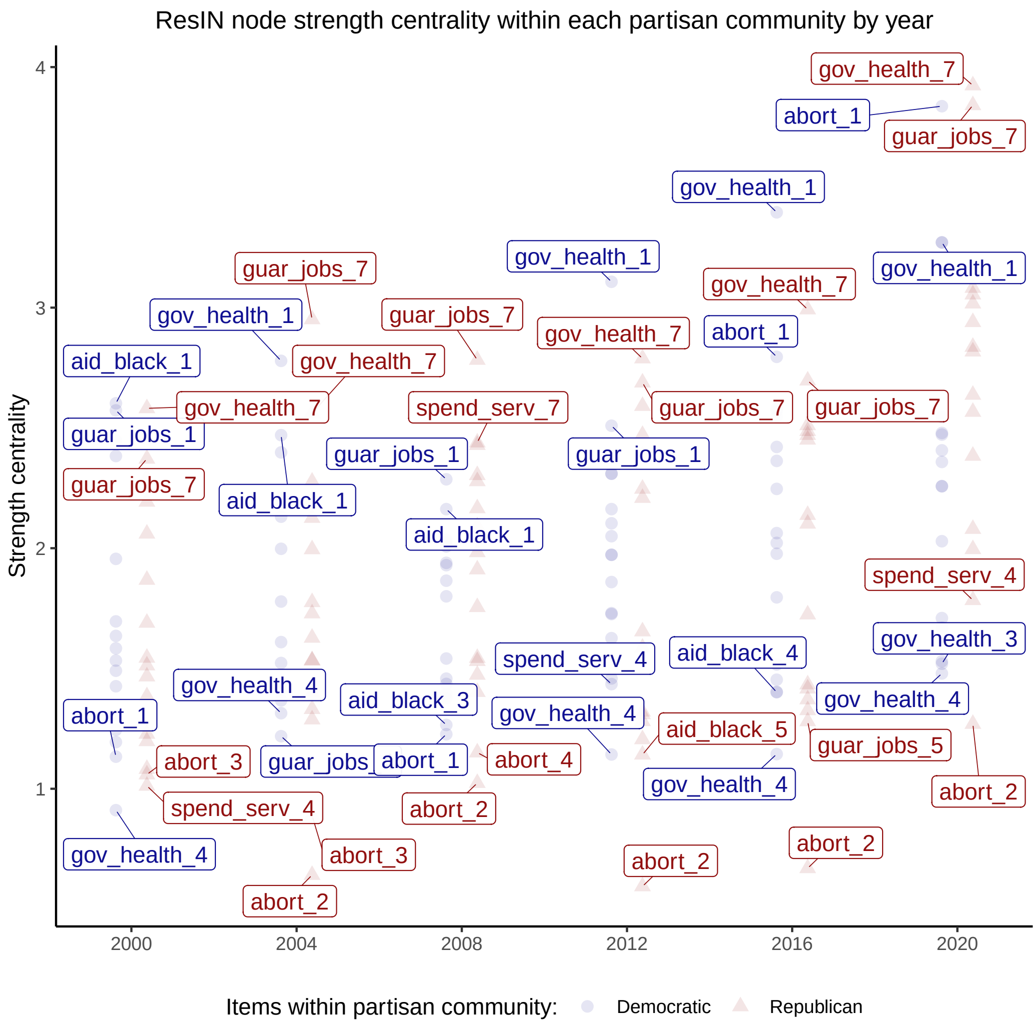

Figure 3.9: ResIN node strength centrality statistics within each partisan sub-cluster and presidential election year. For clarity, we only labeled the top, runner-up, and bottom two attitude nodes within each cluster. Clusters memberships were assigned based on whether more Democrats or more Republicans endorsed a particular issue position in a given year. Source: ANES cumulative file.

3.5 Limitations and future research

While our study adds valuable insights to the field of BNA and belief system structuration, it has important limitations. First, both the structure of ResIN and the derived polarization measures depend on the selection of issue items. In this study, we carefully select political issues that we assume to be salient and relevant to the U.S. political context; however,

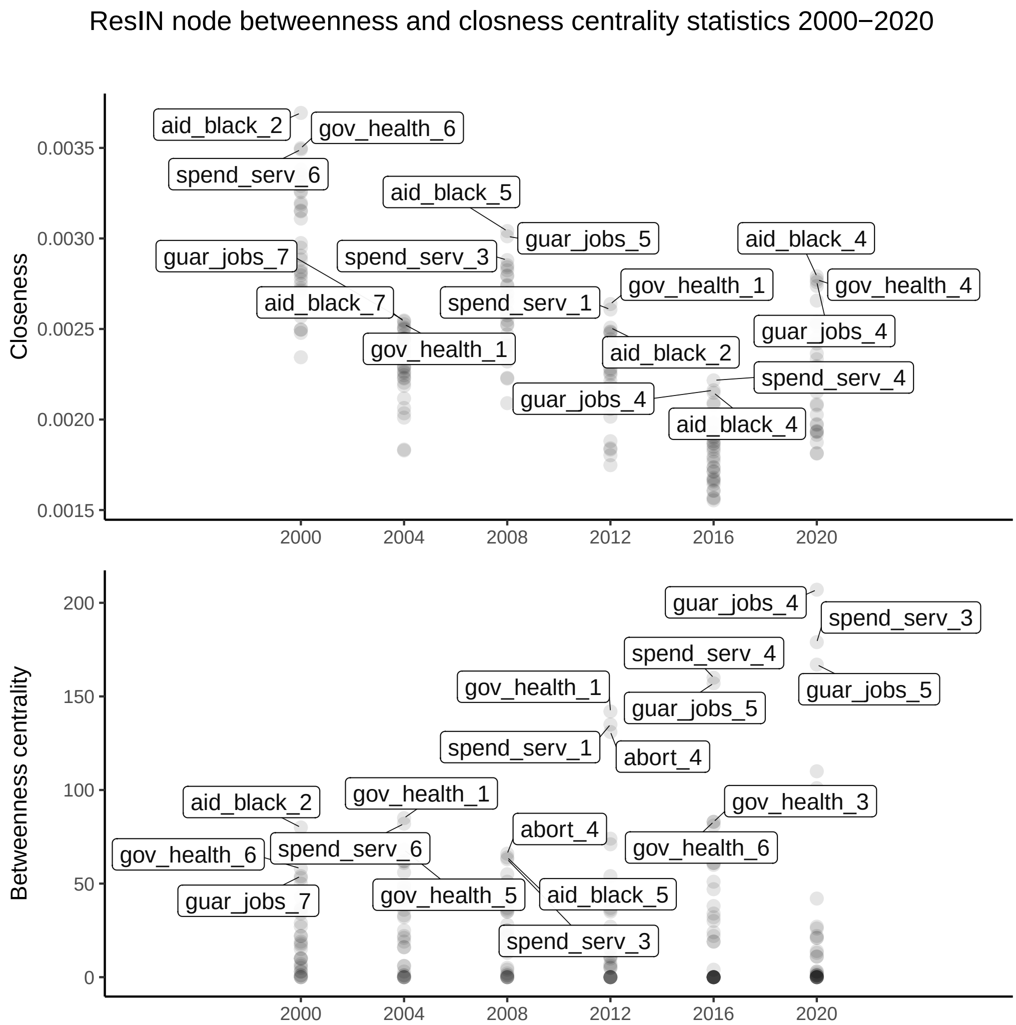

Figure 3.10: ResIN node closeness and betweenness centrality statistics for each presidential election year. Only labeled the top three nodes based are labeled based on each centrality statistic. Source: ANES cumulative file.

our choices are limited by data availability. Because the network structure largely depends on how selected issues and their corresponding attitude items are interconnected, including or excluding particular issues can substantially alter the observed ResIN shape and, consequently, the pattern of polarization. For instance, omitting a key bridging issue during 2008–2012 (a period when ResIN begins to exhibit clear belief modularity) would accelerate the network's collapse, potentially resulting in a flattened network structure by 2012. Conversely, including more cross-cutting issues, such as government spending on education, crime control, or social welfare, would slow down this collapsing process. Future research could explore richer datasets with more diverse issues and evaluate the generalizability of ResIN under different issue selections.

Second, since the ResIN-based polarization measures we propose are novel, there is much left to explore beyond our current analyses. This includes testing their unique predictive power for different forms of attitude structuration processes–such as belief constraint and polarization–as well as assessing their performance on other datasets or through simulations. We hope our findings will encourage researchers to further investigate these proposed measures in the future.

Third, although we have demonstrated that ResIN provides novel insights into polarization in the U.S., our model still needs to be tested and validated in comparative contexts. While other work has applied ResIN across European countries using European Social Survey (ESS) and showcased the varying structural properties of European belief systems [13, 222], the current analysis on the case study on ANES may not be directly applicable in scenarios where the belief system remain largely unstructured. Moreover, the two-dimensional space used in our visualizations may be sufficient to describe the structure of belief networks in a two-party system, but it may oversimplify the ideological space in other contexts where belief networks might follow a higher-dimensional structure. Although ResIN is capable of capturing multidimensional belief networks, it may lose detail when being projected onto a low-dimensional latent space. Therefore, the extent to which ResIN can offer similar insights in cases where belief networks are higher-dimensional remains uncertain and should be explored through further research applying ResIN to other datasets and contexts.

Chapter 4

Biases in online polarization measurements: what visible engagement fails to capture in online political communication

Through analysis of political attitudes in belief systems, we observe how political polarization unfolds in collective mental states with increasing alignment across multiple ideological dimensions. Yet, political polarization also manifests in segregated behavioral patterns online, which, with the rise of digitization and especially social media, has been a growing area of research attracting interdisciplinary researchers.

Building on the summary in Section 1.3.2, this chapter zooms into polarization measurements via the behavioral perspective, addressing the limitations and biases in existing measurement outcomes introduced by the incomplete data points collected through public APIs. Specifically, this chapter categorizes various forms of political engagement (e.g., viewing or commenting under political videos, subscribing to political channels) into the "visible" and "invisible" categories, depending on whether a given form of engagement has been well documented in the platform APIs and thus more publicly available and retrievable to researchers. Given that measurements of political polarization online have so far been largely focused on visible engagement while neglecting invisible engagement, this chapter aims to investigate the measurement bias that arises from such a restricted lens, highlighting the benefits of exploring novel data collection approaches to access richer and multidimensional data.

4.1 Introduction

Political polarization, the phenomenon of individuals diverging into distinct ideological camps, or becoming antagonistic towards out-group members, has become a critical global concern [9]. The measurement of this phenomenon, which traditionally relies on survey responses, now increasingly incorporates online behavioral data from social media platforms, re-conceptualized in ways such as the extent to which users opt into ideologically congruent content that align with their preexisting beliefs (i.e., selective exposure) [82], and how the overall user activities are segregated along the ideological line [223]. While online datasets enable researchers to observe facets of polarization absent in self-reported data, it is essential to critically examine the observational perspective these datasets afford. When using social media datasets to measure polarization, most researchers focus solely on activities that leave visible traces and are accessible by platform application programming interfaces (APIs), i.e., visible engagement, while neglecting activities that are harder to retrieve and remain hidden, i.e., invisible engagement [224, 225]. However, can the analysis of visible traces alone provide a valid measurement of political polarization?

We address this question through a case study of political communication on YouTube, focusing on Hungary, one of the most severely polarized countries in Europe [226] that has so far received disproportionately little attention from researchers [11]. From a sample of Hungarian Internet users ($N = 758$), we compile an individually-linked dataset of survey responses and digital traces. Our analysis centers on three forms of political engagement on YouTube: viewing, subscribing, and commenting, examining potential biases introduced when only visible engagement (i.e., commenting) is analyzed while invisible engagement (i.e., viewing and subscribing) is overlooked. For each form of engagement, we identify the corresponding user group–viewers, subscribers, and commenters–who interact with Hungarian political content. Depending on how data is aggregated and what analysis unit is used, measurement bias can result in incomplete or deviated descriptions at different levels. Here, our study elaborates on two analytical levels: the user and the channel level, which represent the platform's demand and supply sides. This focus aligns with most prior research that analyzes either a specific user group [e.g., 227, 131] or a set of channels [e.g., 228, 229]. More specifically, we define user- and channel-level biases as follows:

- **User-level bias** refers to disparities in measurement outcomes across user groups that engage in different forms or at different levels.
  - We examine how users' political attributes (e.g., ideological leaning) correlate with their form and level of engagement. For example, if right-leaning users are more likely to comment, studies based solely on comment data may misrepresent the distribution of public opinion.
  - We also assess how the ideological distribution of consumed content varies by

engagement type. For example, if users comment primarily on ideologically aligned videos but view a more ideologically diverse set, reliance on comment data alone would yield a skewed picture of selective exposure.

- **Channel-level bias** refers to disparities in measurement outcomes across audience landscapes built on user groups that engage in different forms.
  - We explore how patterns of audience overlap between channels vary depending on the observed form of engagement. For example, if left- and right-leaning channels attract distinct commenters but share many viewers, comment-based analyses would overstate the degree of polarization.

By examining the measurement differences between visible and invisible digital traces, this chapter centers on the bias stemming from trace selection error, while interpreting results within a broader framework of digital data collection error [224]. To clarify, we do not systematically examine other biases that may arise during the data collection phase (e.g., sampling bias), although in our later discussion of API-based and user-centric data collection these biases may become relevant. Our work contributes to the studies of online political communication as follows. Firstly, we reveal the relationship between political attributes and forms of online political engagement, helping researchers to reflect on their observed ideological space given the engagement form included in the analysis. Secondly, we connect the discussion of political engagement forms with reflections on the measurement processes that presume the selection of certain forms of engagement. By evaluating how excluding invisible engagement may impact downstream observations, we highlight the importance of articulating the scope of observations (e.g., only users who comment) when measuring polarization based on a limited dataset (e.g., only including the comments). Third, our analysis of a combined dataset showcases how obtaining richer and multidimensional user data can capture variations in ideologies and behaviors across different lenses of observation.

While our case study on Hungarian YouTube usage is unique, we caution readers about its generalizability, and hope to encourage explorations of more responsible approaches for collecting comprehensive datasets as well as analyses across different platforms and countries.

4.2 Biases from selecting digital traces of online political engagement

Regarding ways of online political engagement, existing works have distinguished categories such as passive and active political Internet use, and discussed ways in which they correlate with offline political engagement of various types [e.g., 230, 231, 232]. Yet, few have examined whether different categories of engagement lead to varying outcomes of polarization, ambiguating the ongoing debate about the relationship between social media usage and polarization [233]. Among the few that have considered this perspective, work by Matthes et al. theorizes about how active and passive social media use can shape affective polarization differently, and demonstrates that affective polarization is linked with active, but not passive political engagement [233]. Furthermore, when assessing online polarization through the lens of selective exposure–defined as a tendency for users to selectively consume ideologically congruent content [34], researchers have drawn on behavioral traces of passive consumption (e.g., browsing [234]) and active consumption (e.g., liking and commenting [235]), but not yet jointly consider how different categories of engagement may encode varying implications for selective exposure.

In this chapter, we connect the categorization of active/passive consumption with visible/invisible engagement, not only because they are practically overlapping (e.g., passive engagement such as viewing is usually invisible to researchers and to other users), but also because the visible/invisible framework helps articulate how limitations of data sources could impact findings derived from observational analysis. Given the scarce discussion on how various categories of engagement are accessible through platforms' APIs, it remains unclear how the visible and invisible engagement data would capture different polarization patterns. We now address these open tasks by leveraging a combined dataset of survey and digital traces on YouTube, focusing three forms of engagement with varying visibility (i.e., viewing, subscribing, commenting).

As one of the most popular video-sharing platforms, YouTube has been instrumental in facilitating online political communication through disseminating political content generated by politicians, news organizations, and grassroots content producers [236]. Studies collecting data from YouTube, mostly relying on its public API, have used three types of seeds–channels, videos and users–as the starting point for data queries. Many studies begin with retrieving relevant content, such as channels or videos grouped under a certain theme. Some borrow a pre-defined list of channels or videos [e.g., 236, 228, 227], while others expand these lists through snowballing [237, 238, 239]. A few also collect YouTube links sourced from external sites [131, 240]. For works starting with users or utilizing user endpoints to expand samples, obtaining a random user sample appears impractical, and snowball sampling has been applied in this case as well [229, 241]. Very few have explored alternative pathways such as user-centric data collections [242]. One study that explores this pathway collects web-browsing data from a representative sample in the US to extract YouTube URLs viewed by the participants [243].

Compared to Facebook or Twitter/X studies where users are usually the elementary units of collection and analysis, YouTube studies are more content-centered and content-driven. Investigations of the information landscape on YouTube typically begin by identifying the relevant content and, if needed, proceed to collect data on users who have engaged with it.

As pointed out by Heft et al., content-centered approaches rely heavily on the dictionaries used to query for relevant content [244], which may introduce sampling bias if certain users deliberately avoid using dictionary terms [245].

Because most of the aforementioned studies collect data via the YouTube API, the type of data accessible to researchers is almost always determined by what is available in the API scheme. Researchers have access to channel-level metadata (e.g., title, category, counts of views and likes) and video-level metadata (e.g., title, description, upload time), but not detailed subscriber or viewer lists for these contents. For videos, researchers can further retrieve user comments; and for users, researchers can retrieve their subscription lists only if the user has made them public [241]. Thus, among all forms of user engagement, commenting has a more accessible–in our case, more visible–source of data than viewing and subscribing, making it a more common basis for measuring polarization. For instance, work by Bessi et al. quantifies polarization as the proportion of comments a user leaves on content supporting one ideological side [131]. This underscores the importance of assessing whether, and how, a reliance on commenting-based engagement data introduces measurement bias in polarization research on YouTube.

4.3 Data and Methods

In this section, we present our pipeline for data collection and analysis. We start with broadly introducing the user-centric approaches for data collections in Section 4.3.1, followed by our data donation procedure in Section 4.3.2, where we outline the sampling strategies and describe the datasets we obtained. We then detail how we identify and label political content in Section 4.3.3, and explain how we filter respondents during preprocessing and define respondent groups engaging via different forms in Section 4.3.4. Finally, we describe our analytical steps to address user-level and channel-level biases in Section 4.3.5 and 4.3.6, respectively.

4.3.1 User-centric approaches of data collection

Among the various ways to access online political engagement data, retrieving data through platform APIs was the most common method in the 2010s. However, novel methods are needed as many popular platforms (e.g., Facebook, Twitter/X) increase their API restrictions [246, 247]. One promising alternative in recent years is user-centric approaches [247], where participants are invited to voluntarily share access to their digital traces. There are two main user-centric approaches [242]. The first involves implementing tools that monitor participants' online presence, such as software that tracks browsing history or records content encountered on social media platforms [248]. The second is data donation, which utilizes Data Download Packages (DDPs) that users can manually download from platforms [249]. Due to the General Data Protection Regulation (GDPR) law, technology giants such as Google, Meta, and Netflix are required to allow users to access and download their personal data stored on the platform. Through data donation, researchers invite participants to voluntarily donate their DDPs to legally and ethically retrieve their digital traces.

Compared to API-based access, the data donation approach is not bound by the restrictions of data availability from APIs, and provides richer–though not always complete–views of users' online activities. However, due to the complexity of the data donation procedure (e.g., recruiting participants, designing instructions and filtering out ineligible donations), sample sizes in these studies tend to be much smaller than those in API-based research. Additionally, ensuring sample representativeness can be challenging and is not easily addressed through survey design alone [250]. One goal of our study is thus to showcase how to work through these challenges and obtain valuable insights from such datasets.

4.3.2 Collecting data through data donation

Now we detail our data collection procedure and the datasets we obtained from participants. Our data collection was conducted from February to June 2023[1]. After quality checks, we were able to obtain a combined dataset of survey responses and DDPs from a non-probability sample of Hungarian Internet users ($N = 758$). Participants who consented to donate were asked to upload their DDPs from YouTube, and complete a survey questionnaire with demographic and political ideology questions. Details of recruitment and collection are provided in Appendix C.6.

The population of interest is defined as Hungarian Internet users aged 16 and older who use the Internet for communication via chat or email. To mitigate potential sampling biases, we applied individual weights to align the sample more closely with population distributions based on socio-demographic factors. These weights were generated using iterative proportional fitting [251] to adjust for discrepancies. The weighting factors include gender, age, education level, type of settlement, and geographical region. Additional analysis indicates that the data donation process does not introduce non-response biases in terms of respondents' political interests or ideological leanings (see Appendix C.6 for details).

Through data donation, we obtain two sets of data from each participant: survey responses and digital traces on YouTube. In the survey, participants self-report their basic demographics (e.g., gender, age, education) and various political attributes (e.g., interest in politics, ideological positions). These data points are later used in user-level analyss, as we discuss later in Section 4.3.5 and 4.3.5.

[1]The data collection project has been approved by the HUN-REN Centre for Social Science Ethical Board (1-FOIG/130-37/2022).

For digital traces, we consider both visible (i.e., accessible via YouTube API, commenting) and invisible (i.e., inaccessible via YouTube API, viewing and subscribing) engagement that are included in the DDPs. Given the significant shifts in Hungarian's media environment under Orbán's evolving media policies [252], we restrict our analysis to user activities from the most recent five years of our dataset–May 2018 to May 2023.

4.3.3 Identifying (Hungarian) political content and assigning ideological labels

Given our focus on political engagement, we limit our analysis to political content within the donated dataset. To efficiently identify such content, we retrieved metadata via the YouTube API for all channels and videos that the respondents have engaged with. We first retrieved profile metadata for all 787,945 channels that appear in respondents' viewing, subscribing, or commenting activities. This metadata contains aggregated statistics (e.g., viewer and subscriber count) and YouTube-assigned topic tags. We classified channels as political if they included "politics" among their topic categories–approximately 1% of all channels). At the video level, we also collected API metadata for over 2.4 million videos viewed or commented by respondents, again using the presence of the "politics" tag to identify political videos.

After narrowing down to political channels and videos, we further focus on one specific context–Hungarian politics–that delivers consistent ideological implications. We manually labeled a set of political channels and channels containing political videos (i) to distinguish domestic (Hungarian) channels from international ones, and (ii) to assign political leaning labels for domestic channels along the anti-/pro-government spectrum, which is the most pronounced cleavage in today's Hungarian media system[2] [253]. The detailed labeling process is described in Appendix C.7. Among the 11,065 channels with political videos or tagged as "politics", we identified 626 Hungarian political channels, of which 139 were classified as pro-government, 276 as anti-government and 149 as neutral.

[2] Although YouTube is non-traditional media, the most popular sites offering political content are either linked to politicians or to media outlets who have "traditional" non-social media platforms (e.g. online news sites, TV channels).

Filter level (group abbreviation)	Size
All respondents who participate in data donation (ALL)	758
Respondents who have some level of YouTube activity (YTB)	735
Respondents who have engaged with political channel(s) or video(s) (POL)	700
Respondents who have engaged with Hungarian political channel(s) (POL-HU)	668

Table 4.1: Groups of respondent obtained in each filtering step.

Respondent group	Viewing	Subscribing	Commenting
# of unique respondents	640	299	72
# of unique channels	545	210	108
# of unique videos	57400	-	993

Table 4.2: Number of unique respondents, channels and videos for three forms of engagement (i.e., viewing, subscribing, and commenting) with Hungarian political channels.

4.3.4 Filtering and extracting groups of respondents

Through the preprocessing above, our respondent sample has been incrementally narrowed: from the initial sample of 758 respondents, to the subgroup of 735 with some level of YouTube activities, to the smaller subgroup of 700 who have engaged with any political channels or videos via viewing, subscribing, or commenting, and finally to our focal group of 668 who have engaged with at least one Hungarian political channel via one of these engagement forms. We include all these groups in comparisons as a preliminary check, so that the biases introduced by selecting different engagement forms are not conflated with those caused by excluding non-active and non-political YouTube respondents. We report the group size for each stage in Table 4.1.

Within the focal group of 668 respondents, we identify three respondent groups–viewers, subscribers, and commenters–based on whether one has engaged with at least one of the Hungarian political channels through viewing, subscribing, or commenting respectively. We report the group size, and the number of unique channels and videos for each group in Table 4.2.

4.3.5 Addressing the user-level bias

We start our analysis with investigating user-level bias–the disparity in measurement outcomes across respondent groups engaging via different forms. First, we outline how we measure bias in self-reported ideologies across respondent groups (Section 4.3.5). We then describe how we identify factors associated with levels of engagement, which may be biased if only highly engaged users are included in the analysis (Section 4.3.5). Lastly, we illustrate how we compare the selective exposure pattern across different forms of engagement (Section 4.3.5).

Analyzing bias in self-reported ideologies across different forms of engagement

First, we capture the bias in polarization measurement based on the self-reported ideological leanings from the survey. Focusing on respondents' anti-/pro-government position (see detailed description in Section 4.3.5), we compare the polarization degree using two long-standing polarization metrics: variance and kurtosis [77]. Significant variations in these metrics across respondent groups indicate that studies focusing only on visible groups (e.g., commenters) risk measurement bias and should avoid generalizations about YouTube users. Further details of statistical tests used for the comparison are provided in Appendix C.1.

Analyzing bias in political attributes across varying levels of engagement

Besides comparing respondents who engage or do not engage in a certain form, we also examine the variation across different levels of engagement. To assess bias in political attributes, we compare these attributes across engagement levels using Negative Binomial Regression[3] to account for over-dispersion in the data. For each form of engagement, we construct a model comprising respondents who have engaged via this form with *any* content on YouTube[4] (i.e., from the respondent group YTB in Table 4.1) during the five-year period. The dependent variable (DV) is the level of engagement, quantified by the number of recorded activities (e.g., the number of videos viewed, the number of channels subscribed) for the corresponding engagement form with Hungarian political content (see Table 4.3). For respondents who have engaged with some YouTube content but not Hungarian political content, their DVs would be zero.

[3] We tested four different models: Poisson, Negative Binomial, and the zero-inflated version of these models. Based on fit statistics, the negative binomial model was the best to use in this study.

[4] For a robustness check we tested different sample compositions, see Section 4.4.1 and Appendix C.4.

Model	DV	Size
view	number of Hungarian political videos viewed by a respondent	690
subscribe	number of Hungarian political channels subscribed by a respondent	680
comment	number of Hungarian political videos commented by a respondent	314

Table 4.3: Dependent variables and sample sizes (i.e., number of respondents who have engaged in a certain form) for Negative Binomial models for each engagement form.

IV	Encoding
gender	binary (1: male, 2: female)
age	continuous (divided by 10)
education	3-level (1: low, 3: high)
timespan	continuous (percentage of active months on YouTube; range from 0 to 1)
interest in politics	5-level (1: not interested, 5: very interested)
pro-gov	binary (1: belong to the pro-gov group, 0: does not belong)
anti-gov	binary (1: belong to the anti-gov group, 0: does not belong)
neutral*	binary (1: belong to the neutral group and NA, 0: does not belong)
left	binary (1: belong to the left-leaning group, 0: does not belong)
right	binary (1: belong to the right-leaning group, 0: does not belong)
center*	binary (1: belong to the center group and NA, 0: does not belong)

Table 4.4: Independent variables (IVs) for Negative Binomial Models. The star sign (*) marks the reference variables.

For independent variables (IVs), we control for basic demographics (i.e., gender, age, and education) and timespan. Timespan quantifies the amount of time a respondent is potentially active on YouTube in the five-year period, ranging from 0 (not active at all) and 1 (active all the time). We compute this by calculating the number of overlapping months between the respondent's activity interval (from the earliest to the latest activity of a respondent) and the five-year interval, and normalizing it by the maximum length (i.e., 60 months).

The main focus of this analysis is to explore how political attribute varies across different engagement levels. To address this, we include the following political variables as IVs. Political interest is measured on a five-point scale (1 = not interested, 5 = very interested). Ideological variables are constructed using respondents' anti-/pro-government and left/right positions. The anti-/pro-government position is measured on a 0-10 attitude scale toward Fidesz, the incumbent party in Hungary (0 = strongly dislike, 10 = strongly like)[5]. As we

[5]In the sample, 46% of the respondents lean towards anti-government, while only 15% lean towards pro-

do not expect a simple linear relationship between this variable and engagement levels, we recode it into three binary indicators corresponding to ideological groups: anti-government (0-2), neutral (3-7), and pro-government (8-10). The neutral group serve as the reference in the regressions. We impute the missing values (see Appendix C.5), and adopt these binary indicators to preserve the statistical power given the limited sample size. Similarly, the left/right positions are also recoded into three categories: left (1-3), 4 center (4), right (5-7), with the center group used as the reference. Additionally, we include extremity measures for anti-/pro-government and left/right to capture the effect of individual polarity. Extremity value equals the absolute distance between the center (5 for anti-/pro-government, 4 for left/right) and the actual value. Descriptive statistics for the main variables are available in Appendix C.9.

For brevity, we present the results with the anti-/pro-government variables in the main text, since this dimension is the most salient one in Hungarian politics [255], and is correlated with left/right (Pearson's r = 0.59, $p \sim 0.00$), producing similar results downstream. Supplementary analyses using left/right variables are provided in the Appendix C.2.

Analyzing bias in observed selective exposure across different forms of engagement

Next, we investigate the variation in selective exposure–a tendency of users to selectively consume ideologically congruent content [34]–across viewers, subscribers and commenters. Selective exposure has been shown to be causally linked to political polarization [34], with social media potentially facilitating this process [82]. Evaluating differences in selective exposure across engagement forms helps identify potential biases in polarization-related findings based solely on visible engagement.

We characterize selective exposure by assessing the average and diversity of the content leanings for each respondent, and compare the distribution of these measures across respondent groups. For a given respondent i, we extract three sets of political content: videos they have viewed (V_i), channels they have subscribed to (S_i), and videos they have commented (C_i). For each set, we calculate the average and variance of content leanings[6], which indicate the overall orientation and the ideological diversity of consumed content.

To assess whether selective exposure patterns in visible traces (i.e., commenting) differ from those in invisible traces (i.e., viewing and subscribing), we compare the distribution of leaning average and variance using the Kolmogorov–Smirnov (KS) test. We benchmark the empirical D statistics against a distribution of statistics derived from within-group resam-

government. Please note that such an imbalance reflects the nature of Hungarian Internet users being generally more anti-government than the general population. Pro-government people use the Internet less frequently and are under-represented in all online samples [254].

[6] As discussed in Section 4.3.3, each channel is labeled as one of the following: pro-government (1), neutral (0), or anti-government (-1); and each video carries the label of its channel.

pling (see details in Appendix C.8), so that noises (i.e., within-sample fluctuation) are distinguished from what we hope to examine (i.e., between-sample difference).

4.3.6 Addressing the channel-level bias

Next, we turn to channel-level analysis. We show that relying on a convenient form of engagement (e.g., commenting) and using only visible traces to characterize the user bases of political channels, researchers risk overlooking structural properties of channel landscapes captured through less visible traces such as viewing and subscribing. We illustrate this bias using both graphical presentations of channel networks, and quantitative evidence of varying community structures of Hungarian political channels across different engagement forms.

Analyzing bias in degree of segregation in channel networks

We construct network projections for Hungarian political channels, based on three sets of engagement data from respondents' DDPs: viewing, subscribing, and commenting. These channel networks illustrate how channels are interconnected via shared user bases, with individual channels represented as nodes and links formed between channels that share similar audiences. To quantify audience similarity, we define a user vector $\vec{u}_i$ of length U (i.e., the number of unique users who engage in a certain form) for each channel i; the nth element of $\vec{u}_i$ equals the total number of times user n has engaged with channel i in a certain form. For any pair of channel i and j, we assess the audience overlap based on cosine similarity between $\vec{u}_i$ and $\vec{u}_j$.

$$\text{Similarity}(i,j) = \cos(\theta) = \frac{\vec{u}_i \cdot \vec{u}_j}{||\vec{u}_i||\,||\vec{u}_j||}$$

Edges in the network are weighted by these similarity scores. We set a non-zero threshold to filter out weak links for visualization puroses, and retain all links with positive weights in the assortativity analysis.

Measuring channel assortativity

Besides network visualizations, we also quantitatively evaluate the degree of segregation by assessing the extent to which ideologically similar channels share overlapping audiences. High segregation indicates that channels cater to distinct ideological audience groups, reflecting siloed consumption patterns within Hungary's political landscape.

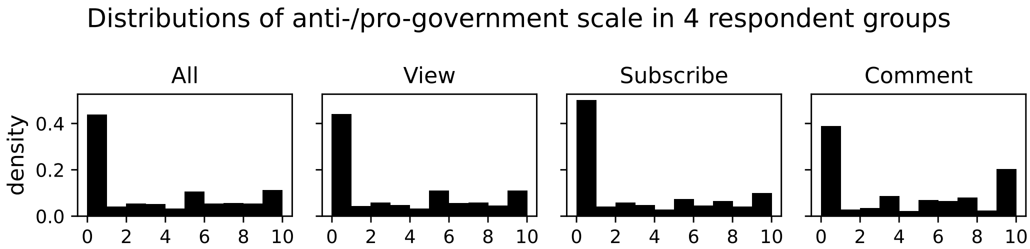

Figure 4.1: Distributions of respondent positions along the anti-/pro-government scale for four respondent groups (i.e., the entire sample, viewers, subscribers, and commenters).

We first compute an overall assortativity score of the leaning attributes for the entire network, then compute the EI homophily index [256] at the node level. The overall assortativity score equals the weighted correlation between the leaning labels of two connected nodes (i.e., channels) with a positive edge weight. The higher the correlation, the more likely two channels with the same leaning have a similar user base, hence the more segregated the entire network. At the node level, the EI index measures the extent to which a given channel shares a similar user base (and thus is connected) with others on the same leaning side (i.e., internal links), versus others on different leaning side (i.e., external links). A lower EI index reflects stronger homophily, contributing to a more segregated network overall.

$$\text{EI index} = \frac{\sum \text{external link weights} - \sum \text{internal link weights}}{\sum \text{external link weights} + \sum \text{internal link weights}}$$

As with the user-level metrics, we compare the distribution of EI indices across different engagement forms using KS-test. To test robustness, we conduct 1,000 bootstrap rounds and report two p-values for the D statistics, accounting for within-sample fluctuations (see details in Appendix C.8).

4.4 Results

4.4.1 User-level disparity

We first present our results from the user-level analysis. In this section, we report how engagement forms interact with ideological distribution (Section 4.4.1), how engagement levels associate with political attributes (Section 4.4.1), and how selective exposure patterns differ across engagement forms (Section 4.4.1).

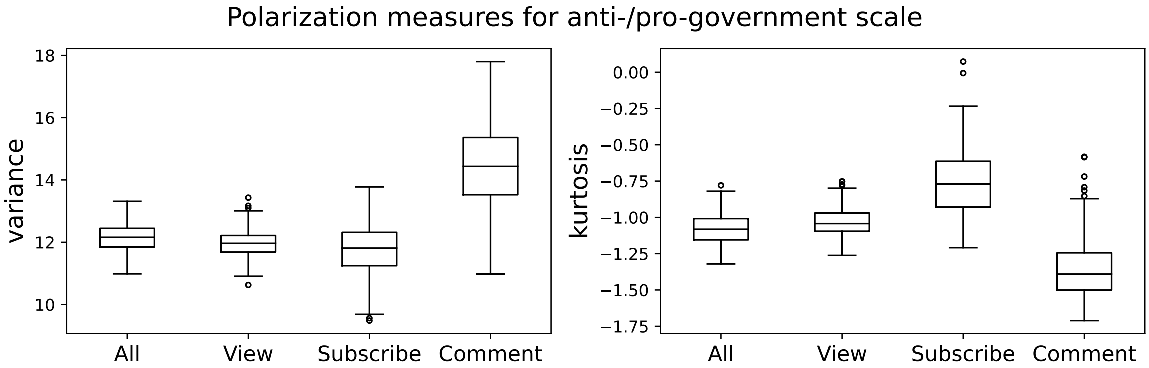

Figure 4.2: Variance and kurtosis of users' anti-/pro-government scale across four respondent groups (i.e., the entire sample, viewers, subscribers, and commenters). The value range comes from 200 rounds of bootstrapping with survey weights applied.

Engagement forms and ideological polarization

We start with comparing respondent groups in terms of self-reported ideologies[7]. We see slight visible variations in the ideological distribution between viewers/subscribers and commenters, but not between all respondents and viewers (see Figure 4.1). Anti-government respondents constitute the majority across all three groups, with their dominance being especially pronounced among subscribers. Pro-government respondents, in contrast, are more dominant among commenters compared to subscribers and viewers. These differences are reflected in the variation in average anti-/pro-government positions across three groups (3.24 for viewers, 2.85 for subscribers, and 4.07 for commenters). The tendency for commenters to concentrate less on the center and spread further to the ideological extreme suggets that the commenters of Hungarian political content likely represent a more polarized fraction of the Internet users.

As expected, commenters have a significantly higher level of polarization, with greater variance and lower kurtosis than viewers, subscribers, and all respondents (see Figure 4.2). Moreover, the higher proportion of anti-government and left-leaning respondents in subscribers also leads to subscribers having a higher kurtosis than viewers. This finding, combined with the above observations, suggests that subscribers exhibit a more decentralized ideological spread, albeit in an imbalanced way. Statistical tests using Mann-Whitney confirm these differences (see Appendix C.8 Table C.1).

[7] Since we do not see significant ideological variations in respondent groups in different filtering stages, we include the analysis for all respondents and respondent groups across engagement forms in the main text, and provide the results comparing respondent groups during filtering stages in Appendix C.1.

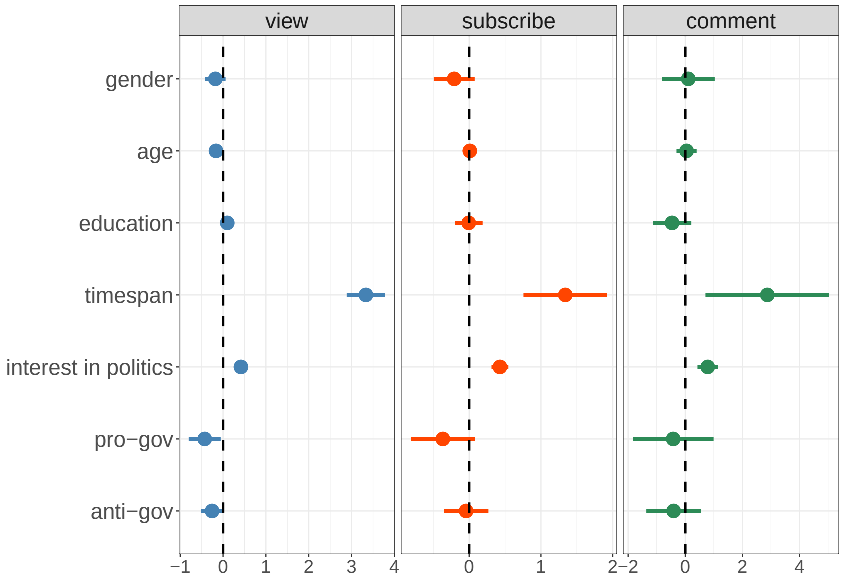

Figure 4.3: Negative binomial regression coefficients with 95% confidence intervals for viewing, subscribing and commenting activities on Hungarian political content. The DVs are the number of recorded activities for viewing, subscribing, and commenting from left to right for each panel.

Engagement levels and ideological characteristics

The previous analysis compares users who do and do not engage in a certain form but does not assess the difference among users who engage at different levels. Thus, we further examine how different levels of each engagement form associate with individuals' political attributes. As shown in Figure 4.3, we find that the only persisting salient factor is interest in politics. Individuals with greater political interest exhibit higher levels of engagement with Hungarian political content across all forms.

Controlling for basic demographics and timespan, we find that pro-government respondents are likely to engage in lower levels of viewing, as indicated by a significant negative coefficient (-0.429, p = 0.025); anti-government respondents, while showing a similar trend, do not exhibit statistically significant predictive power (coefficient = -0.255, p = 0.052). This asymmetry might stem from pro-government respondents being more more averse to ideologically incongruent content. When breaking down DVs into different channel categories, pro-government respondents would view significantly less neutral channels, while anti-government respondents would not (see Appendix C.3 Figure C.6). Models with left-/right variables deliver similar findings (see Appendix C.2). Interestingly, when we assess the joint effect of political interest and anti/pro-government variable, we see that highly-

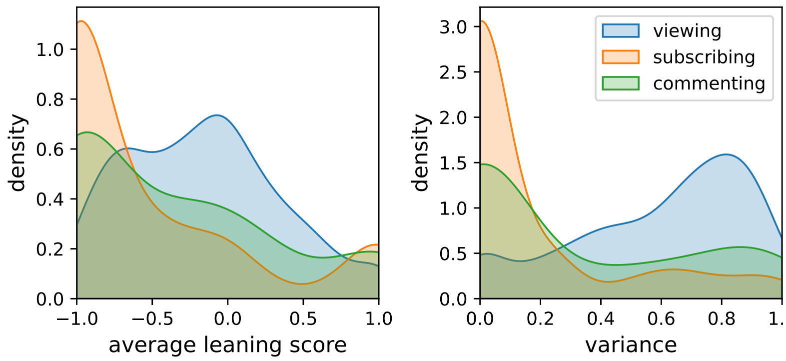

Figure 4.4: Average leaning score (left) and the leaning variance (right) for the Hungarian political content the respondents have engaged with through viewing, subscribing *or* commenting. An average leaning score of -1 (+1) means consuming only anti-government (pro-government) content, and 0 means consuming balanced or neutral content. Density functions are generated by kernel density estimate (KDE) methods.

interested people view much more political content and those uninterested, regardless of their political leaning (Appendix C.2 Figure C.5). This further confirms that interest in politics, instead of ideology, is the main driver of higher engagement levels.

If belonging to one side of the ideology generally cannot predict the engagement level, can the extremity of ideological positions do so? The answer is also no, as we do not see any strong coefficients from the extremity variables included in separate models (Appendix B Figure C.4).

For a robustness check, we re-run the models with an alternative filtering approach[8], and a different imputation method[9]. Both approaches yield results consistent with our main findings.

Engagement form and selective exposure

We now turn to respondents' behavioral aspect and explore the variation in selective exposure across different engagement forms. For each respondent, we compute the average and variance of the leaning of the content viewed, subscribed, and commented, and show the distribution for respondents with the corresponding engagement records in Figure 4.4. In the left panel, the average leaning score for viewing is more centralized and less skewed than subscribing and commenting, with its mean falling around -0.166 (-0.540 for subscrib-

[8]Instead of including only those who have engaged in a certain form (e.g., commenting), we also add those who have not engaged with a 0-value DV (see Appendix C.4).

[9]We used multiple imputation methods and analyzed pooled regression results (see Appendix C.5).

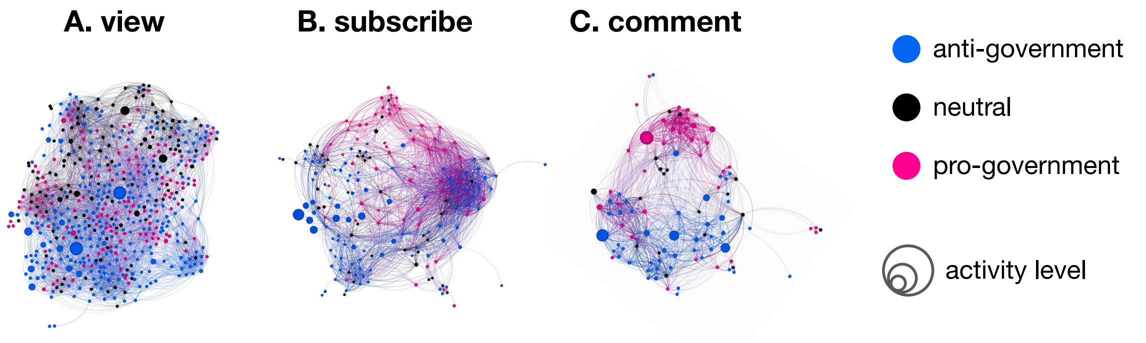

Figure 4.5: Networks of Hungarian political channels based on different types of user engagement. Each node represents one unique channel. Two nodes are connected if the users who engage with the two corresponding channels in a certain form overlap to a certain extent. Node color indicates the leaning labels for each Hungarian political channel, and node size is a function of activity levels for a given engagement type. Edges are filtered based on weights for visualization purposes (edge density from A to C: 0.02, 0.05, 0.1).

ing and -0.374 for commenting). This suggests that the selective exposure characterized through commenting or subscribing behaviors appears to be more severe than through viewing. This finding is supported by the right panel: content viewed by respondents tends to be more ideologically dispersed, with a higher variance than content subscribed or commented by respondents. Kolmogorov-Smirnov (KS) tests confirm that these visual differences are statistically significant (see Appendix C.8 Table C.2).

Building on this observation, we further ask if the disparity in selective exposure outcomes stems from differences in sample composition or in forms of engagement. In other words, can we conclude that users who engage via viewing content are more likely to encounter mixed content than those who engage via subscribing or commenting, or is it that the content one views differs systematically from what one chooses to subscribe to or comment? To disentangle these possibilities, we re-generate Figure 4.4 using only the subset of respondents who engaged in all three forms (N = 57, see Appendix C.10 Figure C.12). The resulting pattern remains broadly consistent: the content viewed remains more mixed than content subscribed or commented, as indicated by a more central mean and a higher variance. However, these differences are less pronounced than those in Figure 4.4, suggesting that the disparities observed in Figure 4.4 are shaped by differences in both sample composition and engagement form.

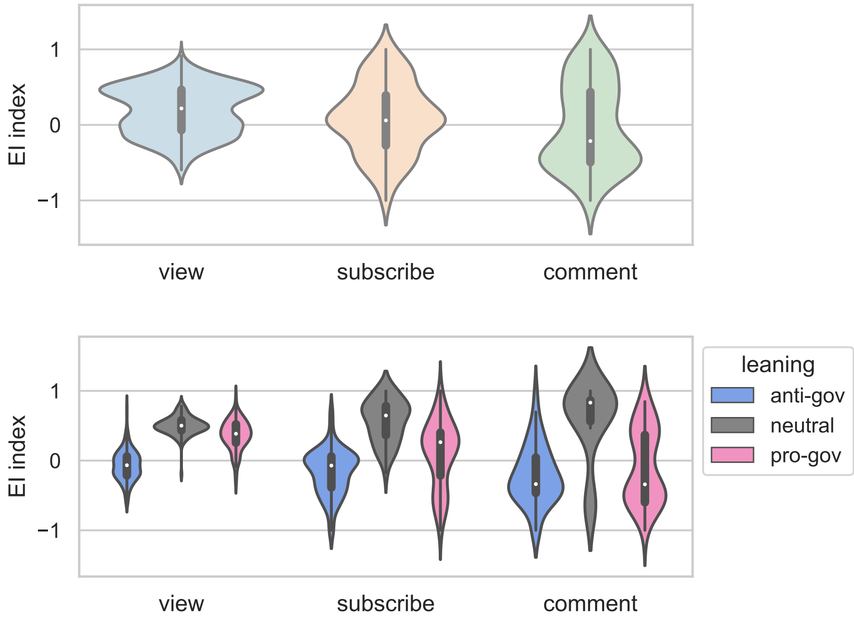

Figure 4.6: Node-level EI homophily indices across channel networks based on different forms of engagement. The upper figure shows the overall distribution; the bottom figure breaks down by channel leanings and shows the group-wise variation.

4.4.2 Channel-level disparity

Finally, we turn to channel-level disparity and show how the channel audience overlap depicted based on different forms of engagement can display varying levels of segregation.

Engagement forms and audience segregation

Many studies that do not aim to infer a user-level landscape of political communication on YouTube, adopt a descriptive approach at the channel level instead, using channels as the basic unit of analysis. Accordingly, we examine channel-level biases that may arise from focusing only on comments when characterizing channels.

We first generate three channel networks based on viewing, subscribing, and commenting data from DDPs (see Figure 4.5). Compared to the viewing network, the subscribing and commenting ones appear more segregated along the anti-/pro-government cleavage, where ideologically-aligned channels sharing more common users and clustering more closely together. We then assess these variations quantitatively. The network-level assortativity scores confirm that the viewing network has the lowest level of segregation, and that the commenting network is the most segregated among all (see Table 4.5). At the node level, the distribution of EI indices from networks built on commenting data differs significantly from others: nodes in the commenting network generally have lower EI indices than those in the viewing and subscribing network. Again, we perform KS-test to formally test these differences and confirm that they are statistically significant (see Appendix C.8 Table C.3).

Additionally, when the distributions are segmented by leaning classes (see Figure 4.6), neutral channels show the highest degree of mixing with channels of different leanings, while anti-government channels are generally more segregated than pro-government channels.

Graph	View	Subscribe	Comment
Assortativity	0.0833	0.1289	0.3154
Average EI	0.1909	0.0801	-0.0298

Table 4.5: Assortativity scores and average EI index for channel networks based on different forms of engagement.

4.5 Discussion

4.5.1 User-level: the incomplete picture of observed polarization

At the user level, we investigate whether and how engagement forms (i.e., how people engage–viewing, subscribing, or commenting) and levels (i.e., how much people engage) correlate with users' political attributes (e.g., ideological leaning, political interest) and patterns of selective exposure. Our findings suggest that, polarization measurements based solely on visible engagement (e.g., commenting) or excluding less politically active users can be biased in terms of ideological distribution and selective exposure.

First, for respondents who engage in different forms, we show that respondents who engage through commenting represent a subset more polarized in their self-reported ideologies compared to those who engage through viewing or subscribing. This finding is in line with insights from a larger-scale study on Facebook by González-Bailón et al. [257], where liberal and conservative users are more deeply segregated through visible engagement (i.e., clicks, reactions, likes, reshares and comments) compared to invisible engagement (i.e., views) during the U.S. 2020 election. We therefore caution researchers who rely on YouTube public comments for user-level analysis, that such visible samples cannot represent all politically active users on YouTube, and are likely a more polarized fraction of the population.

However, among respondents who differ in their level of engagement, we find no significant differences in ideological leaning. Individuals who view more videos, subscribe to more channels, or comment under more videos, are not significantly more pro- or anti-government, or ideologically extreme, yet they are more politically interested. Thus, excluding users who engage at low levels risks omitting individuals who are less politically interested.

The above two observations jointly reveal the nuances of where and how user-level bias may occur. Those who only engage through commenting form a biased subsample in terms of ideological composition. However, this bias does not extend to differences between those who comment little and those who comment substantially. For levels of engagement, the clearest associated factor is political interest rather than ideology. While these outcomes diverge on whether self-reported ideologies is a primarily biased factor, they are not directly comparable due to discrepancies in the model design, such as sample variation, the presence or absence of control variables (e.g., political interest) and different binning of pro-/anti-government variables (see Section 4.3.5 and 4.3.5).

In addition, our analysis of respondent's selective exposure suggests that selective exposure is more pronounced in commenting and subscribing than viewing. Although respondents selectively comment under videos and subscribe to channels, they are fairly open to viewing ideologically neutral or diverse content on YouTube. This pattern aligns with Guess's finding that people have a moderate media diet when browsing news content online [258]. Thus, characterizing selective exposure based solely on the commenting data [e.g., 131] can overstate the extent to which users opt into ideologically homogeneous content, consequently exaggerating the existence of echo-chamber on YouTube.

4.5.2 Channel-level: shared views but segregated comments

At the channel level, we find that audience overlap among Hungarian political channels depends on how *audience* is defined. Works that collect data via the YouTube API typically define audience as users who have made comments under videos of interest [e.g., 131, 227, 259]. We argue that this approach can yield biased measurements of polarization, if polarization is quantified based on the degree to which certain channels form distinct communities and fragment the politically active users.

Indeed, our channel-level results show that the network of channels conditioned on commenting data does differ structurally from that conditioned on viewing data, which are not captured in YouTube API. Specifically, viewers of Hungarian political content show substantial overlap among channels with different leanings, while commenters are more segregated into ideological silos. This suggests that observations based on visible engagement (e.g., commenting), reflect a tilted, and potentially exaggerated picture of polarization. Therefore, when using only the commenting dataset, researchers should interpret evidence of "echo chamber" on YouTube [e.g., 260] as reflective of commenting behavior rather than general content consumption, and refrain from extending findings to broader patterns of engagement.

4.5.3 Sources of biases

We contextualize our findings using Total Error Framework for Digital Traces of Human Behavior on Online Platforms (TED-On) [224], which provides a standard vocabulary to describe the source of measurement biases. By examining the differences between visible and invisible digital traces, this chapter centers on measurement bias coming from trace selection errors, although we have dealt with trace reduction error and trace augmentation errors on the sideline.

The primary contribution of our paper is to quantify trace selection errors stemming from considering only visible political engagement data on YouTube. Trace selection error refers to the measurement error caused by selectively analyzing some type of digital traces and overlooking others. Our paper shows that when researchers select visible forms of engagement (i.e., comment) to construct polarization measures, the outcome can be incomplete, often reflecting a greater degree of polarization compared to the outcome yielded by other invisible engagement forms.

Although our analysis does not focus on data pre-processing (e.g., labeling content), it is worth noting as a potential source of error in polarization metrics. Sen et al. define two error sources during pre-processing: trace reduction error and trace augmentation error [224]. The former occurs when relevant traces are excluded or irrelevant traces are included, while the latter refers to errors caused by inaccurate annotations of digital traces.

In this study, we control these errors by covering as many popular channels as possible when filtering for political content (i.e., trace reduction) and assigning ideological labels (i.e., trace augmentation). While automatic tagging methods are improving, off-the-shelf solutions for context-specific data remain limited. We opted for manual labeling, verified for accuracy, but given the time and resource restrictions, we had to limit the amount of content to be labeled. Although we inspected channels not explicitly tagged as "politics" but containing videos with "politics" tags, we did not examine the vast amount of channels without "politics" videos. Preliminary analysis indicates that most of this excluded portion consists of music content. We also excluded political videos viewed by only one or two users, as these channels tend to have a significantly lower proportion of political content compared to channels with more viewers. To mitigate the risk of missing out widely viewed political content, we ensured that our labeled dataset covers more than 90% of total viewership.

Nevertheless, we acknowledge a potential source of trace augmentation error stemming from our channel-level (instead of video-level) labeling. Since not all videos uploaded by political channels are political, this may lead to an overestimation of user engagement with political content. This error could be better controlled by video-level annotation using context-specific classifiers.

4.5.4 Addressing biases at multiple levels

So far in this chapter, we have shown that by selecting data points of visible political engagement, measurement bias can arise at both the user and channel levels, depending on what aggregation units are adopted when measuring polarization. This multi-level perspective is not only necessary for understanding how such bias may skew measurement outcomes, but also valuable for tracing the potential source of imbalance–for example, the disparity in commenting behavior between politically interested and politically uninterested participants–that shapes the observations. This section briefly discusses supporting work of mine that uses a different dataset and investigates patterns of visible engagement at both the user and post levels. In this study, we show that focusing on visible engagement such as liking, retweeting, quoting, and replying can yield a skewed picture at the post level but not necessarily the user level. For technical details and complete results, please refer to Baqir et al. [261].

Based on a dataset of online discussions related to providing military aid to Ukraine on Twitter/X, we leverage counts of visible engagement (i.e., likes, retweets, quotes, and replies) and impressions (i.e., the number of times a post was viewed) across 17M+ posts collected from 5.2M+ users. We analyze the rate of visible engagement[10] at the user and the post level, that is, how likely a user or a post can attract visible engagement per impression.

The results reveal that, at the user level, anti-aid users do not differ significantly from pro-aid users in terms of their visible engagement rate. At the post level, however, we notice that posts sharing highly partisan news sources (i.e., those classified as extreme right, right, or left) tend to have higher median and more variable visible engagement rates than those sharing moderate sources (i.e., right-center, center, and left-center). Alarmingly, posts that link to low or very low credibility news sources consistently show higher visible engagement rates. These results highlight a risk of skewing engagement data toward posts that share highly partisan and low-credibility sources, if analyses only consider visible engagement metrics. For polarization measurement, this skewness may lead to an exaggerated picture of polarized content engagement, as it disproportionately captures activity concentrated at the political extremes with more radical stances, rather than within the moderate center.

[10] In the original paper this is termed as "active engagement", but is conceptually aligned with what I define as "visible engagement" in this chapter. To avoid terminological confusion, I retain the term "visible engagement" throughout.

4.6 Limitations and future work

Back to the main study using survey responses and YouTube digital traces, we recognize several limitations and encourage future researchers to explore these under-explored areas.

First, our analysis concentrates on political content identified using YouTube's topic tags. While convenient and commonly used in prior research [e.g., 236, 262], these tags may not serve as a definitive ground truth. Although we manually verified the list of Hungarian political channels for precision, the recall remains unclear–we may have missed political content not tagged as "politics" by YouTube. Additionally, seemingly apolitical content on YouTube may still be relevant to political communication, given the increasing politicization of pop culture on social media [263].

Second, we use manually assigned ideological labels for political channels and videos using a three-category spectrum (i.e., pro-government, neutral, anti-government). While this captures the most dominant political divide in Hungary, it offers limited granularity. Future work can expand this framework to reflect on additional ideological dimensions or topic cleavages. Even within our simplified categories, annotators have struggled with a low inter-rater agreement ($\kappa = 0.42$). However, weighted κ–adjusted for the number of watched videos–are substantially better (0.83), indicating strong agreement for more-viewed channels, which also carry greater weight in our analysis. We also assume that channel-level ideology aligns with the ideological slant of individual videos–a simplification that future work could test empirically.

Third, our analysis includes only Hungarian political channels, which may not capture the full scope of political content on YouTube. While our case study provides a fresh perspective into a relatively understudied political context, we caution readers about generalizing these findings to other countries. Furthermore, the engagement forms we analyze are specific to YouTube and those stored in DDPs, which can limit cross-platform comparisons. For instance, video likes were unavailable in the DDPs at the time of data collection, so we cannot evaluate the polarization in liking behaviors. Other forms of engagement such as searching and re-posting may reflect different cognitive mechanisms and collective patterns not captured in our study. We hope this work motivates future research using more diverse datasets across other contexts and platforms.

Relevant to a broader exploration of different engagement forms, a closer examination of nuanced engagement mechanism may yield further insights. For instance, strategic commenting on opposition content would actually increase cross-cutting interactions, which would weaken the observed pattern of interactional segregation if researchers only consider who commented under whom without accounting for the textual context. Therefore, polarization studies can benefit from a multi-faceted analysis that integrates both behavioral and textual data of online political engagement. Beyond mapping who talks to whom, examining the context of in-group and out-group dialogues could provide valuable insights. Passive viewing without cognitive engagement is another important behavioral aspect. Relatedly, existing work on "incidental news exposure" – the unintentional or unplanned encounter with (political) news content, often occurring without deeper cognitive engagement – shows that it is negatively associated with online political participation [264], and linked to lower political knowledge when accompanied by discussions with weak ties [265]. For polarization studies using viewing data, this motivates a further breakdown of active viewing and passive viewing activities.

Fourth, our analysis is limited by a small sample (especially in commenting data). For instance, our sample includes only 72 respondents who commented on Hungarian political videos. This small size increases susceptibility to sampling error and volatility in exposure patterns and network analysis. In addition, hyperactive respondents disproportionately influence the structural properties of channel networks. While we have implemented bootstrapping to statistically disentangle the within-sample fluctuation from between-sample variation, we caution researchers when interpreting such a small-sample results.

Fifth, our sample may suffer from coverage bias due to the requirements for data donation. Although our target population is Hungarian internet users aged 16+, we only included those who regularly use Google or Facebook. Despite the fact that both platforms are widely used in Hungary (penetration $> 85\%$ based on our preliminary study, see Appendix C.6), this criterion may exclude relevant subgroups.

Finally, although the data donation approach provides a richer dataset than API-based collections, there can still be missing data if respondents having multiple YouTube accounts or have cleaned viewing history. In our sample, 46 respondents have comments or subscriptions but no views in their DDPs, likely due to disabled watch histories. Thus, we run analyses on alternative samples (see Section 4.4.1 and Appendix C.4) to ensure that our results are robust. However, this highlights the need to consider the methodological consequences of selecting specific traces to measure a construct.

Chapter 5

Conclusions, limitations, and future directions

5.1 Conclusions

This dissertation sets out to examine how political communication unfolds within an increasingly complex sociotechnical system, and how existing research must adapt conceptually and methodologically to capture that complexity. Motivated by the democratic imperative to understand and facilitate meaningful public discourse, I explore how polarization has been conceptualized and measured in the contemporary media environment, while addressing challenges in polarization measurement: (i) moving beyond unidimensional analyses to capture the dynamics of belief systems during the polarization process, and (ii) incorporating both visible and invisible forms of online engagement to provide a more complete and nuanced understanding of polarization.

In Chapter 1, I establish the theoretical foundation for this investigation by broadly describing political communication in today's information ecosystem, and reviewing how political polarization has been defined and operationalized across disciplines. I propose a conceptual distinction between attitudinal and behavioral polarization and identify key limitations in each perspective. This framework laid the groundwork for the empirical chapters that followed, each addressing a specific gap in how the theory of intermedia agenda setting and the phenomenon of political polarization are studied in the digital age.

Chapter 2 serves as a background study of the information landscape shaped by mass media in a high-choice, politically fragmented environment. Based on intermedia agenda-setting (IAS) theory, the case study analyzes how candidate-related news agenda has evolved across the 2016 and 2020 U.S. presidential elections. This study reaffirms the presence of IAS dynamics among media outlets of varying credibility and different ideologies, reveals partisan asymmetries in agenda convergence between coverage of Democratic candidates and that of Republican candidates, and raises concerns about a potential decline in the IAS power of high-credibility media. The analytical pipeline developed for this project showcases a scalable approach to dynamic IAS modeling, illustrating how computational methods can be effectively adapted with manual checks in stages of model design and validation to suit a specific context. Overall, this chapter contextualizes the broader media ecosystem in which political polarization may take shape, serving as a background picture for understanding the shifting terrain of public discourse from the supply end of information.

The later two chapters turn towards the topic of political polarization and delve into its measurement based on attitudinal and behavioral data, respectively. Chapter 3 starts with the attitudinal perspective of political polarization. Here, I emphasize the multidimensional nature of political ideologies and discuss the limitations of traditional measures–including linear measures and existing BNA approaches–in belief structure analysis. Building on Converse's conception of belief constraint [16], I apply a novel belief network analysis approach–response-item network (ResIN)–to visualize the spatial divide of opposing ideologies in the U.S. political belief systems. Using ANES data (2000-2020), ResIN captures the "egg breaking" process of ideological polarization, i.e., the transformation of a loosely organized belief structure into a more constrained, flattened out unidimensional shape. In addition to the graphical presentation, I also propose network-based metrics to quantify polarization at the system and attitude level. Using two system-level measures, the study shows an overall trend of increasing ideological polarization over the two-decade period, with the most dramatic rise occurring between 2016 and 2020. In addition, there is an intriguing partisan asymmetry in belief system constraint, with Democrats' belief systems generally appearing less coherent than those of Republicans during this period–a pattern that only recently reversed in 2020. At the attitude level, the study identifies the different roles that particular attitudes played in the structure of different belief network communities. For instance, abortion appeared as the most central issue for Democrats but not Republicans; government spending on public services and healthcare was one common ground for bridging both of the otherwise highly polarized partisan communities. In sum, Chapter 3 shows that ResIN is a promising tool to assess the state and process of polarization, as demonstrated by the empirical examples using ANES data. ResIN goes beyond the information provided by BNA by intuitively visualizing ideological divide throughout the polarization process and providing multi-level quantitative measures for these dynamics.

Next, Chapter 4 switches into the behavioral perspective of polarization measurements, reflecting on the measurement bias introduced by the sole reliance on visible engagement data–the engagement data retrievable via public platform APIs, such as commenting data on YouTube–from social media platforms. Using individually-linked survey responses and donated digital traces on YouTube, I compare various measurement outcomes of political polarization yielded by data of visible engagement (i.e., commenting) and of invisible engagement (i.e., viewing, subscribing). The study reveals that (1) respondents who engage visibly (i.e., commenters) form a more polarized subset than those who engage invisibly (i.e., subscribers and viewers), (2) individual engagement level does not associate with their ideological leaning or extremity, yet is significantly correlated with the level of political interest; (3) selective exposure is more pronounced in engagement through commenting and subscribing compared to viewing, and (4) while politically divergent channels attract mixed viewership, their subscriber and commenter bases are more ideologically siloed. Overall, commenters and their commenting behaviors–which are more accessible via public APIs and thus more frequently studied–exhibit higher levels of polarization and ideological selectivity. Thus, visible engagement on YouTube portrays an incomplete picture of political communication, potentially leading to inaccurate and often exaggerated estimates of polarization.

This dissertation makes conceptual, methodological, and empirical contributions to the study of political polarization in today's complex media and information environment. Positioned at the intersection of political communication research, computational social science, and network science, my work critically examines how polarization is operationalized and measured across different data types and epistemological traditions. In doing so, it both clarifies foundational concepts and advances innovative tools for future inquiry.

First, a core conceptual contribution of this work lies in the distinction it draws between attitudinal and behavioral polarization. While prior studies have often employed mixed terms without articulating the specific type of polarization being measured [11, 83], this dissertation carefully categorizes existing measures and shows how their emphasized aspects and data sources differ. By explicitly mapping out this distinction, it enables clearer interpretation of empirical findings, reconciliation of discrepancies across studies, and the development of a more coherent analytical vocabulary. This conceptual clarity not only situates past and current research within a structured typology but also lays the groundwork for better-targeted research designs.

Second, methodologically, this work extends belief network analysis (BNA) to the study of political polarization, addressing the challenge of unidimensional analyses that fail to capture the interconnectedness of political belief systems. The proposed ResIN model offers a novel approach for operationalizing political belief systems as statistical networks of interconnected attitudes. Unlike conventional measures that often impose a linear or unidimensional view of political ideology, ResIN captures the multidimensional nature of belief structures, providing visual and quantitative tools to model system-level and attitude-level shifts over time. This contribution not only enriches the methodological toolbox of polarization research but also strengthens the bridge between political polarization research and network science.

Third, this dissertation addresses a frequently overlooked challenge in computational polarization research: measurement bias resulting from a sole reliance on political engagement captured by publicly accessible digital traces. Through the analysis of a combined dataset that integrates survey responses with user-donated YouTube digital traces, our study answers the previously obscure question of the extent to which polarization patterns are reliably captured by limited datasets, showing how focusing exclusively on visible engagement (e.g., commenting) may yield a biased view of online political behaviors and eventually an incomplete picture of political polarization. By foregrounding the distinction between visible and invisible forms of engagement, this work underscores the importance of diversified data sources and reflective scoping in designing behavioral polarization studies. The proposed framework encourages future scholars to ponder and articulate: polarization in terms of what form of engagement, and among whom?

Fourth, the empirical chapters contribute to a more globalized and context-sensitive understanding of polarization. Drawing on datasets from Hungary, the U.S., and other European countries, the research examines the phenomenon of polarization across different political systems and ideological cleavages. In particular, it broadens the field's U.S.-centric lens based on a binary party system, by discussing how polarization can manifest through different structural divides and in multiparty contexts. This perspective also motivates more comparative research to explore the mechanism of polarization beyond the left–right spectrum.

To sum up, these contributions aim to advance the field of political communication in four key directions: by refining the conceptual taxonomy of polarization, by applying a novel network model that accounts for the complexity of belief systems, by addressing the underexplored measurement bias in limited datasets, and by broadening empirical inquiry to account for both methodological and contextual diversity. In an era where political communication unfolds in an increasingly complex ecosystem, and where data is abundant but sometimes supports a restricted view, this dissertation provides a grounded yet forward-looking framework for understanding polarization as both a measurable and meaning-laden social process.

Beyond these scholarly contributions, the findings of this dissertation also speak to the normative stakes of democratic discourse. Polarization measurement is not only a technical matter. It can often shape how the policymakers, media, and the public perceive and engage with the challenges in online participatory democracy. Compared to sweeping claims such as "social media really is undermining democracy" [266] or social media is the "polarization engine" [267], adopting a more precise language in the conceptualization of polarization, as well as encouraging more nuanced interpretations in context-specific studies, can better inform public understanding of democratic disagreement, the design of healthier deliberative systems, and the development of policy interventions that respond to actual rather than perceived divisions. A narrow focus on visible engagement, for example, may obscure how deliberative democracy actually functions on social media–whether these platforms promote or undermine what Young identifies as "inclusion, political equality, reasonableness, and publicity" [268]. By examining these biases, my work highlights how research practices can expand our understanding about polarization, potentially impacting public discourse, media framing, and policy responses. In this sense, it encourages scholars and practitioners alike to approach measurement with reflexivity, recognizing its role in constructing, not merely describing, the social reality of political communication.

Meanwhile, my work has practical implications for platform design and policy. Current platform data-sharing schemes prioritize certain engagement forms while rendering others less visible to researchers and less salient in public discourse. This restricted visibility of online political engagement risks distorting both measurement outcomes and public perceptions. Academics and social media platforms should jointly negotiate a more balanced and transparent data sharing scheme that allows researchers to ethically access multidimensional engagement metrics. Moreover, as visible engagement often overrepresents politically active and hyper-partisan users, recommendation systems, if calibrated primarily to such engagement signals, may unintentionally amplify extremity and antagonism. This calls for more thorough algorithmic auditing at the platform to to constantly review design choices that may exacerbate polarization while safeguarding pluralistic debate.

5.2 Limitations

Below, I summarize the key limitations of each chapter as well as the joint challenges of my work.

In Chapter 2, the intermedia agenda setting (IAS) analysis is subject to three primary limitations. First, although we employ terminology such as "agenda setter" and "Granger causality," the statistical approach relies on correlational associations, which cannot establish strict causality. While the analysis aligns with conventions in the IAS literature, the findings should be interpreted as descriptive associations rather than causal claims. Second, our dictionary-based model relies on context-specific topics and keywords. Although the dictionaries are applicable and comparable across years, they are not easily generalizable beyond the issues for which they were constructed. Third, the scope of our analysis is restricted by the limited coverage of ideological labels provided on Media Bias Fact Check.

In Chapter 3, the ResIN study faces three central limitations. First, the network structure and derived measures are largely dependent on the selection of political issue items. Although carefully chosen, the set of five issues may not capture the full complexity of political belief systems. Second, because ResIN-based polarization measures are novel, their predictive validity and robustness remain underexplored. More work is needed to test their utility across datasets, contexts, and simulations, particularly in modeling processes such as belief constraint or attitudinal structuration. Third, while the ANES-based case study demonstrates ResIN's value in the U.S., the model's applicability across different political contexts remains uncertain. ResIN can, in principle, capture multidimensional ideological structures, but projecting belief networks onto two-dimensional spaces may obscure complexity in systems not organized along a simple liberal–conservative axis.

Finally in Chapter 4, the Hungarian YouTube case study is limited in the following respects. First, the small sample of commenters (N = 72) restricts statistical power and makes the network structure vulnerable to distortion by hyperactive individuals. Although we employed bootstrapping to mitigate these risks, we caution readers to carefully interpret the results. Second, the identification of political content and the further categorization of anti-/pro-government channels is coarse and provides limited granularity. Third, the analysis focuses exclusively on Hungarian political channels, limiting generalizability across countries and platforms. Engagement forms are also constrained by the availability of digital trace data: for instance, likes and search behaviors were not captured in the dataset. Fourth, we encounter missing data at various phases of this project. Given the reliance on data donation from users of Google and Facebook, data from certain subgroups may be systematically excluded; within the DDPs, missing data may also arise when watch histories are disabled or fragmented across multiple accounts, raising questions about trace completeness.

Some common challenges emerge across these empirical chapters. First, establishing causal mechanisms remains difficult with observational data. While the descriptive patterns of IAS and belief system dynamics illustrate what these phenomena look like, they cannot fully explain why they occur. Second, case studies rooted in specific issues, contexts, or platforms should be interpreted with caution, as their external validity is limited. Comparative designs and multi-platform datasets are necessary to assess generalizability. Third, whether relying on curated source lists, survey items, or donated platform traces, each chapter faces potential coverage and selection biases. The proposed models and frameworks must therefore be tested against broader contexts and more diverse datasets.

5.3 Future directions to explore

Building on existing findings of my PhD-level work, I see a few promising directions for future research. In what follows I will discuss three potential avenues for exploration, articulating the unaddressed research gaps and open questions.

5.3.1 Extending ResIN to analysis of polarization in other political contexts

In Chapter 3, I apply ResIN to the study of ideological polarization in the context of U.S. politics. Political polarization, however, is a global challenge faced by many democracies

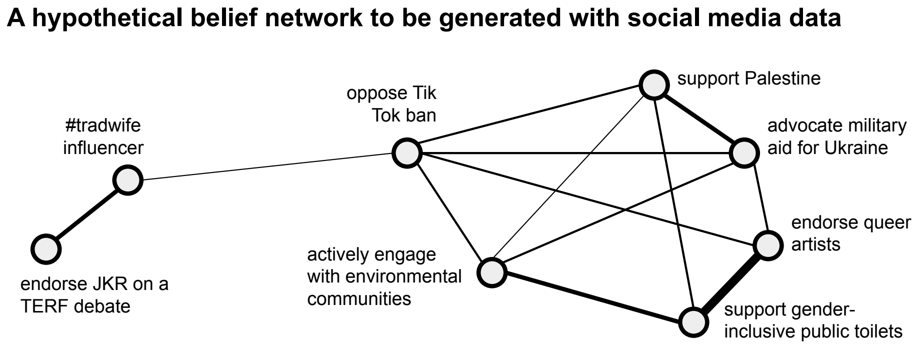

Figure 5.1: A hypothetical belief network to be generated by social media data. Nodes represent individual beliefs, and links are weighted by the extent of overlap between respondents who endorse both beliefs. Panel C illustrates the human-in-the-loop pipeline to build belief networks using social media data.

[28, 70], with its empirical research concentrating disproportionately in the U.S. context [12]. As noted in Section 3.5, it is worth exploring whether ResIN's framework for polarization modeling holds its interpretive value in other countries.

In the book chapter [13], we provide a preliminary outlook on ResIN's capacity to reveal the distinct belief structures across different European countries. Using the data from the 9th wave of European Social Survey (ESS), we generate ResIN snapshots for each participating country, and demonstrate how ResIN captures distinct belief structures through three examples (i.e., Slovakia, Spain and Germany). Following this study, many open questions remain to be explored. For instance, how does the process of polarization unfold in a political system with coalition parties? Does the belief system always collapse into a single liberal-conservative dimension, or can it evolve along multiple ideological axes? How can the polarization measures–afforded by ResIN–that have been developed in the context of the U.S. be adapted to other contexts where the public belief system may form more than two ideological clusters? These questions invite a more nuanced conceptualization of polarization–one that moves beyond binary ideological divides and accommodates higher-dimensional, more intricate belief structures. In this vein, future work can extend the ResIN approach to datasets from other countries and offer a comparative perspective of political polarization across diverse political landscapes.

5.3.2 Examining the effect of media exposure on belief system evolution

Chapter 3 illustrates how the structural organization of U.S. public belief systems shifted from 2000 to 2020. While the present analysis focuses on visualizing these structural patterns and summarizing key polarization trends, an important next step is to investigate the causal mechanisms driving such transformations. One particularly promising direction lies in examining how patterns of media exposure, especially within today's hybrid information environment, shape the evolution of belief systems.

Agenda setting and framing theory suggest that media content not only determine what issues are salient for the audience, but also how attributes are associated with certain issues–the "frame"–in audiences' minds [48, 49]. Repeated exposure to co-occurring issue frames, thematic linkages, and partisan cues can foster the perception that certain political beliefs "belong together" [269], thereby reinforcing ideological consistency across different issues. Furthermore, social media recommendation systems often amplify emotionally charged or ideologically extreme content [270, 271], which may shape users' belief systems by strengthening perceived links among radical beliefs and fostering the formation of radical ideological communities.

Thus, it is reasonable to hypothesize that individuals more frequently exposed to homogeneous media content–whether through mass media or algorithmic curation–are more likely to internalize heuristic signals about how issues are associated, thereby developing more ideologically consistent belief systems. Future research could test this hypothesis by analyzing panel datasets with media-use variables and extending the ResIN approach to include individual-level measures of ideological consistency. Such designs would allow for a more causal assessment of whether, and how, media exposure patterns contribute to the evolution of belief systems.

5.3.3 Identifying bridging attitudes and applying ResIN to analyzing social media data

Confronted with a visible divide and distinct belief structures between liberals and conservatives, how do we facilitate cross-community dialogues to achieve effective democratic deliberation? Building on Chapter 3, another meaningful future direction is to use ResIN to systematically identify and leverage bridges in political beliefs with great potential for depolarization.

Bridging political disagreement requires moving beyond the simplistic left-right binary and capturing the complex, multidimensional nature of political beliefs. For instance, individuals with opposing partisan leanings may still agree on certain policy issues [15], and such partial agreement can be the starting point of mutual understanding. A common ground of political beliefs is therefore the belief(s) that can be shared between two opposing communities (e.g., liberals and conservatives), cross-pressuring the public opinion to not collapse into an all-encompassing divide. Identifying such common ground can inform strategies to reduce the perception of a rigid "us versus them" divide. Indeed, studies on depolarization strategies have shown the effectiveness of emphasizing common-ground beliefs, such as shared national identities and moral similarities [272]. Yet, there is a lack of research that systematically identifies bridges between opposing political communities through data-driven analyses.

To unveil the common ground across political communities, it is essential to consider the interdependency of multiple political beliefs, and a network view of political belief systems is highly valuable. As discussed in Chapter 3, a belief network is a representation of interconnected beliefs based on the co-endorsement patterns in mass public, and ResIN offers a powerful tool for capturing these attitude-level nuances and mapping the structure of public belief systems.

Traditionally reliant on survey data, BNA (as well as ResIN) research has been bound by upstream availability (i.e., which issues are included in the survey) and downstream researchers' selections (i.e., which issues are deemed relevant). To overcome these limitations, I propose extending ResIN to the analysis of social media data to identify common-ground beliefs at scale, using a human-in-the-loop pipeline that leverages unsupervised topic modeling (e.g., BERTopic) and large language models (LLMs) to systematically identify and refine belief units from unstructured social media texts. My key research questions include: What political beliefs are shared between liberal and conservative communities? Under what circumstances do shared beliefs foster meaningful cross-partisan conversations? How can recommendation algorithms and content moderation be designed to promote bridge-building across communities? By addressing these questions, this research enables the analysis of belief systems in open public discourse and expands the topical scope of BNA in polarization research, which I hope can offer actionable insights into avenues for depolarization.

5.3.4 Comparing attitudinal and behavioral measurements of political ideologies

The third potential area for investigation is to compare the ideologies measured based on attitudinal and behavioral data. As summarized in Section 1.3 and further discussed in Chapters 3 and 4, polarization is usually measured using either behavioral or attitudinal data, but it is rarely observed jointly from both perspectives. In this section, I argue that not only is such a joint analysis necessary for defining the applicable scope of existing assumptions adopted in behavioral observations, but it is also meaningful in revealing the variation in attitude-behavior interdependency in people's political participation.

When measuring polarization through online behavioral data, researchers cannot directly obtain individuals' ideologies as in survey responses, and thus normally infer individuals' political ideologies based on their behavioral patterns [e.g., 80, 55, 59, 273]. The linkage between behavioral preference and ideological leaning is thus an important assumption for the latent variable models applied in these studies, which aim to estimate the "ideal point" of ideology based on the behavioral traces–textual or interactional–an individual has left on social media. For instance, the model used in Barberá [67] estimates Twitter/X users' ideological leanings based on their following behavior, assuming the ideological leaning of a given user is similar to the politicians they follow. This method has been extended to the reposting behavior [176], assuming that the users would repost those who share a similar ideology to theirs. For textual behaviors, Vafa et al. [273] utilizes lawmakers' post corpus on Twitter/X to estimate their ideological leanings. The assumption for this text-based ideal points (TBIP) model is that the language one uses on social media reflects their ideological leaning to a reliable extent that such a model would yield reasonable estimates.

Although some of these works have validated the estimated ideologies with offline sociodemographic data [80, 55], the extent to which these assumptions hold across contexts remains obscure. One might ask, for instance, whether such models can be adapted for other forms of engagement such as liking, replying, or viewing; or for other platforms such as YouTube or Reddit. Further, one may also be interested in examining whether the accuracy of such estimation would have a systematic bias towards those politically interested compared to those who are politically indifferent and thus ideologically omnivorous. Similarly, is there a bias towards a certain type of content with left- or right-leaning ideologies? Is there an asymmetry in the estimation accuracy between the left-leaning and right-leaning users? Essentially, the nuances to be uncovered here involve the variation in attitude-behavior association across different analytical units (e.g., individual, content), especially when these units constitute the fundamental building blocks of an online political engagement landscape where polarization is measured and interpreted. Addressing these questions requires a dataset that combines digital traces and offline attitudinal assessment, and we encourage future researchers to explore more novel and responsible data collection pathways (e.g., data donation as mentioned in Chapter 4) to disentangle whether behavioral proxies systematically misrepresent certain ideological segments, and whether these biases are consistent across platforms or forms of engagement. Through such analysis, we can better understand the limitations of current ideological inference models, as well as the nuanced social and cognitive processes underlying political participation online.

Bibliography

[1] Peter Dahlgren. The internet, public spheres, and political communication: Dispersion and deliberation. *Political communication*, 22(2):147–162, 2005.

[2] Amy Gutmann and Dennis F Thompson. *Democracy and disagreement*. Harvard University Press, 1996.

[3] Dennis Friess and Christiane Eilders. A systematic review of online deliberation research. *Policy & Internet*, 7(3):319–339, 2015.

[4] Peter Van Aelst, Jesper Strömbäck, Toril Aalberg, Frank Esser, Claes De Vreese, Jörg Matthes, David Hopmann, Susana Salgado, Nicolas Hubé, Agnieszka Stepińska, et al. Political communication in a high-choice media environment: A challenge for democracy? *Annals of the international communication association*, 41(1):3–27, 2017.

[5] World Bank Group. Individuals using the internet (% of population), 2024. URL https://data.worldbank.org/indicator/IT.NET.USER.ZS.

[6] Statista. Global smartphone penetration rate as share of population from 2016 to 2024, 2024. URL https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/.

[7] María Antonia Paz, Julio Montero-Díaz, and Alicia Moreno-Delgado. Hate speech: A systematized review. *Sage Open*, 10(4):2158244020973022, 2020.

[8] Julie Posetti and Alice Matthews. A short guide to the history of 'fake news' and disinformation. *International Center for Journalists*, 7(2018):2018–07, 2018.

[9] Thomas Carothers and Andrew O'Donohue. *Democracies divided: The global challenge of political polarization*. Brookings Institution Press, 2019.

[10] Yphtach Lelkes. Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1):392–410, 2016.

[11] Emily Kubin and Christian Von Sikorski. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*, 45(3):188–206, 2021.

[12] Max Falkenberg, Matteo Cinelli, Alessandro Galeazzi, Christopher A Bail, Rosa M Benito, Axel Bruns, Anatoliy Gruzd, David Lazer, Jae K Lee, Jennifer McCoy, et al. Towards global equity in political polarization research. *arXiv preprint arXiv:2504.11090*, 2025.

[13] Philip Warncke, Yijing Chen, Anne Speer, Bart de Bruin, Adrian Lüders, and Dino Carpentras. Resin: A new method to analyze socio-political attitude systems. *OSF preprint*, 2025. URL https://osf.io/preprints/osf/p3x2n_v2.

[14] Yijing Chen, Yaguang Liu, Lisa Singh, and Ceren Budak. Intermedia agenda setting during the 2016 and 2020 us presidential elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 254–275, 2024.

[15] Lilliana Mason. "i disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American journal of political science*, 59(1):128–145, 2015.

[16] Philip E Converse. The nature of belief systems in mass publics (1964). *Critical review*, 18(1-3):1–74, 2006.

[17] Yijing Chen, Anne Speer, Bart de Bruin, Dino Carpentras, and Philip Warncke. A "broken egg" of us political beliefs: Using response-item network (resin) to measure ideological polarization. 2024. URL https://osf.io/preprints/osf/autkb.

[18] Yijing Chen, Zoltán Kmetty, Gerardo Iñiguez, and Elisa Omodei. The public that engages invisibly: What visible engagement fails to capture in online political communication. *OSF preprint*, 2024. URL https://osf.io/preprints/socarxiv/mv24c.

[19] Gianpietro Mazzoleni and Winfried Schulz. " mediatization" of politics: A challenge for democracy? *Political communication*, 16(3):247–261, 1999.

[20] Maxwell McCombs. A look at agenda-setting: Past, present and future. *Journalism studies*, 6(4):543–557, 2005.

[21] Brian McNair. *An introduction to political communication*. Routledge, 2017.

[22] Brian D Loader and Dan Mercea. Networking democracy? social media innovations and participatory politics. *Information, communication & society*, 14(6):757–769, 2011.

[23] Andrew Chadwick. *The hybrid media system: Politics and power*. Oxford University Press, 2017.

[24] James G Webster. *The marketplace of attention: How audiences take shape in a digital age*. MIT press, 2014.

[25] Peter Dahlgren. Reconfiguring civic culture in the new media milieu. In *Media and the restyling of politics: Consumerism, celebrity and cynicism*, pages 151–170. SAGE Publications Ltd, 2003.

[26] Zizi Papacharissi. *A private sphere: Democracy in a digital age*. Polity, 2010.

[27] Markus Prior. Media and political polarization. *Annual review of political science*, 16(1):101–127, 2013.

[28] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*, 2018.

[29] Kathleen Hall Jamieson and Joseph N Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.

[30] Cass R Sunstein. *Going to extremes: How like minds unite and divide*. Oxford University Press, 2009.

[31] Paolo Mancini. Media fragmentation, party system, and democracy. *The International Journal of Press/Politics*, 18(1):43–60, 2013.

[32] Desiree Steppat, Laia Castro Herrero, and Frank Esser. Selective exposure in different political information environments–how media fragmentation and polarization shape congruent news use. *European Journal of Communication*, 37(1):82–102, 2022.

[33] Auksė Balčytienė and Kristina Juraitė. Systemic media changes and social and political polarization in europe. In *European Media in Crisis*, pages 20–42. Routledge, 2015.

[34] Natalie Jomini Stroud. Polarization and partisan selective exposure. *Journal of communication*, 60(3):556–576, 2010.

[35] David O Sears and Jonathan L Freedman. Selective exposure to information: A critical review. *Public opinion quarterly*, 31(2):194–213, 1967.

[36] Markus Prior. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press, 2007.

[37] Elliot Panek. High-choice revisited: An experimental analysis of the dynamics of news selection behavior in high-choice media environments. *Journalism & Mass Communication Quarterly*, 93(4):836–856, 2016.

[38] David P Baron. Persistent media bias. *Journal of Public Economics*, 90(1-2):1–36, 2006.

[39] Maxwell E McCombs and Donald L Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.

[40] Lei Guo and Maxwell McCombs. Network agenda setting: A third level of media effects. In *annual conference of the International Communication Association, Boston, MA*, 2011.

[41] Michael Chan and Francis LF Lee. Selective exposure and agenda setting: Exploring the impact of partisan media exposure on agenda diversity and political participation. *Asian Journal of Communication*, 24(4):301–314, 2014.

[42] Frank Mangold, David Schoch, and Sebastian Stier. Ideological self-selection in online news exposure: Evidence from europe and the us. *Science Advances*, 10(37): eadg9287, 2024.

[43] Joseph T Klapper. *The effects of mass communication.* Free press, 1960.

[44] Andrew M Guess, Pablo Barberá, Simon Munzert, and JungHwan Yang. The consequences of online partisan media. *Proceedings of the National Academy of Sciences*, 118(14):e2013464118, 2021.

[45] Maxwell McCombs and Sebastian Valenzuela. *Setting the agenda: Mass media and public opinion*. John Wiley & Sons, 2020.

[46] Natalie Jomini Stroud. *Niche news: The politics of news choice*. Oxford University Press, 2011.

[47] Shanto Iyengar and Kyu S Hahn. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication*, 59(1):19–39, 2009.

[48] Patricia Moy, David Tewksbury, and Eike Mark Rinke. Agenda-setting, priming, and framing. *The international encyclopedia of communication theory and philosophy*, 1 (1):1–13, 2016.

[49] Dennis Chong and James N Druckman. Framing theory. *Annu. Rev. Polit. Sci.*, 10(1): 103–126, 2007.

[50] Pablo Barberá. schoscschoscho. *Social media and democracy: The state of the field, prospects for reform*, pages 34–55, 2020.

[51] Jay J Van Bavel, Steve Rathje, Elizabeth Harris, Claire Robertson, and Anni Sternisko. How social media shapes polarization. *Trends in Cognitive Sciences*, 25(11): 913–916, 2021.

[52] Cass R Sunstein. Republic: Divided democracy in the age of social media. 2018.

[53] Peter H Ditto and David F Lopez. Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of personality and social psychology*, 63(4):568, 1992.

[54] R Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of computer-mediated communication*, 14(2): 265–285, 2009.

[55] Pablo Barberá, 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(10):1531–1542, 2015.

[56] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 89–96, 2011.

[57] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

[58] Isaac Waller and Ashton Anderson. Quantifying social organization and political polarization in online platforms. *Nature*, 600(7888):264–268, 2021.

[59] Max Falkenberg, Alessandro Galeazzi, Maddalena Torricelli, Niccolò Di Marco, Francesca Larosa, Madalina Sas, Amin Mekacher, Warren Pearce, Fabiana Zollo, Walter Quattrociocchi, et al. Growing polarization around climate change on social media. *Nature Climate Change*, 12(12):1114–1121, 2022.

[60] Matthew Barnidge. Exposure to political disagreement in social media versus face-to-face and anonymous online settings. *Political communication*, 34(2):302–321, 2017.

[61] Richard Fletcher and Rasmus Kleis Nielsen. Are people incidentally exposed to news on social media? a comparative analysis. *New media & society*, 20(7):2450–2468, 2018.

[62] Laura Silver, Christine Huang, and Kyle Taylor. In emerging economies, smartphone and social media users have broader social networks. *Pew Research Center*, 2019.

[63] Matthew Gentzkow and Jesse M Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

[64] Gianmarco De Francisci Morales, Corrado Monti, and Michele Starnini. No echo in the chambers of political interactions on reddit. *Scientific reports*, 11(1):2818, 2021.

[65] Ro'ee Levy. Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3):831–870, 2021.

[66] Kyle A Heatherly, Yanqin Lu, and Jae Kook Lee. Filtering out the other side? cross-cutting and like-minded discussions on social networking sites. *New Media & Society*, 19(8):1271–1289, 2017.

[67] Pablo Barberá. How social media reduces mass political polarization. evidence from germany, spain, and the us. *Job Market Paper, New York University*, 46:1–46, 2014.

[68] Raffael Heiss, Christian von Sikorski, and Jörg Matthes. Populist twitter posts in news stories: Statement recognition and the polarizing effects on candidate evaluation and anti-immigrant attitudes. *Journalism Practice*, 13(6):742–758, 2019.

[69] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, 2018.

[70] Andreas Schedler. Rethinking political polarization. *Political science quarterly*, 138 (3):335–359, 2023.

[71] Shadi Hamid. Polarization isn't just a phase. it's our new default setting. *The Washington Post*. URL https://www.washingtonpost.com/opinions/2024/10/24/polarization-democrats-republicans-trump/.

[72] Jamil Anderlini. Europe's growing polarization spells trouble. *Politico*. URL https://www.politico.eu/newsletter/politico-confidential/europes-growing-polarization-spells-trouble/.

[73] Barack Obama. President obama's remarks from the 2024 democracy forum, 12 2024. URL https://www.obama.org/democracy-forum-2024/president-obama-remarks/.

[74] James A Piazza. Political polarization and political violence. *Security Studies*, 32(3): 476–504, 2023.

[75] Anton Gollwitzer, Cameron Martel, William J Brady, Philip Pärnamets, Isaac G Freedman, Eric D Knowles, and Jay J Van Bavel. Partisan differences in physical distancing are linked to health outcomes during the covid-19 pandemic. *Nature human behaviour*, 4(11):1186–1197, 2020.

[76] Milan W Svolik. Polarization versus democracy. *Journal of democracy*, 30(3):20–32, 2019.

[77] Paul DiMaggio, John Evans, and Bethany Bryson. Have american's social attitudes become more polarized? *American journal of Sociology*, 102(3):690–755, 1996.

[78] Alan I Abramowitz and Kyle L Saunders. Is polarization a myth? *The Journal of Politics*, 70(2):542–555, 2008.

[79] Morris P Fiorina and Samuel J Abrams. Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11(1):563–588, 2008.

[80] Pablo Barberá. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis*, 23(1):76–91, 2015.

[81] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[82] Dominic Spohr. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business information review*, 34(3):150–160, 2017.

[83] Moran Yarchi, Christian Baden, and Neta Kligler-Vilenchik. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 38(1-2):98–139, 2021.

[84] Almog Simchon, William J Brady, and Jay J Van Bavel. Troll and divide: the language of online polarization. *PNAS nexus*, 1(1):pgac019, 2022.

[85] Marcela Suárez Estrada, Yulissa Juarez, and CA Piña-García. Toxic social media: Affective polarization after feminist protests. *Social Media+ Society*, 8(2): 20563051221098343, 2022.

[86] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[87] Angus Campbell. *The american voter*. University of Chicago Press, 1980.

[88] Morris P Fiorina, Samuel J Abrams, and Jeremy Pope. *Culture war? The myth of a polarized America*. New York: PearsonLongman, 2004.

[89] Shanto Iyengar, Gaurav Sood, and Yphtach Lelkes. Affect, not ideology: A social identity perspective on polarization. *Public opinion quarterly*, 76(3):405–431, 2012.

[90] Mia Costa. Ideology, not affect: What americans want from political representation. *American Journal of Political Science*, 65(2):342–358, 2021.

[91] Robert S Erikson and Kent L Tedin. *American public opinion: Its origins, content, and impact*. Routledge, 2019.

[92] Christopher Ellis and James A Stimson. *Ideology in america*. Cambridge University Press, 2012.

[93] Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114 (2):408–446, 2008.

[94] Yphtach Lelkes and Sean J Westwood. The limits of partisan prejudice. *The Journal of Politics*, 79(2):485–501, 2017.

[95] R Kelly Garrett, Shira Dvir Gvirsman, Benjamin K Johnson, Yariv Tsfati, Rachel Neo, and Aysenur Dal. Implications of pro-and counterattitudinal information exposure for affective polarization. *Human communication research*, 40(3):309–332, 2014.

[96] Matthew Levendusky. *How partisan media polarize America*. University of Chicago Press, 2013.

[97] Aaron Bramson, Patrick Grim, Daniel J Singer, William J Berger, Graham Sack, Steven Fisher, Carissa Flocken, and Bennett Holman. Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of science*, 84(1):115–159, 2017.

[98] Alan Abramowitz and Kyle Saunders. Why can't we all just get along? the reality of a polarized america. In *The Forum*, volume 3, page 0000102202154088841076. De Gruyter, 2005.

[99] Morris P Fiorina and Matthew S Levendusky. Disconnected: The political class versus the people. *Red and blue nation*, 1:49–71, 2006.

[100] Erik Knudsen. Affective polarization in multiparty systems? comparing affective polarization towards voters and parties in norway and the united states. *Scandinavian Political Studies*, 44(1):34–44, 2021.

[101] Matthew S Levendusky. Americans, not partisans: Can priming american national identity reduce affective polarization? *The Journal of Politics*, 80(1):59–70, 2018.

[102] Matthew Levendusky and Neil Malhotra. Does media coverage of partisan polarization affect political attitudes? *Political Communication*, 33(2):283–301, 2016.

[103] Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood. The origins and consequences of affective polarization in the united states. *Annual review of political science*, 22(1):129–146, 2019.

[104] Markus Wagner. Affective polarization in europe. *European Political Science Review*, 16(3):378–392, 2024.

[105] Sara Vissers and Dietlind Stolle. The internet and new modes of political participation: online versus offline participation. *Information, Communication & Society*, 17 (8):937–955, 2014.

[106] Ofcom. Light-hearted news on social media drawing gen z away from traditional sources. *Ofcome*, . URL https://www.ofcom.org.uk/media-use-and-attitudes/attitudes-to-news/light-hearted-news-social-media-drawing-gen-z/.

[107] Leticia Batista Cabanas. Young europeans face rising threat from misinformation as social media becomes main news source. *Euronews*. URL https://www.euronews.com/my-europe/2025/02/19/young-europeans-face-rising-threat-from-misinformation-as-social-media-becomes- ~:text=According%20to%20the%20latest%20European,about%20politics%20and%20social%20issues.

[108] Ofcom. “i saw it on social”’: How digital life affects political participation. *Springtide Research Institute*, . URL https://springtideresearch.org/post/politics/i-saw-it-on-social-how-digital-life-affects-political-participation?srsltid=AfmBOoqX0DPMRSTgFekcvgNPtBAGLlzLGAcnjCfXVfdPgiVl2skRCHDL.

[109] Su Jung Kim. The role of social media news usage and platforms in civic and political engagement: Focusing on types of usage and platforms. *Computers in Human Behavior*, 138:107475, 2023.

[110] Laura Tölle and Matthias Trier. Polarization in online social networks: A review of mechanisms and dimensions. In *ECIS*, 2023.

[111] AllSides. AllSides Media Bias Chart, year = 2024, url = https://www.allsides.com/media-bias/media-bias-chart, urldate = 2025-04-14.

[112] Michael D Conover, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Partisan asymmetries in online political activity. *EPJ Data science*, 1:1–19, 2012.

[113] Natalie Jomini Stroud. Media use and political predispositions: Revisiting the concept of selective exposure. *Political behavior*, 30:341–366, 2008.

[114] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[115] Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

[116] Engin Bozdag. Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15:209–227, 2013.

[117] Dan Jervis-Bardy. Clare o'neil blames social media 'echo chambers' for fracturing social cohesion amid gaza conflict. *The Nightly*. URL `https://thenightly.com.au/politics/australia/clare-oneil-blames-social-media-echo-chambers-for-fracturing-social-cohesion-am`

[118] Peter Suciu. Whether harris or trump wins, social media ensures we remain divided. *Forbes*. URL `https://www.forbes.com/sites/petersuciu/2024/11/02/whether-harris-or-trump-wins-social-media-ensures-we-remain-divided/`.

[119] Mostafa EI-Bermawy. Your filter bubble is destroying democracy. *Wired*. URL `https://www.wired.com/2016/11/filter-bubble-destroying-democracy/`.

[120] Seth J Hill and Chris Tausanovitch. A disconnect in representation? comparison of trends in congressional and public polarization. *The Journal of Politics*, 77(4): 1058–1075, 2015.

[121] Delia Baldassarri and Peter Bearman. Dynamics of political polarization. *American sociological review*, 72(5):784–811, 2007.

[122] Martin Gestefeld, Jan Lorenz, Nils Tobias Henschel, and Klaus Boehnke. Decomposing attitude distributions to quantify mass polarization in europe. *Preprint]. SocArXiv. https://doi. org/10.31235/osf. io/3pn5w*, 2021.

[123] Daniel DellaPosta. Pluralistic collapse: The "oil spill" model of mass opinion polarization. *American Sociological Review*, 85(3):507–536, 2020.

[124] James Adams, Lawrence Ezrow, and Zeynep Somer-Topcu. Is anybody listening? evidence that voters do not respond to european parties' policy statements during elections. *American Journal of Political Science*, 55(2):370–382, 2011.

[125] James Adams, Lawrence Ezrow, and Debra Leiter. Partisan sorting and niche parties in europe. In *Assessing Political Representation in Europe*, pages 56–78. Routledge, 2020.

[126] Levi Boxell, Matthew Gentzkow, and Jesse M Shapiro. Cross-country trends in affective polarization. *Review of Economics and Statistics*, 106(2):557–565, 2024.

[127] Lilliana Mason. Ideologues without issues: The polarizing consequences of ideological identities. *Public Opinion Quarterly*, 82(S1):866–887, 2018.

[128] Andrei Boutyline and Stephen Vaisey. Belief network analysis: A relational approach to understanding the structure of attitudes. *American journal of sociology*, 122(5): 1371–1447, 2017.

[129] Nic Fishman and Nicholas T Davis. Change we can believe in: Structural and content dynamics within belief networks. *American Journal of Political Science*, 66(3):648–663, 2022.

[130] David Lazer, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. Meaningful measures of human society in the twenty-first century. *Nature*, 595(7866):189–196, 2021.

[131] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. Users polarization on facebook and youtube. *PloS one*, 11(8):e0159641, 2016.

[132] Steven H Chaffee and Miriam J Metzger. The end of mass communication? *Mass communication & society*, 4(4):365–379, 2001.

[133] Matthew Gentzkow, Jesse M Shapiro, and Daniel F Stone. Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1, pages 623–645. Elsevier, 2015.

[134] Pamela J Shoemaker and Stephen D Reese. *Mediating the message in the 21st century: A media sociology perspective*. Routledge, 2013.

[135] Matthew A Baum and Tim Groeling. New media and the polarization of american political discourse. *Political Communication*, 25(4):345–365, 2008.

[136] Ashley Muddiman, Natalie Jomini Stroud, and Maxwell McCombs. Media fragmentation, attribute agenda setting, and political opinions about iraq. *Journal of Broadcasting & Electronic Media*, 58(2):215–233, 2014.

[137] Scott Maier. All the news fit to post? comparing news content on the web to newspapers, television, and radio. *Journalism & Mass Communication Quarterly*, 87(3-4): 548–562, 2010.

[138] Jae Kook Lee. The effect of the internet on homogeneity of the media agenda: A test of the fragmentation thesis. *Journalism & Mass Communication Quarterly*, 84(4): 745–760, 2007.

[139] Scott A Eldridge. Thank god for deadspin": Interlopers, metajournalistic commentary, and fake news through the lens of "journalistic realization. *New Media & Society*, 21(4):856–878, 2019.

[140] Stephen P Nicholson. *Voting the agenda: Candidates, elections, and ballot propositions*. Princeton University Press, 2021.

[141] Lei Guo and Chris Vargo. "fake news" and emerging online media ecosystem: An integrated intermedia agenda-setting analysis of the 2016 us presidential election. *Communication Research*, 47(2):178–200, 2020.

[142] Stephen D Reese and Lucig H Danielian. Intermedia influence and the drug issue: Converging on cocaine. In *Communication campaigns about drugs*, pages 29–45. Routledge, 2012.

[143] Warren Breed. Newspaper 'opinion leaders' and processes of standardization. *Journalism Quarterly*, 32(3):277–328, 1955.

[144] Chris J Vargo and Lei Guo. Networks, big data, and intermedia agenda setting: An analysis of traditional, partisan, and emerging online us news. *Journalism & Mass Communication Quarterly*, 94(4):1031–1055, 2017.

[145] Chris J Vargo, Lei Guo, and Michelle A Amazeen. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New media & society*, 20(5):2028–2049, 2018.

[146] Chris J Vargo, Ekaterina Basilaia, and Donald Lewis Shaw. Event versus issue: Twitter reflections of major news, a case study. In *Communication and information technologies annual*, volume 9, pages 215–239. Emerald Group Publishing Limited, 2015.

[147] Peter Hermann, Susan Svrluga, and Michael E Miller. Alleged gunman tells police he wanted to rescue children at dc pizza shop after hearing fictional internet accounts. *The Washington Post*, 5, 2016.

[148] Maxwell McCombs and Amy Reynolds. How the news shapes our civic agenda. In *Media effects*, pages 17–32. Routledge, 2009.

[149] Lei Guo. The application of social network analysis in agenda setting research: A methodological exploration. *Journal of Broadcasting & Electronic Media*, 56(4): 616–631, 2012.

[150] Spiro Kiousis, Michael Mitrook, Xu Wu, and Trent Seltzer. First-and second-level agenda-building and agenda-setting effects: Exploring the linkages among candidate news releases, media coverage, and public opinion during the 2002 florida gubernatorial election. *Journal of Public Relations Research*, 18(3):265–285, 2006.

[151] Maxwell McCombs, Juan Pablo Llamas, Esteban Lopez-Escobar, and Federico Rey. Candidate images in spanish elections: Second-level agenda-setting effects. *Journalism & Mass Communication Quarterly*, 74(4):703–717, 1997.

[152] Spiro Kiousis, Philemon Bantimaroudis, and Hyun Ban. Candidate image attributes: Experiments on the substantive dimension of second level agenda setting. *Communication Research*, 26(4):414–428, 1999.

[153] Guy Golan and Wayne Wanta. Second-level agenda setting in the new hampshire primary: A comparison of coverage in three newspapers and public perceptions of candidates. *Journalism & Mass Communication Quarterly*, 78(2):247–259, 2001.

[154] Jisu Kim and Young Min. An issue attention cycle analysis of the network agenda setting model. *The power of information networks: New directions for agenda setting*, page 132, 2015.

[155] Ceren Budak, Natalie Jomini Stroud, Ashley Muddiman, Caroline C Murray, and Yujin Kim. The stability of cable and broadcast news intermedia agenda setting across the covid-19 issue attention cycle. *Political Communication*, pages 1–21, 2023.

[156] Lia Bozarth and Ceren Budak. Market forces: Quantifying the role of top credible ad servers in the fake news ecosystem. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 83–94, 2021.

[157] Andrew Couts and Andrew Wyrich. Here are all the 'fake news' sites to watch out for on facebook, 2016. URL https://www.dailydot.com/debug/fake-news-sites-list-facebook/.

[158] Dave Van Zandt. Media bias / fact check, 2015. URL https://mediabiasfactcheck.com/.

[159] Joshua Gillin. Politifact's guide to fake news websites and what they peddle, 2017. URL https://www.politifact.com/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.

[160] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2):2053168019848554, 2019.

[161] Melissa Zimdars. My "fake news list" went viral. but made-up stories are only part of the problem. *Washington Post*, 2016.

[162] Hearings. The policy agendas project at the university of texas at austin. 2017. www.comparativeagendas.net, Accessed September 26, 2017.

[163] Leticia Bode, Ceren Budak, Jonathan M Ladd, Frank Newport, Josh Pasek, Lisa O Singh, Stuart N Soroka, and Michael W Traugott. *Words that matter: How the news and social media shaped the 2016 Presidential campaign*. Brookings Institution Press, 2019.

[164] Jennifer Agiesta. Read: The methodology behind new polling project the breakthrough. *CNN*, 2020. URL https://edition.cnn.com/2020/08/16/politics/election-2020-polls-the-breakthrough-methodology/index.html.

[165] Robert Churchill, Lisa Singh, Rebecca Ryan, and Pamela Davis-Kean. A guided topic-noise model for short texts. In *Proceedings of the ACM Web Conference 2022*, pages 2870–2878, 2022.

[166] Kaye D Sweetser, Guy J Golan, and Wayne Wanta. Intermedia agenda setting in television, advertising, and blogs during the 2004 election. *Mass Communication & Society*, 11(2):197–216, 2008.

[167] Hans-Bernd Brosius and Hans Mathias Kepplinger. The agenda-setting function of television news: Static and dynamic views. *Communication research*, 17(2):183–211, 1990.

[168] Sharon Meraz. Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks. *Journalism & mass communication quarterly*, 88(1):176–194, 2011.

[169] Jacob Groshek and Megan Clough Groshek. Agenda trending: Reciprocity and the predictive capacity of social network sites in intermedia agenda setting across issues over time. *Available at SSRN 2199144*, 2013.

[170] Ceren Budak, Lia Bozarth, Robert M Bond, Drew Margolin, Jason J Jones, and R Kelly Garrett. Bursts of contemporaneous publication among high- and low-credibility online information providers. *New Media & Society*, page 14614448231183617, 2023.

[171] Ashley Muddiman, Ceren Budak, Caroline Murray, Yujin Kim, and Natalie Jomini Stroud. Indexing theory during an emerging health crisis: how us tv news indexed elite perspectives and amplified covid-19 misinformation. *Annals of the International Communication Association*, 46(3):174–204, 2022.

[172] Ceren Budak. What happened? the spread of fake news publisher content during the 2016 us presidential election. In *The World Wide Web Conference*, pages 139–150, 2019.

[173] Andrew M Guess, Brendan Nyhan, and Jason Reifler. Exposure to untrustworthy websites in the 2016 us election. *Nature human behaviour*, 4(5):472–480, 2020.

[174] Stuart N Soroka. *Negativity in democratic politics: Causes and consequences*. Cambridge University Press, 2014.

[175] Ceren Budak, Sharad Goel, and Justin M Rao. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271, 2016.

[176] James Flamino, Alessandro Galeazzi, Stuart Feldman, Michael W Macy, Brendan Cross, Zhenkun Zhou, Matteo Serafino, Alexandre Bovet, Hernán A Makse, and Boleslaw K Szymanski. Political polarization of news media and influencers on twitter in the 2016 and 2020 us presidential elections. *Nature Human Behaviour*, 7(6): 904–916, 2023.

[177] W Lance Bennett and Steven Livingston. The disinformation order: Disruptive communication and the decline of democratic institutions. *European journal of communication*, 33(2):122–139, 2018.

[178] David Protess and Maxwell E McCombs. *Agenda setting: Readings on media, public opinion, and policymaking*. Routledge, 2016.

[179] Guy Golan. Inter-media agenda setting and global news coverage: Assessing the influence of the new york times on three network television evening news programs. *Journalism studies*, 7(2):323–333, 2006.

[180] Lia Bozarth, Aparajita Saraf, and Ceren Budak. Higher ground? how groundtruth labeling impacts our understanding of fake news about the 2016 us presidential nominees. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 48–59, 2020.

[181] Mark J Brandt, Chris G Sibley, and Danny Osborne. What is central to political belief system networks? *Personality and Social Psychology Bulletin*, 45(9):1352–1364, 2019.

[182] Philip Warncke. What explains country-level differences in political belief system coherence? *Political Behavior*, pages 1–24, 2025.

[183] Elmer E Schattschneider. Political parties and the public interest. *The annals of the American academy of political and social science*, 280(1):18–26, 1952.

[184] Lloyd A Free and Hadley Cantril. The political beliefs of americans: A study of public opinion. *(No Title)*, 1967.

[185] Seymour Martin Lipset, Stein Rokkan, et al. *Cleavage structures, party systems, and voter alignments: An introduction*, volume 2. Free Press New York, 1967.

[186] John T Jost, Brian A Nosek, and Samuel D Gosling. Ideology: Its resurgence in social, personality, and political psychology. *Perspectives on Psychological Science*, 3(2):126–136, 2008.

[187] Liesbet Hooghe and Gary Marks. Cleavage theory meets europe's crises: Lipset, rokkan, and the transnational cleavage. *Journal of European public policy*, 25(1): 109–135, 2018.

[188] Christopher M Federico. Ideology and public opinion. In *New directions in public opinion*, pages 75–98. Routledge, 2019.

[189] Dino Carpentras, Adrian Lüders, and Michael Quayle. Response item network (resin): A network-based approach to explore attitude systems. *Humanities and Social Sciences Communications*, 11(1):1–14, 2024.

[190] Dino Carpentras, Adrian Lüders, and Michael Quayle. Mapping the global opinion space to explain anti-vaccine attraction. *Scientific reports*, 12(1):6188, 2022.

[191] Adrian Lüders, Dino Carpentras, and Michael Quayle. Attitude networks as intergroup realities: Using network-modelling to research attitude-identity relationships in polarized political contexts. *British Journal of Social Psychology*, 63(1):37–51, 2024.

[192] Kevin Durrheim and Michael Quayle and. Human murmuration: Group polarisation as compression in interaction-language dynamics captured by large language models. *European Review of Social Psychology*, 0(0):1–40, 2025. doi: 10.1080/10463283.2025.2499332. URL https://doi.org/10.1080/10463283.2025.2499332.

[193] Kati Kish Bar-On, Eugen Dimant, Yphtach Lelkes, and David G Rand. Navigating polarization: Insights from behavioral science. *Available at SSRN 4784194*, 2024.

[194] Femke Roosma, John Gelissen, and Wim Van Oorschot. The multidimensionality of welfare state attitudes: A european cross-national study. *Social indicators research*, 113:235–255, 2013.

[195] John Zaller. The nature and origins of mass opinion. *Cambridge University*, 1992.

[196] James H Kuklinski and Buddy Peyton. Belief systems and political decision making. 2007.

[197] Mark J Brandt and Willem WA Sleegers. Evaluating belief system networks as a theory of political belief system dynamics. *Personality and Social Psychology Review*, 25(2):159–185, 2021.

[198] Eric Groenendyk, Erik O Kimbrough, and Mark Pickup. How norms shape the nature of belief systems in mass publics. *American Journal of Political Science*, 67(3):623–638, 2023.

[199] Donald R Kinder and Nathan P Kalmoe. *Neither liberal nor conservative: Ideological innocence in the American public*. University of Chicago Press, 2017.

[200] Christopher D Johnston and Trent Ollerenshaw. How different are cultural and economic ideology? *Current Opinion in Behavioral Sciences*, 34:94–101, 2020.

[201] Angus Campbell, Philip E Converse, Warren E Miller, and Donald E Stokes. The american voter. *Oxford: John Wiley*, 1960.

[202] Donald P Green, Bradley Palmquist, and Eric Schickler. *Partisan hearts and minds: Political parties and the social identities of voters*. Yale University Press, 2004.

[203] Alan I Abramowitz. Transformation and polarization: The 2008 presidential election and the new american electorate. *Electoral Studies*, 29(4):594–603, 2010.

[204] Morris P Fiorina, Samuel J Abrams, and Jeremy C Pope. *Culture war?* Longman, 2011.

[205] Daniel DellaPosta, Yongren Shi, and Michael Macy. Why do liberals drink lattes? *American Journal of Sociology*, 120(5):1473–1511, 2015.

[206] Matthew S Levendusky. The microfoundations of mass polarization. *Political Analysis*, 17(2):162–176, 2009.

[207] Hahrie Han. The disappearing center: Engaged citizens, polarization, and american democracy, 2011.

[208] Nicholas T Davis and Johanna L Dunaway. Party polarization, media choice, and mass partisan-ideological sorting. *Public Opinion Quarterly*, 80(S1):272–297, 2016.

[209] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[210] James A Stimson. Belief systems: Constraint, complexity, and the 1972 election. *American Journal of Political Science*, pages 393–417, 1975.

[211] Fumiko Samejima. The general graded response model. In *Handbook of polytomous item response theory models*, pages 87–118. Routledge, 2011.

[212] LC Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 1977.

[213] Austin C Kozlowski and James P Murphy. Issue alignment and partisanship in the american public: Revisiting the 'partisans without constraint'thesis. *Social Science Research*, 94:102498, 2021.

[214] Andrew Kohut, Carroll Doherty, Michael Dimock, and Scott Keeter. Partisan polarization surges in bush, obama years. *Pew Research Center for the People and the Press RSS*, 2012.

[215] Matt Grossmann and David A Hopkins. Ideological republicans and group interest democrats: The asymmetry of american party politics. *Perspectives on Politics*, 13 (1):119–139, 2015.

[216] Pew Research Center. Political polarization in the american public. *Pew Research Center*, 2014. URL https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/.

[217] James N Druckman and Jeremy Levy. Affective polarization in the american public. In *Handbook on politics and public opinion*, pages 257–270. Edward Elgar Publishing, 2022.

[218] Pew Research Center. Partisan antipathy: More intense, more personal. *Pew Research Center*, 2019.

[219] Shanto Iyengar. Identity politics, party polarization, and the rise of donald trump. In *The Changing Character of the American Right, Volume I: Ideology, Politics and Policy in the Era of Trump*, pages 79–94. Springer, 2025.

[220] Rachel Kleinfeld. Polarization, democracy, and political violence in the united states: What the research says. 2023.

[221] Philip Warncke, Dino Carpentras, and Adrian Lüders. *ResIN: Response Item Networks*, 2024. URL https://CRAN.R-project.org/package=ResIN. R package version 2.0.0.

[222] Jochem Van Noord, Felicity M Turner-Zwinkels, Rebekka Kesberg, Mark J Brandt, Matthew J Easterbrook, Toon Kuppens, and Bram Spruyt. The nature and structure of european belief systems: Exploring the varieties of belief systems across 23 european countries. *European Sociological Review*, 41(1):143–161, 2025.

[223] Jonathan Bright. Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication*, 23(1):17–33, 2018.

[224] Indira Sen, Fabian Flöck, Katrin Weller, Bernd Weiß, and Claudia Wagner. A total error framework for digital traces of human behavior on online platforms. *Public Opinion Quarterly*, 85(S1):399–422, 2021.

[225] Claudia Wagner, Markus Strohmaier, Alexandra Olteanu, Emre Kıcıman, Noshir Contractor, and Tina Eliassi-Rad. Measuring algorithmically infused societies. *Nature*, 595(7866):197–204, 2021.

[226] Federico Vegetti. The political nature of ideological polarization: The case of hungary. *The ANNALS of the American Academy of Political and Social Science*, 681(1): 78–96, 2019.

[227] Siqi Wu and Paul Resnick. Cross-partisan discussions on youtube: Conservatives talk to liberals but liberals don't talk to conservatives. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 808–819, 2021.

[228] Adrian Rauchfleisch and Jonas Kaiser. The german far-right on youtube: An analysis of user overlap and user comments. *Journal of Broadcasting & Electronic Media*, 64 (3):373–396, 2020.

[229] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, 2007.

[230] Daniela V Dimitrova, Adam Shehata, Jesper Strömbäck, and Lars W Nord. The effects of digital media on political knowledge and participation in election campaigns: Evidence from panel data. *Communication research*, 41(1):95–118, 2014.

[231] Sanne Kruikemeier, Guda Van Noort, Rens Vliegenthart, and Claes H de Vreese. Unraveling the effects of active and passive forms of political internet use: Does it affect citizens' political involvement? *New Media & Society*, 16(6):903–920, 2014.

[232] Cristian Vaccari, Augusto Valeriani, Pablo Barberá, Rich Bonneau, John T Jost, Jonathan Nagler, and Joshua A Tucker. Political expression and action on social media: Exploring the relationship between lower-and higher-threshold political activities among twitter users in italy. *Journal of Computer-Mediated Communication*, 20(2):221–239, 2015.

[233] Jörg Matthes, Andreas Nanz, Ruta Kaskeleviciute, Franz Reiter, Isabelle Freiling, Ariadne Neureiter, Marlis Stubenvoll, Sebastian E Sherrah, Sarah Juricek, Atika Aisyarahmi Munzir, and Others. The way we use social media matters: A panel study on passive versus active political social media use and affective polarization. *International Journal of Communication (19328036)*, 17, 2023.

[234] Erik Peterson, Sharad Goel, and Shanto Iyengar. Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign. *Political science research and methods*, 9(2):242–258, 2021.

[235] Matteo Cinelli, Emanuele Brugnoli, Ana Lucia Schmidt, Fabiana Zollo, Walter Quattrociocchi, and Antonio Scala. Selective exposure shapes the facebook news diet. *PloS one*, 15(3):e0229129, 2020.

[236] Kevin Munger and Joseph Phillips. Right-wing youtube: A supply and demand perspective. *The International Journal of Press/Politics*, 27(1):186–219, 2022.

[237] Raphael Ottoni, Evandro Cunha, Gabriel Magno, Pedro Bernardina, Wagner Meira Jr, and Virgílio Almeida. Analyzing right-wing youtube channels: Hate, violence and discrimination. In *Proceedings of the 10th ACM conference on web science*, pages 323–332, 2018.

[238] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio AF Almeida, and Wagner Meira Jr. Auditing radicalization pathways on youtube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 131–141, 2020.

[239] Galen Stocking, Patrick Van Kessel, Michael Barthel, Katerina Eva Matsa, and Maya Khuzam. Many americans get news on youtube, where news organizations and independent producers thrive side by side. *Pew Research Center*, 28, 2020.

[240] Angela Lai, Megan A Brown, James Bisbee, Joshua A Tucker, Jonathan Nagler, and Richard Bonneau. Estimating the ideology of political youtube videos. *Political Analysis*, pages 1–16, 2022.

[241] Sam Clark and Anna Zaitsev. Understanding youtube communities via subscription-based channel embeddings. *arXiv preprint arXiv:2010.09892*, 2020.

[242] Jakob Ohme, Theo Araujo, Laura Boeschoten, Deen Freelon, Nilam Ram, Byron B Reeves, and Thomas N Robinson. Digital trace data collection for social media effects research: Apis, data donation, and (screen) tracking. *Communication Methods and Measures*, 18(2):124–141, 2024.

[243] Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, Markus Mobius, David M Rothschild, and Duncan J Watts. Examining the consumption of radical content on youtube. *Proceedings of the National Academy of Sciences*, 118(32):e2101967118, 2021.

[244] Annett Heft, Kilian Buehling, Xixuan Zhang, Dominik Schindler, and Miriam Milzner. Challenges of and approaches to data collection across platforms and time: Conspiracy-related digital traces as examples of political contention. *Journal of Information Technology & Politics*, 21(3):323–339, 2024.

[245] Adrienne Massanari. # gamergate and the fappening: How reddit's algorithm, governance, and culture support toxic technocultures. *New media & society*, 19(3):329–346, 2017.

[246] Rebekah Tromble. Where have all the data gone? a critical reflection on academic digital research in the post-api age. *Social Media+ Society*, 7(1):2056305121988929, 2021.

[247] Johannes Breuer, Zoltán Kmetty, Mario Haim, and Sebastian Stier. User-centric approaches for collecting facebook data in the 'post-api age': Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 26(14):2649–2668, 2023.

[248] Mario Haim and Angela Nienierza. Computational observation: Challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in. *Computational Communication Research*, 1(1): 79–102, 2019.

[249] Laura Boeschoten, Jef Ausloos, Judith E Möller, Theo Araujo, and Daniel L Oberski. A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, 4(2):388–423, 2022.

[250] Valerie Hase and Mario Haim. Can we get rid of bias? mitigating systematic error in data donation studies through survey design strategies. *Computational Communication Research*, 6(2), 2024.

[251] Yvonne M Bishop, Stephen E Fienberg, and Paul W Holland. *Discrete multivariate analysis: Theory and practice*. Springer Science & Business Media, 2007.

[252] Attila Bátorfy and Ágnes Urbán. State advertising as an instrument of transformation of the media market in hungary. *East European Politics*, 36(1):44–65, 2020.

[253] Péter Bajomi-Lázár. Hungary's clientelistic media system. In *The Routledge Companion to Political Journalism*, pages 103–110. Routledge, 2021.

[254] Ádám Stefkovics. *The role of mixed-mode surveys in political surveying and election studies*. Phd thesis, Eotvos Lorand University, 2022.

[255] Paweł Matuszewski and Gabriella Szabó. Are echo chambers based on partisanship? twitter and political polarity in poland and hungary. *Social media+ society*, 5(2): 2056305119837671, 2019.

[256] David Krackhardt and Robert N Stern. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, pages 123–140, 1988.

[257] Sandra González-Bailón, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M Guess, et al. Asymmetric ideological segregation in exposure to political news on facebook. *Science*, 381(6656):392–398, 2023.

[258] Andrew M Guess. (almost) everything in moderation: new evidence on americans' online media diets. *American Journal of Political Science*, 65(4):1007–1022, 2021.

[259] Giang TC Tran, Luong Vuong Nguyen, Jason J Jung, and Jeonghun Han. Understanding political polarization based on user activity: a case study in korean political youtube channels. *Sage Open*, 12(2):21582440221094587, 2022.

[260] Niccolò Di Marco, Matteo Cinelli, and Walter Quattrociocchi. Infodemics on youtube: Reliability of content and echo chambers on covid-19. *arXiv preprint arXiv:2106.08684*, 2021.

[261] Anees Baqir, Yijing Chen, Fernando Diaz-Diaz, Sercan Kiyak, Thomas Louf, Virginia Morini, Valentina Pansanella, Maddalena Torricelli, and Alessandro Galeazzi. Unveiling the drivers of active participation in social media discourse. *Scientific Reports*, 15(1):4906, 2025.

[262] Mathias Bärtl. Youtube channels, uploads and views: A statistical analysis of the past 10 years. *Convergence*, 24(1):16–32, 2018.

[263] Morten Bay. Weaponizing the haters: The last jedi and the strategic politicization of pop culture through social media manipulation. *First Monday*, 2018.

[264] Jörg Matthes, Alon Zoizner, Andreas Nanz, David Nicolas Hopmann, Yannis Theocharis, and Selina Noetzel. The relationship between incidental news exposure and political participation: A cross-country, multilevel analysis. *Digital Journalism*, pages 1–20, 2025.

[265] Saifuddin Ahmed, Teresa Gil-Lopez, Sangwon Lee, and Muhammad Masood. Pathways from incidental news exposure to political knowledge: Examining paradoxical effects of political discussion on social media with strong and weak ties. *new media & society*, page 14614448241287763, 2024.

[266] Jonathan Haidt. Yes, social media really is undermining democracy. *The Atlantic*, 2022. URL https://www.theatlantic.com/ideas/archive/2022/07/social-media-harm-facebook-meta-response/670975/.

[267] Carole Cadwalladr. 'a polarisation engine': how social media has created a 'perfect storm' for uk's far-right riots. *The Guardian*, 2024. URL https://www.theguardian.com/media/article/2024/aug/03/a-polarisation-engine-how-social-media-has-created-a-perfect-storm-for-uks-far-

[268] Iris Marion Young. *Inclusion and democracy*. OUP Oxford, 2002.

[269] Stephen J Flusberg, Kevin J Holmes, Paul H Thibodeau, Robin L Nabi, and Teenie Matlock. The psychology of framing: How everyday language shapes the way we think, feel, and act. *Psychological Science in the Public Interest*, 25(3):105–161, 2024.

[270] Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. Algorithmic amplification of politics on twitter. *Proceedings of the national academy of sciences*, 119(1):e2025334119, 2022.

[271] Smitha Milli, Micah Carroll, Yike Wang, Sashrika Pandey, Sebastian Zhao, and Anca D Dragan. Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS nexus*, 4(3):pgaf062, 2025.

[272] Jan G Voelkel, Michael N Stagnaro, James Y Chu, Sophia L Pink, Joseph S Mernyk, Chrystal Redekopp, Isaias Ghezae, Matthew Cashman, Dhaval Adjodah, Levi G Allen, et al. Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity. *Science*, 386(6719):eadh4764, 2024.

[273] Keyon Vafa, Suresh Naidu, and David M Blei. Text-based ideal points. *arXiv preprint arXiv:2005.04232*, 2020.

[274] Klaus Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[275] Fabienne Lind, Maria Gruber, and Hajo G Boomgaarden. Content analysis by the crowd: Assessing the usability of crowdsourcing for coding latent constructs. *Communication methods and measures*, 11(3):191–209, 2017.

[276] Teresa Gil-López, Clara Christner, Ernesto de León, Mykola Makhortykh, Aleksandra Urman, Michaela Maier, and Silke Adam. Do (not!) track me: Relationship between willingness to participate and sample composition in online information behavior tracking research. *Social science computer review*, 41(6):2274–2292, 2023.

[277] Zoltán Kmetty, Ádám Stefkovics, Júlia Számely, Dongning Deng, Anikó Kellner, Edit Pauló, Elisa Omodei, and Júlia Koltai. Determinants of willingness to donate data from social media platforms. *Information, Communication & Society*, pages 1–26, 2024.

Appendix A

Intermedia agenda setting during the 2016 and 2020 U.S. presidential elections

A.1 Collecting human labels from MTurk

Our quality-control pipeline involves (a) selecting workers who have acquired a *Masters* qualification and reside in the U.S., and (b) blocking workers who fail to correctly label any of the two screening texts in a single HIT (see the full pipeline in SM Figure A.2). We include detailed instructions on the top of the task page, where we describe the data sources, the purpose of our study, and the quality assurance steps we take to decide whether to accept a submission (see the task interface in SM Figure A.1). Workers are compensated 1.0 USD per HIT, achieving an hourly rate of 15 USD at a relaxing speed of 4 minutes per HIT.

The entire task costs 407 USD, providing labels for 240 news headlines, 240 survey responses, and 240 tweets sampled from our data[1], with each text being read by three MTurk workers. In total, 40 workers are involved in our study, all of whom have contributed at least one qualified assignment. 19 out of these workers were blocked from further submitting for failing the screening questions. Out of 284 total submitted assignments, we use 263 qualified ones (93.36%) to generate labels for 718 text inputs. We assess the reliability of the workers by computing inter-rater reliability (Krippendorff's alpha $= 0.4385$ for 2016 and 0.4251 for 2020) for the primary topic, a commonly used measure in the literature [274]. The reliability, while low by traditional content analysis standards, is significantly higher than accepted levels for crowd-sourced approaches [275].

[1] The topic labeling task is designed for three datasets of short texts relevant to presidential candidates: (1) news headlines, (2) tweets that mentioned at least one candidate's last name, and (3) survey responses to the question "what have you read/seen/heard about candidate X?". Because we only analyze headline data for this paper, we skip the discussions of the other two datasets.

A.2 Evaluating model performance

Our evaluation proceeds as follows. For each text input, we use the model plus a random human labeler as the model-human pair and randomly select two human labelers as the human-human pair. For the single-label case (i.e., model option (a)), the agreement score equals the number of times we receive two identical labels divided by the total number of text inputs evaluated; for multi-label cases (i.e., model option (b) and (c)), we compute the Jaccard similarity for each pair of labels and obtain the average. We compare agreement scores between model-human and human-human for nine model variants with (1) different keyword filtering/weighting strategies and (2) varying numbers of topics per text, and choose the one producing the highest and closest agreement score for the model-human pair compared to the human-human pair.

Please read instructions carefully before you start :)

Hi, we are a group of researchers from the University of Michigan, School of Information. Our current project is trying to understand **the topics covered in the news, social media, and responses to survey questions about the presidential candidates during the 2016 and 2020 elections.**

- Participation is voluntary.
- The University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board has determined that this research is exempt from IRB oversight.
- Please contact **contact removed for blind review** for any questions or concerns.

Your task: to assign topic labels for 10 candidate-related short texts we sampled from

1. news headlines,
2. tweets, and
3. survey responses to *"what have you recently heard/read/seen about [CANDIDATE NAME]"* (if/when such text samples are included in the HIT, they will be pre-fixed with [about CANDIDATE NAME]).

Please assign the following three types of topic labels:

Topic Type	Guidance
Primary Topic	the most relevant topic (if not listed, select "not applicable")
Secondary Topic	the second most relevant topic (if not listed, select "not applicable")
Relevant Topic(s)	all other topic(s) relevant (if any)

*NOTES:

- If there are multiple topics and you cannot decide which is more relevant, please rank topic relevance based on the order of keywords appearance, i.e., the first (second) keyword that shows up in the short text would link to the primary (secondary) topic.
- **[IMPORTANT] Label quality assurance:** A small portion of the texts are pre-labeled with the correct topics. You won't be able to participate in future labeling task if your labeling accuracy falls below a certain threshold for these pre-labeled texts, but you will still receive full payment for the HITs that you have *attentively* worked on.

Some examples with suggested answers:

Example Type	Example(s)	Suggested Answers
One dominant topic	Biden's immigration pick yields outrage on left	Primary: immigration Secondary: not applicable Relevant: blank
One dominant topic	[about Trump] He doesn't want to support renewable energy .	Primary: energy Secondary: not applicable Relevant: blank
Multiple relevant topics (with a clear focus)	Canada open to renegotiating free trade with Trump.	Primary: foreign trade Secondary: international affairs Relevant: blank
Multiple relevant topics (with no clear ranking)	Biden's covid-19 taskforce recommends withholding food stamps , rent assistance , healthcare from vaccine refusers	Primary: healthcare Secondary: social welfare Relevant: housing
No relevant topic (candidate-related but no specific topics are involved; expressing pure sentiment)	[About Trump] He's very good at being Donald Trump. [about Hillary] I don't trust her.	Primary: not applicable Secondary: not applicable Relevant: blank
No relevant topic (not candidate-related at all)	BLM invades Trader Joe's to protest lack of black access to grocery stores.	Primary: not applicable Secondary: not applicable Relevant: blank

Please assign topic labels to the following 10 short texts.

(Hover over topic descriptions to view some examples of short texts.)

****{text1}****

Examples for civil rights:

1. Teens, pranksters spam Trump's voter fraud hotline

2. [about Biden] Heard of Biden's speech about racial justice

3. Donald Trump's hotel in D.C. vandalized with "black lives matter" graffiti

Primary Topic	Secondary Topic	Relevant Topic(s)	Topic Description (guidance, or a few example sub-categories for each topic)
○	○	☐	**not applicable:** none of the following topics is relevant; pure sentiment (e.g., like/dislike) w/o specific topics
○	○	☐	**agriculture:** agriculture policy, trade & marketing; farmers; fisheries & fishing; animal & crop disease
			civil rights: racial equality; gender equality; voting rights; freedom of speech; gun rights; right to privacy; age discrimination; anti-government activities
			crime: law enforcement agencies; crimes & crime control; police; prisons; court administration; child abuse & family issues
			culture: cultural policy; culture & entertainment
			defence: defence alliance & agreement; military intelligence; nuclear arms; military aid; military procurement; domestic security responses; foreign military operations
			economy: banking; small businesses; disaster relief; tax policies; consumer finance; insurance regulation; bankruptcy; corporate management; securities & commodities
○	○	☐	**education:** education policy; elementary & primary schools; vocational education; higher education, student loans; education of underprivileged students
○	○	☐	**election campaign:** campaign-related events: conventions; debates; townhalls & rallies; running mate nomination
○	○	☐	**energy:** energy policy; nuclear; electricity; natural gas & oil; coal; alternative & renewable; conservation & efficiency; research & development
○	○	☐	**environment:** environmental policy; drinking water; waste disposal; hazardous waste; air pollution; recycling; species & forest; land and water conservation
○	○	☐	**foreign trade:** trade agreements; exports; private investments; tariff & imports; exchange rates; competitiveness; trade policy
○	○	☐	**government operation:** general governmental operations; intergovernmental relations; bureaucracy; census & statistics; postal service; procurement & contractors
○	○	☐	**healthcare:** public health and candidates' health conditions; coronavirus spread & control; healthcare reform; insurance; medical facilities; disease prevention; healthcare research & development; mental health; drug and alcohol abuse
○	○	☐	**housing:** community development; urban development; rural housing; low-income assistance; housing for veteran, the elderly & the homeless
○	○	☐	**immigration:** immigration issues & policies; refugees; citizenship
○	○	☐	**international affairs:** international affairs & foreign aid; resources exploitation; developing countries; international finance; human rights issues; terrorism; international organizations
○	○	☐	**labor:** labour, employment & pensions; employee benefits; labor unions; fair labor standards; worker safety; employment training; youth employment
○	○	☐	**religion:** general religious issues; religous groups; church activities; religious freedom
○	○	☐	**social welfare:** social welfare policy; low-income/elderly/disabled assistance; volunteer associations; child care
○	○	☐	**space, science, technology, & communications:** issues related to general space, science, technology & communications: mass/social media presence, space programs, telecommunication regulation
○	○	☐	**transportation:** mass transportation construction; highways, air & railroad travel; maritime transportation; infrastructure
○	○	☐	**trump controversies:** controversial topics related to Trump, such as family or personal scandal, health condition speculations and disputable remarks
○	○	☐	**biden controversies:** controversial topics related to Biden, such as family or personal scandal, health condition speculations and disputable remarks
○	○	☐	**general controversies:** general controversial topics with no main targeting candidate

****{text2}****

Figure A.1: Screenshots of labeling task interface we launched on Amazon Mturk.

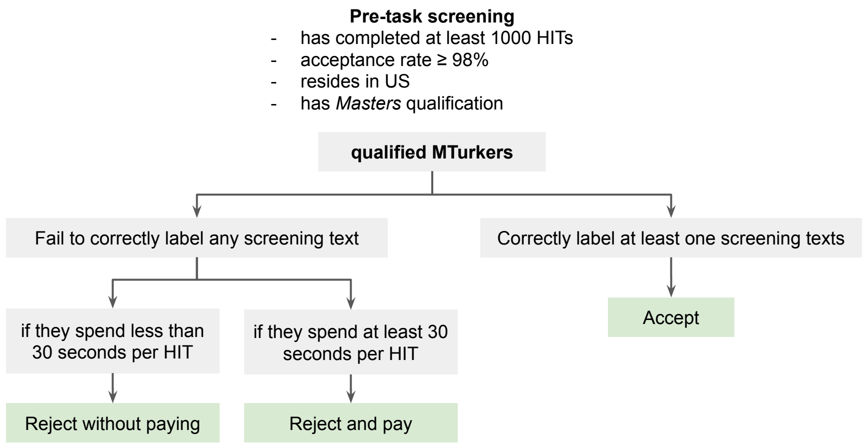

Figure A.2: Diagram of our quality-control pipeline on MTurk.

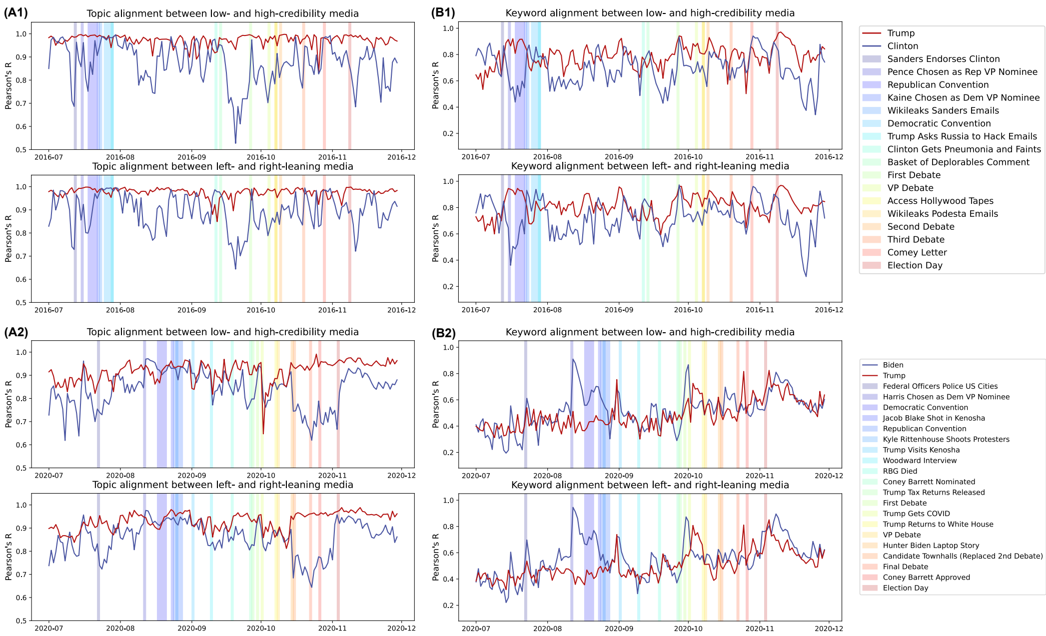

Figure A.3: (A) Topic and (B) keyword alignment over time in 2016 (A1 and B1) and 2020 (A2 and B2). Two pairings of media have been applied: (i) between low- and high-credibility news media, and (ii) between left- and right-leaning news media.

Between low- and high-credibility media											
Trump 2016						Clinton 2016					
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
1	clinton_controversies	0.0818	-0.3985	0.9783	0.0004	1	clinton_controversies	0.3771	-0.8581	0.9519	0.0000
2	healthcare	0.0463	0.3867	0.9734	0.0078	2	government_ops	0.3415	0.7204	0.9767	0.0000
3	government_ops	0.0348	-0.0975	0.9739	0.0215	3	healthcare	0.3292	-1.0742	0.9188	0.0000
4	religion	0.0347	0.5750	0.9738	0.0215	4	civil_rights	0.1625	1.0886	0.8809	0.0000
5	energy	0.0187	0.8224	0.9741	0.0925	5	defence	0.0485	1.3085	0.8693	0.0064
6	immigration	0.0136	-0.2392	0.9739	0.1532	6	foreign_trade	0.0439	4.3758	0.8706	0.0096
7	environment	0.0130	-0.5692	0.9731	0.1618	7	environment	0.0377	4.9955	0.8741	0.0165
8	sstc	0.0126	0.1448	0.9739	0.1694	8	trump_controversies	0.0297	0.6729	0.8750	0.0336
9	intl_affairs	0.0115	0.1368	0.9724	0.1891	9	economy	0.0177	0.6136	0.8706	0.1022
10	foreign_trade	0.0108	0.5925	0.9723	0.2035	10	energy	0.0077	-2.2671	0.8713	0.2826
11	defence	0.0027	0.1100	0.9731	0.5243	11	labour	0.0063	0.8403	0.8731	0.3298
12	trump_controversies	0.0026	0.0285	0.9746	0.5341	12	sstc	0.0038	-0.2669	0.8725	0.4481
13	economy	0.0025	0.0698	0.9729	0.5430	13	intl_affairs	0.0030	0.3211	0.8734	0.5031
14	education	0.0021	0.2084	0.9734	0.5721	14	religion	0.0028	0.6867	0.8712	0.5182
15	civil_rights	0.0007	-0.0266	0.9734	0.7446	15	education	0.0010	-0.5998	0.8709	0.6997
16	labour	0.0003	0.0644	0.9733	0.8464	16	crime	0.0010	0.0950	0.8699	0.6969
17	social_welfare	0.0001	-0.0875	0.9734	0.9063	17	social_welfare	0.0006	-0.8843	0.8714	0.7598
18	crime	0.0000	0.0007	0.9733	0.9945	18	immigration	0.0002	0.2195	0.8722	0.8507
Trump 2020						Biden 2020					
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
1	healthcare	0.4583	0.4737	0.9459	0.0000	1	biden_controversies	0.4331	-1.3262	1.0412	0.0000
2	government_ops	0.2056	-0.4172	0.9039	0.0000	2	healthcare	0.1132	0.5269	0.8452	0.0000
3	sstc	0.1400	-0.8924	0.9735	0.0000	3	sstc	0.1115	-1.5748	0.9071	0.0000
4	intl_affairs	0.0631	-0.8651	0.9339	0.0018	4	economy	0.0608	1.0088	0.8434	0.0022
5	labour	0.0593	-2.5242	0.9155	0.0025	5	social_welfare	0.0288	3.3612	0.8529	0.0367
6	education	0.0584	1.5050	0.9247	0.0027	6	defence	0.0245	1.3161	0.8400	0.0541
7	energy	0.0389	-2.5416	0.9152	0.0149	7	immigration	0.0152	-1.7457	0.8364	0.1303
8	social_welfare	0.0379	4.8195	0.9195	0.0162	8	energy	0.0120	0.7488	0.8432	0.1798
9	trump_controversies	0.0332	-0.2389	0.9193	0.0246	9	foreign_trade	0.0107	1.3465	0.8413	0.2040
10	immigration	0.0315	1.7172	0.9131	0.0286	10	civil_rights	0.0092	0.2771	0.8371	0.2405
11	environment	0.0266	-1.2478	0.9129	0.0446	11	environment	0.0072	-0.5412	0.8334	0.2992
12	biden_controversies	0.0221	-1.6594	0.9462	0.0676	12	crime	0.0067	-0.2425	0.8412	0.3143
13	civil_rights	0.0211	-0.2589	0.9133	0.0742	13	intl_affairs	0.0034	0.2468	0.8402	0.4745
14	foreign_trade	0.0201	-3.0306	0.9143	0.0818	14	trump_controversies	0.0022	0.1638	0.8371	0.5654
15	defence	0.0107	-0.4614	0.9224	0.2055	15	education	0.0012	0.4482	0.8399	0.6753
16	economy	0.0019	0.1600	0.9190	0.5917	16	government_ops	0.0011	0.0406	0.8460	0.6850
17	crime	0.0013	0.0634	0.9177	0.6551	17	labour	0.0005	-0.2058	0.8399	0.7870
18	religion	0.0000	-0.1049	0.9183	0.9344	18	religion	0.0002	-0.1412	0.8390	0.8626
Between left- and right-leaning media											
		Trump 2016						Clinton 2016			
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
1	trump_controversies	0.3421	0.2847	0.9884	0.0000	1	clinton_controversies	0.5109	-0.8366	0.9702	0.0000
2	government_ops	0.2855	-0.2427	0.9773	0.0000	2	government_ops	0.4927	0.7247	0.9977	0.0000
3	sstc	0.0922	0.3410	0.9772	0.0001	3	healthcare	0.2963	-0.8535	0.9295	0.0000
4	clinton_controversies	0.0615	-0.3001	0.9797	0.0021	4	civil_rights	0.1551	0.8905	0.8996	0.0000
5	healthcare	0.0538	0.3623	0.9760	0.0040	5	defence	0.0303	0.8668	0.8905	0.0319
6	environment	0.0267	-0.7085	0.9757	0.0443	6	trump_controversies	0.0237	0.5026	0.8945	0.0585
7	immigration	0.0231	-0.2711	0.9766	0.0618	7	environment	0.0229	3.2566	0.8937	0.0630
8	religion	0.0224	0.4008	0.9763	0.0660	8	foreign_trade	0.0093	1.6879	0.8918	0.2368
9	economy	0.0209	-0.1762	0.9772	0.0756	9	economy	0.0072	0.3279	0.8916	0.2983
10	education	0.0074	-0.3383	0.9759	0.2905	10	intl_affairs	0.0021	0.2250	0.8933	0.5754
11	foreign_trade	0.0047	0.3403	0.9754	0.4012	11	education	0.0012	0.5600	0.8933	0.6672
12	labour	0.0033	0.2042	0.9758	0.4789	12	religion	0.0010	0.3518	0.8919	0.6928
13	energy	0.0031	0.2888	0.9763	0.4982	13	social_welfare	0.0009	0.8794	0.8928	0.7166
14	civil_rights	0.0022	-0.0409	0.9761	0.5634	14	immigration	0.0008	-0.3322	0.8919	0.7338
15	crime	0.0005	-0.0260	0.9762	0.7785	15	crime	0.0002	-0.0345	0.8930	0.8660
16	defence	0.0005	-0.0400	0.9761	0.7898	16	labour	0.0001	0.0661	0.8924	0.9272
17	social_welfare	0.0002	0.1137	0.9759	0.8604	17	energy	0.0000	0.1051	0.8923	0.9526
18	intl_affairs	0.0001	-0.0114	0.9761	0.9003	18	sstc	0.0000	-0.0246	0.8923	0.9336
		Trump 2020						Biden 2020			
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
1	healthcare	0.4520	0.3817	0.9539	0.0000	1	biden_controversies	0.7565	-1.5721	1.0982	0.0000
2	government_ops	0.2913	-0.4029	0.9178	0.0000	2	sstc	0.1202	-1.4664	0.9219	0.0000
3	sstc	0.2464	-0.9607	0.9911	0.0000	3	healthcare	0.1155	0.4772	0.8642	0.0000
4	education	0.1291	1.8162	0.9395	0.0000	4	foreign_trade	0.0656	2.9863	0.8640	0.0014
5	social_welfare	0.0806	5.7017	0.9332	0.0004	5	social_welfare	0.0512	4.0195	0.8752	0.0051
6	labour	0.0713	-2.2454	0.9293	0.0009	6	government_ops	0.0315	0.1947	0.8931	0.0286
7	immigration	0.0548	1.8365	0.9261	0.0037	7	economy	0.0224	0.5495	0.8608	0.0656
8	foreign_trade	0.0403	-3.4842	0.9272	0.0132	8	defence	0.0201	1.0680	0.8593	0.0817
9	energy	0.0353	-1.9664	0.9293	0.0204	9	immigration	0.0156	-1.5855	0.8561	0.1255
10	intl_affairs	0.0338	-0.5139	0.9409	0.0233	10	crime	0.0079	-0.2360	0.8606	0.2747
11	environment	0.0244	-0.9693	0.9275	0.0547	11	energy	0.0074	0.5273	0.8614	0.2928
12	civil_rights	0.0232	-0.2202	0.9275	0.0611	12	civil_rights	0.0071	0.2192	0.8570	0.3008
13	economy	0.0227	0.4459	0.9340	0.0640	13	religion	0.0039	0.5572	0.8571	0.4460
14	biden_controversies	0.0168	-1.1736	0.9514	0.1116	14	environment	0.0034	-0.3347	0.8549	0.4745
15	trump_controversies	0.0092	-0.1019	0.9320	0.2405	15	labour	0.0025	0.4178	0.8558	0.5407
16	crime	0.0003	-0.0262	0.9317	0.8198	16	intl_affairs	0.0016	0.1527	0.8592	0.6218
17	religion	0.0002	0.1607	0.9311	0.8765	17	trump_controversies	0.0006	0.0763	0.8575	0.7653
18	defence	0.0001	0.0442	0.9311	0.8814	18	education	0.0001	0.1030	0.8585	0.9146

Table A.1: Summary table of OLS results to model temporal alignment between low- and high-credibility media (top), and between left- and right-leaning media (bottom). The dependent variable is the time series of temporal alignment (Pearson's R) in the overall topic distribution between low- and high-credibility media, or between left- and right-leaning media; the independent variable is the time series of the temporal difference in the proportional attention devoted to a certain topic (high-credibility subtracted by low-credibility, or right-leaning subtracted by left-leaning).

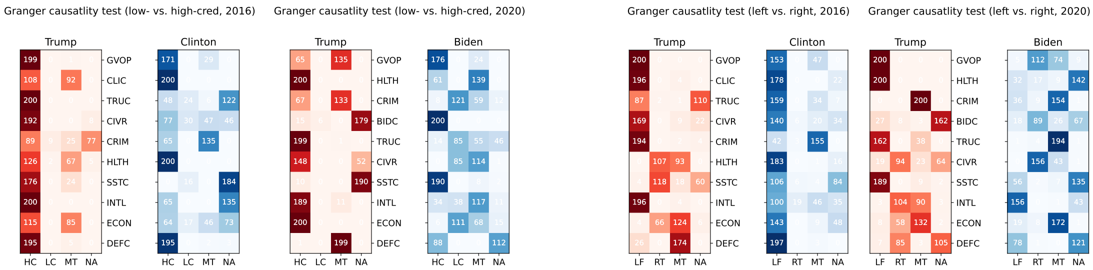

Figure A.4: Granger causality between low- and high-credibility media (the left four columns) and between left- and right-leaning media (the right four columns). The cell values are the number of times a given type of IAS relationship appears significant out of 200 bootstrapping runs (sampling 80% of the data). Types of IAS relationship (along the X-axis) are displayed in abbreviations: HC means led by high-credibility media; LC means led by low-credibility media; LF means led by left-leaning media; RT means led by right-leaning media; MT means we found significant results (i.e., mutual interaction) in both directions; NA means we found no significant results in either direction. Significance threshold for p-value is 0.05. We include results for the top 10 topics (in descending order) that show frequently in all news headlines for a given year.

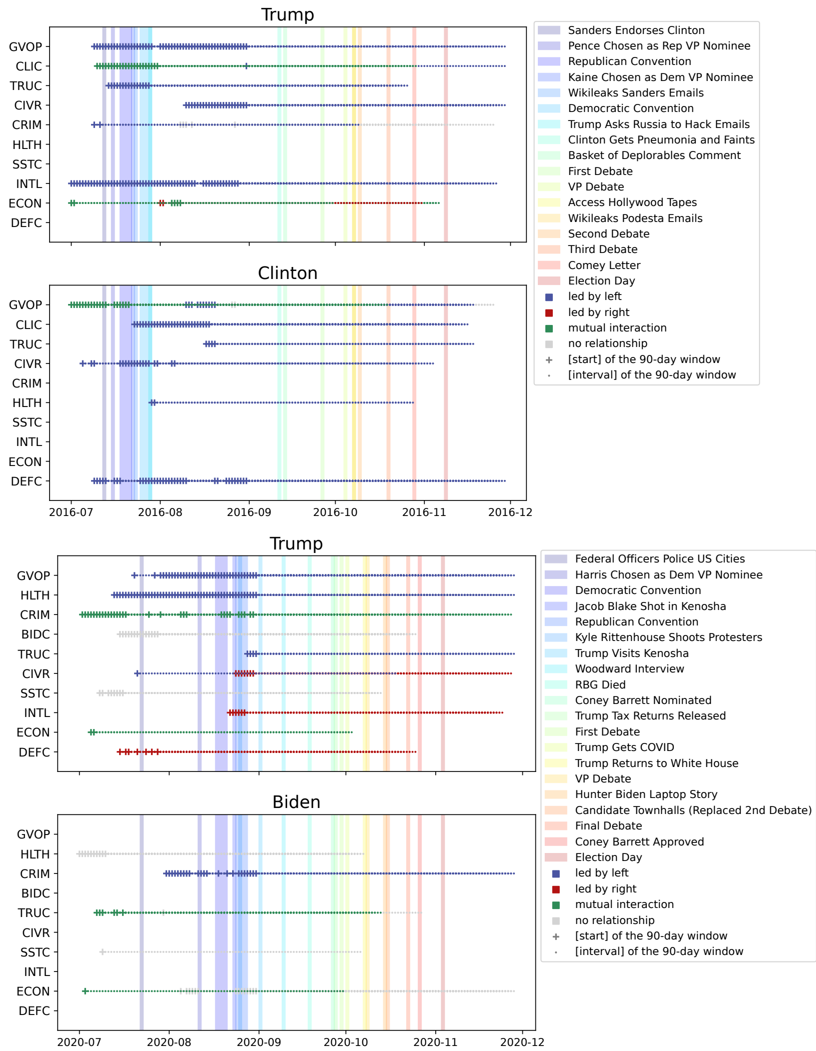

Figure A.5: Granger causality results between low- and high-credibility media for 2016 (the first and second figures) and 2020 (the third and fourth figures). We test Granger causalities in a sliding window of 90 days and display robust results that appear in more than 95% of the bootstrapping runs. Each plus sign marks the starting point of the 90-day sliding window with a significant result. We include results for the top 10 topics that show frequently in all news headlines for a given year.

Model variant			Headline			Survey			Tweets			Average across three sources		
year	keywords filtering/weighting	# of topics per text	model-human	human-human	diff.	model-human	human-human	diff.	model-human	human-human	diff.	model-human	human-human	diff.
2016	**strongly only**	**primary only**	**0.4881**	**0.5653**	**-0.0771**	**0.5661**	**0.6311**	**-0.0650**	**0.3564**	**0.4591**	**-0.1028**	**0.4702**	**0.5519**	**-0.0816**
2016	strongly + weakly (equal weights)	primary only	0.4566	0.5689	-0.1123	0.5537	0.6287	-0.0751	0.3068	0.4677	-0.1609	0.4390	0.5551	-0.1161
2016	strongly + weakly (different weights)	primary only	0.4876	0.5680	-0.0804	0.5419	0.6269	-0.0850	0.3118	0.4697	-0.1579	0.4471	0.5549	-0.1077
2016	strongly only	primary + secondary	0.2116	0.5993	-0.3877	0.3056	0.6337	-0.3282	0.2197	0.5026	-0.2829	0.2456	0.5785	-0.3329
2016	strongly + weakly (equal weights)	primary + secondary	0.1956	0.6015	-0.4059	0.2768	0.6331	-0.3563	0.1627	0.5021	-0.3394	0.2117	0.5789	-0.3672
2016	strongly + weakly (different weights)	primary + secondary	0.1989	0.6002	-0.4014	0.2739	0.6346	-0.3607	0.1646	0.5019	-0.3374	0.2124	0.5789	-0.3665
2016	strongly only	all included	0.2018	0.6619	-0.4601	0.3096	0.7023	-0.3927	0.2239	0.5914	-0.3675	0.2451	0.6519	-0.4068
2016	strongly + weakly (equal weights)	all included	0.1834	0.6646	-0.4813	0.2751	0.7036	-0.4285	0.1683	0.5896	-0.4213	0.2089	0.6526	-0.4437
2016	strongly + weakly (different weights)	all included	0.1846	0.6641	-0.4794	0.2748	0.7032	-0.4283	0.1679	0.5914	-0.4234	0.2091	0.6529	-0.4437
2020	**strongly only**	**primary only**	**0.4570**	**0.5149**	**-0.0580**	**0.5426**	**0.6021**	**-0.0595**	**0.3725**	**0.4700**	**-0.0975**	**0.4573**	**0.5290**	**-0.0717**
2020	strongly + weakly (equal weights)	primary only	0.4533	0.5116	-0.0583	0.5467	0.6015	-0.0549	0.3628	0.4704	-0.1077	0.4542	0.5279	-0.0736
2020	strongly + weakly (different weights)	primary only	0.4649	0.5133	-0.0484	0.5572	0.6039	-0.0467	0.3652	0.4716	-0.1065	0.4624	0.5296	-0.0672
2020	strongly only	primary + secondary	0.2252	0.5232	-0.2980	0.2650	0.6308	-0.3658	0.2226	0.4965	-0.2739	0.2376	0.5502	-0.3126
2020	strongly + weakly (equal weights)	primary + secondary	0.2105	0.5252	-0.3147	0.2457	0.6267	-0.3811	0.2024	0.4928	-0.2904	0.2195	0.5482	-0.3287
2020	strongly + weakly (different weights)	primary + secondary	0.2174	0.5272	-0.3099	0.2472	0.6263	-0.3791	0.1996	0.4944	-0.2948	0.2214	0.5493	-0.3279
2020	strongly only	all included	0.2272	0.6337	-0.4065	0.2657	0.6933	-0.4276	0.2235	0.5878	-0.3643	0.2388	0.6383	-0.3995
2020	strongly + weakly (equal weights)	all included	0.2198	0.6357	-0.4159	0.2407	0.6948	-0.4541	0.1899	0.5898	-0.3999	0.2168	0.6401	-0.4233
2020	strongly + weakly (different weights)	all included	0.2202	0.6339	-0.4137	0.2391	0.6944	-0.4552	0.1905	0.5891	-0.3986	0.2166	0.6391	-0.4225

Table A.2: Complete agreement scores for model-human and human-human comparisons.

Appendix B

Using response-item network (ResIN) to measure ideological polarization

B.1 ANES included items

Party identification of respondent

(VCF0301) Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?

1. Strong Democrat
2. Weak Democrat
3. Independent - Democrat
4. Independent - Independent
5. Independent - Republican
6. Weak Republican
7. Strong Republican

Attitudes regarding five issues

ANES code	Abbreviation	Description	Levels of attitudes
VCF0839	spend_serv	Government service-spending scale	7
VCF0806	gov_health	Government health insurance scale	7
VCF0809	guar_jobs	Guaranteed jobs and income scale	7
VCF0830	aid_black	Aid to black scale	7
VCF0838	abort	By law, when should abortion be allowed	4

Table B.1: Selected issues from ANES to include in ResIN.

Government service-spending scale

Some people think the government should provide fewer services, even in areas such as health and education, in order to reduce spending. Suppose these people are at one end of a scale, at point 1. Other people feel that it is important for the government to provide many more services even if it means an increase in spending. Suppose these people are at the other end, at point 7. And of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Government health insurance scale

There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan which would cover all medical and hospital expenses for everyone. Suppose these people are at one end of a scale, at point 1. Others feel that medical expenses should be paid by individuals, and through private insurance plans like Blue Cross. Suppose these people are at the other end, at point 7. And of course, some people have opinions somewhere in between at points 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Guaranteed jobs and income scale

Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. Suppose these people are at one end of a scale, at point 1. Others think the government should just let each person get ahead on his/their own. Suppose these people are at the other end, at point 7. And of course, some other people have opinions somewhere in between, at pints 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Aid to black scale

Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks. Suppose these people are at one end of a scale, at point 1. Others feel that the government should not make any special effort to help blacks because they should help themselves. Suppose these people are at the other end, at point 7. And of course, some other people have opinions somewhere in between, at points 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about it?

Legal abortion scale

There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view? You can just tell me the number of the opinion you choose.

1. By law, abortion should never be permitted.
2. The law should permit abortion only in case of rape, incest, or when the woman's life is in danger.
3. The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.
4. By law, a woman should always be able to obtain an abortion as a matter of personal choice.

B.2 Simulation of a "broken egg" via IRT

We use IRT to set up a polarized attitude system. Under the IRT paradigm, every respondent has (i) a value θ for a certain latent variable, and (ii) a so-called item characteristic curve $f(\theta)$ for every issue position. $f(\theta)$ determines the probability of this respondent selecting a certain issue position [211]. For instance, let us consider a case where the latent variable is the left-right spectrum and a small (large) value of θ indicates the left-wing (right-wing) position. For an issue position "support for gun control", because left-leaning respondents are more likely to support gun control in the U.S., we expect $f(\theta)$ to be greater for smaller θ (i.e. higher probability for left-leaning respondents to select this issue position), and smaller for larger θ. Usually, IRT-based probability curves have a bell-shape resembling normal distributions. For a given issue, if the peak of probability functions for issue positions are well separated (i.e., they overlap only minimally), people with similar values of θ would concentrate their responses on the same issue position, while people with different values of θ would have distinct issue positions. The random model, in contrast, supposes no latent variable nor any association between issue positions. For each respondent, it assumes that all possible positions are equally probable. Following the previous example, within the random model, a right-leaning respondent is equally likely to select any position for a given issue, regardless of the ideological characteristics of the position. One simple implementation of this model is given by setting all item characteristic curves to be constant and identical one another.

To control the level of polarization, we interpolate between a well-polarized IRT model and the random model using a "randomness" parameter r. Therefore, the final item characteristic curve for a given r is:

$$f_r(\theta) = (1-r)f(\theta) + r/L$$

where $f(\theta)$ is the "classic" item-characteristic curve, as obtained from Samejima [211], and L is the number of issue positions for a certain issue (e.g., corresponding to the number of response options in a survey). In each simulation round, we fix the number of respondents ($N = 1000$), the number of issues ($K = 8$) and the number of positions per issue ($L = 7$), and vary the randomness parameter r. Each respondents has a random value for the latent variable θ. Based on the issue position probability computed via $f_r(\theta)$, each respondent selects one position from each of the 8 issues. In this manner, we are able to generate synthetic datasets resembling survey responses. From here, we repeat the same procedure generating ResINs and measure linearization for each simulated network.

We display the simulation results in Figure B.1. Panel (a) displays three ResIN snapshots obtained from $r = 0$, 0.5, and 1 and panel (b) shows the monotonically negative relationship between r and linearization. When the system is fully polarized (i.e. $r = 0$), the corresponding network is stretched flat and fully elongated, exhibiting a high level of linearization. As we increase the noise by tuning up the randomness parameter, the elongated structure folds on itself and starts to take a more oval shape, and the linearization level drops rapidly. Finally, in the case of pure randomness when $r = 1$ and all issue positions are equally probable[1], the system displays a round shape with a linearization level approximating 1.

[1] Ideally, this will result in all issue positions having 0 correlation to each other, but, due to the finite number of people, correlations are not exactly 0, but present some random deviation from it. As a result, every issue position is connected to other random positions, resulting in the circular pattern that we observe in the figure.

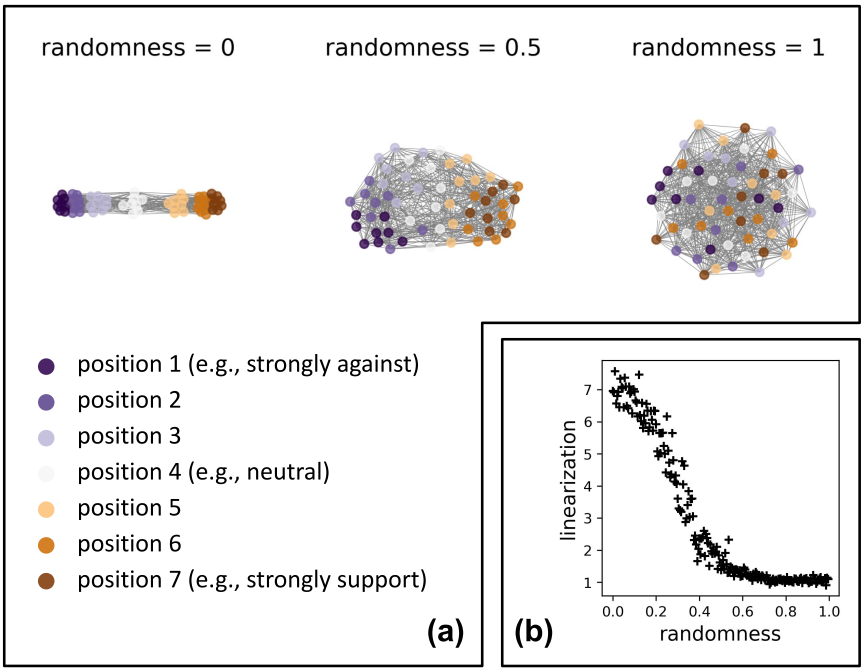

Figure B.1: "Broken egg" simulations with varying randomness parameters r. Subfigure (a) shows the ResIN output generated by $r = 0, 0.5, 1$; subfigure (b) shows the negative relationship between randomness and linearization level.

Appendix C

What visible engagement fails to capture in online political communication

C.1 Engagement form and ideological polarization

To compare the ideological distribution of the different respondent groups, we adopt two longstanding polarization metrics, variance and kurtosis [77], to quantify if and to which extent, the measurement of polarization across different respondent groups would yield varying results. We use variance to measure the statistical dispersion of respondents' ideological positions, and kurtosis to measure the overall "tailedness" of these distributions, i.e., how concentrated a given distribution is around its mean. A more polarized population is expected to display higher variance and lower kurtosis, with people having more dispersed ideological positions and a lower level of consensus around the average. If these summary statistics vary greatly across respondent groups, studies that observe only the visible group (i.e., commenters) will be evidenced to suffer from measurement bias and should refrain from making general conclusions about YouTube users. We use Mann-Whitney U test to compare the distribution of variance and kurtosis of different user groups. We incorporate survey weights among the respondents by bootstrapping the respondent samples with the survey weights for 100 rounds, and computing variance and kurtosis for samples obtained in each round. We use Mann-Whitney U test to compare the distribution of variance and kurtosis per 100 rounds of bootstrapping. Given the limited sample size of commenting data, we examine if the between-sample variation is significantly larger than the within-sample fluctuation through rerunning the whole process for 200 rounds to assess the robustness of the Mann-Whitney U test results (see details in Appendix C.8).

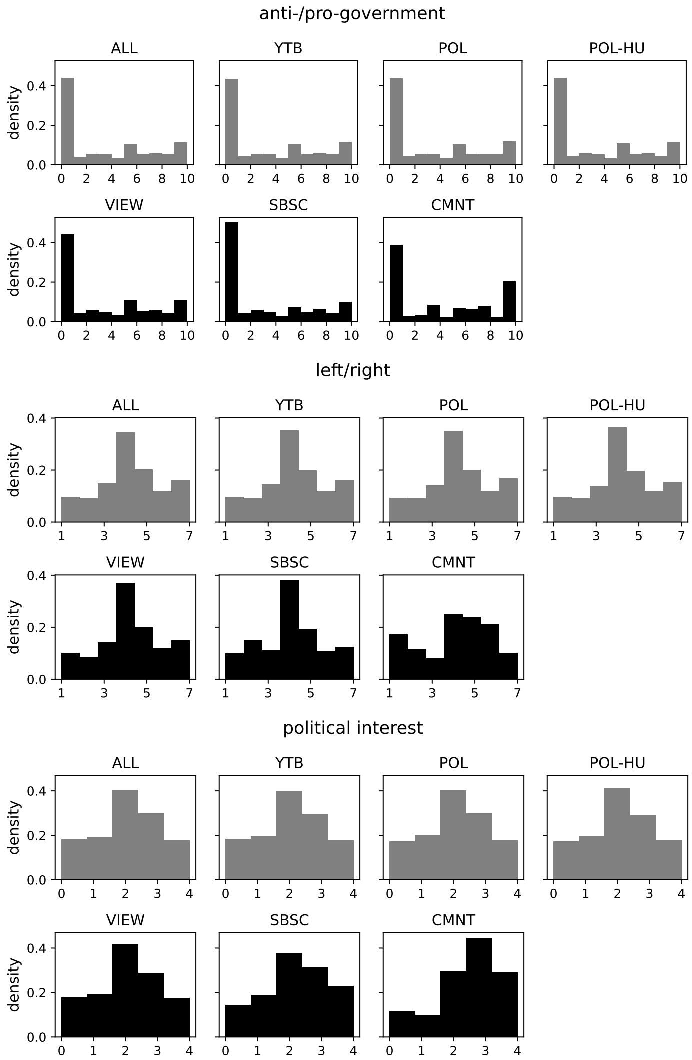

Figure C.1: Distribution of anti-/pro-government scale, left/right scale, political interest across respondent groups throughout different filtering stage (i.e., the entire sample, the subset with YouTube activities, the subset that has engaged with political content on YouTube, the subset that has engaged with Hungarian political content on YouTube), and with different engagement forms (i.e., viewers, subscribers, commenters).

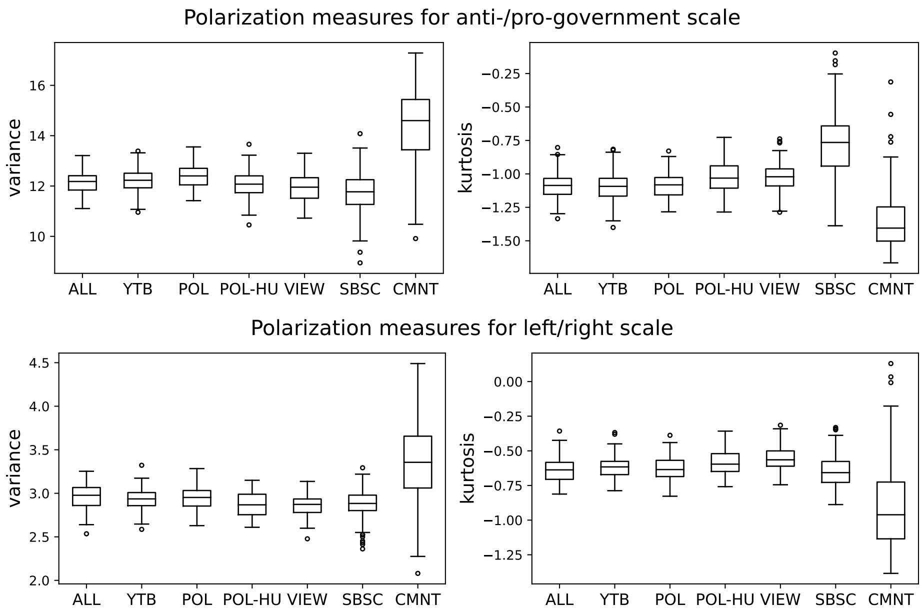

Figure C.2: Distribution of two polarization metrics (i.e., variance and kurtosis) along anti-/pro-government scale and left/right scale, across respondent groups throughout different filtering stage (i.e., the entire sample, the subset with YouTube activities, the subset that has engaged with political content on YouTube, the subset that has engaged with Hungarian political content on YouTube), and with different engagement forms (i.e., viewers, subscribers, commenters).

C.2 Supplementary regression results with left/right ideological groups and extremity level and joint prediction for political interest and political leaning

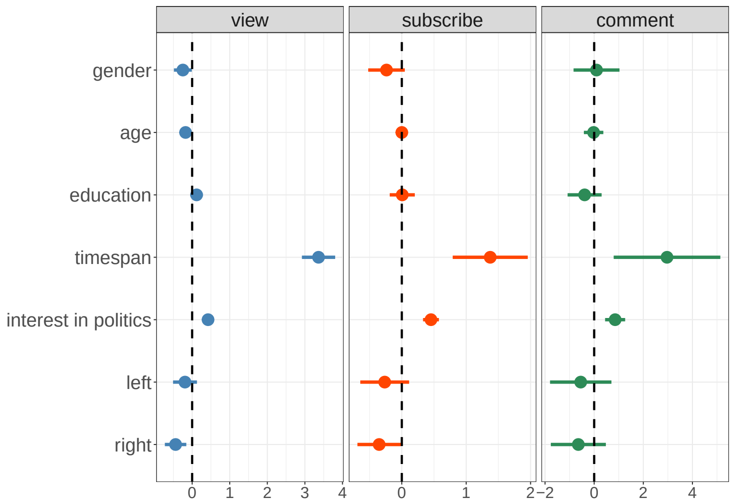

Figure C.3: Negative binomial regression coefficients with 95% confidence intervals for viewing, subscribing and commenting activities on Hungarian political content. The DVs are the number of recorded activities for viewing, subscribing, and commenting from left to right for each panel. The IVs contains the left/right ideological groups.

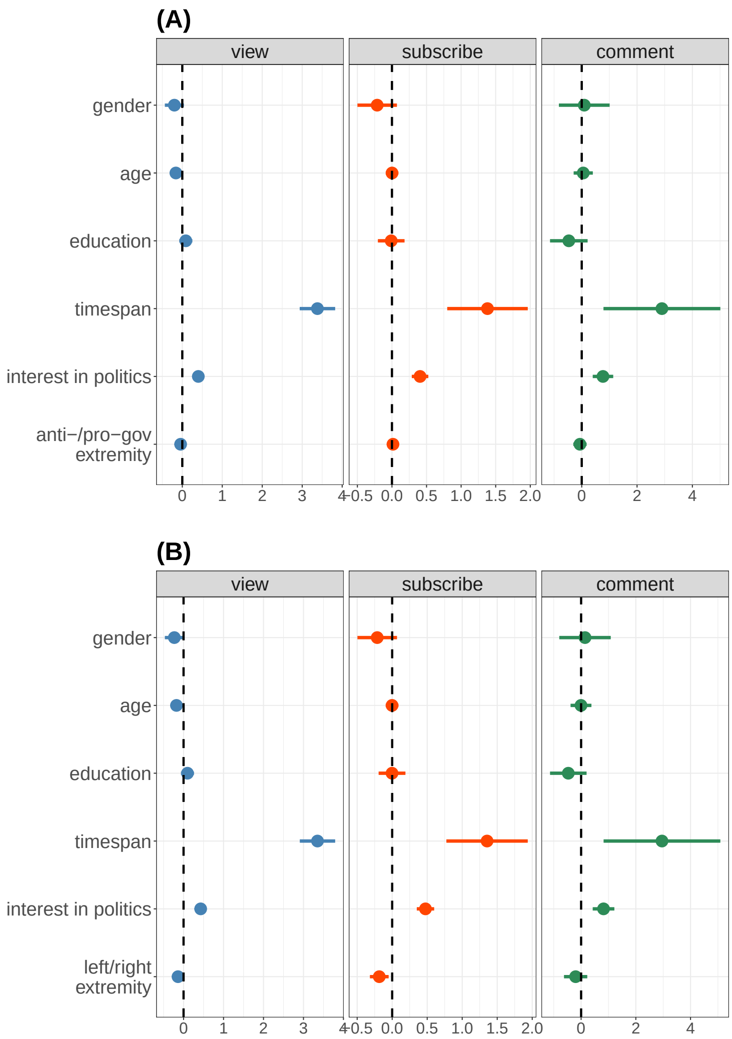

Figure C.4: Negative binomial regression results for viewing, subscribing and commenting activities on Hungarian political content. The DVs are the number of recorded activities for viewing, subscribing, and commenting from left to right for each panel. Panel A shows the model with the anti-/pro-government extremity as IVs, Panel B shows the model with the left/right extremity as IVs.

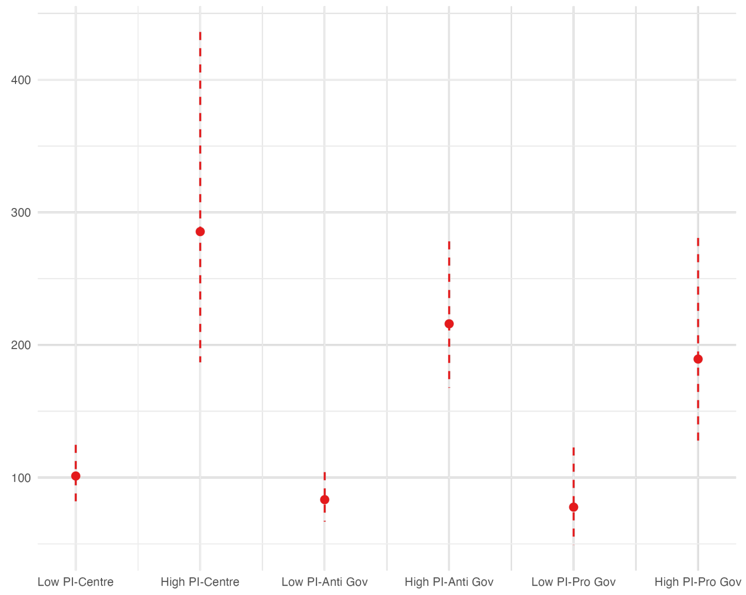

Figure C.5: Marginal prediction of the joint effect of political interest and anti/pro-government sympathy on the number of viewed political videos

C.3 Supplementary regression results with DVs breakdown by channel category

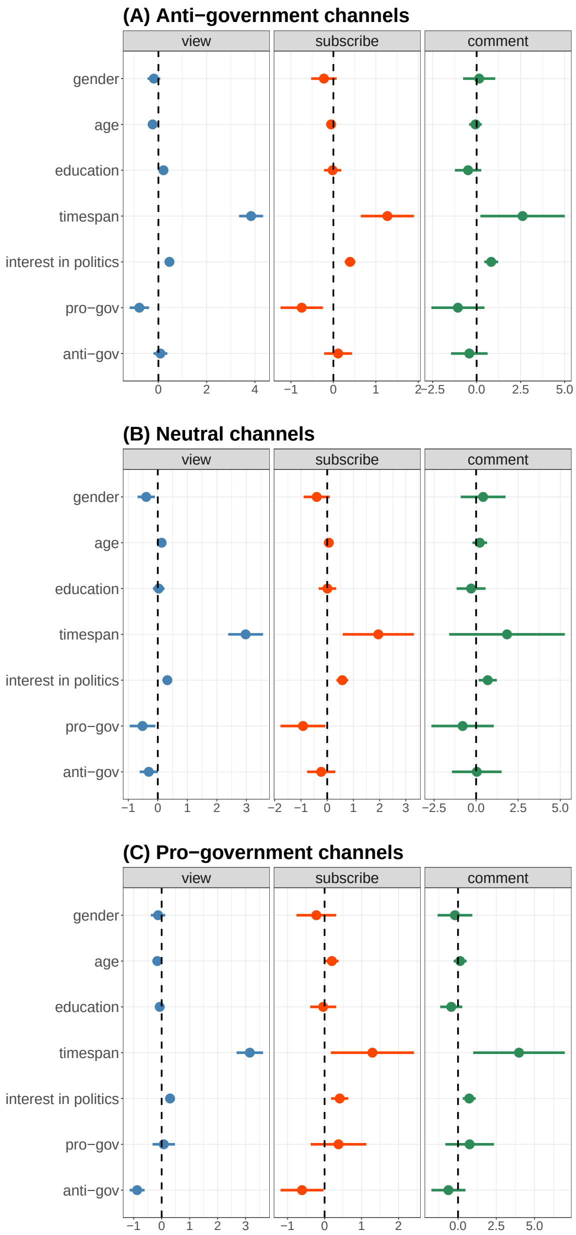

Figure C.6: Negative binomial regression results for viewing, subscribing and commenting activities on Hungarian political content, breaking down the channel categories, and using anti-/pro-government group belongings as IVs.

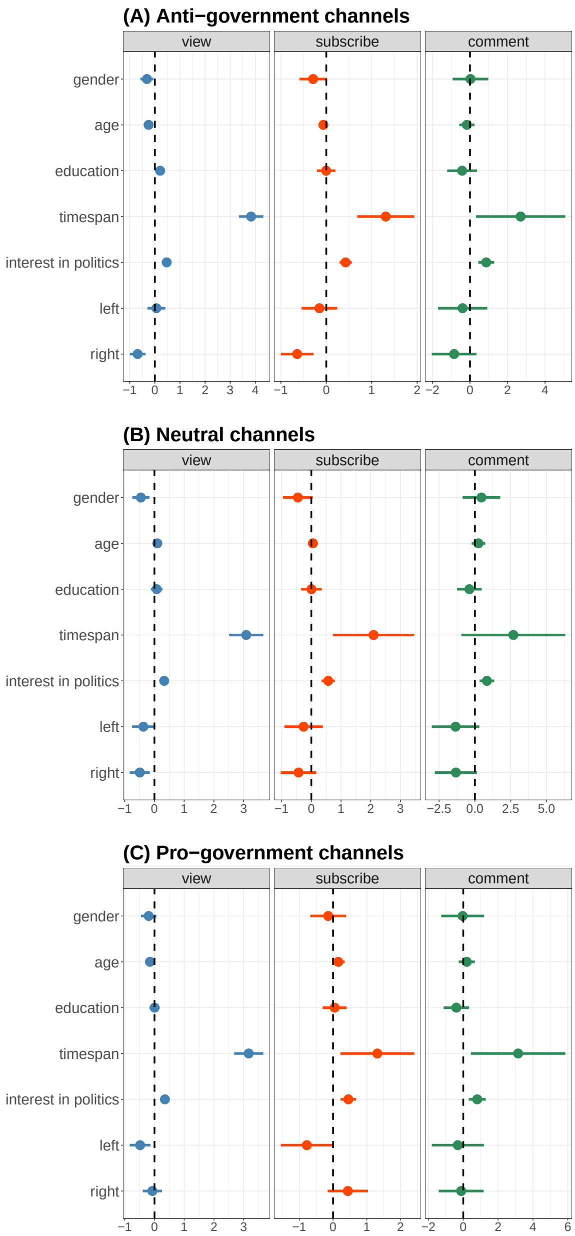

Figure C.7: Negative binomial regression results for viewing, subscribing and commenting activities on Hungarian political content, breaking down the channel categories, and using left/right group belongings as IVs.

C.4 Supplementary regression results with alternative filtering approaches.

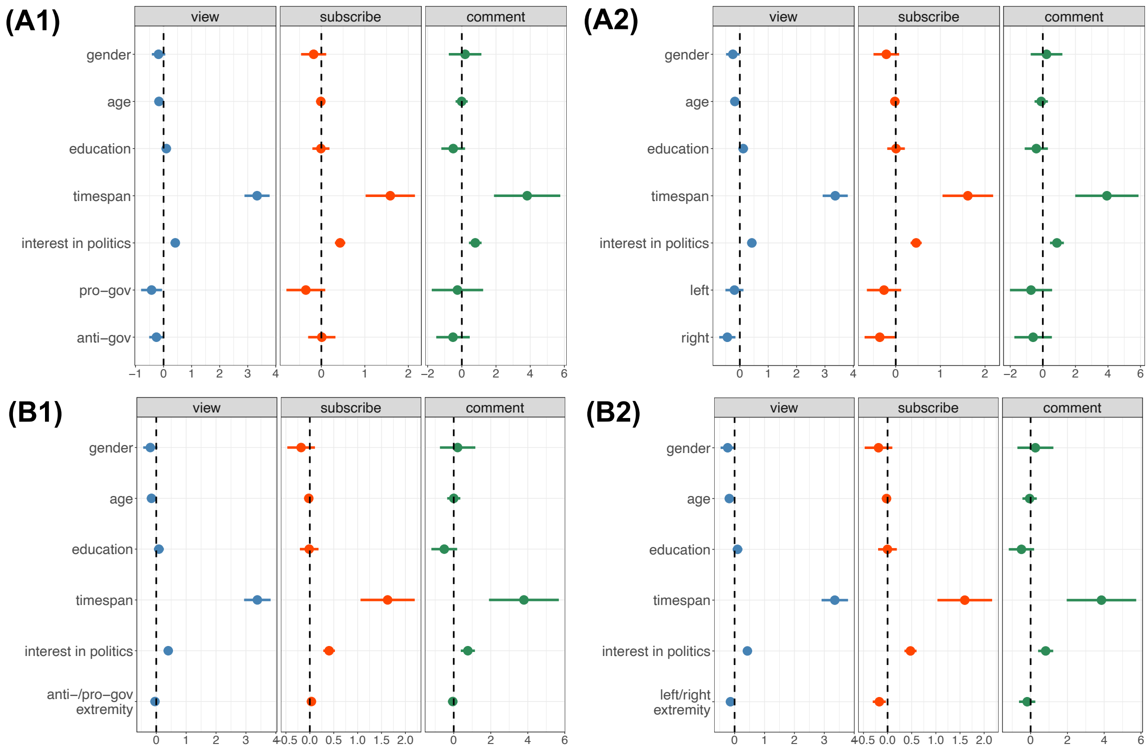

Figure C.8: Negative binomial regression results for viewing, subscribing and commenting activities on Hungarian political content, including respondents who have not engaged with any content on YouTube (i.e., from the group ALL in Table 4.1). Panel A shows the model with the position of anti-/pro-government (A1) and left/right (A2) as IVs. Panel B and B2 shows the model with the extremity of anti-/pro-government (B1) and left/right (B2) as IVs.

C.5 Imputing missing values for ideological variables and supplementary regression results with multiple imputation methods.

For the anti-/pro-government scale, 12% of the responses are Not Applicable / Don't Know (NA/DK). We do not think that NA would be anti-government due to self-censorship, as the average level of interest in politics for NA respondents (2.02) is lower than the average level of interest in politics for anti-government respondents (3.20), and relatively closer to the average level of interest for neutral respondents (2.69). Therefore, an intuitive approach to impute these missing values is to replace them with the central position with no strong position on either direction (i.e., 5 for the anti-/pro-government scale, and 4 for the left/right scale).

However, we also used a multiple imputation approach to make our results robust across different imputation methods. We included all the independent variables in the imputation process and used predictive mean matching (PMM). We created five imputed datasets and calculated the pool results in the regression models, using the "mice" R package. We highlight the results with the simple imputation in the main text and added that with the multiple imputation in this section. Here, instead of odds ratios, we use Average Marginal Effects (AME) to display the pooled regression results, as odds ratios can not be averaged over different regression runs.

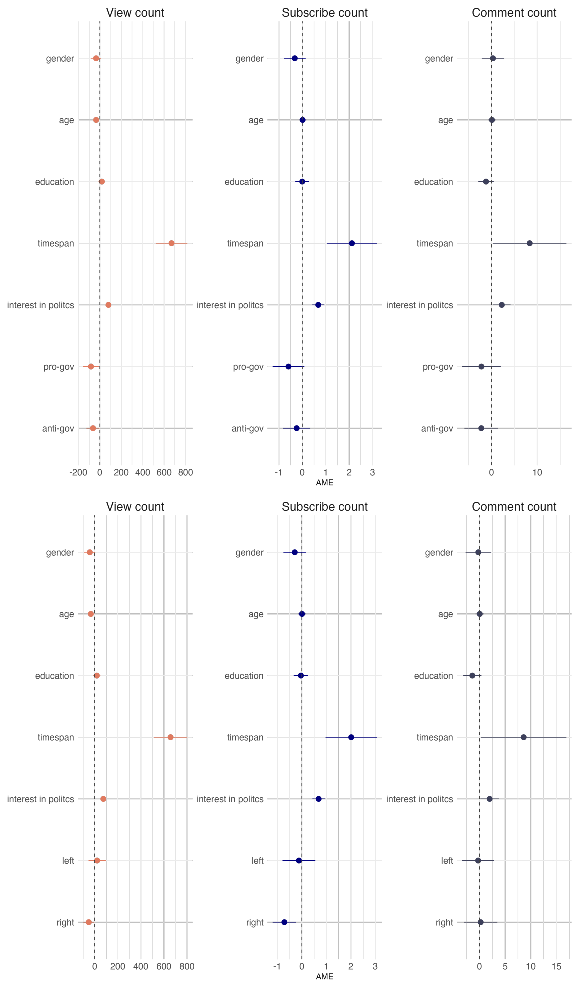

Figure C.9: Pooled regression results for multiple imputation displaying Average Marginal Effects (AME). The top subfigure shows the model with pro-/anti-government positions as IVs, and the bottom subfigure shows the model with left/right positions as IVs.

C.6 Data donation steps: recruitment and collection

We recruit participants through NRC[1], an online polling company, using their online access panel. Recruitment began with NRC sending an invitation email to potential participants. This email detailed the project's goals, outlined the incentives available, and provided a link to the project's webpage.

[1] www.nrc.hu

Initially, participants were asked to complete a brief online survey that included a consent form and screening questions to determine eligibility. This preliminary screening was designed to exclude individuals without regular Google and Facebook profiles. To proceed, participants needed to review a comprehensive description of the study and consent to participate by agreeing to the terms. Without this consent, they were unable to move forward in the process. Eligible participants were then directed to a webpage that included detailed instructions for exporting and downloading data from several platforms, including Google, Facebook, Instagram, Twitter/X, and TikTok. Once participants had downloaded their data, they uploaded the unaltered files to the project's website. Providing data from Google and Facebook was mandatory, while additional incentives were offered for data from Instagram, TikTok, and Twitter/X. After submitting their data, participants were required to complete a 30-minute questionnaire that covered a range of topics, including basic demographics, socioeconomic status, and more relevant to our study, political ideologies. This individually-linked dataset of survey responses and digital traces allows us to observe the sample from two perspectives–attitude and behavior, and to capture the bias present in studies that collect data *only* from online sources.

Sample bias is a crucial question in studies using user-donated data packages or web-tracking data, as the sampling process in these scenarios is usually less controlled and far more complicated than a standard survey. Previous studies have mixed results on how biased these samples were [250, 276, 277]. The variables of our primary interest, political ideology and level of interest in politics, correlate significantly with the decision to participate in [250], but not in all cases in [276].

Some of the participants in our data donation study (162 people) had also participated in a previous study with other online panel members who were also approached to participate in the donation study. With this sample of 769 people, we could test the difference between participating and non-participating panel members, as we had detailed information about them from the preliminary research. Comparing users who actually donate (N = 162) and users who are eligible to donate (N = 769), we find that neither political interest nor political ideology correlates significantly with the decision to participate (see Figure C.10). The most substantial factors here are demographic variables (i.e., education and age), for which we apply weighting to adjust these biases (see details in [277]).

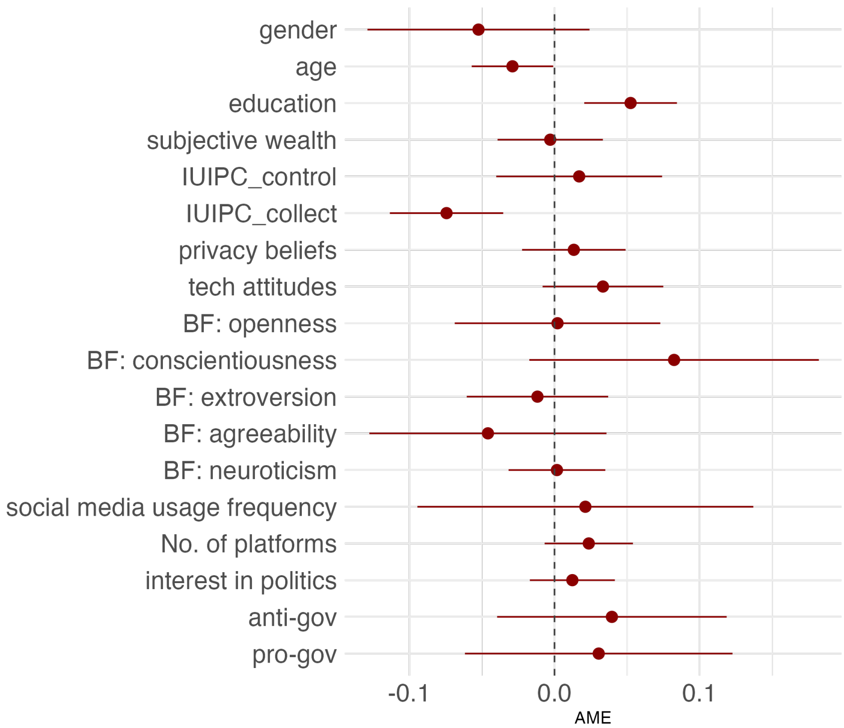

Figure C.10: Logistic regression model result displaying AME for predictors of actual data donation population.

C.7 Labeling Hungarian political channels

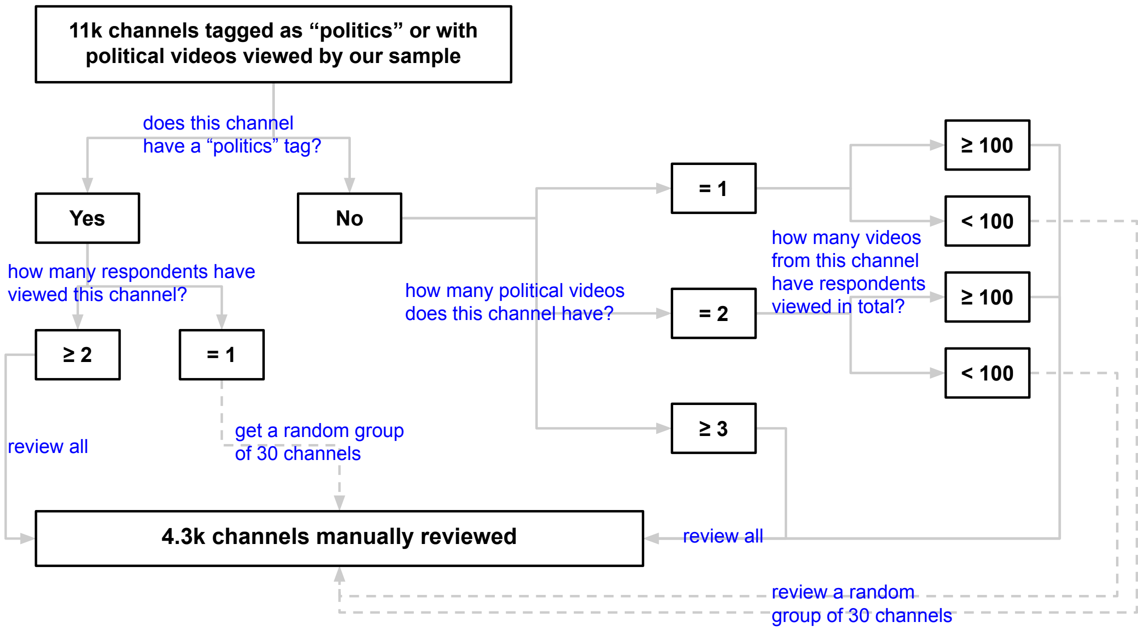

Figure C.11: Decision process of determining which relevant channels to manually review from the 11k YouTube channels that are tagged as "politics" or have uploaded political videos viewed by at least one respondent in our sample.

We rely on YouTube's topic tagging for channels and videos to have an initial set of political channels (i.e., channels with the topic tag "politics") and political videos (i.e., videos with the topic tag "politics"). Among the 11,065 channels that are tagged as "politics" or have uploaded political videos viewed by our sample, we narrow down to a smaller subset of channels for manual labeling. We perform three rounds of classification to include as much politically-relevant content as we can in the limited time frame. In each round, we go through the following channels that can be relevant (see a decision tree illustration in Figure C.11:

- Round 1: political channels with at least two viewers from our sample
- Round 2: channels that have uploaded at least three political videos viewed by our sample; and channels that have uploaded one or two political videos viewed by our sample and in total over 100 videos viewed by our sample

- Round 3: (from the rest unclassified channels) a random group of 30 channels with two political videos viewed by our sample (views ¡ 100); a random group of 30 channels with one political video viewed by our sample (views ¡ 100); and a random group of 30 channels with the “politics” tag and only one viewer from our sample

In total, we reviewed 4,270 channels, from which we identified 626 Hungarian political channels. Among these, 139 were classified as pro-government, 276 as anti-government and 149 as neutral since we focused on the anti-/pro-government scale, the main political and media cleavages in Hungary. We excluded extreme and conspiracy-focused channels that did not fit into the pro-government/anti-government dichotomy. This meant leaving out 62 channels, which typically had very low reach in our sample.

Two annotators performed the channel labeling. For validation, we first calculated the annotator agreement for the labeling process. We selected 400 channels[2] from the annotated ones. First, for classifying whether a channel is a Hungarian political channel or not, the annotator agreement was 0.79, and the kappa value[**?**] was 0.56. The cross-validation kappa value weighted by the number of videos watched was 0.67, so annotators were more likely to encode the same for larger, more watched channels. In addition, we examined the label match for Hungarian political channels regarding their anti-/pro-government leaning. The base kappa value for these channels was 0.42, but the weighted kappa value was 0.83. This difference between the unweighted and weighted kappa meant that the annotators gave the same labels for popular channels with a larger amount of audience, and they only struggled with smaller channels, which have relatively minimal impact on downstream analysis. This result is important to us from the validity side, as the bigger channels have more weight in the user-level analysis.

C.8 Comparing two distributions with unbalanced sample sizes

When we compare the distributions of user-level metrics (i.e., the mean and variance of the content ideology) and channel-level metrics (i.e., the EI index of nodes in channel networks), the sample sizes of two distributions can vary greatly. This makes the one-time comparison result in doubt, as the sample fluctuation within-group may produce spuriously significant results that overstate the between-group difference. In order to disentangle this within-group effect from our between-group comparison, we perform multiple rounds of KS-test between

[2]Hungarian political channels were under-represented among all channels. In order to be able to assess the coding accuracy of Hungarian channels in the coding evaluation, we over-represented these channels in the sample of 400 (200 such channels were included in the sample of 400).

bootstrapping samples and compare the between-group statistics against the within-group ones.

Let's say we want to compare the two distributions, d_1 and d_2, of sample size N_1 and N_2 respectively. In our case, usually $N_1 \neq N_2$, with one sample being much larger than the other. Therefore, to account for fluctuations in distribution due to small sample size, we perform 1,000 rounds of bootstrapping to assess whether the difference between two distributions are significantly larger than the fluctuations within each sample. To operationalize this, we first perform KS-test for d_1 and d_2, obtain the empirical test statistics T_e. We then bootstrap d_1 and d_2 with survey weights for 1,000 rounds; in each round, we obtain a bootstrapped sample d_1' (d_2') of size $N_3 = min(N_1, N_2)$, perform KS-test for d_1 and d_1' (d_2 and d_2') and obtain the test statistics T_1' (T_2'). With the distribution of T_1' and T_2', we can compute two p-values for T_e, p_1 and p_2, which indicate whether the difference between d_1 and d_2 is significantly larger than within-sample fluctuations of d_1 and d_2. Only when both p_1 and p_2 are below the significance threshold (e.g., 0.05), we can reject both null hypotheses and conclude that the difference between d_1 and d_2 is significantly larger than the sample fluctuations within d_1 and d_2.

Variable	Metric	Sample 1	Sample 2	U statistics 95% CI	p-value 95% CI	post-hoc power %
left/right	variance	ALL	YTB	[19106.4, 22668.3]	[0.02, 0.89]	10
left/right	variance	ALL	POL	[19650.6, 22991.0]	[0.01, 0.89]	21.5
left/right	**variance**	**ALL**	**POL-HU**	**[22733.3, 26180.1]**	**[0.0, 0.02]**	**97.5**
left/right	variance	YTB	POL	[18578.8, 22406.7]	[0.03, 0.92]	8
left/right	variance	YTB	POL-HU	[21357.3, 25349.2]	[0.0, 0.24]	88
left/right	variance	POL	POL-HU	[21012.9, 24889.7]	[0.0, 0.38]	72
left/right	variance	VIEW	SBSC	[19719.8, 23529.0]	[0.0, 0.89]	23.5
left/right	**variance**	**VIEW**	**CMNT**	**[3747.8, 6008.0]**	**[0.0, 0.0]**	**100**
left/right	**variance**	**SBSC**	**CMNT**	**[3839.1, 5996.3]**	**[0.0, 0.0]**	**100**
left/right	kurtosis	ALL	YTB	[16701.2, 20069.6]	[0.0, 0.8]	26.5
left/right	kurtosis	ALL	POL	[16538.7, 19973.6]	[0.0, 0.77]	27.5
left/right	**kurtosis**	**ALL**	**POL-HU**	**[12831.3, 15957.2]**	**[0.0, 0.0]**	**99.5**
left/right	kurtosis	YTB	POL	[18081.4, 21847.4]	[0.07, 0.95]	3.5
left/right	**kurtosis**	**YTB**	**POL-HU**	**[14162.6, 17496.6]**	**[0.0, 0.03]**	**96**
left/right	kurtosis	POL	POL-HU	[13997.4, 17885.9]	[0.0, 0.07]	93.5
left/right	**kurtosis**	**VIEW**	**SBSC**	**[25549.6, 29051.2]**	**[0.0, 0.0]**	**100**
left/right	**kurtosis**	**VIEW**	**CMNT**	**[34703.6, 36837.4]**	**[0.0, 0.0]**	**100**
left/right	**kurtosis**	**SBSC**	**CMNT**	**[32500.4, 34987.2]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	variance	ALL	YTB	[15545.6, 18945.7]	[0.0, 0.36]	66
anti-/pro-gov	variance	ALL	POL	[13905.6, 18097.1]	[0.0, 0.1]	93
anti-/pro-gov	**variance**	**ALL**	**POL-HU**	**[20601.6, 23912.8]**	**[0.0, 0.6]**	**50**
anti-/pro-gov	variance	YTB	POL	[16838.8, 20510.6]	[0.01, 0.89]	23
anti-/pro-gov	**variance**	**YTB**	**POL-HU**	**[23293.4, 26780.1]**	**[0.0, 0.0]**	**98**
anti-/pro-gov	**variance**	**POL**	**POL-HU**	**[24195.4, 27632.6]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	variance	VIEW	SBSC	[22189.4, 25894.4]	[0.0, 0.06]	93
anti-/pro-gov	**variance**	**VIEW**	**CMNT**	**[1089.6, 2504.6]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	**variance**	**SBSC**	**CMNT**	**[1085.6, 2319.4]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	kurtosis	ALL	YTB	[19865.3, 23623.4]	[0.0, 0.88]	32
anti-/pro-gov	kurtosis	ALL	POL	[18361.2, 22271.2]	[0.04, 0.94]	7.5
anti-/pro-gov	**kurtosis**	**ALL**	**POL-HU**	**[12770.8, 16268.9]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	kurtosis	YTB	POL	[16635.5, 20729.8]	[0.0, 0.84]	20
anti-/pro-gov	**kurtosis**	**YTB**	**POL-HU**	**[11342.6, 14931.4]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	**kurtosis**	**POL**	**POL-HU**	**[12485.8, 16440.7]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	**kurtosis**	**VIEW**	**SBSC**	**[4621.0, 7042.3]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	**kurtosis**	**VIEW**	**CMNT**	**[35405.4, 37324.0]**	**[0.0, 0.0]**	**100**
anti-/pro-gov	**kurtosis**	**SBSC**	**CMNT**	**[38050.0, 39189.2]**	**[0.0, 0.0]**	**100**

Table C.1: Results of Mann-Whitney U test that compares the distribution of variance and kurtosis yielded from 200 bootstrapped samples. We rerun the entire boostrapping and Mann-Whitney U testing for 200 rounds to obtain the 95% confidence intervals of U statistics and p values. In this table, each row reports the results of 200-round Mann-Whitney U tests for a polarization metric (e.g., variance) of certain ideological variable (e.g., anti-/pro-government) between two groups of respondents (i.e., sample 1 and sample 2). We highlight the rows with bolded texts when the difference is statistically significant, that is, when the entire 95% CI of p-value falls below the threshold of 0.05. We estimated post-hoc power by computing the proportion of 200 runs with a p-value below 0.05.

Metric	Sample 1	N1	Sample 2	N2	Bootstrapped size	p1	p2
mean	viewers	640	subscribers	299	299	0.000	0.000
mean	viewers	640	commenters	72	72	0.000	0.000
mean	subscribers	299	commenters	72	72	0.004	0.006
variance	viewers	640	subscribers	299	299	0.000	0.000
variance	viewers	640	commenters	72	72	0.000	0.000
variance	subscribers	299	commenters	72	72	0.000	0.002

Table C.2: Results of Kolmogorov–Smirnov test to compare the distribution of mean and variance of content ideology consumed by viewers, subscribers and commenters. A post-hoc power analysis based on 200 repeated runs revealed 100% power, with all p-values below 0.05.

Sample 1	N1	Sample 2	N2	Boostrapped size	p1	p2	Post-hoc power (%)
View	545	Comment	105	105	0.000	0.000	
Subscribe	197	Comment	105	105	0.000	0.000	

Table C.3: Results of Kolmogorov–Smirnov test to compare the distribution of node-level EI index across channel networks based on visible (i.e., commenting) and invisible (i.e., viewing and subscribing) engagement forms. A post-hoc power analysis based on 200 repeated runs revealed 100% power, with all p-values below 0.05.

C.9 Descriptive statistics for main variables

Activity	N	N (at least 1 activity)	Mean	Median	SE
Viewing	690*	690	9998	4700	477
Subscribing	735	680	57.9	15	4.8
Commenting	735	314	20.2	0	4.7
Viewing political channel	690*	640	192.7	43.5	18
Subscribing political channel	735	299	1.5	0	0.2
Commenting political channel	735	72	1.4	0	0.4
Timespan (2018 May - 2023 May)	707**	-	49.1	59	0.7

* People without any viewing likely turned down tracking of viewing activity or deleted their view data.
** People with only subscription activity do not have timespan data, as YouTube does not timestamp subscriptions.

Table C.4: Descriptive statistics for main variables.

Category	Viewing	Subscribing	Commenting	Timespan
Gender				
Male	10548 (806)	50.8 (5.8)	30.2 (10.1)	50.5 (1.0)
Female	9501 (575)	64.3 (7.4)	11.3 (2.9)	47.8 (0.9)
Age				
Below 35	15524 (967)	63.1 (6.49)	33.6 (13.9)	47.2 (1.3)
35-49	10782 (788)	64 (7.3)	16.3 (4.4)	52.2 (0.9)
50-64	5428 (784)	33.2 (6.4)	10.6 (5.1)	48 (1.6)
65+	3986 (640)	72.4 (27.9)	17.9 (6.8)	47 (2.4)
Education				
Low	10778 (1175)	73.7 (11.1)	33.8 (18.4)	46.1 (1.8)
Middle	11378 (763)	61.6 (8.9)	21.4 (4.6)	50.5 (1.0)
High	7498 (655)	37.1 (5.3)	5.2 (0.9)	50.4 (1.0)

Table C.5: Descriptive statistics–mean and (standard error)–for viewing, subscribing, commenting, and timespan by gender, age, and education.

Category	Pol. Viewing	Pol. Subscribing	Pol. Commenting
Gender			
Male	252 (33.2)	1.9 (0.3)	1.4 (0.4)
Female	139 (18.8)	1.2 (0.2)	1.4 (0.6)
Age			
Below 35	235 (33.8)	1.4 (0.2)	1.2 (0.4)
35-49	188 (29.0)	1.4 (0.2)	1.1 (0.4)
50-64	135 (22.5)	1.2 (0.2)	1.3 (1.1)
65+	211 (80.9)	2.5 (0.9)	2.8 (2.2)
Education			
Low	128 (29.1)	1.0 (0.2)	0.8 (0.3)
Middle	245 (32.3)	2.1 (0.4)	2.8 (1.0)
High	196 (29.7)	1.3 (0.2)	0.3 (0.1)

Table C.6: Descriptive Statistics–mean and (standard error)–for political viewing, subscribing, commenting, and timespan by gender, age, and education

C.10 Selective exposure for the same sample group

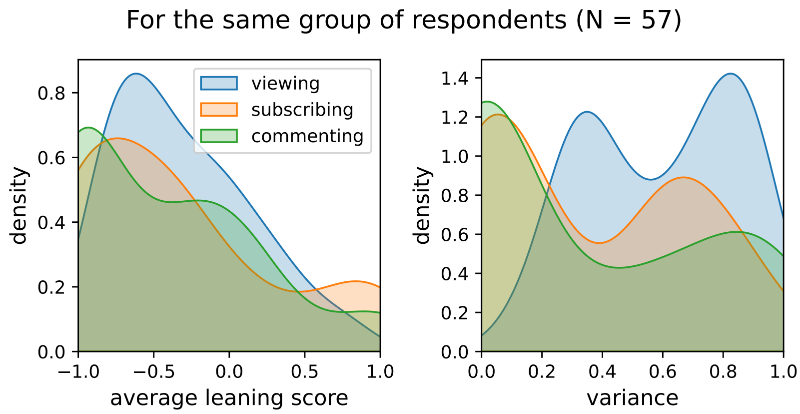

Figure C.12: Average leaning score (left) and the leaning variance (right) for the Hungarian political content the respondents have engaged with through viewing, subscribing or commenting. Here we present the result generated from the same group of respondents who have engaged in all three forms. An average leaning score of -1 (+1) means consuming only anti-government (pro-government) content, and 0 means consuming balanced or neutral content. Density functions are generated by kernel density estimate (KDE) methods.