A Framework for the Unsupervised Modeling and Extraction of Polarization Knowledge from News Media

DEMETRIS PASCHALIDES, Computer Science, University of Cyprus, Nicosia, Cyprus
GEORGE PALLIS, Computer Science, University of Cyprus, Nicosia, Cyprus
MARIOS DIKAIAKOS, Computer Science, University of Cyprus, Nicosia, Cyprus

Polarization poses global concerns for social cohesion and stability, making its understanding crucial for effective mitigation measures. In this paper, we introduce an unsupervised, domain-agnostic framework for computationally modeling, extracting, and measuring polarization in digital media. By leveraging Natural Language Processing and Graph Analysis techniques, the proposed framework creates a Polarization Data Model (PDM) that encompasses key elements of Polarization Knowledge (PK), such as entities, fellowships, dipoles, and discussion topics. To evaluate the effectiveness of the framework, we propose a multi-level PK evaluation methodology that assesses its ability to: (i) capture entities' attitudes toward various topics, (ii) align politically cohesive fellowships with their respective party manifestos, and (iii) identify domain-specific topics along with their degree of polarization. We applied this evaluation methodology to the use cases of Abortion, Immigration, and Gun Control. The results demonstrate our framework's robust performance across these case studies, yielding promising outcomes compared to state-of-the-art and baseline methods.

CCS Concepts: • **Computing methodologies** → **Information extraction**; **Semantic networks**; • **Networks** → *Social media networks*; • **Information systems** → *Sentiment analysis*; *Clustering and classification;*

Additional Key Words and Phrases: Polarization, Multi-level Polarization, Polarization Modeling, Polarization Extraction, Polarization Computational Evaluation

ACM Reference Format:
Demetris Paschalides, George Pallis, and Marios Dikaiakos. 2025. A Framework for the Unsupervised Modeling and Extraction of Polarization Knowledge from News Media. *ACM Trans. Soc. Comput.* 8, 1-2, Article 5 (January 2025), 38 pages. https://doi.org/10.1145/3703594

1 Introduction

Polarization is a growing concern globally, with potentially severe implications for social cohesion and stability [7]. It can be observed in the way recent events and public debates unfold and take shape, for example during the 2016 US presidential elections [72], the refugee crisis in Germany [78], the Brexit referendum [73], the storming of the US Capitol in 2021 [15], and the COVID-19 pandemic [70]. The increasing prevalence of polarization is evident in how individuals and

This research is funded in part by the EU Commission via the ATHENA 101132686 project (HORIZON-CL2-2023-DEMOCRACY-01).
Authors' Contact Information: Demetris Paschalides, Computer Science, University of Cyprus, Nicosia, Nicosia, Cyprus; e-mail: paschalides.demetris@ucy.ac.cy; George Pallis, Computer Science, University of Cyprus, Nicosia, Nicosia, Cyprus; e-mail: pallis@ucy.ac.cy; Marios Dikaiakos, Computer Science, University of Cyprus, Nicosia, Nicosia, Cyprus; e-mail: mdd@ucy.ac.cy.

ACM 2469-7818/2025/01-ART5
https://doi.org/10.1145/3703594

groups are becoming more entrenched in their beliefs and opinions, often resulting in heightened hostility towards those holding different views. Polarization can result in a variety of negative outcomes, such as a decrease in social trust [48], a rise in extremist ideologies, and a decline in democratic institutions [6, 42]. Therefore, understanding the causes and dynamics of polarization is crucial for maintaining the stability of modern democracies.

With the rise of digital platforms, polarization is increasingly manifested in online content. This phenomenon is closely linked to the spread of misinformation, hate speech, and disinformation, particularly on social media [59, 68]. These challenges create opportunities for computational approaches, particularly through the use of **Natural Language Processing (NLP)**, to analyze and monitor polarization effectively. However, the vast scale, unstructured nature, and inherent noise in online data make it difficult to accurately measure and model polarization [31].

To address this, we present an open-source computational framework designed to identify polarization as it emerges within the narratives of news media. This framework leverages an unsupervised pipeline that combines NLP techniques with graph analysis to model and extract polarization effectively. News articles are a valuable resource for analyzing societal polarization as they often report on contentious issues, highlighting conflicting viewpoints and the stances of various entities [34]. Analyzing the narratives in news media can reveal how polarization emerges in public discourse, making news articles a valuable resource for its study.

A core component of this framework is the **Polarization Data Model (PDM)**, structured as a **Knowledge Graph (KG)** to represent key elements such as entities (e.g., individuals, organizations), their attitudes, fellowships (sub-groups with shared views), dipoles (opposing fellowships), and polarizing topics. This structured representation allows for a detailed and dynamic understanding of these elements and their interconnections, providing insights into polarization mechanisms. Our framework integrates both content and structural analysis—combining network-based relationships and content-based attitude and topic analysis—to offer a more comprehensive view of polarization.

A major challenge in polarization research is evaluating the accuracy and robustness of the extracted knowledge due to the complexity of political discourse and the scarcity of annotated ground truth data. To tackle this, we introduce a multi-level evaluation methodology that assesses polarization at the entity, fellowship, and topic levels. Our evaluation compares the results of our framework against ground truth data, baseline models, and state-of-the-art methods, ensuring a comprehensive assessment of its effectiveness. We demonstrate the utility of our approach through case studies in three key domains: Abortion, Immigration, and Gun Control.

Our approach offers researchers a tool for analyzing polarization that goes beyond existing methods. By integrating entities, attitudes, and polarizing topics into a unified KG, we provide a granular view of the mechanisms driving polarization. This enables deeper insights into the affiliations and oppositions among entities, as well as the specific topics that underlie these relationships, which are often overlooked in traditional analyses [33].

The rest of the paper is structured as follows: Section 2 provides an overview of polarization and related computational studies. Section 3 details the Polarization Data Model and the proposed framework. Section 4 outlines the method used to evaluate the polarization knowledge, along with the datasets and results. The paper concludes in Section 5.

2 Background and Related Work

2.1 Defining and Theorizing Polarization

Polarization is a multifaceted phenomenon by which a society or group divides into distinct sub-groups with conflicting positions, attitudes, and beliefs on various topics, leading to an increasingly wide gap between these factions over time [26, 29, 44, 74]. Ideological polarization involves

the growing divergence in political or social beliefs and policy preferences among individuals or groups [1, 52]. Affective polarization, on the other hand, captures the emotional hostility between opposing groups, even when ideological differences may be minor [43, 44]. Some studies view affective polarization as distinct from ideological polarization, suggesting that emotional hostility can intensify even without increasing ideological extremity [43], while others argue that as ideological divides grow, they reinforce affective polarization, leading to greater emotional animosity [46].

Building on these definitions of ideological and affective polarization, we observe that polarization manifests and intensifies through interconnected processes at both the individual and collective levels, illustrating a dynamic interplay between personal beliefs and group identities. At the individual level, polarization influences how people perceive, interpret, and respond to information, shaping their attitudes toward others. At the collective level, individuals align with like-minded peers, forming sub-groups with distinct and often opposing stances. This alignment strengthens in-group identities and exacerbates divisions with out-groups, widening the societal gap [46, 75].

Several psychological mechanisms drive polarization, contributing to both its ideological and affective forms. Motivated reasoning and confirmation bias play key roles by reinforcing individuals' existing beliefs while disregarding contradictory evidence [16, 56, 87]. Group dynamics, such as social reinforcement and persuasive argumentation, further push group members toward adopting more extreme positions [12, 35, 74]. Naïve realism, where people believe their own perspective is the absolute truth, exacerbates affective polarization by deepening distrust and hostility toward opposing groups [64]. Tribalism intensifies this divide by fostering strong loyalty to in-groups and antagonism toward out-groups, heightening social divisions [13].

These processes align with social identity theory, where group membership shapes an individual's self-concept [75]. Strong identification with a group, such as a political party, increases loyalty to the in-group and heightens negative attitudes toward the out-group [44]. Media dynamics and partisan communication amplify these divides by reinforcing group narratives and framing opposing groups negatively [11, 43]. Inter-group conflict theory offers a framework for understanding these dynamics [75]. It outlines three key processes:

(1) **Social Categorization**: Individuals categorize themselves and others into groups (e.g., "*us*" vs. "*them*").
(2) **Social Identity**: People adopt the identity of their categorized group, internalizing its norms and values.
(3) **Social Comparison**: Group members compare their group to others, often viewing theirs more favorably.

By applying the inter-group conflict theory, we can see how the processes of social categorization, identification, and comparison contribute to both ideological and affective polarization. As individuals align with their in-group, they reinforce shared beliefs and grow more hostile toward out-groups [46]. This process occurs at both individual and collective levels, creating a feedback loop that deepens polarization. Understanding these dynamics highlights the need to model polarization by capturing the interplay between personal attitudes, group identities, and divisive topics.

2.2 Polarization Modeling, Extraction, and Quantification

2.2.1 Computational Properties of Polarization. Computational approaches have become increasingly important in capturing various aspects of polarization, providing nuanced insights into how it develops and intensifies within societies [46]. Existing computational studies on polarization can be distinguished by the definitions they adopt and the properties they incorporate into their methodologies. Key characteristics of computational frameworks for modeling polarization include:

Polarization Definition: Computational approaches operationalize polarization in different ways, typically focusing on either structural polarization or content-based polarization. Structural Polarization represents the separation of groups within a network. Studies adopting this definition analyze the topology of social networks to identify distinct clusters or communities that correspond to polarized groups [2, 5, 31, 32, 37]. Methods used include measuring network properties such as frustration [8], random walks [31], betweenness [31], and community boundaries [31, 37]. While these studies may not explicitly address affective polarization, the detection of group separations and limited interactions between groups can be indicative of social distance and mutual hostility, which are key aspects of affective polarization [22]. Content-based Polarization focuses on the divergence in opinions, attitudes, and sentiments expressed in textual content, capturing differences in beliefs and perspectives [10, 24, 40, 54, 65, 84]. Techniques include topic modeling to identify key issues driving polarization [40, 65] and sentiment analysis to gauge emotional tones [10, 53, 54, 84]. Although the primary focus might be on content divergence, these methods can provide insights into ideological polarization by highlighting divergent topics and stances.

Data Format: The data used for modeling polarization can be *structured* or *unstructured*, influencing the methods applied. Structured data, like social media metadata (e.g., user mentions, hashtags, and follower relationships) or tabular data (e.g., U.S. Congress votes), facilitates network-based analysis of structural polarization [5, 22, 31]. Examples include constructing mention or retweet networks from X (formerly known as Twitter) data or hyperlink networks from HTML pages [2, 37]. In contrast, unstructured data, such as news articles and opinion pieces, require NLP techniques to extract meaningful insights, capturing content-based polarization through expressed opinions and sentiments [53, 65]. Depending on the format—network or text—the analytical approach differs. Networks, derived from structured data, are used for relationship analysis, while text data enables NLP methods like topic modeling and sentiment analysis to reveal polarized views and emotional expressions [10, 40].

Group Definition: The method of defining groups within the data is critical for analyzing polarization. These groups can be either *pre-defined*, through the use of *manual labeling* [2, 3, 5, 33, 40, 53, 65] or *automated labeling* [24, 32, 84], or *partitioned* using unsupervised methods such as community detection [22, 31, 37, 54, 63]. A significant number of studies apply pre-defined political groups, such as Democrats and Republicans, on the political spectrum between Left and Right. This approach often leads to a conflation of group identity with specific ideological stances, as these parties are typically emblematic of broader political ideologies — Democrats aligning with left-leaning (Liberal) views and Republicans with right-leaning (Conservative) views.

To demonstrate how these computational properties are applied in existing research, we present a selection of studies in Table 1, with each study characterized according to the specified properties. In the following subsections, we delve deeper into these works, categorizing them based on their approaches to polarization modeling and extraction—specifically, structural and content-based methods. We discuss how each study models polarization, the techniques they use for extraction, and metrics employed for polarization quantification. This detailed analysis allow us to highlight the strengths and limitations of existing approaches, particularly in relation to capturing both ideological and affective dimensions of polarization across multiple levels.

2.2.2 Polarization Modeling and Extraction. Existing computational studies on modeling and extracting polarization leverage various computational properties to capture the dynamics of group divisions and opposing viewpoints. These approaches can be broadly categorized into two main types: *structural* and *content-based* methods.

Table 1. Overview of the Properties of the State-of-the-art Computational Polarization Methods

Work	Input Data Type	Data Format		Polarization Definition		Group Definition	
		Network	Text	Structural	Content	Pre-Defined	Partitioned
Adamic and Glance 2005 [2]	Structured	✓	–	✓	–	✓	–
Conover et al. 2011 [22]	Structured	✓	–	✓	–	–	✓
Waugh et al. 2011 [82]	Structured	✓	–	✓	–	–	✓
Balasubramanyan et al. 2012 [10]	Unstructured	–	✓	–	✓	✓	–
Weber et al. 2013 [84]	Structured	✓	–	–	✓	✓	–
Guerra et al. 2013 [37]	Structured	✓	–	✓	–	–	✓
Mejova et al 2014 [53]	Unstructured	–	✓	–	✓	✓	–
Akoglu 2014 [3]	Structured	✓	–	✓	–	✓	–
Garimella et al. 2015 [31]	Structured	✓	–	✓	–	–	✓
Morales et al. 2015 [54]	Structured	✓	–	–	✓	✓	–
Andris et al. 2015 [5]	Structured	✓	–	✓	–	✓	–
Garimella and Weber 2017 [33]	Structured	✓	–	✓	–	✓	–
Demszky et al. 2018 [24]	Structured	–	✓	–	✓	✓	–
Roy and Goldwasser 2020 [65]	Unstructured	–	✓	–	✓	✓	–
Yun Chen et al. 2021 [19]	Structured	✓	–	✓	–	✓	–
He et al. 2021 [40]	Unstructured	–	✓	–	✓	✓	–
Garimella et al. 2022 [32]	Structured	✓	–	✓	–	✓	✓
Sinno et al. 2022 [71]	Unstructured	–	✓	–	✓	✓	–
Ramaciotti Morales et al. 2023 [63]	Structured	✓	–	✓	–	–	✓
Our Framework	Unstructured	✓	✓	✓	✓	–	✓

Our framework is denoted with bold and italic lettering.

Structural approaches focus on analyzing network interactions among users, such as retweets, mentions, hyperlinks, or shared hashtags, to model how individuals cluster into polarized groups. [2] examine political polarization during the 2004 US Presidential Election by modeling a hyperlink network between Liberal and Conservative blogs, quantifying polarization based on their separation. [31] present a method for identifying controversial topics on social media by combining network analysis with message content, partitioning networks of conversations to measure polarization. Similarly, [33] use X data to track polarization over time by constructing networks of interactions through retweets and hashtags, while [3] and [5] apply network analysis to roll-call votes in the US Congress, highlighting divisions between Liberal and Conservative groups. In contrast, [22] and [37] use partitioning algorithms to detect conflicting subgroups in X networks, capturing polarization without predefined group labels. More recent work, such as [63] and [32], utilized network embeddings and co-browsing graphs to explore ideological polarization and polarized news consumption in social media and online news platforms, respectively, further illustrating how structural methods reveal patterns of division in digital spaces.

Content-based approaches focus on analyzing textual data, such as social media posts, news articles, and online discourse, to capture ideological divisions based on opinions, sentiments, and language use. [53] examine how Liberal and Conservative news outlets portray controversial topics, using sentiment analysis to compare emotional expression and biased language, thus identifying ideological polarization. [65] introduce a semi-supervised approach to detect nuanced subframes in news articles, embedding Liberal and Conservative coverage into a shared space to capture ideological differences more effectively. [40] present a method for detecting polarized topics using PaCTE, which measures polarization through the cosine distance between partisan news sources' contextualized topic embeddings. [10] propose MCR-LDA to model how political topics evoke different responses from sub-communities, emphasizing content-based polarization. [24] build on this with a topic modeling method that identifies salient topics independent of specific events, while [71] analyze annotated media outlets, revealing ideological divergence based on word-level language use.

2.2.3 Polarization Quantification. Polarization quantification methods vary depending on the computational properties being modeled, whether structural or content-based, and are often designed to capture polarization as it manifests in different contexts. Structural metrics are predominantly employed to measure how groups divide within networks, while content-based approaches capture ideological or affective divisions reflected in textual data [10, 24, 40].

Structural Metrics: A common structural polarization measure is modularity, which assesses how well a network is divided into communities [55]. Modularity calculates the strength of division between groups within a network [2, 22]. A higher modularity score typically indicates stronger polarization, as it reflects dense intra-group connections and sparse inter-group ties. However, modularity has been shown to suffer from a resolution limit, making the detection of small communities difficult [30], thus, deeming modularity an indirect polarization measure [37]. Beyond modularity, other structural metrics offer deeper insights into network polarization. The E-I Index [47] measures the ratio of inter- to intra-group ties, helpful in evaluating group isolation. Garimella and Weber 2017 [33] propose **Random Walk Controversy (RWC)**, which assesses echo chambers by measuring the likelihood of a random walk staying within the same group. Salloum et al. 2022 [67] adapt these metrics with the Adaptive E-I Index and Adaptive RWC, adjusting for group size and degree distribution, making them more accurate in networks with uneven group sizes. Metrics like **Betweenness Centrality Controversy (BCC)** and **Boundary Polarization (BP)** focus on boundary nodes connecting communities [31, 37]. BCC assesses how few central individuals bridge groups, while BP measures boundary node distribution, with fewer boundary nodes indicating higher polarization [33]. [54] introduce the Polarization Index to quantify polarization between two distinct communities, such as Left vs. Right. It measures ideological distance and interaction frequency, with greater separation and lower interaction indicating higher polarization, making it particularly useful in political contexts to capture division and engagement between opposing sides.

Content-based Metrics: While structural approaches are valuable, they often overlook the ideological and affective dimensions of polarization that emerge in the content shared between groups. [10] address this gap with a content-based approach that combines sentiment analysis and topic modeling to gauge the degree of polarization within political discourse on platforms like Reddit [10]. This approach reveals not just the structural separation of groups but also the divergence in the content of their discussions. [40] models ideological polarization in news articles using PaCTE, calculating the cosine distance between topic embeddings in order to quantify how differently polarized groups frame the same issues.

The reliance on framing, sentiment, and ideological embeddings in these methods offers a more direct measure of how polarized groups differ not only in structure but in the actual content of their communications. By focusing on textual data, content-based metrics provide insights into how polarization manifests through both ideological disagreements and emotional hostility, capturing the affective aspect that structural measures may miss.

2.2.4 Comparison and Discussion. Reviewing the computational approaches to polarization modeling, extraction, and quantification reveals that most existing methods conceptualize polarization primarily as a structural phenomenon. These methods typically frame polarization as the opposition between predefined groups—such as Liberals and Conservatives—by focusing on network structures derived from social interactions. While structural approaches have provided valuable insights into how polarization manifests within networks, they often exhibit limitations, particularly in capturing the full spectrum of polarization, which includes both ideological and affective dimensions [46].

A major limitation of structural methods is their reliance on predefined groups [31, 60]. By assuming specific polarized groups *a-priori*, these approaches constrain their ability to detect emergent or nuanced forms of polarization that may not align with established categories. This reliance can prevent the identification of evolving group dynamics and obscure the complexity of polarization in different contexts [31, 63]. Furthermore, structural methods often offer limited consideration of ideological content. By focusing primarily on patterns of connections within a network, these approaches may overlook the ideological differences expressed through language and discourse. As a result, they may inadequately capture ideological polarization—the divergence in beliefs and policy preferences—which is essential for fully understanding the extent and nature of divisions within a polarized society.

Conversely, content-based approaches focus on the textual content shared between groups, analyzing language, sentiments, and topics to capture ideological polarization. While these methods adeptly reveal differences in beliefs and attitudes, they often overlook affective structures that reinforce polarization [40, 65]. By neglecting the network of interactions among individuals, content-based methods may miss how social relationships and group affiliations contribute to affective polarization—the emotional and social distances between groups. Additionally, content-based methods typically offer limited integration with interaction dynamics. Without considering how individuals interact within a network, these methods may not fully capture the interplay between ideology and social relationships that drive polarization. This lack of integration reduces their ability to explain how polarized opinions spread and intensify through social and group influence.

Both structural and content-based methods often employ supervised approaches, where the research environment is controlled, and polarization is predetermined. This supervised nature limits their ability to capture the nuances and complexities of polarization, especially in different domains and contexts where new forms of polarization might emerge. To address these limitations, it is essential to establish unsupervised and domain-independent methods for modeling and extracting polarization. Such methods should be capable of capturing both structural and content-based aspects of polarization, integrating ideological and affective dimensions across multiple levels. To this end, we propose an unsupervised and domain-independent framework for modeling and extracting polarization knowledge, based on the inter-group conflict theory [75]. This framework is designed to process large volumes of news articles and generate comprehensive polarization knowledge using NLP and graph analysis methods. A core component of the proposed framework is the **Polarization Data Model (PDM)**, which offers a structured representation of polarization, encapsulating key entities, their fellowships, fellowship dipoles, and primary polarizing topics.

This work builds upon and extends our previous POLAR system, initially developed for the unsupervised extraction of polarizing topics from news articles [60]. Key extensions include a more detailed elaboration of the methodology, and an expansion of the PDM to adopt a more structured format. A significant extension in this work relates to the evaluation methodology. In our earlier work, the evaluation was limited to measuring topic-level polarization, primarily due to the inherent challenges of assessing polarization. In this paper, we address those challenges by proposing a multi-level evaluation framework that assesses polarization across entities, fellowships, and topics. This new evaluation method allows us to comprehensively benchmark the framework's performance, comparing extracted polarization knowledge against ground truth data, baseline models, and state-of-the-art approaches. By enhancing both the PDM and the evaluation framework, we provide a more robust tool for studying polarization across multiple dimensions.

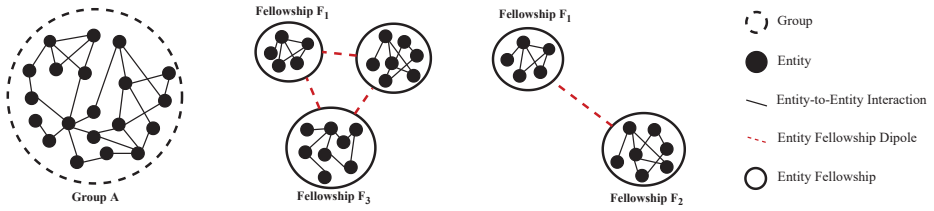


Fig. 1. An illustration of the process which a group of entities, namely *Group A* (see Figure (a)), can be segregated into fellowships F_1, F_2, and F_3, which collide with each other (see Figure (b)), thus forming fellowship dipoles, similar to D_{12} (see Figure (c)). Figure (d) represents the legend for the different structures in Figures (a), (b), and (c).

3 Modeling and Extracting Polarization Knowledge

3.1 Modeling the Polarization Phenomenon

In our study, we adopt the definition of polarization as outlined by Sunstein 1999 [74], which characterizes it as a "*social process where a social or political group becomes divided into two or more opposing sub-groups with conflicting beliefs*". Dissecting this definition, the term "*group*" refers to a set of two or more entities that relate to one another, share common characteristics and have a collective sense of unity [80]. A relationship is defined as a number of recurring interactions between two or more *entities*, and is considered as the basis of various social structures (e.g., groups). Despite the commonness of a group, inter-group conflict [75] results to the division of the group into two or more conflicting sub-groups i.e., *fellowships*. Similarly, the conflict between a pair of fellowships can be viewed as a *dipole*. An illustrative example of this process is depicted in Figure 1.

3.2 Polarization Data Model

Computationally, we define the Polarization Data Model (PDM) as a heterogeneous, directed, and weighted knowledge graph $G(V, E)$, where each vertex $v \in V$ and each edge $e \in E$ is characterized by a type $\tau(v) : V \rightarrow \{Entity, Fellowship, Topic\}$ and $\lambda(e) : E \rightarrow \{Relationship, Member, Attitude, Conflict\}$. A pair of *Entity* type vertices v_i and v_j can be connected via a bi-directional edge $v_i \leftrightarrow v_j$ of type *Relationship*, representing an interaction between them. A *Relationship* is characterized by the weight w_{ij}, which indicates the status of the relation between v_i and v_j (i.e., *positive* or *negative*), with values ranging from -1 (extremely negative) to 1 (extremely positive). The value of w_{ij} is determined by the attitude of one entity towards the other, and vice versa. The subgraph of $PDM[v_i \in V \mid \tau(v_i) = Entity]$ is defined as a ***Sentiment Attitude Graph* (SAG)**. An entity fellowship is identified as a dense subgraph of SAG, which contains predominantly positive relationships among its vertices. For this information to exist within the PDM, the vertex type *Fellowship* is defined. Hence, a vertex v_i of type *Entity* can be connected with a vertex v_k of type *Fellowship*, via an edge $v_i \leftrightarrow v_k$ of type *Member*, indicating that entity v_i is a member of fellowship v_k. Subsequently, a dipole describes the conflict between a pair of fellowships and is characterized by their (mostly) negative entity inter-connections. Thus, a pair of vertices v_k and v_z of type *Fellowship*, can be connected via a bi-directional edge $v_k \leftrightarrow v_z$ of type *Conflict*. A *Conflict* is characterized by the weight w_{kz}, which indicates the extent of the conflict between the fellowship pair, with values ranging from 0 (no polarization) to 1 (extreme polarization). Vertices of type *Topic*, represent the topics of discussion between entities, thus, bear *supportive* or *oppositional* attitudes. Hence, a vertex v_i of type *Entity* can be connected with a vertex v_x of type *Topic* via edge $v_i \leftrightarrow v_x$ of type *Attitude*. Similarly with *Relationship*, *Attitude* is characterized by w_{ix},

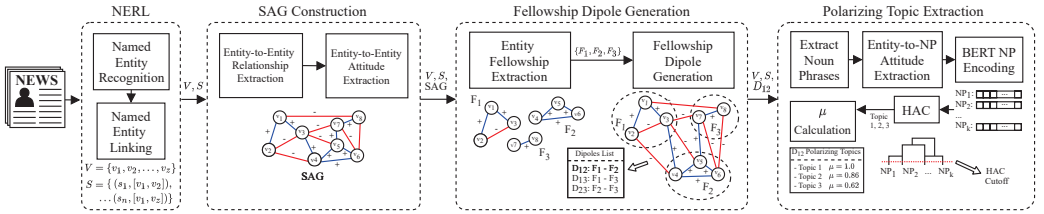


Fig. 2. An overview of the proposed framework for the unsupervised and domain-agnostic polarization knowledge extraction from news articles.

which indicates the attitude of the entity v_i towards the topic v_x with values ranging from -1 (extremely opposed) to 1 (extremely supportive). We refer to entities, fellowships, dipoles, topics, and attitudes as "***Polarization Knowledge* (PK)**."

PDM and Polarization Knowledge: The Polarization Data Model (PDM) integrates both structural and content-based dimensions of polarization. Structurally, the model represents relationships between entities through *Relationship* and *Conflict* edges, which capture interactions and opposition between individuals or groups, key elements in analyzing group formation and structural polarization. Content-wise, the model leverages *Attitude* edges to connect entities with specific topics, reflecting their stance towards those topics, revealing ideological differences in opinions on key issues. Additionally, by modeling negative relationships and conflicts, it captures the tensions and hostilities that arise between polarized groups [43]. The PDM also spans the individual-collective spectrum [46], as it models both individual entities and fellowships/dipoles, making it versatile for studying polarization at various societal levels.

3.3 Unsupervised Extraction of Polarization Knowledge from News Articles

3.3.1 Framework Overview. Our framework is designed to facilitate the study of polarization in news media across diverse domains, making it a valuable tool for researchers, analysts, and scholars in fields such as political and social science, as well as computer science. It unsupervisedly extracts polarization information in line with the Polarization Data Model (PDM). This process involves a series of transformation procedures that allow users to (i) analyze polarization in specific use cases, and (ii) integrate the concept of polarization into their ongoing research. This approach ensures that our framework is both versatile and applicable in diverse research contexts.

The onboarding begins with users submitting the configuration parameters through our framework's programming interface, initializing the overall pipeline as depicted in Figure 2. Once initialized, our framework activates parallel *News Collectors*, which retrieve relevant news articles from the GDELT database, filtering the articles based on the user-defined parameters. Following article retrieval, the ***Named Entity Recognition and Linking (NERL)*** process identifies and disambiguates key entity mentions. This step produces a set of entities (V) and the sentences they appear in (S). The pipeline then proceeds to generate the Sentiment Attitude Graph (SAG) through the *Entity-to-Entity Relationship Extraction* and *Entity-to-Entity Attitude Extraction* components. Subsequently, our framework identifies polarizing fellowship dipoles by extracting fellowships (F) and generating dipoles (D). For each dipole, discussion topics are extracted, and the polarization level of each topic is quantified. Intermediate results are stored for reproducibility, optimization, error recovery, and in-depth analysis. As an output, our framework generates the polarization information for the specified study subject, which includes the: (i) SAG, (ii) fellowships, (iii) polarized dipoles, and *(iv)* polarizing topics.

Technical Details: To enhance accessibility and facilitate collaboration, we have implemented the proposed framework as a Python package.[1] The developed package can seamlessly integrate with Jupyter notebooks to increase reproducibility and ease-of-use. Our framework operates as a modular system, allowing users to customize their analysis by selectively enabling or disabling specific components. This modular design empowers researchers to tailor the framework to their specific needs. For instance, users can bypass the news collection step when working with pre-existing datasets or exclude topic extraction when focusing on predefined thematic areas. The proposed framework is designed for efficiency, employing a multi-process approach to execute each pipeline step in parallel, maximizing processing speed. Additionally, it can be extended to operate within distributed computing environments, making it suitable for handling large-scale tasks and resource-intensive analyses. More details are presented in the subsequent sections.

3.3.2 Retrieval of News Articles. News articles are the primary data source for our framework. While most existing polarization studies rely on messages circulating in **Online Social Networks (OSNs)** [2, 7, 23, 31, 87], which are typically short, noisy, and informal [31], news articles offer longer, more formal, and descriptive content. This richness makes them a more reliable source for: (i) Extracting detailed knowledge about polarizing topics, their semantics, and structure; and (ii) Identifying factors that may contribute to or mitigate polarization. Furthermore, evidence suggests that attitudes expressed in news articles (e.g., bias or hyper-partisanship) play an important role in instigating polarization [2, 7], influencing both elite discourse and potentially affecting mass public opinion [11, 46]. By analyzing news articles, our framework captures how polarization is constructed and propagated through media narratives, offering insights into the dynamics of polarization at different societal levels.

Our framework requires two user-defined parameters to run: a theme, which sets the general topic of the study, and a time-span, which limits the focus of the inquiry to a particular time frame. These parameters guide the granularity and scope of the analysis but do not pre-define specific categories or labels, as in supervised approaches. The framework operates in an unsupervised manner, meaning it does not rely on pre-labeled datasets or controlled case studies. Instead, it allows users to specify the general area of interest, and the system autonomously extracts relevant entities, fellowships, and topics from large-scale news data without the need for prior supervision or case-specific training. For instance, to study polarization during the COVID-19 pandemic in the US, users can define the theme as "*US COVID-19 Pandemic*" and specify the period between 2020 and 2021. Alternatively, for a more confined study, users might set the parameters to "*US COVID-19 Vaccine Candidates*" and the period from March to July 2021. Parallel News Collectors are activated to retrieve relevant news articles from the GDELT Project—a large, open database of global news articles. The News Collectors automatically map locales using gazetteers (e.g., converting "*US*" to "*United States*") and filter content using relevant keywords associated with the theme (e.g., "*COVID-19*" and "*pandemic*"). The system also ensures that articles align with the specified theme by checking the presence of keywords in the article's title or URL and applying date filters to fit the specified time-span.

3.3.3 Extracting Entities and their Relationships. Entities are fundamental components of the proposed framework as they populate the social and political groups being studied [80]. In order to populate the Sentiment Attitude Graph (SAG), the location and classification of entities mentioned in the input corpus is required. To address this, we apply **Named Entity Recognition (NER)**, which is a subtask of **Information Extraction (IR)** that aims to identify named entities mentioned in unstructured text and categorize them into pre-defined categories [86]. We utilize a

[1] https://pypi.org/project/XXXXXX - The URL is omitted due to double-blind reviewing.

pre-trained BERT-based transformer model [25] for NER. This model, trained on the CoNLL-2003 annotated dataset [79], is adept at identifying and classifying entities within text as sequences of tokens, each associated with specific entity types. When processing a news article, our method identifies mentions of entities and their corresponding sentences. For each sentence s_i, the model identifies a set of entities V_{s_i} mentioned in that sentence. The output is thus a collection of annotated sentences, represented as $S = (s_i, V_{s_i}), \ldots$, effectively pairing each sentence with its set of identified entities.

A challenge that arises from the implementation of NER is the large dimensionality resulting from the fact that mentions of the same entity are not recognized as such. For example, references to "*Trump*", *President Trump*", and "*Donald Trump*", are treated as if they refer to distinct entities. A solution to this is the assignment of a unique identifier to each entity mention via an existing knowledge source, such as Wikipedia,[2] which can be accomplished through **Named Entity Linking (NEL)** [69]. The NEL task consists of two steps: (i) finding entity candidates; and (ii) collective disambiguation.

NEL using Wikidata: To identify the entity candidates, we utilize a snapshot of the **Wikidata Knowledge Graph (WKG)**[3] with approximately 1 billion item entries and 1.1 TB of data. For efficient access and retrieval, we index the Wikidata entries in Elasticsearch.[4] To obtain the entity mention candidates, we first extract a small set of potential entities from the World Knowledge Graph (WKG) for each entity mention in the dataset S. This retrieval is achieved through a string similarity query executed over Elasticsearch. The query utilizes the **Token Sort Ratio (TSR)** measure, a technique that breaks down the entity mentions into individual components, known as tokens. These tokens are then sorted alphabetically before being compared. The comparison is based on a ratio mechanism, which calculates the similarity between the sorted tokens of the entity mention and the entities in the WKG. Candidates are selected based on their TSR score; any KG vertex with a TSR score of at least 0.5 is deemed a candidate, implying that there is at least a 50% string similarity between the mention and the potential entity. We then perform collective disambiguation over the selected candidate entities. Traditional Named Entity Linking (NEL) methods utilize machine learning models that require large training corpora and extensive feature engineering. In contrast, we employ a more optimal and domain-agnostic solution by encoding the WKG vertices in vectors of low-dimensional space, called vertex embeddings [58]. Vertex embeddings refer to dense vectors that are able to capture the structural properties and relationships of the vertex within the graph. These embeddings enable complex graphs like the WKG to be represented in a format suitable for machine learning algorithms. By reducing the high-dimensional data of the graph into a more manageable form, vertex embeddings facilitate more efficient processing and analysis, allowing for the application of various data-driven techniques that can discern patterns and insights from the graph's structure. We train the WKG vertex embeddings using the DeepWalk algorithm [61] with the suggested configuration by the authors [58]. The DeepWalk algorithm was selected as it efficiently learns a latent representation of adjacency matrices in WKG, can be executed in an online manner, and is scalable and parallelizable. To determine the most suitable candidate vertex for each mention, we calculate the semantic similarity between entities. This is done by assessing the cosine similarity between their respective vertex (i.e., entity) embeddings. To simplify this process and make it more computationally efficient, we implement a greedy optimization algorithm [58]. This algorithm aids in refining and updating the annotated set of entities, denoted as S.

[2] https://www.wikipedia.org/
[3] https://www.wikidata.org/
[4] https://www.elastic.co/

3.3.4 Entity-to-Entity Relationship. Within the context of the proposed PDM, the identification of entities and their pair-wise relationships is essential in order to ensure the accuracy and relevance of the Sentiment Attitude Graph. Thus far, we have successfully identified and disambiguated entities within each article by employing Named Entity Recognition (NER) and Named Entity Linking (NEL) techniques. However, a significant challenge arises in the computational modeling and identification of relationships between these entities to guarantee a strong likelihood of their real-life cohesion. To address the challenge of identifying real-life relationships between entities, we apply a method based on quantifying the frequency of entity-pair co-occurrences within our corpus, as in Eckhardt et al. 2014 [27]. Typically, the distribution of these frequencies is left-skewed, meaning that while most entity pairs co-occur infrequently, a few pairs do so much more regularly. To distinguish meaningful relationships, we set a frequency threshold. Only entity pairs that exceed this threshold are included in the Sentiment Attitude Graph (SAG), ensuring that their co-occurrences are likely indicative of genuine relationships. This method also helps keep the SAG at a manageable size for further analysis. In our framework, we specifically select entity pairs that fall within the top 95_{th} percentile of all pair frequencies, thereby focusing on the most relevant and significant relationships in the dataset.

3.3.5 Calculation of the Nature of Entity Relationships. The nature of an entity relationship can be described as *positive*, *neutral*, or *negative*. A positive relationship indicates a possible friendship and supportiveness between the entities, whereas a negative relationship indicates opposition and hostility. The lack of these characteristics indicate a neutral relationship. For instance, an entity representing an individual, such as a political leader, can have a distinctly positive or negative relationship with another entity, like a specific policy or organization.Similarly, collective entities like political parties, countries, or religions, can also engage in relationships that are characterized by these sentiment attributes. The nature of the relationship, whether positive, negative, or neutral, is thus a function of the entities' interactions and stances towards each other, encompassing a wide array of entity types and contexts. A simple method to determine the nature of an entity pair relationship from text, is to capture the contextual sentiment of the sentences that the entity pair co-occurs within. While this method offers initial insights, it has limitations, such as not accounting for grammatical dependencies between words or the presence of multiple entities within a sentence [4]. To address these limitations, we propose a method for *sentiment attitude identification* [20, 66], a task that concentrates on discerning the positive or negative attitudes directed from one element in the text to another. Specifically, we identify the explicit syntactical dependency paths between entity pairs and calculate their sentiment scores. This method effectively addresses the limitations of simpler sentiment analysis approaches by focusing specifically on the syntactical relationships between entities within a sentence, providing a more targeted assessment of their interactions.

Given a sentence s_i, we identify entity-pairs (v_x, v_y) within SAG, where v_x is the *attitude holder*, and v_y is the *attitude target*. Figure 3 illustrates a descriptive example of the proposed method. The sentiment attitude from v_x towards v_y is calculated as $att(s_i, v_x, v_y) \in \{positive, neutral, negative\}$. Our method employs syntactical dependency paths to ascertain the direction of sentiment between entities in a text. These paths are constructed using dependency labels that define the grammatical relationship between words in a sentence [57]. Such dependency labels include *nsubj* (nominal subject), representing the subject of a verb; *dobj* (direct object), indicating the object that the verb is directly acting upon; *ccomp* (clausal complement), which connects a verb to a complement clause; and *nmod* (nominal modifier), used for modifiers of nouns or clausal predicates. Based on these labels, we apply specific syntactical rules for sentiment attitude identification [20]: (i) the dependencies between the subject *nsubj* and direct object *dobj* of

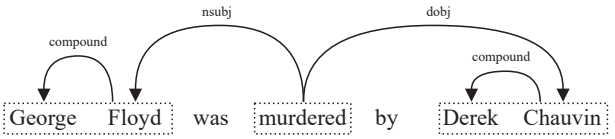


Fig. 3. The holder of attitude is "George Floyd", and the target is "Derek Chauvin". The dependency path between the two contains only the word "murdered", which satisfies the first dependency pattern of being between *nsubj* and *dobj*. *nsubj* stands for "nominal subject" and refers to the noun phrase that functions as the subject of a clause. *dobj* stands for "direct object" and refers to the noun phrase that receives the action of the verb in a sentence. Due to the word "murdered", the overall sentiment attitude is *negative*.

the sentence s_i; (ii) the dependency pattern of $(nsubj, ccomp, nsubj)$ of s_i; and (iii) an indicator of $nmod : against$, a negative relation (nominal modifier) between the two entities within s_i. Once these dependency paths are established, we calculate a sentiment score for each path. This score quantifies the sentiment attitude, positive or negative, between the connected entities. The pseudocode of the sentiment attitude extraction is available in the Appendix A. This approach offers flexibility, allowing the sentiment calculation component to be adapted to the domain, such as using the MPQA lexicon [85] for news or the Loughran-McDonald lexicon [50] for finance.

Taking into account that *SAG* is an undirected graph, the bi-directional entity relationships are considered. For each entity pair, both $att(s_i, v_x, v_y)$ and $att(s_i, v_y, v_x)$ are calculated. An average aggregation of the sentiment attitudes between the entity pair v_x and v_y is computed as w_{xy}, populating the edges (v_x, v_y, w_{xy}) of *SAG*.

3.3.6 Identifying the Entity Fellowships: In conventional graph partitioning, node fellowships are defined as densely-connected endorsement subgraphs, which can be computed by community detection algorithms. However, subgraph density alone is not indicative of supportive relationships amongst subgraph nodes; consequently, typical community detection algorithms are not adequate for computing fellowships in our modeling framework. Instead, we need to identify densely-connected subgraphs, which contain predominantly *positive edges*. Therefore, we model the computation of SAG fellowships as a **signed-network clustering (SNC)** problem [77]. SNC analyzes undirected graphs with both positive and negative edges and aims to identify clusters with dense internal connections, primarily positive, and inter-cluster connections that are predominantly negative. Several algorithms that have been proposed in the literature to solve the SNC problem require as input a predefined number of clusters k, which is undesirable in our case as the size of SAG and the number of its fellowships are not known in advance. Furthermore, SNC algorithms that do not require a predefined cluster count rely on modularity, which has been shown to suffer from a resolution limit, making the detection of small communities difficult [30]. To overcome these limitations, the authors in [28] introduced SiMap, which is an extension of the **Constant Potts Model (CPM)** to be applicable on signed networks. CPM overcomes the modularity resolution limit by utilizing an objective function denoted as:

$$H(G, C) = -\sum_{c} (w_c - \lambda N_c^2)$$

where G is a graph and C is a clustering scheme with clusters $c \in C$. The value of w_c corresponds to the sum of weights of c, and N_c is the number of vertices in c. The parameter $\lambda \in [0, 1]$ is the constant used to configure the resolution that the objective function is applied to. By sliding λ from $0 \rightarrow 1$, CPM can generate smaller and denser clusters, thus, overcoming modularity's resolution limit. SiMap method extents CPM to signed graphs, with the extension denoted as:

$$H(G, C) = -\alpha \sum_{c} (w_c^+ - \lambda N_c^2) + (1 - \alpha) \sum_{c} (w_c^- - \lambda^- N_c^2)$$

where λ^- is the resolution constant of the negative ties in the clusters, α denotes the contribution of the positive edges against the negative ones, and w_c^+ and w_c^- denote the sum of positive weights and absolute sum of negative weights respectively for cluster c. The value of λ^- is set to 0 in order to produce dense positive and negative-free clusters. α is set to 0.5 to equally weigh positive and negative edges in clusters, making the objective function:

$$H(G, C) = -0.5 \sum (w_c - \lambda N_c^2)$$

CPM is optimized iteratively, using a local update formula denoted as $\Delta H = w_{ck} - w_{c'k} + 2\lambda N_k(N_{c'} - N_c)$. The formula is applied on the transfer of k vertices from cluster c to c', if this transfer maximizes the objective value. The value of w_{ij} equals to the sum of weights from cluster i to j. For the optimization of signed CPM, positive (G^+) and negative (G^-) graphs are treated separately, resulting in:

$$\Delta H(G, C) = \alpha \Delta H(G^+, C) - (1 - \alpha)\Delta H(G^-, C)$$

As a result, SiMap accepts a resolution parameter λ instead of the number of clusters k, and is able to produce smaller and denser partitions, thus overcoming the resolution limit. Therefore, we apply SiMap for the identification of the set of entity fellowships F from SAG, setting $\lambda = 0.05$ as suggested by [28].

3.3.7 Extracting the Fellowship Dipoles: In our framework, we introduce the concept of dipoles to capture the structural conflict and polarization between pairs of fellowships. To identify dipoles, however, we need to define and quantify some metric of the polarization between all possible pairs of fellowships. We approach this by viewing the Sentiment Attitude Graph (SAG) as a signed network and applying ***structural balance theory*** **(SB)** [17] to define structural polarization. According to SB, a signed graph is balanced if and only if (i) all its edges are positive or (ii) its nodes can be partitioned into two disjoint sets such that positive edges exist only within clusters and negative edges are only present across clusters. The concept of SB is relevant because balanced structures in a network, characterized by consistent positive or negative relationships, have been linked to polarization [76]. A balanced graph, therefore, represents two distinct and opposing fellowships, indicating a state of high polarization. To measure the SB of a signed graph, the majority of approaches utilize triads, sub-graphs consisting of three interconnected nodes within the larger graph. SB theory dictates that a network's balance can be gauged by inspecting these triads for an odd number of negative edges, as outlined by Cartwright and Harary, 1956 [17]. Such a configuration in a triad indicates underlying tension or conflict, and points to an imbalance in the larger graph. Within our framework, a dipole is deemed structurally balanced if it does not contain any such imbalanced triads. To measure balance, we employ the frustration index [8], a metric that identifies the minimum number of edge sign alterations required to achieve balance, thereby quantifying the structural imbalance in a dipole and, by extension, the polarization between the associated fellowships.

Frustration Index: Given a graph $G(V, E)$ with vertices V and edges E, the set $E*$ represents the minimum deletion set of edges that result in a balanced graph. The frustration index L is determined by the size of the smallest set of edges in G that need to be deleted for balance, denoted as $L(G) = |E*|$. The exact algorithmic computation of $L(G)$ is closely related to NP-hard graph problems, such as *k-coloring*. There exist several estimation approaches for the calculation of $L(G)$, which model it as a global optimization problem. These methods approximate the frustration index, $L(G)$, by using its upper bound, expressed as $L(G) \leq m^-$, where m^- represents the number of negative edges in the graph. This implies that by removing all negative edges, the graph becomes balanced with only positive connections. $L(G)$ generates an arbitrary count of edges, which varies

depending on the graph G. To standardize this number, we utilize the normalized frustration index [9], denoted as $L'(G) = 1 - (L(G)/(m/2))$, which adjusts the values into a consistent scale from 0 to 1, with 0 being totally imbalanced, and 1 perfectly balanced.

Given a dipole D_{ij} between fellowships F_i and F_j, we calculate its frustration index as $L'(D_{ij})$. Dipoles with higher frustration index indicate a higher probability of a polarized state. To identify polarized dipoles, we apply a threshold, considering only those with a frustration index ≥ 0.7.

3.3.8 Extracting Dipole-specific Discussion Topics. The identification of topics within a given corpus is referred to as topic modeling. This process typically involves the use of statistical methods, such as LDA [45], on large collections of texts. In the context of fellowship dipoles, the identification of topics is required to be done from a limited number of sentences where the dipole entities co-occur. Given a dipole D_{ij} between fellowships F_i and F_j, the topics are to be identified from sentences $S_{D_{ij}} \subseteq S$. An example sentence between entities of Republican Party and Democratic Party in the context of the COVID-19 pandemic in the US[5] is presented below:

Sentence: "*The massive gap between Republicans and Democrats on vaccinating kids.*"

An observation that we can make on the example above, is that a large portion of the sentence's information pertains to the identified entity mentions, which are used within the SAG. Therefore, to avoid redundancy and focus on uncovering new topics in the dipole topic identification process, we mask these entities within the sentence, effectively excluding them from further analysis:

Excluded Entities: "*The massive gap between* $< E_1 >$ *and* $< E_2 >$ *on vaccinating kids.*"

As these sentences are significantly smaller than the complete articles, the task of identifying discussion topics from entity co-occurring sentences is not achievable using conventional topic modeling techniques. From the example sentence, the topic of discussion seems to be the vaccination of children in the context of COVID-19. Linguistically, this is represented from the **noun phrase (NP)** of "*vaccinating kids*". Grammatically, an NP functions as a noun in a sentence. One way to identify the NPs of a sentence is using constituency parsing [86], the task of segmenting a text into sub-phrases or constituents. By applying constituency parsing in the previous example, the NPs retrieved are "*massive gap*", "*Republicans*", "*Democrats*", and "*vaccinating kids*". The NPs of "*Republicans*" and "*Democrats*" coincide with the identified entities and therefore are ignored. The NP of "*massive gap*" is the linguistic component that represents the negative sentiment attitude between the two entities. To identify the dipole-specific discussion topics, we propose the use of semantic clustering of sentence-level NPs. This approach has demonstrated the ability to capture the underlying topics of discussion in text and is commonly employed in techniques such as topic modeling [36] and keyword extraction [62]. Further details are elaborated in the subsequent sections.

Noun Phrase Representation: In order to cluster the NPs identified from the dipoles, a method of representation that captures their semantic meaning must be employed. Within the field of NLP, this is typically achieved through the use of a language model [25]. Common language models utilized include N-Grams, **Bag-of-Words (BoW)**, and TF-IDF, which are used to create informative and similar textual clusters. However, it has been found that such approaches fail to effectively capture the semantic meaning of phrases, resulting in sparse results with multiple clusters representing semantically similar topics. An example of this can be observed in the noun phrase "*vaccinating kids*", which bears the same semantic meaning as "*inoculating children*" and "*immunization of youngsters*", yet has a significantly different lexical representation from the rest.

[5] https://www.washingtonpost.com/politics/2021/12/10/massive-gap-between-republicans-democrats-vaccinating-kids/

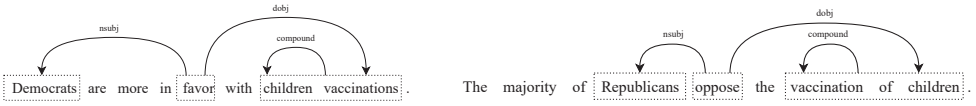


Fig. 4. The syntactical dependency diagram of two sentence examples.

A novel approach to language modeling that addresses this issue is the use of *word embeddings*, namely distributed representations for text that allow words with similar meaning to have similar representations. The distributed representation is learned based on the usage of words, resulting in words that are used in similar ways having similar representations and naturally capturing their meaning. This can be contrasted with the representation in a BoW model, where different words have different representations regardless of how they are used. The approach is grounded in linguistic theory, specifically the "distributional hypothesis" [38], which posits that words that have similar context will have similar meanings. As a result, word embeddings associate each word with a real-valued vector with several hundreds of dimensions, in contrast to the thousands (or millions) of dimensions required for sparse word representations such as a BoW. One notable method of word embedding is **BERT (Bidirectional Encoder Representations from Transformers)**, developed by researchers at Google AI Language [25]. BERT has received significant attention in the NLP and ML community for its state-of-the-art results in a wide variety of NLP tasks, including language interpretation and similarity.

Specifically, for each dipole D_{ij}, we first extract Noun Phrases (NPs) from $S_{D_{ij}}$ and convert them into BERT word vectors. These vectors are then grouped into clusters based on semantic coherence. To identify these clusters without predefining their number, we utilize **Hierarchical Agglomerative Clustering (HAC)** [51]. HAC replaces the need for a predefined number of clusters with a cutoff threshold, which we set at 0.20. This threshold, based on cosine distance, ensures that the NPs in each cluster share at least 80% semantic similarity. The clusters formed by this process represent the collection of discussion topics within the dipole, denoted as $T_{D_{ij}}$.

3.3.9 Quantifying Topic Polarization. As opposed to structural polarization (see Section 2), content or topical polarization refers to the degree of attitude disagreement between entities of a dipole on a particular topic. In order to determine the polarization of a given dipole topic, we must identify the sentiment attitudes of the dipole's entities towards the topic. In our framework, topics are identified as clusters of Noun Phrases (NPs) that are semantically similar and appear in sentences where entities co-occur. We then calculate the sentiment attitude of these entities towards the NPs associated with each topic. This is feasible by adapting the sentiment attitude approach outlined in Section 3.3.5, in which, instead of a target entity v_y within a given sentence, we define a target NP denoted as np_y. We then calculate the attitude as the sentiment score $att(s_i, v_x, np_y)$ for each dependency path between the entity and the NP. An example of this can be seen in the two sentences depicted in Figure 4. The first dependency rule from Section 3.3.5 applies in both instances, generating the path of {"*Democrats*", "*favor*", "*children vaccinations*"}, and {"*Republicans*", "*oppose*", "*vaccination of children*"} respectively. The paths indicate a positive attitude originating from the entity of "*Democratic Party*" towards the NP of "*children vaccinations*", and a negative attitude originating from "*Republican Party*" towards "*vaccination of children*", with both NPs representing the topic of "*children vaccination*". Consequently, given a dipole's D_{ij} discussion topic $t_z \in T_{D_{ij}}$, the set sentiment attitudes A_{t_z} expressed from dipole fellowship entities $v_x \in F_i \cup F_j$ towards t_z is quantified as $att(s_i, v_x, np_y)$, where s_i is a sentence where the entity v_x and topical NP np_y co-occur. The step-by-step process of topic polarization quantification is available in Appendix A.

Quantifying Polarization Index: Having obtained the set of entity sentiment attitudes for a given dipole topic, we can quantify the degree of attitude disagreement as a measure of the topic's

polarization. Various methods of calculating the attitude disagreement [14] include the use of spread (or range), which measures polarization as the distance between the minimum and maximum sentiment attitude values, and dispersion (or variance), which identifies polarization as the presence of a bimodal distribution in the set of attitudes. However, the spread approach may not accurately capture polarization in cases where the minimum and maximum attitude values are outliers and not representative. Therefore, dispersion-based approaches are more suitable for this purpose. In our framework, we employ the polarization index, which was proposed by Morales et al. 2015 [54]. As stated by the authors, "*a population is perfectly polarized when divided into two groups of the same size and with opposite attitudes*." In the context of our work, "population" is defined as the set of topic sentiment attitudes A_{t_z}. The polarization index, denoted as μ, is defined as:

$$\mu = (1 - \Delta_A)\delta_A$$

where $\Delta_{A_{t_z}}$ is the normalized difference in set sizes between the positive and negative sentiment attitudes, A^+ and A^-, respectively. δ_A is the attitude difference and is calculated as $\delta = |gc^+ - gc^-|/2$, with gc^+ and gc^- equal to the average attitude values of A^+ and A^-, respectively. The values of μ range from 0 to 1, with $\mu = 1$ indicating perfect polarization and $\mu = 0$ indicating no polarization.

4 Experiments and Evaluation

Evaluating polarization is inherently challenging due to the complex and evolving nature of political discourse, as well as the diversity of political ideologies and attitudes. Additionally, the scarcity of annotated data adds another layer of difficulty to the evaluation process. In response to these challenges, we have enhanced the current framework by structuring the evaluation around key aspects of Polarization Knowledge (PK), as defined in the Polarization Data Model (PDM). Specifically, our methodology evaluates polarization at the *entity*, *fellowship*, and *topic* levels, providing a robust and multi-faceted assessment.

To ensure the validity of the evaluation results, we conduct a separate evaluation process for each level of PK. This evaluation process involves the use of external sources and the manual annotation of data by a team of three annotators with a CS background, including one experienced annotator, and two MSc students. We define a series of questions to guide the evaluation:

- Q1. What is the effectiveness of our framework in capturing entity attitudes towards various discussion topics, particularly in comparison to other methods, when taking into account known attitudes from external sources?
- Q2. What is the extent of alignment between politically cohesive fellowships identified and their official party manifestos in terms of their overall attitudes towards different discussion topics?
- Q3. What is the accuracy in extracting the discussion topics of a domain and how effective is it in capturing the per-topic polarization degree compared to other methods?

To evaluate entity-to-topic attitudes (Q1), we compare the output of our framework with ground-truth data collected from external platforms, such as OnTheIssues and BallotPedia.[6] We use standard metrics of precision, recall, and F1-score to measure its performance. In evaluating politically cohesive fellowships (Q2), we assess the ability of our framework to identify and extract politically cohesive groups and measure their alignment with official party positions on specific policy issues. We use metrics of precision, recall, F1-score, and AUC to measure the level of agreement of politically cohesive groups regarding entity-members' political ideologies and attitudes towards different topics, as compared to their respective party manifestos. To evaluate topic-level polarization (Q3), we rank discussion topics in terms of their level of polarization as calculated by our

[6] https://ballotpedia.org

framework, compared to a ground-truth ranking constructed using an annotation methodology from the literature [40]. The evaluation is carried out by measuring the ranking agreement between the two lists using the **Ranked-biased Overlap (RBO)** metric [83].

We conduct the proposed evaluation on three case studies of Abortion, Immigration, and Gun Control. We first present the statistics on the application of our framework on each case study, including the extraction of their respective PDMs. To prepare the ground-truth datasets for the evaluation of each polarization level, we outline our annotation procedures. We evaluate each level by comparing our framework's performance against those of existing methods.

4.1 Evaluating on Abortion, Immigration, and Gun Control

The issues of Abortion, Immigration, and Gun Control are chosen as case studies for the evaluation. These issues have been acknowledged as historically contentious and politically divisive, as they are relevant to a wide range of policy areas, including social welfare, human rights, and healthcare, as well as national security and public safety [65]. In the US political context, distinct and polarized positions on these topics have been taken by the Democratic and Republican parties [18]. The Democratic party is known to support a woman's right to choose and have access to safe abortion, as well as a more liberal approach to immigration and gun control. On the other hand, the Republican Party is generally known for supporting pro-life policies, advocating for restrictions on abortion, and taking a more restrictive approach to immigration. Regarding gun control, the party opposes restrictive regulations, favoring broader firearm access and protecting Second Amendment rights.

To alleviate the limitations of available datasets, our evaluation focuses specifically on the political conflict between the Democratic and Republican parties in the United States. This is due to the abundance of information and prior knowledge that exist for these groups regarding the case studies. The availability of data on these issues makes it possible to conduct a thorough analysis and evaluation of polarization knowledge, and to gain insights into the ways in which polarization is shaping public opinion and influencing political outcomes.

4.1.1 Dataset Description. For the evaluation of the Abortion, Immigration, and Gun Control case studies, we utilize the datasets from the work of Roy and Goldwasser 2020 [65], consisting of 16,475 news articles annotated with the political bias of their sources and topics and subframes that are points of polarization. The political bias of sources is annotated on a Left-Right political spectrum. The datasets have 20, 22, and 19 topics for Abortion, Immigration, and Gun Control, respectively. The authors focus on identifying polarizing subframes within news articles by building topic-specific lexicons using repeating expressions and grouping them into subframes, resulting in three lexicons for each topic. The subframes were then identified through the use of human knowledge by extracting talking points from Wikipedia[7] and OnTheIssues,[8] resulting in multiple subframes for each case study (see Appendix A). The methodology and definition of subframes by Roy and Goldwasser 2020 are similar to the approach used in our framework, which utilizes Noun Phrases (NPs) for the identification of polarizing topics. In the context of the dataset, subframes can be considered equivalent to topics for evaluating polarization knowledge.

4.1.2 PDM and Polarization Knowledge. In the process of evaluating the proposed framework, we utilized as input the Abortion, Immigration, and Gun Control news corpora. The statistics of the order of the PDM and SAG for each case study are summarized in Table 2. Following the identification of relationships and statuses of the entities, the SAG for each case study was constructed with

[7] http://wikipedia.org
[8] https://www.ontheissues.org

Table 2. Statistics on the PDM for Each of the Case Studies

	Abortion	Immigration	Gun Control
Entities	8,113	18,409	15,217
SAG Nodes	228	459	194
SAG Edges	523	1,440	478
Fellowships	49	156	69
Dipoles	16	34	42
Noun Phrases	107,521	298,918	201,419
Topics	533	2,517	1,262

228 entities and 523 relationships for Abortion, 459 entities and 1,440 relationships for Immigration, and 194 entities and 478 edges for Gun Control. The identification of entity fellowships resulted in 49, 456, and 69 fellowships extracted from each case, respectively. These fellowships participated in 16, 34, and 42 dipoles respectively, filtered based on their structural polarization. In terms of the discussion topics, our framework identified the unique NPs for each case study: 107,521 NPs for Abortion, 298,918 NPs for Immigration, and 201,419 NPs for Gun Control. After applying the semantic clustering process to these NPs, we were able to group them into distinct topics—specifically, 533 topics for Abortion, 2,517 topics for Immigration, and 1,262 topics for Gun Control.

Specifically, in the abortion case study, our framework effectively identified key polarized entities such as the Republican and Democratic parties, Planned Parenthood, and prominent religious organizations, such as the Catholic Church and Christianity. It further revealed the formation of distinct fellowships, grouped by their political ideologies (conservative vs. democratic), stances on abortion (anti-abortion vs. pro-choice), and views on healthcare policies and religious beliefs. The framework also uncovered highly polarized topics like "pro-life vs. pro-choice," "partial-birth abortion," "fetal tissue", and "taxpayer-funded abortion," demonstrating the framework's capability to highlight the focal points within the abortion debate.

4.2 Polarization Knowledge Annotation

4.2.1 Entity-level Polarization Annotation. At the entity level, the evaluation concentrates on assessing the framework's ability to identify the attitudes of entities towards topics and comparing it to the true entity attitudes, captured in the **Ground-Truth (GT)**, which is created by manually annotating data from the OnTheIssues platform. This platform contains quotes and votes from various politically affiliated entities on issues such as Abortion, Immigration, and Gun Control, collected from newspapers, articles, press releases, and speeches. By utilizing this platform, relevant quotes from politically affiliated entities in the SAG are identified and annotated, based on their connection to the topics of each case study, as well as their supportive or oppositional attitude towards the respective topic. The methodology for this process is depicted in Figure 5

For example, to illustrate the process, consider the entity of Joe Biden and his quotes on the topic of Abortion. Initially, the entity "*Joe Biden*" and the case study "*Abortion*" are used to collect the relevant quotes through the Quotes database, which are then fed to the Annotation Platform. An example of a relevant quote from Joe Biden is "*Expand embryonic stem cell research.*" (June 2004). Here, we invite the annotators, through the topics related to Abortion, to determine the extent to which the above quote expresses a positive or negative attitude towards the topics. If the quote has nothing to do with a topic or expresses neither a positive nor a negative attitude, it is considered to be neutral. This particular example is determined to express a positive attitude towards the topic of "*Stem Cell Research*". After all the quotes are annotated, we construct the Entities Topic Attitude GT by aggregating the entity attitudes per topic. Consequently, with the completion of

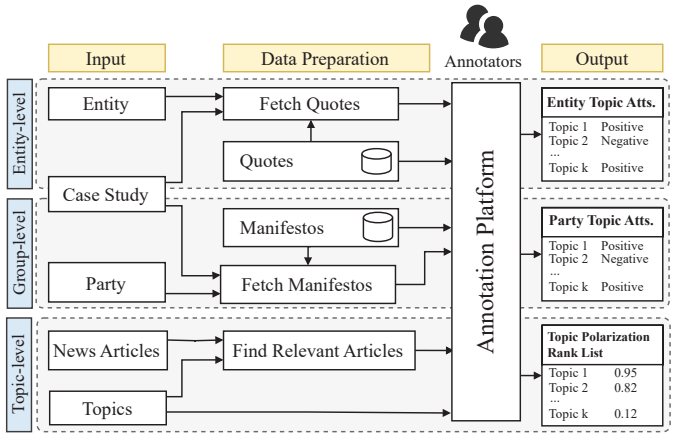


Fig. 5. An overview of the PK annotation methodology.

the entity-level annotation process, we extract the attitudes for 24, 23, and 16 entities for Abortion, Immigration, and Gun Control, respectively.

4.2.2 Fellowship-level Polarization Annotation. At the fellowship level, our evaluation targets the attitudinal alignment between fellowships that are politically cohesive and their respective party manifestos. We define a fellowship as politically cohesive if the majority political leaning of its entities is $\geq 75\%$ towards a specific political party, thereby considering such fellowships as ideologically aligned. As previously mentioned, political groups often adopt official platforms outlining their positions on various issues. These party manifestos can provide insight into the official attitude of the party towards the issue at hand [18]. Therefore, quantifying the degree of alignment between the attitudes of an ideologically cohesive fellowship and the official attitudes of the respective party towards the topics can provide an indication of the effectiveness of our framework in capturing the attitudinal and ideological fellowship characteristics.

To derive representations of official party politicians we collect party manifestos from the Democratic and Republican parties for the time period encompassed by the case study news corpus [81]. Specifically, we fetch the relevant sections of a specific Party's Manifestos and feed them into an annotation platform where annotators note the positive or negative attitude expressed by each statement towards the case study topics. If a statement has no relation to the topic or expresses neither a positive nor negative attitude, it is considered to be neutral. For example, the following statement from the 2016 Democratic Party manifesto relates to the topic of Abortion: "*The President and the Democratic Party believe that women have a right to control their reproductive choices.*". In this case, the annotators would note that the statement expresses a positive attitude towards "*Reproductive rights*" and "*Pro-choice*" and a negative attitude towards "*Anti-abortion*". After the annotation process is completed, the results are aggregated and a list is produced indicating the official party attitudes towards the case study topics. The Democratic and Republican party topic attitudes according to their manifestos for Abortion, Immigration, and Gun Control, are depicted in Appendix A.

4.2.3 Topic-level Polarization Annotation. At the topic-level PK evaluation, the process of annotation is not straightforward as there is no readily available information for the topic polarization ranking, similarly to the entity and group level annotations. Therefore, we create a ground-truth (GT) for the topic polarization ranking through the processing and annotation of each case study's

Table 3. Topic Tuples for Abortion, Immigration, and Gun Control, Annotated as Having Direct Disagreement

Abortion		Immigration		Gun Control	
Anti-Abortion	Pro-Choice	Minimum Wage	Salary Stagnation	Ban on Handguns	2nd Amendment
Pro-Life	Pro-Choice	Wealth Gap	Minimum Wage	Ban on Handguns	Concealed Carry Reciprocity Act
Life Protection	Planned Parenthood	Cheap Labor Availability	Minimum Wage	Gun Control to Restrain Violence	2nd Amendment
Life Protection	Pregnancy Centers	Cheap Labor Availability	Wealth Gap	Gun Control to Restrain Violence	Concealed Carry Reciprocity Act
Sanctity of Life	Women Freedom	Amnesty	Dream Act	White Identity	Person of Color Identity
Late Term Abortion	Roe V. Wade	Amnesty	DACA	Right to Self-Defense	Stop Gun Crimes
Right of Human Life	Reproduction Right	Family Separation Policy	DACA		
Abortion Provider Economy	Abortion Funding	Racial Identity	Born Identity		
		Racial Identity	Racism and Xenophobia		
		Born Identity	Racism and Xenophobia		

news corpus (as illustrated in Figure 5). We apply an annotation methodology similar to that proposed by He et al. 2021 [40], which aimed to study the polarization between two conflicting stances on topics related to US COVID-19 pandemic between Democrats and Republicans. The first step of the annotation process involves the identification of discussion topics through news articles. For our case studies on Abortion, Immigration, and Gun Control, we adopted predefined topics from Roy and Goldwasser 2020 [65], as outlined in Section 4.1.1 and detailed in Appendix A. The authors characterized these topics using specific word vocabularies, which we used to describe each topic in our analysis.

We ask the annotators to identify pairs of topics (i.e., *subframes*) that describe a direct disagreement between the topics. For example, in the case study of Abortion, the topics of "Anti-abortion" and "Pro-Choice" describe a clear disagreement. The same applies with the topics "Pro-Life" and "Pro-Choice". On the contrary, there is no direct opposition between the topics "Pro-life" and "Abortion Provider Economy" as defined by Roy and Goldwasser 2020 [65]. According to the definition provided in their work (which can be found in Appendix A), "Pro-life" refers to anything related to supporting life, including personalities, movements, or legislation. However, this definition does not directly conflict with the topic "Abortion Provider Economy" which concerns statistics, services, and/or profits of abortion providers. As a result, Table 3 presents the annotated topic tuples per case study that express a direct contrast between them. This selection was based on identifying articles that contained three or more keywords from each topic pair's vocabulary. Among these, we chose the 30 articles from each political leaning that had the highest frequency of topic-related keywords. Subsequently, we tasked the annotators to label each article as 0 or 1 to indicate whether the article expresses a supportive attitude towards a particular topic in the tuple, respectively. For example, considering the topic tuple "Pro-life" and "Pro-choice", the annotators label the article as 0 if it supports the first element of the tuple (i.e., Pro-life) and 1 if it supports the second element of the tuple (i.e., Pro-choice). If the article does not express a clear attitude, it is labeled as -1. The overall attitude of the corpus D towards a topic tuple u is quantified by counting the labels of 0 and 1 within the corpus for that specific tuple. The articles labeled with -1 are not counted because they do not display a clear political standing. These counts are represented as $N_D^u(0)$ for the number of 0 labels and $N_D^u(1)$ for the number of 1 labels associated with tuple u. The overall attitude of the corpus on the tuple is quantified as:

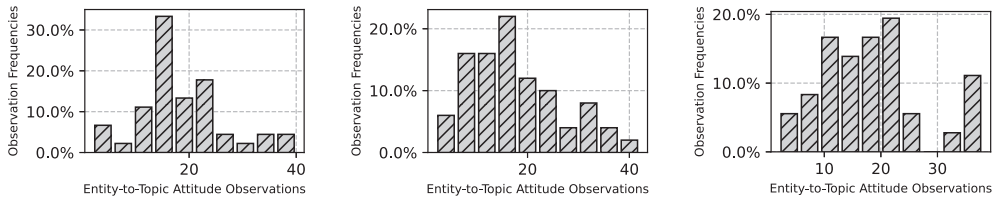
$$le(D, u) = (N_D^u(1) - N_D^u(0))/|D|$$

The resulting value of $le(D, u) \in [-1, 1]$, where a value of -1 indicates a supportive attitude towards the first topic of the tuple, 1 indicates a supportive attitude towards the second of the tuple, and 0 indicates a neutral attitude or lack of clear polarization towards either topics. We compute the ground-truth polarization score for a specific topic tuple u, comparing the liberal corpus with the conservative corpus, denoted as D^L and D^R, respectively. We calculate the polarization score as:

$$\alpha(D^L, D^R, u) = |le(D^L, u) - le(D^R, u)|/2 \in [0, 1]$$

Table 4. Topic Polarization Rank list GT for Abortion

Topic Index	Abortion Topic Tuples		Normalized α
7	Right of Human Life	Reproduction Rights	0.225
3	Life Protection	Planned Parenthood	0.075
2	Pro-Life	Pro-Choice	0.060
8	Abortion Provider Economy	Abortion Funding	0.035
4	Life Protection	Pregnancy Centers	0.020
6	Late Term Abortion	Roe v. Wade	0.010
1	Anti-Abortion	Pro-Choice	0.010
5	Sanctity of Life	Women Freedom	0.005



(a) Abortion (b) Immigration (c) Gun Control

Fig. 6. Distribution of the number of entity-to-topic attitude observations in the cases of Abortion, Immigration, and Gun Control. The figures depict a visible imbalanced between the observation numbers. Each observation is the frequency of entity attitudes expressed toward topics.

A higher value of α signifies higher polarization level. Finally, the ground-truth polarization-based topic ranked list between a Liberal corpus D^L and a Conservative corpus D^R is computed based on the corresponding ground-truth polarization scores $\alpha(D^L, D^R, u)$ for each topic tuple u. We use this ranked list to identify the most polarizing topics in the corpora and to compare the level of polarization between the Liberal and Conservative perspectives on those topics. Table 4 presents the ground-truth Abortion polarization topic ranked list (the rest are available in Appendix A).

4.3 Effectiveness of Capturing Entity Attitudes Toward Topics

In order to evaluate the accuracy of capturing the entity-level PK (Q1), we utilize the metrics of precision, recall, and F1-score for the comparison between the entity-to-topic values generated by our framework and the GT data. However, a limitation that may affect the final performance scores is the presence of a minimum number of comparison examples (24 entities for Abortion, 23 entities for Immigration, and 16 entities for Gun Control), as well as an imbalance in the distribution of attitude observations (i.e., sentences in news articles depicting attitudes from an entity towards a topic). This is illustrated in Figure 6, which presents the unbalanced distribution of entity-to-topic observations for GT entities.

To address the imbalance of observations for entity attitudes toward topics, we apply a weighting scheme during the performance evaluation, based on the standard deviation of observations. Specifically, for each entity $v \in V$ and topic $t \in T$, the weight w is calculated as $w = 1/(1 + std)$, where std represents the standard deviation of entity attitude observations for the given topic. These weights are then incorporated into the performance evaluation process. For each entity e, we calculate the precision, recall, and F1-score for each topic t in the set of topics. However, instead of using true positive, true negative, false positive, and false negative counters, the standard deviation weights for each topic are used to adjust the degree of uncertainty. As a result, topics with more balanced entity observations receive more weight during the evaluation.

Baseline: To assess the effectiveness of the proposed framework, we employ benchmarking against zero-shot classification models. Zero-shot classification is a technique that enables models to make predictions without prior exposure to a particular task. These models leverage the extensive knowledge they acquire during pre-training on large datasets, which empowers them to make informed decisions about tasks for which they were not explicitly trained. By comparing our framework against these zero-shot classifiers, we can gauge its ability to effectively discern stances across a wide range of untrained scenarios. The baseline models chosen for this evaluation are BART [49], DeBERTa [39], and Flan-T5 [21]. These choices are motivated by several key factors. First, these models consistently achieve state-of-the-art results across various language understanding tasks, including sentence and sentiment classification, making them robust benchmarks for assessing our framework's capabilities. Second, they offer diverse architectural and training differences, allowing us to evaluate how our framework performs in comparison to a range of neural models.

BART is a bidirectional encoder and left-to-right decoder, able to comprehend longer sentences. DeBERTa introduces a disentangled attention mechanism and enhanced mask decoder, enabling better understanding of nuanced language. Flan-T5 is a **Large Language Model (LLM)** trained with a focus on prompting and offers task-specific knowledge, providing insights into our framework's performance in task-oriented language understanding. These baseline models come in varying sizes and training paradigms, with BART and DeBERTa as zero-shot models and Flan-T5 as a large language instruction-based model.

Zero-shot models, like BART and DeBERTa, require two inputs to function: the premise and the hypotheses. The premise presents the context on which the questions expressed in the hypotheses will be applied. These models utilize their generalized training and text understanding to provide a complementary probability for the validity of each hypothesis. For example, consider the sentence "*Joe Biden supports women reproductive rights*", in which we would like to assess whether the entity of "Joe Biden" bears a supportive or oppositional stance towards the topic of "*Reproductive Rights*". The process of a zero-shot model to classify the stance is the following:

— **Premise**: "*Joe Biden supports women reproductive rights.*"
— **Hypothesis 1**: "*Joe Biden has a supportive stance towards reproductive rights.*"
— **Hypothesis 2**: "*Joe Biden has an oppositional stance towards reproductive rights.*"
— **Output 1**: Supportive stance: 95%
— **Output 2**: Oppositional stance: 5%

Examining the results, it is evident that the likelihood of "*Joe Biden*" having a "*Supportive*" stance toward the topic of "*Reproductive Rights*" is substantially higher than an "*Oppositional*" stance. Similarly, for LLMs such as Flan-T5, the input consists of an instruction prompt that defines the task, input information, and the desired output. An illustrative example of the prompt employed for the preceding scenario is as follows:

```
Given a SENTENCE, an ENTITY, and a TOPIC, output the stance of the ENTITY
towards the TOPIC in the premise of the SENTENCE as either Supportive,
or Oppositional.

SENTENCE: ''Joe Biden supports women reproductive rights.''
ENTITY: Joe Biden
TOPIC: Reproductive Rights

OUTPUT: Supportive
```

Evaluation Results: Table 5 presents the results of the performance comparison between the proposed framework and the three baseline models (BART, DeBERTa, and Flan-T5) in terms of

Table 5. Values for the Metrics of Precision, Recall, and F1 Score between our Framework (*OUR*), and the Baselines: BART (*BRT*), DeBERTa (*DBRT*), and Flan-T5 ($F-T5$), in Terms of the Entity-level PK Evaluation on Abortion, Immigration, and Gun Control

Use Case	Precision				Recall				F1 Score				Observations
	OUR	BRT	DBRT	F-T5	OUR	BRT	DBRT	F-T5	OUR	BRT	DBRT	F-T5	
Abortion	0.80	0.65	0.41	**0.82**	0.76	0.84	**0.86**	0.70	**0.74**	0.71	0.52	0.73	21
Immigration	**0.78**	0.73	0.31	**0.78**	0.80	0.76	**0.84**	0.75	**0.73**	0.66	0.41	0.71	21
Gun Control	0.79	**0.85**	0.48	**0.85**	0.72	0.74	**0.78**	0.60	0.71	**0.77**	0.56	0.68	23
Average	0.79	0.74	0.40	**0.81**	0.76	0.78	**0.82**	0.68	**0.73**	0.71	0.50	0.71	22

Precision, Recall, and F1 Score, for stance identification of entities towards topics. For the use case of Abortion, our framework achieves a precision of 0.80, which is slightly lower than Flan-T5's precision of 0.82 but higher than both BART (0.65) and DeBERTa (0.41). In terms of recall, our framework attains a score of 0.76, which is lower than BART (0.84) and DeBERTa (0.86) but higher than Flan-T5 (0.70). The F1 Score of our framework is 0.74, competitive with Flan-T5 (0.73) and higher than BART (0.52) and DeBERTa (0.73). For the use case of Immigration, our framework exhibits a precision of 0.78, the highest among all models, including Flan-T5 (0.78), BART (0.73), and DeBERTa (0.31). In recall, our framework achieves 0.80, higher than BART (0.76) and DeBERTa (0.84) and slightly lower than Flan-T5 (0.75). The F1 Score for our framework is 0.73, competitive with Flan-T5 (0.71), higher than BART (0.66), and significantly higher than DeBERTa (0.41). For the use case of Gun Control, our framework demonstrates a precision of 0.79, higher than BART (0.85) and DeBERTa (0.48) but slightly lower than Flan-T5 (0.85). In recall, our framework achieves 0.72, similar to BART (0.74), DeBERTa (0.78), and slightly lower than Flan-T5 (0.60).

Across all three use cases, the average performance of the proposed framework includes an average precision of 0.79, an average recall of 0.76, and an average F1 Score of 0.71. These results suggest that our framework demonstrates competitive performance, with precision and recall scores comparable to or better than state-of-the-art baselines for entity-level polarization stance detection in these specific topics. Moreover, it is important to consider the trade-offs between resource constraints and the desired outcomes when selecting a polarization stance detection method. While methods like Flan-T5, with their larger models and computational demands may offer higher precision, and DeBERTa may provide better recall, our framework aims to strike a balance between precision and recall, making it a practical choice for various scenarios. Our framework's resource-efficient entity-to-topic stance detection component operates effectively without the need for resource-intensive models or GPUs. This lightweight approach not only improves accessibility and integration but also proves advantageous in resource-constrained settings where extensive computational resources may not be readily available or practical.

4.4 Alignment between Cohesive Fellowships and Party Manifestos

For our fellowship-level PK (Q2) evaluation, we concentrated on fellowships with strong ideological ties to the Democratic or Republican parties. To determine the political affiliations of entities within the SAG, we extracted data from Wikipedia's Infobox tab using web scraping. This tab provides structured information, notably the *Political Party* field [41]. Entities were categorized based on their affiliation with either the Democratic or Republican party, extracted from this field. We considered the majority political leaning of the entities, and if it was $\geq 75\%$ Democrat or Republican, we regarded the fellowship as ideologically cohesive. During the voting process, we excluded entities which no clear political party affiliation was available in the Infobox. The resulting fellowships are characterized for each topic t by the following:

— i: Unique identifier for each fellowship.
— p: The fellowships affiliated political party, i.e., Democratic or Republican.

Table 6. Results of the 3-fold Cross-validation for the Evaluation of the Alignment between Fellowships and their Respective Party Manifestos, as Described in the Methodology Section

Case Study	Precision	Recall	F1 Score	AUC	Accuracy
Abortion	0.74	0.74	0.74	0.72	0.79
Gun Control	0.81	0.81	0.81	0.81	0.81
Immigration	0.99	0.99	0.99	0.99	0.99
Average	0.87	0.87	0.87	0.84	0.87

The table shows the performance metrics of precision, recall, F1-score, and accuracy for each of the case studies of Abortion, Gun Control and Immigration, as well as an average score across the case studies.

— coh_{att}: Common attitude percentage for topic t.
— s_t: Number of supportive attitude observations for topic t.
— o_t: Number of oppositional attitude observations for topic t.

In order to evaluate the attitudinal alignment of each fellowship with its respective political party, we employ a supervised approach. An estimator, denoted as $\phi(i, t, p, coh_{att}, s_t, o_t)$, is trained to predict the attitude of fellowship i towards topic t based on its respective party manifestos. The manifesto attitude is represented as a binary value, with 1 indicating a supportive attitude and 0 indicating an oppositional attitude. During the evaluation, estimator ϕ is implemented as a logistic regression model. To evaluate the performance of the model, we employ 3-fold cross validation. This technique is used to evaluate the generalization performance of the model by training it on different subsets of the data and testing it on the remaining subset, with the process being repeated three times. Similarly to the entity-level evaluation, the performance of the model is quantified via the metrics of precision, recall, F1-score, and accuracy, as well as AUC. If the model is able to achieve high performance, it would indicate its ability to accurately predict the attitude of a fellowship towards a topic and that the extracted fellowships have a high degree of alignment between their attitudes and the attitudes of their respective party manifestos. This would suggest that our framework effectively captures the attitudinal and ideological characteristics of the fellowships and aligns them with the official stance of their respective parties.

Evaluation Results: Table 6 depicts the average 3-cross validation results of group-level PK evaluation regarding the alignment of extracted fellowships with their respective party manifestos, for each of the subjects of Abortion, Immigration, and Gun Control. As it is shown, the model performed well in all three case studies, with an average performance of 0.87 for all metrics. The case study of Immigration achieved the highest performance with 0.99 for all metrics, followed by Gun Control with 0.81 and Abortion with 0.74. It is important to notice that the average performance of the model is high and consistent for all the case studies. This indicates that the approach is able to accurately predict the alignment of the fellowships with their respective party manifestos, which is a good indication that the fellowship extraction process, performed by our framework, is able to extract ideologically cohesive groups of entities. The results of the manifesto alignment are also presented in a more detailed manner, by providing the degree of alignment between each of the four largest fellowships (as shown in Table 6) and the manifestos of the Democratic and Republican parties, respectively, for the topics related to Abortion, Immigration, and Gun Control. Table 7 illustrates this information for the topic of Abortion. Similar tables for Immigration and Gun Control can be found in Appendix A.

Regarding the case of Abortion, as depicted in Table 7, fellowships 1 and 2 are labeled as Republican and have a high degree of alignment with the Republican party's manifesto for all the topics, as all their attitudes match those of the Republican manifesto. Fellowships 3 and 4 are labeled

Table 7. Degree of Alignment between the Attitudes of the Four Largest Fellowships and the Manifestos of the Democratic and Republican Parties for the Topic of Abortion

Abortion																
	Fellowship 1				Fellowship 2				Fellowship 3				Fellowship 4			
Topic	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**
Abortion Funding	0.60	−1.00	−1.00	1.00	1.00	−1.00	−1.00	1.00	0.75	1.00	1.00	1.00	0.58	1.00	1.00	1.00
Abortion Pr. Economy	0.50	−1.00	−1.00	1.00	0.50	−1.00	−1.00	1.00	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Anti-Abortion	0.56	−1.00	1.00	0.00	0.54	−1.00	1.00	0.00	0.50	−1.00	−0.93	1.00	0.70	−1.00	−0.93	1.00
Birth Control	0.55	−1.00	−1.00	1.00	1.00	−1.00	−1.00	1.00	0.50	1.00	1.00	1.00	0.51	1.00	1.00	1.00
Health Care	–	–	−0.88	–	–	–	−0.88	–	–	–	1.00	–	1.00	1.00	1.00	1.00
Hobby Lobby	–	–	0.00	–	–	–	0.00	–	–	–	−1.00	–	–	–	−1.00	–
Late-Term Abortion	0.55	−1.00	−1.00	1.00	1.00	−1.00	−1.00	1.00	0.56	1.00	0.00	–	0.53	1.00	0.00	–
Life Protection	–	–	1.00	–	–	–	1.00	–	–	–	0.00	–	–	–	0.00	–
Planned Parenthood	0.68	−1.00	−1.00	1.00	0.55	−1.00	−1.00	1.00	0.81	−1.00	1.00	0.00	0.56	1.00	1.00	1.00
Pregnancy Centers	1.00	−1.00	−1.00	1.00	–	–	−1.00	–	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pro-Choice	0.75	−1.00	−1.00	1.00	0.56	−1.00	−1.00	1.00	0.67	1.00	1.00	1.00	0.70	1.00	1.00	1.00
Pro-Life	1.00	1.00	1.00	1.00	–	–	1.00	–	1.00	1.00	0.00	–	1.00	1.00	0.00	–
Reproduction Rights	1.00	−1.00	−1.00	1.00	1.00	−1.00	−1.00	1.00	0.71	1.00	1.00	1.00	0.67	1.00	1.00	1.00
Right of Human Life	0.56	−1.00	1.00	0.00	–	–	1.00	–	0.95	1.00	0.00	–	1.00	−1.00	0.00	–
Roe v. Wade	–	–	0.00	–	–	–	0.00	–	–	–	1.00	–	–	–	1.00	–
Sale of Fetal Tissue	–	–	−1.00	–	–	–	−1.00	–	0.50	1.00	0.00	–	–	–	0.00	–
Sanctity of Life	–	–	1.00	–	–	–	1.00	–	1.00	1.00	0.00	–	–	–	0.00	–
Sexual Assault Victims	0.72	−1.00	0.00	–	0.63	−1.00	0.00	–	0.56	1.00	0.00	–	0.57	−1.00	0.00	–
Stem Cell Research	0.65	−1.00	−0.82	1.00	1.00	−1.00	−0.82	1.00	1.00	1.00	0.00	–	0.56	−1.00	0.00	–
Women Freedom	–	–	−1.00	–	–	–	−1.00	–	–	–	1.00	–	–	–	1.00	–
Average	0.70			0.83	0.78			0.89	0.75			0.88	0.74			1.00

as Democratic and also have a high degree of alignment with the Democratic party's manifesto for most of the topics. However, for some topics such as *Abortion Funding* and *Anti-Abortion*, the alignment is not perfect as the attitudes of fellowship 4 are not in line with the Democratic party's manifesto. This indicates that the fellowships are not entirely homogeneous, and some members may have different views than the majority of the group. Overall, the results suggest that the fellowships are highly ideologically cohesive with their respective parties, with high alignment between their attitudes and manifestos.

In the case of Immigration (see Appendix A), fellowships 1 and 2 are both labeled as Republican and have a high degree of alignment with their party manifesto. Similarly, fellowships 3 and 4 are both labeled as Democratic. A common knowledge is that, the Republican party tends to be against *Amnesty*, *Asylum*, and *Birthright Citizenship & 14th Amendment* (*Birth. Citiz. & 14th Am.*) which the fellowships reflect the same attitude. On the other hand, Democratic party tends to be in favor of the above topics, which is also the case for the Democratic-labeled fellowships. The degree of manifesto alignment for the fellowships indicates that they are highly ideologically cohesive with their respective parties.

Consequently, for Gun Control, Republican-labeled fellowships (fellowships 1 and 2) have a high degree of alignment with the Republican manifesto for most of the topics (see Appendix A). For example, for the topic of *Assault Weapon*, both fellowships have a supportive attitude, which is aligned with the Republican manifesto. Similarly, for the topic of *Background Check*, both fellowships have an oppositional attitude, also aligned with the Republican manifesto. On the other hand, the Democratic fellowships (fellowships 3 and 4) have a varying degree of alignment with the Democratic manifesto. For some topics, both fellowships have an attitude that aligns with the manifesto (e.g., *Background Check*). However, for other topics, the fellowships have an attitude that contradicts with the manifesto (e.g., *Assault Weapon*). Overall, the values of Table 12 suggests that the fellowships extracted by our framework align well with their respective party manifestos, indicating that the fellowships are highly ideologically cohesive. This is consistent with the assumption that the fellowships were filtered based on high ideological cohesiveness.

4.5 Topic Identification and Polarization Ranking

For assessing the performance of the proposed framework at the topic-level, we employ two evaluation types. The first type is Topic Identification Accuracy, which assesses how accurately the framework is able to identify the topics present in the articles, as identified by Roy and Goldwasser

2020 [65]. The second type is Topic Polarization Ranking, which evaluates the framework's ability to accurately depict the ranking of polarization for the topics identified in the first evaluation step.

Topic Identification Accuracy: To assess the accuracy of Topic Identification in our framework, we start by aligning the topical clusters it generates with predefined topics (subframes) for each case study. This alignment process involves converting both the NPs from our clusters and the vocabulary of the predefined topics into semantic word vectors using the BERT model. We then calculate the centroid of each cluster and topic by averaging the vectors of their respective NPs and n-grams. The relationship between our clusters and the predefined subframes is determined by computing the cosine similarity between their centroids. If this similarity score for a cluster and a subframe is 0.80 or higher, we consider the cluster to be directly related to that subframe. This approach enables us to map our clusters to specific subframes and calculate the percentage accuracy of our framework in correctly identifying these topics.

Topic Polarization Ranking Agreement: The evaluation of topic-level PK involves measuring the agreement between the GT list of polarizing topics for each case study and the list created by our framework. We determine the ranking of polarizing topics using the GT data, which was created during the annotation process. The highly ideologically cohesive fellowships are taken into account in the creation of the polarizing topic list by our framework, using the polarization index ranking function for each topic to reflect the conflict between the fellowships in the dipoles.

Performance Metric: Traditionally, to measure the similarity between two rank lists, ranking agreement metrics are used. For the topic polarization ranking agreement, we use **Ranked Bias Overlap (RBO)** [83]. RBO is an intersection-based ranking agreement measure, compared to the traditional correlation-based measures (e.g., Spearman ρ and Kendall τ). We employ RBO as the consecutive differences between the ranked topics suggest that the rankings are prone to small changes. RBO takes values from 0 to 1, with 1 indicating a full overlap (practically the same), and 0 indicating that the ranked lists are disjoint. RBO accepts a parameter ρ, which corresponds to the top-k elements in the list that contribute the most to the scoring. We compare the different lists with $p = 0.3$, meaning that we consider the top-3 elements, similar to what He et al. 2021 used in their evaluation [40].

Baseline: In order to establish a baseline for evaluating the accuracy of depicting topic polarization rankings, we employ two state-of-the-art methodologies, **PaCTE (Partisanship-aware Contextualized Topic Embeddings)**, developed by He et al. 2021 [40], and **LOE (Leave-Out Estimator)** developed by Demszky et al. 2019 [24]. PaCTE uses a transformer-based model to learn embeddings of news sources from Left and Right perspectives, and trains the model to capture topic-specific representations of political leaning. LOE, on the other hand, uses the frequency of topic tokens in Left and Right news article sources to calculate polarizing topics, under the assumption that Left or Right tokens are drawn from a multinomial logit model. The estimated partisanship is produced as the polarization score of a topic between Left and Right. In comparison with topic modeling [36] and keyword extraction [62] methods, PaCTE and LOE are able to capture intricate aspects of polarization inherent in the topics discussed within a news corpus. This makes PaCTE and LOE particularly well-suited for our evaluation objectives.

Evaluation Results: The topic-level PK evaluation results are depicted in Tables 8 and 9. As it appears in Table 9, our framework performs well in the Topic Polarization Ranking Agreement evaluation, with an average score of 0.8140 across the three case studies of Abortion, Immigration, and Gun Control. Specifically, it achieves the highest score for the Gun Control case study at 0.9100, followed by scores of 0.7389 for Immigration and 0.7929 for Abortion. However, in the **Topic Identification Accuracy (TIA)** evaluation (see Table 8), our framework performs relatively poor

Table 8. Results for the Topic Identification Accuracy

Case Study	Our Framework	PaCTE	LOE
Abortion	0.88	**1.00**	**1.00**
Immigration	0.60	**0.80**	**0.80**
Gun Control	**0.83**	**0.83**	**0.83**
Average	0.77	**0.87**	**0.87**

Table 9. RBO Measurements for the Topic Polarization Ranking Agreement

Case Study	Our Framework	PaCTE	LOE
Abortion	**0.7929**	0.6833	0.6196
Immigration	**0.7389**	0.7205	0.3860
Gun Control	**0.9100**	0.8433	0.4933
Average	**0.8140**	0.7490	0.4996

compared to PaCTE and LOE, with an average score of 0.77 across the three case studies. Its TIA scores are 0.88 for Abortion, 0.60 for Immigration, and 0.83 for Gun Control.

The reason for the poorer performance in TIA is likely due to the nature of the data utilized by our framework to determine the topics. Unlike PaCTE and LOE, which look at a larger corpus of texts, our framework only considers Noun Phrases (NPs) where entities express positive or negative attitudes, thus limiting its scope for topic identification. Despite the low TIA, the overall results indicate that our framework is able to accurately rank the polarization of topics.

4.5.1 Discussion. The Polarization Knowledge evaluation results indicate that our framework is able to effectively capture and analyze the polarization knowledge in news media. At the entity-level, the model performed well overall, with a balanced performance across the three case studies of Abortion, Immigration, and Gun Control. The use of a more sophisticated and direct approach to capturing entity-to-topic sentiment attitudes, as well as the weighting scheme applied during performance evaluation, likely contributed to the improved performance. At the fellowship-level, the model performed consistently well with an average performance of 0.87 across all metrics, indicating that the approach is able to accurately predict the alignment of the fellowships with their respective party manifestos. This suggests that the fellowship extraction process performed is able to extract ideologically cohesive groups of entities. At the topic-level, the model performed well in the Topic Polarization Ranking Agreement evaluation, with an average score of 0.8140 across the three case studies. However, the results in the Topic Identification Accuracy evaluation were relatively poor, with an average score of 0.77. This is likely due to the limited scope of the data used for topic identification. Despite this, the overall results indicate that our framework is able to accurately rank the polarization of topics and it can be effectively used by the research community for the analysis of political polarization in text data.

5 Conclusion

In this study, we introduce a framework for the unsupervised and domain-agnostic modeling, extraction, and quantification of Polarization Knowledge (PK) from digital news media. By combining structural and content-based analyses, our framework effectively captures the nuanced dynamics of polarization within specific domains. Additionally, we propose a multi-level evaluation methodology that assesses PK at the entity, fellowship, and topic levels, offering a comprehensive view of polarization across different dimensions.

Our case studies on Abortion, Immigration, and Gun Control demonstrate the framework's effectiveness. At the entity level, the system outperforms traditional classifiers by accurately linking key entities to polarizing topics. At the fellowship level, it successfully identifies group alignments and their respective ideological stances, revealing the underlying structures of polarized communities. Although the framework efficiently ranks topics by their polarization levels, we recognize the potential for further refinement in topic identification to enhance accuracy.

By making the framework available as an open-source project, we provide researchers with a valuable tool for systematically analyzing polarization without relying on predefined labels or groupings. This resource can be utilized to explore societal divisions and investigate related phenomena such as hate speech, misinformation, and media manipulation. Ultimately, the proposed framework facilitates the further advancement of polarization analysis in digital media, offering a structured approach for understanding this complex phenomenon.

References

[1] Alan I. Abramowitz and Steven Webster. 2016. The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies* 41 (2016), 12–22. https://doi.org/10.1016/j.electstud.2015.11.001

[2] Lada A. Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*. ACM, 36–43.

[3] L. Akoglu. 2014. Quantifying political polarity based on bipartite opinion networks. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* (2014), 2–11.

[4] Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management* 58, 4 (2021), 102597.

[5] Clio Andris, David Lee, Marcus Hamilton, Mauro Martino, Christian Gunning, and John Selden. 2015. The rise of partisanship and super-cooperators in the U.S. House of Representatives. *PLOS ONE* 10 (2015).

[6] Sinan Aral. 2020. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health–and How We Must Adapt*. Currency.

[7] Sinan Aral and Dean Eckles. 2019. Protecting elections from social media manipulation. *Science* (2019).

[8] Samin Aref and Zachary Neal. 2020. Detecting coalitions by optimally partitioning signed networks of political collaboration. *Scientific Reports* 10 (2020), 1–10.

[9] Samin Aref and Mark Wilson. 2019. Balance and frustration in signed networks. *Journal of Complex Networks* 7 (2019).

[10] Ramnath Balasubramanyan, William Cohen, Doug Pierce, and David Redlawsk. 2012. Modeling polarizing topics: When do different political communities respond differently to the same news? *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media* (2012).

[11] Pablo Barberá. 2020. *Social Media, Echo Chambers, and Political Polarization*. Cambridge University Press, 34–55. https://doi.org/10.1017/9781108890960.004

[12] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. 2019. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters* (2019).

[13] Lori D. Bougher. 2017. The correlates of discord: Identity, issue alignment, and political hostility in polarized America. *Political Behavior* 39 (2017), 731–762. https://api.semanticscholar.org/CorpusID:152043932

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

[15] Sam Cabral. 2021. Capitol riots: Panel of Americans 'shocked' and 'disgusted'. *BBC* (2021).

[16] Drew Calvert. 2017. The psychology behind fake news. *Northwestern Kellogg* (2017).

[17] D. Cartwright and F. Harary. 1956. Structural balance: A generalization of Heider's theory. *Psychological Review* (1956).

[18] Pew Research Center. 2016. *Views of Parties' Positions on Issues, Ideologies, Political Animosity Report*.

[19] Ted Hsuan Yun Chen, Ali Salloum, Antti Gronow, Tuomas Ylä-Anttila, and Mikko Kivelä. 2021. Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere. *Global Environmental Change* 71 (2021), 102348. https://doi.org/10.1016/j.gloenvcha.2021.102348

[20] Eunsol Choi, Hannah Rashkin, Luke Zettlemoyer, and Yejin Choi. 2016. Document-level sentiment inference with social, faction, and discourse context. In *Proc. of 54th ACL*.

[21] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2022. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416* (2022).

[22] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on Twitter. *Fifth International AAAI Conference on Weblogs and Social Media.*

[23] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113, 3 (2016), 554–559.

[24] Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*

[25] J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *North American Chapter of the Assoc. for Computational Linguistics* (2019).

[26] James N. Druckman and Samara Klar. 2021. Affective polarization, local contexts and public opinion in America. *Nature Human Behaviour* 5, 1 (2021), 28–38.

[27] Alan Eckhardt, Juraj Hreško, Jan Procházka, and Otakar Smrf. 2014. Entity linking based on the co-occurrence graph and entity probability. In *Proceedings of the First International Workshop on Entity Recognition & Disambiguation.* 37–44.

[28] Pouya Esmailian and Mahdi Jalili. 2015. Community detection in signed networks: The role of negative ties in different scales. *Scientific Reports* (2015).

[29] Morris P. Fiorina and Samuel J. Abrams. 2008. Political polarization in the American public. *Annual Review of Political Science* 11, 1 (2008), 563–588.

[30] Santo Fortunato and Marc Barthelemy. 2007. Resolution limit in community detection. *Proceedings of the National Academy of Sciences* 104, 1 (2007), 36–41.

[31] Kiran Garimella, Gianmarco F. Morales, Aristides Gionis, and Michael Mathioudakis. 2015. Quantifying controversy in social media. *Trans. Soc. Comput.* (2015).

[32] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. 2021. Political polarization in online news consumption. *ICWSM* (2021).

[33] Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media.*

[34] Matthew Gentzkow, Jesse M. Shapiro, and Daniel F. Stone. 2015. Media bias in the marketplace: Theory. In *Handbook of Media Economics.* Vol. 1. Elsevier, 623–645.

[35] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, my echo chamber, and I: Introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference.* International World Wide Web Conferences Steering Committee, 823–831.

[36] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[37] P. H. Guerra, Wagner Meira Jr., Claire Cardie, and R. Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013* (01 2013), 215–224.

[38] Zellig S. Harris. 1954. Distributional structure. *WORD* 10, 2-3 (1954), 146–162.

[39] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. DeBERTa: Decoding-enhanced BERT with disentangled attention. *CoRR* abs/2006.03654 (2020). arXiv:2006.03654 https://arxiv.org/abs/2006.03654

[40] Zihao He, Negar Mokhberian, Antonio Camara, Andres Abeliuk, and Kristina Lerman. 2021. Detecting polarized topics in COVID-19 news using partisanship-aware contextualized topic embeddings. *EMNLP* (2021).

[41] M. Herrmann and H. Döring. 2021. Party positions from Wikipedia classifications of party ideology. *Political Analysis* (2021).

[42] Philip N. Howard. 2020. *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives.* Yale University Press.

[43] Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. The origins and consequences of affective polarization in the United States. *Annual Review of Political Science* 22, Volume 22, 2019 (2019), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

[44] Shanto Iyengar, Gaurav Sood, and Yphtach Lelkes. 2012. Affect, not ideology a social identity perspective on polarization. *Public Opinion Quarterly* 76 (2012), 405–431. https://api.semanticscholar.org/CorpusID:8592287

[45] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications* 78, 11 (2019), 15169–15211.

[46] Kati Kish Bar-On, Eugen Dimant, Yphtach Lelkes, and David G. Rand. 2024. Unraveling polarization: Insights into individual and collective dynamics. *SSRN* (2024).

[47] David Krackhardt and Robert N. Stern. 1988. Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly* (1988), 123–140.
[48] Amber Hye-Yon Lee. 2022. Social trust in polarized times: How perceptions of political polarization affect Americans' trust in each other. *Political Behavior* 44, 3 (2022), 1533–1554.
[49] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. (2019).
[50] Tim Loughran and Bill McDonald. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54, 4 (2016), 1187–1230.
[51] Ulrike von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing 17(4)* (2007).
[52] Lilliana Mason. 2015. "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science* 59, 1 (2015), 128–145. https://doi.org/10.1111/ajps.12089 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12089
[53] Yelena Mejova, Amy X. Zhang, Nicholas Diakopoulos, and Carlos Castillo. 2014. Controversy and sentiment in online news. *J'14: Computational Journalism Symposium (2014)* (2014).
[54] Alfredo Morales, J. Borondo, J. Losada, and Rosa Benito. 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos* 25 (2015).
[55] M. E. J. Newman. 2016. Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *Phys. Rev. E* (2016).
[56] Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2 (1998), 175 – 220. https://api.semanticscholar.org/CorpusID:8508954
[57] Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. European Language Resources Association (ELRA), Portorož, Slovenia, 1659–1666. https://aclanthology.org/L16-1262
[58] Alberto Parravicini, Rhicheek Patra, Davide Bartolini, and Marco Santambrogio. 2019. Fast and accurate entity linking via graph embedding. In *Proc. of GRADES-NDA*. ACM.
[59] Demetris Paschalides, Chrysovalantis Christodoulou, Kalia Orphanou, Rafael Andreou, Alexandros Kornilakis, George Pallis, Marios D. Dikaiakos, and Evangelos Markatos. 2021. Check-It: A plugin for detecting fake news on the web. *Online Social Networks and Media* 25 (Sep. 2021), 100156. https://doi.org/10.1016/j.osnem.2021.100156
[60] D. Paschalides, G. Pallis, and M. Dikaiakos. 2021. POLAR: A holistic framework for the modelling of polarization and identification of polarizing topics in news media. *ASONAM* (2021).
[61] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online learning of social representations. In *Proc. of the 20th ACM SIGKDD*.
[62] Jakub Piskorski, Nicolas Stefanovitch, Guillaume Jacquet, and Aldo Podavini. 2021. Exploring linguistically-lightweight keyword extraction techniques for indexing news articles in a multilingual set-up. In *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*. 35–44.
[63] Pedro Ramaciotti Morales, Jean-Philippe Cointet, Gabriel Zolotoochin, Antonio Fernandez Peralta, Gerardo Iñiguez, and Armin Pournaki. 2022. Inferring attitudinal spaces in social networks. *Social Network Analysis and Mining* 13 (12 2022). https://doi.org/10.1007/s13278-022-01013-4
[64] Lee Ross and Andrew Ward. 1996. Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and Knowledge* 103 (1996), 103–135.
[65] Shamik Roy and Dan Goldwasser. 2020. Weakly supervised learning of nuanced frames for analyzing polarization in news media. In *Proc. of EMNLP*. ACL.
[66] Nicolay Rusnachenko, Natalia Loukachevitch, and Elena Tutubalina. 2019. Distant supervision for sentiment attitude extraction. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. INCOMA Ltd., Varna, Bulgaria, 1022–1030.
[67] Ali Salloum, Ted Hsuan Yun Chen, and Mikko Kivelä. 2022. Separating polarization from noise: Comparison and normalization of structural polarization measures. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 115 (April 2022), 33 pages. https://doi.org/10.1145/3512962
[68] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM Trans. Intell. Syst. Technol.* (2019).
[69] Wei Shen, Jianyong Wang, and Jiawei Han. 2015. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering* 27, 2 (2015), 443–460.
[70] John Sides, Chris Tausanovitch, and Lynn Vavreck. 2020. The politics of COVID-19: Partisan polarization about the pandemic has increased, but support for health care reform hasn't moved at all. *Harvard Data Science Review* (2020).

[71] Barea Sinno, Bernardo Oviedo, Katherine Atwell, Malihe Alikhani, and Junyi Jessy Li. 2022. Political ideology and polarization: A multi-dimensional approach. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, WA, USA, 231–243. https://doi.org/10.18653/v1/2022.naacl-main.17
[72] Olivia Solon. 2016. Facebook's failure: Did fake news and polarized politics get Trump elected? *The Guardian* (2016).
[73] Marianna Spring and Lucy Webster. 2019. European elections: How disinformation spread in Facebook groups. *BBC* (2019).
[74] Cass R. Sunstein. 1999. The law of group polarization. *University of Chicago Law School* 91 (1999).
[75] H. Tajfel and J. Turner. 1979. An integrative theory of intergroup conflict. *The Social Psych. of Intergroup Rel.* 33 (1979).
[76] Szymon Talaga, Massimo Stella, Trevor Swanson, and Andreia Teixeira. 2023. Polarization and multiscale structural balance in signed networks. *Communications Physics* 6 (12 2023). https://doi.org/10.1038/s42005-023-01467-8
[77] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. 2016. A survey of signed network mining in social media. *ACM CSUR* 49, 3 (2016), 1–37.
[78] Amanda Taub and Max Fisher. 2018. Facebook fueled anti-refugee attacks in Germany, new research suggests. *NYTimes* (2018).
[79] Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 142–147.
[80] John C. Turner. 1981. Towards a cognitive redefinition of the social group. *Current Psychology of Cognition* (1981).
[81] Andrea Volkens, Judith Bara, and Ian Budge. 2009. Data quality in content analysis. The case of the comparative manifestos project. *Historical Social Research/Historische Sozialforschung* (2009), 234–251.
[82] Andrew Scott Waugh, Liuyi Pei, James H. Fowler, Peter J. Mucha, and Mason A. Porter. 2011. Party polarization in congress: A network science approach. *arXiv preprint arXiv:0907.3509* (2011).
[83] William Webber, Alistair Moffat, and Justin Zobel. 2010. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.* 28, 4 (2010), 38 pages.
[84] Ingmar Weber, Venkata Garimella, and Alaa Batayneh. 2013. Secular vs. Islamist polarization in Egypt on Twitter. 290–297.
[85] Theresa Wilson, Janyce Wiebe, and Claire Cardie. 2017. MPQA opinion corpus. *Handbook of Linguistic Annotation*, Springer, (2017), 813–832.
[86] Vikas Yadav and Steven Bethard. 2018. A survey on recent advances in named entity recognition from deep learning models. In *COLING*.
[87] Fabiana Zollo, Alessandro Bessi, Michela Del Vicario, Antonio Scala, Guido Caldarelli, Louis Shekhtman, Shlomo Havlin, and Walter Quattrociocchi. 2017. Debunking in a world of tribes. *PLoS ONE* 12 (2017).

A Appendix

A.1 Entity-pair Sentiment Attitude Calculation

ALGORITHM 1: Entity-pair Sentiment Attitude Calculation

Input: s_i sentence, v_x and v_y entities
Output: $attitude_{xy} \in \{POSITIVE, NEUTRAL, NEGATIVE\}$
function ATT(s_i, v_x, v_y)
    $attitude_{xy} \leftarrow NEUTRAL$
    $dep\_paths \leftarrow$ extract_dependency_paths(s_i, v_x, v_y) ▷ Identify the paths connecting v_x and v_y in s_i
    for all *path* in *dep_paths* **do**
        $sentiment\_score \leftarrow$ calculate_sentiment_score(*path*)
        $attitude_{xy} \leftarrow$ update_attitude($attitude$, $sentiment\_score$)
    end for
    return $attitude_{xy}$
end function

A.2 Quantifying Dipole Topic Polarization

ALGORITHM 2: Quantifying Dipole Topic Polarization

Input: D_{ij} dipole, $t_z \in T_{D_{ij}}$ set of topical NPs, S_{ij} sentences where D_ij entities and t_z NPs occur
Output: μ polarization index

$A_{t_z} \leftarrow \{\}$ ▷ Initialize set of sentiment attitudes for topic t_z
for all $v_x \in D_{ij}$ **do** ▷ Loop through entities in dipole
  $attitude_x \leftarrow NEUTRAL$
  for all $np_y \in t_z$ **do** ▷ Loop through the NPs of topic
    $v\_np\_sentences \leftarrow get\_relevant\_sentences(S_{ij}, v_x, np_y)$ ▷ Filter sentences where v_x and np_y co-occur
    for all $s_u \in v\_np\_sentences$ **do**
      $sentiment\_score \leftarrow \text{ATT}(s_u, v_x, np_y)$ ▷ Calculate sentiment attitude from entity v_x to np_y
      $attitude_x$.update_attitude($attitude_x$, $sentiment\_score$)
    end for
  end for
  A_{t_z}.add($attitude_x$)
end for
$\mu \leftarrow calculate\_polarization\_index(A_{t_z})$ ▷ Calculate polarization index of A_{t_z}
return μ

A.3 Abortion Subframe Descriptions

- **Health Care**: Affordable Care Act, healthcare facilities, health insurance, their coverage, etc.
- **Abortion Provider Economy**: Statistics, services, profits of abortion providers like Planned Parenthood.
- **Abortion Funding**: Source of funding; granting or cutting funding for abortion providers like Planned Parenthood.
- **Reproduction Right**: Reproduction rights and women's access to reproductive healthcare.
- **Right of Human Life**: Fetus in the womb has the same right of life as a grown human.
- **Hobby Lobby**: Court's exemption for corporations to provide contraceptives if it conflicts with their religious belief.
- **Late Term Abortion**: Discuss ban and regulation on abortion after later stages of pregnancy.
- **Roe v. Wade**: Implications of the 1973 landmark decision of the U.S. Court that ensures the right to choose.
- **Stem Cell Research**: Implications using stem cell, embryonic cell and fetal tissue.
- **Sale of Fetal Tissue**: Abortion providers donation or selling of the fetal tissue and body parts from aborted babies.
- **Sexual Assault Victims**: Any kind of sexual offense against women and pregnancies resulted from that.
- **Birth Control**: Birth control measures and access to those.
- **Sanctity of Life**: The holiness of life from a religious and moral perspective and the evil of abortion.
- **Women Freedom**: Advocating women freedom or talking about oppression on women, from a moral perspective.
- **Planned Parenthood**: Abortion services provided by Planned Parenthood.

— **Pregnancy Centers**: Pregnancy services provided by pregnancy care centers, pregnancy crisis centers, etc.
— **Life Protection**: Abortion kills human being and they should be protected.
— **Pro-Life**: Addressing of any personality, movement or legislation as supporting life.
— **Anti-Abortion**: Addressing of any personality, movement or legislation as opposing abortion instead of as pro-life.
— **Pro-Choice**: Addressing of any personality, movement or legislation as supporting abortion and the right to choose.

A.4 Gun Control Subframe Descriptions

— **Gun Buyback Program**: Gun buyback program and its effects.
— **Gun Business**: Licensed gun store owners; gun business industry.
— **Gun Research**: Research on gun violence and how to control it; funding on gun research.
— **Mental Health**: Mental illness; importance of providing mental health care.
— **Gun Homicide**: Statistics on deaths due to gun violence.
— **Ban on Handgun**: Banning handgun and its effects.
— **2nd Amendment**: 2nd Amendment which ensures the right to self-defense and allows citizens to carry guns.
— **Concealed Carry Reciprocity Act**: Concealed carry reciprocity act and its effects and implications.
— **Gun Control to Restrain Violence**: Violence-restraining gun control measures.
— **Illegal Gun**: Illegal possession of gun; gun trafficking, etc.
— **Gun Show Loophole**: Loophole in the gun shows that allows criminals to get guns.
— **Background Check**: Necessity of background check and ways to ensure it while selling guns.
— **Terrorist Attack**: Threats of terrorist attack.
— **Assault Weapon**: Debate over the definition of assault weapon and which ones are needed to be banned.
— **White Identity**: Focusing on white racial identity of a person; white supremacy, etc.
— **Person of Color**: Identity focusing on a person of color racial identity.
— **School Safety Measures**: To ensure school safety; arming teachers; control guns to reduce violence in schools, etc.
— **Right to Self-Defense**: God given right to self defense; necessity of carrying guns for self-defense, etc.
— **Stop Gun Crime**: Urge to stop gun violence; expression of solidarity with mass shooting victims, etc.

A.5 Immigration Subframe Descriptions

— **Minimum Wage**: Wage inequality and discussion on raising the minimum wage.
— **Salary Stagnation**: Reasons of salary stagnation and how to overcome those.
— **Wealth Gap**: Wealth gap among the classes in the society; profits by large organizations, etc.
— **Cheap Labor Availability**: Cheap labor availability and its effects.
— **Taxpayer Money**: Taxpayer money and the facilities they get or are deprived of, such as social security.
— **Racism and Xenophobia**: Addressing of someone/something racist and xenophobic in a discussion.
— **Merit-based Immigration**: Discussion on merit based immigration system.

— **Human Right**: Necessity of protecting human and civil rights; their violations.
— **Asylum**: Implications of granting asylum to the asylum seeking migrants.
— **Refugee**: Political refugees from various countries.
— **Birth Citizenship and 14th Amendment**: Birthright citizenship; 14th Amendment; citizenship granting procedure.
— **Deportation: Illegal Immigrants**: Necessity of deportation of the illegal immigrants.
— **Deportation: General**: Procedure, policy and way to deport the undocumented immigrants.
— **Detention**: Detention facilities; detention procedure and the state of the detainees.
— **Terrorism**: Threats of terrorism by foreign nationals.
— **Border Protection**: Border wall; border patrol and other measures to secure the border.
— **Amnesty**: Implications and procedure of granting amnesty to the undocumented immigrants.
— **DREAM Act**: DREAM Act, its implications; DREAMers and procedure of their path to citizenship.
— **Family Separation Policy**: Family separation policy and its effects; separation of children from their families.
— **DACA**: DACA policy that protects the individuals from deportation who came to the USA as children.
— **Racial Identity**: Discussion on a topic by focusing on the race.
— **Born Identity**: Discussion on a topic by addressing the born identity, such as, "foreign born".

A.6 Topic Polarization Rank List

Table 10. Topic Polarization Rank List GT for the Cases of Immigration

Topic Index	Immigration Topic Tuples		Normalized α
7	Racial Identity	Born Identity	0.0111
9	Born Identity	Racism and Xenophobia	0.0072
5	Amnesty	DACA	0.0060
2	Wealth Gap	Minimum Wage	0.0040
8	Racial Identity	Racism and Xenophobia	0.0040
1	Minimum Wage	Salary Stagnation	0.0036
4	Cheap Labor Availability	Wealth Gap	0.0036
3	Cheap Labor Availability	Minimum Wage	0.0004
6	Family Separation Policy	DACA	0.0003

Table 11. Topic Polarization Rank List GT for the Cases of Gun Control

Topic Index	Gun Control Topic Tuples		Normalized α
6	Right to Self-Defense	Stop Gun Crime	1.000
4	Gun Control to Restrain Violence	Concealed Carry Reciprocity Act	0.917
3	Gun Control to Restrain Violence	2nd Amendment	0.617
2	Ban on Handguns	Concealed Carry Reciprocity Act	0.300
5	White Identity	Person of Color Identity	0.117
1	Ban on Handguns	2nd Amendment	0.017

A.7 Entity-to-Topic Std. Weighting Algorithm

ALGORITHM 3: Calculate Std. Weights

Input: $entities$, $topics$
Output: $weights_{std} \leftarrow \{\}$ ▷ Initialize the dictionary to hold the weights per topic.
for t in $topics$ **do**:
    $observations \leftarrow []$ ▷ For topic t initialize a list to hold values.
    for e in $entities$ **do**:
        $o \leftarrow get\_observations(t, e)$ ▷ Retrieve the number of attitude observations of e towards t.
        $observations$.append(o)
    end for
    $std \leftarrow standard\_deviation(observations)$ ▷ Calculate the std. of the retrieved observations.
    $weights_{std}[t] \leftarrow 1/(1 + std)$ ▷ Calculate the weight of t considering the std. values of the entities.
end for
return $weights_{std}$

ALGORITHM 4: Evaluate Entity-level Performance

Input: $entities$, $topics$, $attitudes_{True}$, $attitudes_{Pred}$, $weights_{std}$
Output: $results \leftarrow \{\}$
for e in $entities$ **do**:
    $tp, tn, fp, fn \leftarrow 0, 0, 0, 0$ ▷ Initialize: $f = false$, $t = true$, $p = positive$, $n = negative$.
    for t in $topics$ **do**:
        $pr\_att \leftarrow attitudes_{pred}[e][t]$ ▷ Retrieve the attitudes of e towards t.
        $tr\_att \leftarrow attitudes_{true}[e][t]$ ▷ Retrieve the attitudes of e towards t.
        if $pr\_att = tr\_att$ **and** $tr\_att = Positive$ **then**
            $tp+ = weights_{std}[t]$
        else if $pr\_att = tr\_att$ **and** $tr\_att = Negative$ **then**
            $tn+ = weights_{std}[t]$
        else if $pr\_att! = tr\_att$ **and** $tr\_att = Positive$ **then**
            $fp+ = weights_{std}[t]$
        else
            $fn+ = weights_{std}[t]$
        end if
    end for
    $precision \leftarrow calculate\_precision(tp, fp)$
    $recall \leftarrow calculate\_recall(tp, fn)$
    $f1 \leftarrow calculate\_f1(precision, recall)$
    $results[e] \leftarrow (precision, recall, f1)$ ▷ Store the results as a tuple.
end for

A.8 Degree of Fellowship Alignment with Party Manifestos

Table 12. Degree of Alignment between the Attitudes of the Four Largest Fellowships and the Manifestos of the Democratic and Republican Parties for the Topic of Gun Control

Gun Control																
	Fellowship 1				Fellowship 2				Fellowship 3				Fellowship 4			
Topic	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**
Assault Weapon	0.69	1.00	1.00	1.00	0.60	1.00	1.00	1.00	0.56	−1.00	−0.75	1.00	1.00	−1.00	−0.75	1.00
Background Check	0.68	−1.00	−1.00	1.00	0.50	1.00	−1.00	0.00	0.73	1.00	1.00	1.00	0.67	1.00	1.00	1.00
Ban on Handguns	0.52	1.00	−1.00	0.00	1.00	−1.00	−1.00	1.00	0.67	−1.00	0.00	–	1.00	1.00	0.00	–
CCRA	0.67	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.50	−1.00	0.00	–	0.54	−1.00	0.00	–
Gun Business Industry	0.63	1.00	1.00	1.00	0.81	1.00	1.00	1.00	0.54	−1.00	−1.00	1.00	0.62	−1.00	−1.00	1.00
Gun Buyback Program	–	–	0.00	–	–	–	0.00	–	1.00	−1.00	0.00	–	–	–	0.00	–
GC to Restrain Violence	1.00	−1.00	−1.00	1.00	–	–	−1.00	–	–	–	1.00	–	1.00	−1.00	1.00	0.00
Gun Homicide	1.00	1.00	0.00	–	1.00	1.00	0.00	–	0.62	−1.00	−1.00	1.00	–	–	−1.00	–
Gun Research	1.00	1.00	0.00	–	0.75	1.00	0.00	–	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Gun Show Loophole	0.57	1.00	0.75	1.00	1.00	1.00	0.75	1.00	1.00	−1.00	−1.00	1.00	0.50	−1.00	−1.00	1.00
Illegal Guns	0.56	1.00	0.00	–	1.00	1.00	0.00	–	0.56	−1.00	−1.00	1.00	1.00	−1.00	−1.00	1.00
Mental Health	0.71	1.00	0.00	–	0.59	1.00	0.00	–	0.50	−1.00	−1.00	1.00	0.67	−1.00	−1.00	1.00
School Safety	1.00	1.00	0.00	–	1.00	1.00	0.00	–	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00
2nd Amendment	0.90	1.00	1.00	1.00	0.69	1.00	1.00	1.00	0.53	1.00	1.00	1.00	0.52	1.00	1.00	1.00
Stop Gun Crime	0.57	1.00	0.00	–	1.00	1.00	0.00	–	0.75	1.00	1.00	1.00	0.80	1.00	1.00	1.00
Terrorist Attack	0.63	−1.00	0.00	–	0.72	−1.00	0.00	–	0.51	−1.00	−1.00	1.00	0.81	−1.00	−1.00	1.00
White Identity	1.00	−1.00	0.00	–	1.00	−1.00	0.00	–	–	–	0.00	–	1.00	−1.00	0.00	–
Average	0.76			0.88	0.82			0.86	0.66			1.00	0.80			0.92

Table 13. Degree of Alignment between the Attitudes of the Four Largest Fellowships and the Manifestos of the Democratic and Republican Parties for the Topic of Immigration

Immigration																
	Fellowship 1				Fellowship 2				Fellowship 3				Fellowship 4			
Topic	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**	coh_{att}	**Att.**	**Manifesto**	**Align.**
Amnesty	0.52	−1.00	−1.00	1.00	0.75	−1.00	−1.00	1.00	0.56	1.00	1.00	1.00	0.52	1.00	1.00	1.00
Asylum	1.00	1.00	0.75	1.00	0.57	1.00	0.75	1.00	–	–	1.00	–	–	–	1.00	–
Birth. Citiz. & 14th Am.	0.60	1.00	0.00	–	1.00	1.00	0.00	–	–	–	0.00	–	–	–	0.00	–
Border Protection	0.71	1.00	1.00	1.00	0.63	−1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00
Born Identity	–	–	0.00	–	0.75	−1.00	0.00	–	–	–	−1.00	–	0.00	−1.00	−1.00	–
Cheap Labor Availability	0.60	−1.00	−1.00	1.00	–	–	−1.00	–	1.00	−1.00	−1.00	1.00	1.00	−1.00	−1.00	1.00
DACA	–	–	−1.00	–	0.50	−1.00	−1.00	1.00	1.00	1.00	1.00	1.00	–	–	1.00	–
Deportation	0.65	−1.00	−0.75	1.00	0.61	−1.00	−0.75	1.00	0.84	−1.00	−1.00	1.00	0.53	−1.00	−1.00	1.00
Dep. of Ill. Immigrants	1.00	1.00	0.71	1.00	–	–	0.71	–	1.00	−1.00	−0.89	1.00	1.00	−1.00	−0.89	1.00
Detention	0.67	1.00	1.00	1.00	0.84	1.00	1.00	1.00	–	–	−1.00	–	0.50	−1.00	−1.00	1.00
DREAM Act	–	–	0.00	–	–	–	0.00	–	–	–	0.00	–	–	–	0.00	–
Family Sep. Policy	1.00	1.00	0.00	–	0.59	1.00	0.00	–	0.50	−1.00	−1.00	1.00	–	–	−1.00	–
Human Rights	0.54	1.00	1.00	1.00	0.65	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00
Merit-Based Immigration	–	–	0.67	–	–	–	0.67	–	–	–	1.00	–	–	–	1.00	–
Minimum Wage	1.00	1.00	1.00	1.00	–	–	1.00	–	0.50	1.00	1.00	1.00	–	–	1.00	–
Racial Identity	1.00	1.00	−1.00	0.00	0.50	1.00	−1.00	0.00	1.00	−1.00	−0.83	1.00	0.67	−1.00	−0.83	1.00
Racism and Xenophobia	1.00	1.00	0.00	–	1.00	1.00	0.00	–	1.00	−1.00	−1.00	1.00	0.58	−1.00	−1.00	1.00
Refugee	1.00	−1.00	−0.67	1.00	0.94	−1.00	−0.67	1.00	1.00	1.00	1.00	1.00	–	–	1.00	–
Salary Stagnation	–	–	0.00	–	1.00	−1.00	0.00	–	–	–	0.00	–	–	–	0.00	–
Taxpayer Money	1.00	−1.00	−1.00	1.00	–	–	−1.00	–	1.00	−1.00	0.00	–	1.00	−1.00	0.00	–
Terrorism	0.77	−1.00	−1.00	1.00	0.52	−1.00	−1.00	1.00	0.82	−1.00	−1.00	1.00	0.59	−1.00	−1.00	1.00
Wealth Gap	–	–	−1.00	–	–	–	−1.00	–	–	–	−1.00	–	–	–	−1.00	–
Average	082			0.92	0.74			0.78	0.88			1.00	0.65			1.00

Table 14. Democratic (Dem.) and Republican (Rep.) Topic Attitudes for Abortion, Immigration, and Gun Control after the Annotation of their Respective Manifestos

	Abortion			**Gun Control**			**Immigration**		
Index	Topic	Dem.	Rep.	Topic	Dem.	Rep.	Topic	Dem.	Rep.
1	Abortion Funding	1.00	−1.00	Assault Weapon	-0.75	1.00	Amnesty	-1.00	1.00
2	Abortion Provider Economy	1.00	−1.00	Background Checks	1.00	−1.00	Asylum	0.75	1.00
3	Anti-Abortion	-0.93	1.00	Ban on Handguns	0.00	−1.00	Birthright Citizenship and 14th Amendment	0.00	0.00
4	Birth Control	1.00	−1.00	Concealed Carry Reciprocity Act	0.00	1.00	Border Protection	1.00	1.00
5	Health Care	1.00	−0.88	Gun Business Industry	-1.00	1.00	Born Identity	0.00	−1.00
6	Hobby Lobby	-1.00	0.00	Gun Buyback Program	0.00	0.00	Cheap Labor Availability	-1.00	−1.00
7	Late-Term Abortion	0.00	−1.00	Gun Control to Restrain Violence	1.00	−1.00	DACA	-1.00	1.00
8	Life Protection	0.00	1.00	Gun Homicide	−1.00	0.00	Deportation	−0.75	−1.00
9	Planned Parenthood	1.00	−1.00	Gun Research	1.00	0.00	Deportation of Illegal Immigrants	0.71	−0.89
10	Pregnancy Centers	1.00	−1.00	Gun Show Loophole	-1.00	0.75	Detention	1.00	−1.00
11	Pro-Choice	1.00	−1.00	Illegal Guns	−1.00	0.00	DREAM Act	0.00	0.00
12	Pro-Life	0.00	1.00	Mental Health	−1.00	0.00	Family Separation Policy	0.00	−1.00
13	Reproduction Rights	1.00	−1.00	Person of Color Identity	0.00	0.00	Human Rights	1.00	1.00
14	Right of Human Life	0.00	1.00	Right to Self-Defense	0.00	0.00	Merit-Based Immigration	0.67	1.00
15	Roe v. Wade	1.00	0.00	School Safety	1.00	0.00	Minimum Wage	1.00	1.00
16	Sale of Fetal Tissue	0.00	−1.00	Second Amendment	1.00	1.00	Racial Identity	-1.00	−0.83
17	Sanctity of Life	0.00	1.00	Stop Gun Crimes	1.00	0.00	Racism and Xenophobia	0.00	−1.00
18	Sexual Assault Victims	0.00	0.00	Terrorist Attack	−1.00	0.00	Refugee	−0.67	1.00
19	Stem Cell Research	0.00	−0.82	White Identity	0.00	0.00	Salary Stagnation	0.00	0.00
20	Women Freedom	1.00	−1.00				Taxpayer Money	−1.00	0.00
21							Terrorism	−1.00	−1.00
22							Wealth Gap	−1.00	−1.00

Received 14 April 2023; revised 12 October 2024; accepted 24 October 2024