Intermedia Agenda Setting during the 2016 and 2020 U.S. Presidential Elections

Yijing Chen[1], Yaguang Liu[2], Lisa Singh[2], Ceren Budak[3]

[1] Central European University
[2] Georgetown University
[3] University of Michigan
chen_yijing@phd.ceu.edu, lisa.singh@georgetown.edu, cbudak@umich.edu

Abstract

Intermedia agenda setting (IAS) theory suggests that different news sources can influence each other's agenda. While this theory has been well-established in existing literature, whether it still holds in today's high-choice media environment that includes news producers of different credibility and ideology dispositions, is an open question. Through two case studies–the 2016 and 2020 U.S. presidential elections–we show that media are still largely aligned, especially in broad topics they choose to cover, and that the level of alignment along the credibility dimension is comparable to that along the ideology dimension. Comparing agendas across different media types, we find that the coverage of the Republican candidate is better aligned than the coverage of the Democratic candidate, and that agenda divergence has increased along both dimensions from 2016 to 2020. Finally, we demonstrate that high-credibility media still plays a dominant role in the IAS process, yet with a cautious warning of its declining IAS power for the Democratic candidate over the course of four years.

Introduction

How do news media select the coverage they present to their audience? Intermedia agenda setting (IAS) theory identifies one important force setting the agenda of a given news producer, that is, other news producers. While this theory is well-established, with a significant amount of early theoretical and empirical support (McCombs and Valenzuela 2020), its stability is in question in today's high-choice media environment (Chaffee and Metzger 2001). Theoretically, one can make a case for either divergence or convergence. News organizations might diverge in their coverage by catering to different audience segments (Gentzkow, Shapiro, and Stone 2015). However, commonalities in journalistic training and the broader social context (e.g., events happening in the real world) can lead to convergence despite economic pressures (Shoemaker and Reese 2013). Empirical evidence is similarly mixed, with support for both divergence (e.g., Baum and Groeling 2008; Stroud 2011; Muddiman, Stroud, and McCombs 2014) and convergence (e.g., Maier 2010; Lee 2007) of media agendas.

Past work has examined agenda alignment across various media categories such as distribution channels (e.g., TV, newspapers, online news) (Lee 2007) and ideology (Baum and Groeling 2008; Stroud 2011; Muddiman, Stroud, and McCombs 2014). Investigations related to ideology are particularly pertinent and common, given the significant role ideology plays in the U.S. political system and audience preferences. Yet, ideology is no longer the only element that activates the selective news coverage adapted for a segmented audience base. Today's high-choice media environment includes low-credibility news producers that deviate from traditional journalistic standards, at times explicitly providing a "critical meta-discourse on traditional journalism" (Eldridge 2019). This might pose a more fundamental threat to the stability of IAS. It is this threat that motivates our study. We ask: To what extent does the news agenda between low- and high-credibility media diverge? Furthermore, is IAS more significant along the credibility dimension than the ideology dimension? We answer these questions by re-examining IAS across media with varying credibility levels and different partisan leanings.

In this paper, we present two important case studies, the 2016 and 2020 U.S. presidential elections. We determine the degree to which different media types (low-credibility vs. high-credibility and left-leaning vs. right-leaning) align in terms of the candidate attributes they focus on. We examine two types of attributes: keywords (e.g., how often the word "liar" is associated with Clinton) and topics (e.g., how often the topic "healthcare" is associated with Trump). We focus on the 2016 and 2020 presidential elections for three reasons. First, these case studies are consequential. During nationally pertinent events such as presidential elections, the news agenda can shape the public political discourse, potentially impacting voting behaviors and, ultimately, the election outcome (Nicholson 2021). Second, the similar nature of these case studies helps us determine the degree to which findings from one IAS analysis generalize to other similar contexts. Finally, the four-year course from 2016 to 2020 has witnessed fundamental shifts in the news ecosystem with the growing prominence of and public attention on low-credibility media (Guo and Vargo 2020), on top of the longstanding partisan division in the contemporary U.S. media environment (Levendusky 2013). Up-to-date studies are needed to refresh our understanding of the impact brought

by these shifts.

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

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

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

Related Work

Intermedia Agenda Setting

Agenda setting theory suggests that news media shape public opinions on issue salience through their coverage–the more media cover a topic, the more important that topic becomes in the public agenda (McCombs and Shaw 1972). Alongside the inquiry of agenda flows between media and the public, intermedia agenda setting (IAS) theory looks into the agenda dynamics *among* media and suggests that different news sources can influence one another.

Previous studies have explored the IAS process with these questions: who takes the lead, on what specific issues, and in what time frames? Regarding the agenda leader/follower, researchers have identified the powerful role of elite news media in setting the agenda for others (Reese and Danielian 2012; McCombs 2005), the tendency of junior newspapers to follow the lead of senior ones (Breed 1955), and more recently, the potential of emerging online media to participate in IAS (Vargo and Guo 2017). The rising prevalence of fake and partisan news media has motivated research efforts to examine IAS through lenses of credibility and ideology. Most relevant to our study, Vargo, Guo, and Amazeen (2018) found a reciprocal relationship in the network issue agenda between fake news and fact-based news, as well as between fake news and partisan news from 2014 to 2016; Guo and Vargo (2020) further pointed out the difference in IAS dynamics between two presidential candidates in 2016, that (a) compared to attributes associated with Clinton, those associated with Trump were tied closer between fake news and fact-based news, and that (b) partisan media were able to lead the agenda for fake news media on attributes associated with Trump, whereas in Clinton's case, the interaction between partisan and fake news media was much weaker. In terms of the temporality of the IAS process, researchers have distinguished between breaking stories and ongoing debates (Vargo, Basilaia, and Shaw 2015), discussed cases of breaking news being manipulated by false reporting (Hermann, Svrluga, and Miller 2016), and called for future work to address the nuances in the time scale of the IAS process (Vargo and Guo 2017). Because IAS can happen through linked temporary spikes and correlated ongoing fluctuations, understanding these dynamics requires us to examine the temporal aspect of convergence or divergence with flexible time scales.

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

Second-level Agenda Setting and Candidate Attributes

Both agenda setting and intermedia agenda setting can be examined at three different levels, each corresponding to distinct units for comparisons of agenda. The first level focuses on broad issues; the second level examines attributes used to describe issues (McCombs and Reynolds 2009; Muddiman, Stroud, and McCombs 2014); and the third level investigates the linkages, or co-occurrences, among various issues or attributes (Guo 2012). Here, we focus on the second-level agenda setting; that is, we take each presidential candidate as a single issue and ask whether and how the attributes associated with these candidates are aligned and flow between different media types.

In previous studies that also conceptualize political figures as issues, scholars have explored various dimensions of

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

Table 1: Number of domains included in our analysis and group size for each domain category.

their attributes, the very basic application being the shaping of "candidate image" during political campaigns (e.g., Kiousis et al. 2006; Guo and Vargo 2020). Among dimensions of attributes compositing such a "candidate image", McCombs et al. (1997) specified two fundamental dimensions: the substantive dimension and the affective dimension. The former dimension organizes the candidate image with a set of relevant subtopics (e.g., personality, issue positions); and the latter dimension focuses on sentimental elements (i.e., positive, negative, or neutral) linked to the candidate. As the varying salience of linkages between attributes and a given candidate provides a cognitive frame through which a candidate is portrayed, researchers have connected framing theory with the second-level agenda setting (McCombs et al. 1997; Kiousis, Bantimaroudis, and Ban 1999; Golan and Wanta 2001). In our study, we inherit this theoretical connection and model the candidate frame using three main groups of substantive attributes: (i) attributes that discuss general government operations (including election campaigns), (ii) attributes that describe a particular policy-making aspect, and (iii) attributes that mention candidate-related controversies.

In terms of the "granularity" of the frame, scholars have distinguished between two types of attributes–aspects and central themes–when investigating the second and the third level of agenda setting (McCombs 2005). An aspect is "a micro attribute with a lower level of abstractness" and a central theme is a "macro-level attribute" that "describes a more abstract conceptual category" (Kim and Min 2015). With existing studies suggesting a higher level of fragmentation at the aspect level than at the central theme level (Budak et al. 2023b; McCombs and Valenzuela 2020), we consider it necessary to keep incorporating both levels of granularity when examining IAS. Our work measures the degree to which different media types are aligned in terms of both the central themes (e.g., how much do low- and high-credibility media align when associating Trump with various themes?) and the specific aspects of those central themes (e.g., how much do left- and right-leaning media align when associating Biden with various aspects?). We operationalize this dual-level measurement using a dictionary-based approach for text coding. Specifically, we capture aspects by detecting phrases (i.e., keywords) that occur in texts (e.g., "vote", "bank", "tax"), and capture central themes by identifying bundles of phrases (i.e., topics) that correspond to a particular candidate attribute (e.g., the topic "civil rights" includes keywords such as "vote" and "discrimination").

Dataset and Preprocessing

We start our examination by identifying a set of low- and high-credibility online news outlets. We borrow the list

of news domains from Bozarth and Budak (2021) that combines five sources of domain credibility labels (Couts and Wyrich 2016; Van Zandt 2015; Gillin 2017; Allcott, Gentzkow, and Yu 2019; Zimdars 2016). In our study, a domain belongs to the low-credibility class if it is explicitly marked as "fake" or "low-credibility" by any of the five sources. We then filter out domains that contain satire or mixed-factual content and group the remaining domains into the high-credibility class. We also assign ideology labels (i.e., left-leaning, right-leaning) to these news domains based on the bias rating tags assigned by Media Bias Fact Check (mediabiasfactcheck.com). Note that ideology labels only cover approximately 60% of all domains, with an imbalanced distribution across low- and high-credibility categories (the left-to-right ratio is 0.28 for the low-credibility group and 2.67 for the high-credibility group). Thus, dimensions of ideology and credibility present two overlapping but not entirely aligned grouping structures.

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

Given that the snapshotting frequency varies greatly across domains and across days, we assign a multiplication index to each snapshot, which will be used as a weight when we aggregate topic and keyword counts. The multiplication index equals the inverse of the number of snapshots for a given domain in a given day. This allows us to make sure that each domain will have at most one "average snapshot" per day that describes its overall agenda. We also filter out domains without sufficient snapshot coverage [3] to avoid temporal patterns being distorted by exogenous factors related to the Wayback scraping jobs (e.g., file size, number of parallel ongoing crawls). We report the number of qualified domains and the group sizes for each category in Table 1).

---

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

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

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

Methods

In this section, we elaborate on specific steps of the dictionary-based topic modeling: how we construct the topic dictionary, how the model computes topic and keyword vectors for each input, and how we validate the model output with crowd-sourced labels collected from Amazon Mechanical Turk (MTurk). Then, we explain how we utilize the model output for downstream analysis, first to measure agenda alignment and then to identify agenda leaders and followers.

Dictionary-based Topic Modeling

Constructing Topic Dictionaries For each of the 2016 and 2020 election seasons, we construct a customized dictionary by merging (1) a highly reliable [4] base dictionary (Budak et al. 2023b) that uses the Comparative Agendas Project (CAP) taxonomy (Hearings 2017), and (2) an extended dictionary customized for each election season. The extended dictionaries are curated by political communication experts through consensus labeling (Bode et al. (2019) for 2016 and Agiesta (2020) for 2020). These dictionaries contain context-specific keywords (e.g., catchphrases, names of political elites) that not only improve keyword coverage in the headline data but also better capture the election-related expressions in both years. The dictionary merging proceeds as follows. We preserve the topic taxonomy established by the CAP codebook, match topic categories from the extended dictionary to the base dictionary, and create a new topic if there exists no reasonable match. After the initial merge, the 2016 dictionary includes 1340 phrases from 26 topics, and the 2020 dictionary contains 1405 phrases from 26 topics. The topics include "government operation" that encompass general administrations and election campaigns, policy-related categories such as "healthcare", "economy" and "international affairs" (included in the CAP taxonomy), as well as scandals-related categories such as "Trump controversies", "Biden controversies" and "Clinton controversies" (added after merging the extended dictionary).[5]

To further enrich the dictionary, we use a semi-supervised topic model, Guided Topic-noise Model (GTM) (Churchill et al. 2022), to identify additional keywords and topics that are salient in the context of presidential elections. GTM utilizes an input dictionary that contains keywords for topics of interest to guide the topic-generation process. It expands the provided lists of keywords using a generalized generative model called Generalized Polya Urn (GPU) (Churchill et al. 2022) to iteratively enhance existing topics and generate new topics containing new keywords and associated weights. Based on previous practices and preliminary runs

---

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

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

on our data samples, we set the number of topics to 50; we then inspect all new keyword-topic pairs generated by GTM and score the degree of relevance for each pair (i.e., 0 for non-relevant, 1 for weakly-relevant, and 2 for highly-relevant), given the possible contexts of a keyword in our dataset. The complete inspection is performed by one author, after two authors reach an acceptable level of inter-rater reliability on their independently-assigned relevance scores for dictionary samples from both years (Krippendorff's alpha = 0.81 for 2016 and 0.7 for 2020). After the inspection, we drop non-relevant keyword-topic pairs and evaluate three strategies of keyword filtering/weighting: (i) only including the highly relevant phrases, (ii) including both the highly and weakly relevant phrases, (iii) including both while giving higher weight to highly relevant phrases. We discuss how we evaluate these strategies in the subsection *Validating Output*. The final model uses only the highly relevant phrases, consisting of 1426 keywords in 2016 and 1453 keywords in 2020.

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

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

Validating Output We finalize and validate our model output by comparing model-human agreement against human-human agreement. Our evaluation rests on the following premise: If the dictionary-based model performs as reliably as a human labeler, the extent to which the model agrees with a random human labeler should be comparable to the extent to which two random human labelers agree with each other. To perform the aforementioned evaluation, we collect human labels through a topic-labeling task on Amazon Mechanical Turk (MTurk), in which we ask MTurk workers to select the primary, secondary and all relevant topics applicable for each of the 10 texts displayed per Human Intelligence Task (HIT). We describe details of the MTurk task in Supplementary Materials (SM) *Collecting Human Labels from MTurk*.

Our evaluation proceeds as follows. For each text input, we use the model plus a random human labeler as the model-
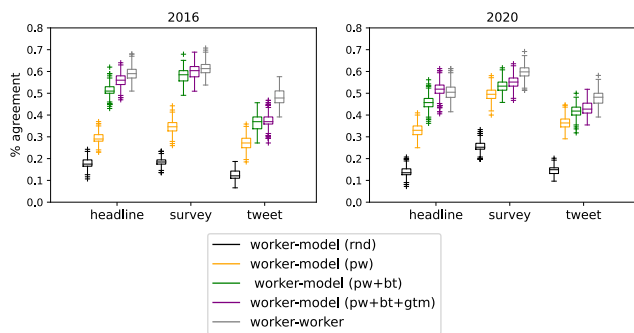
Figure 1: Agreement scores for model-human and human-human pairs. The random baseline adopts the topic distribution generated by the best-performing topic model and shuffles the topic label per text. "rnd" refers to the baseline model that randomly assigns a topic label; "pw" refers to the model using the original CAP dictionary; "pw + bt" refers to the version after the initial merge with the extended dictionary; "pw + bt + gtm" refers to the final version that included GTM keywords.

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

We find that the best performance is achieved when using only the strongly relevant keywords and limiting our attention to the primary topic. Details of the overall evaluation are given in SM Table 4. Using this model version, we show the progression of model-human agreement in Figure 1 as we update the topic dictionary. The figure shows that the worker-model agreement scores are comparable to worker-worker agreement scores, especially for the news media and survey data. We also see that the progression is consistent across the two elections, allowing us to make reliable comparisons between the two elections. Although the suboptimal level of overall human-human agreement reflects the inherent difficulty of the labeling task itself, the model does identify meaningful topical cues from the text that make sense to humans in a considerable proportion of texts. In circumstances of conflicts between human-model pairs, roughly 42% of the texts have at least one matching pair of human and model labels, and 57% have at least one human including the primary model topic in their expanded topic list (i.e., primary, secondary or relevant topics). For topics that have low agreement levels between the model and

workers and occur rarely in our data[6], we drop them in the downstream analysis.

Downstream Analysis

Measuring Agenda Alignment We first assess agenda alignment by comparing the aggregated attention distribution between different media types. The attention distribution can be described at both the keyword level and the topic level. At the keyword level, we compute an aggregated keyword vector $\vec{x}_A$ for media type A by adding up[7] keyword vectors $\vec{x}_i$ for each text i from media type A, and normalizing the output by its sum. Similarly, at the topic level, we obtain an aggregated topic vector $\vec{y}_A$ for media type A, which sums up all topic vectors of texts from media type A and normalizes the output by its sum. An aggregated topic or keyword vector is essentially a probability vector that sums up to 1. With these aggregated vectors, we use Pearson correlation coefficient, a widely-adopted metric in previous agenda-setting studies (e.g., Sweetser, Golan, and Wanta 2008; Guo 2012) to quantify the degree to which the priorities of candidate attributes align between media type A and B, i.e., $\rho(\vec{x}_A, \vec{x}_B)$ for the alignment at the aspect (keyword) level and $\rho(\vec{y}_A, \vec{y}_B)$ at the central theme (topic) level. The higher the correlation, the better the agendas align.

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

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

We perform Granger causality tests for daily time series of attribute proportions. Granger causality analysis is a classic approach to evaluate the (intermedia) agenda-setting power using time series (e.g., Brosius and Kepplinger 1990; Meraz 2011; Groshek and Clough Groshek 2013; Vargo, Guo, and Amazeen 2018; Guo and Vargo 2020). It allows us to statistically assess the temporal "causation" between two time series with varying time lags[8]. Let's say we focus on attribute k (e.g., topic "healthcare" for Trump), and extract the time series of its attention proportion in media type A (e.g., low-credibility media) and B (e.g., high-credibility media), $X_{A,k}$

---

[6]Topics excluded downstream: "forestry", "land water management", "agriculture", "housing", "transportation", "culture".

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

[8]We apply time lags of 1, 2, 3, 4, 5 days for individual topic time series.
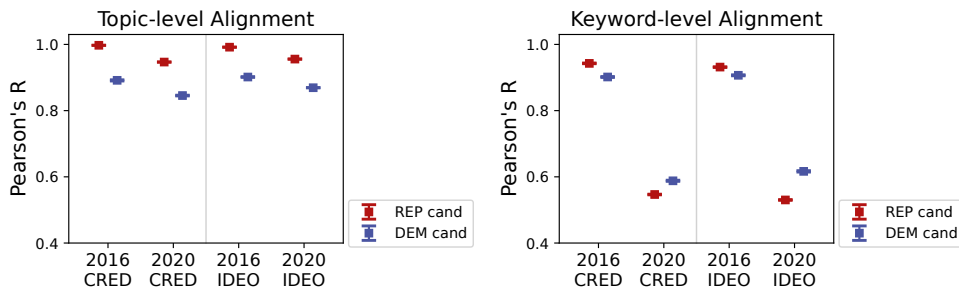
Figure 2: Agenda alignment for the 2016 and 2020 election seasons, at the topic level (left) and at the keyword level (right). We keep the top 18 topics that occur frequently and meaningfully in the data, and keep the top 500 keywords with high frequencies. CRED refers to media types by credibility (i.e., low- and high-credibility media), and IDEO refers to media types by ideology (i.e., left- and right-leaning media).

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

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

Results and Discussions

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

Static Alignment: Divergence or Convergence of Media Agenda?

We start by presenting the aggregate alignment in Figure 2. The first striking finding is the similarity in patterns observed for ideology and credibility. Past work that compares the role ideology and credibility play in news production has found that credibility plays a more significant media

fragmentation role (Budak et al. 2023a). Based on this, we would have expected the media to be more fragmented along the credibility dimension. However, surprisingly, we see that media are no more divided along the credibility line compared to ideology. This could be because ideology is one of many factors that shape broader news production, while it is *the* decisive factor in election campaign coverage. This highlights the enduring role ideology plays in election coverage.

At the topic level, agendas of different media types are still largely aligned in both years ($r > 0.8$), with a slight decrease from 2016 to 2020 (average $\Delta r = -0.041$). In contrast, at the keyword level, the correlations for both candidates have dropped dramatically (average $\Delta r = -0.350$), reaffirming previous findings of a more severe fragmentation at the aspect level than the central theme level (Stroud 2011; McCombs and Valenzuela 2020; Budak et al. 2023b). The downtrend in keyword alignment is more pronounced for the Republican candidate (average $\Delta r = -0.399$ per media pairing) than for the Democratic candidate (average $\Delta r = -0.302$ per media pairing), evincing partisan asymmetries in agenda fragmentation at the keyword level.

Candidate-wise, the coverage of the Republican candidate is generally better aligned than that of the Democratic candidate. As shown in Figure 2, at the topic level, low- and high-credibility media share highly similar priorities for Trump's attributes in both years ($r = 0.997$ in 2016 and $r = 0.946$ in 2020); so do left- and right-leaning media ($r = 0.991$ in 2016 and $r = 0.956$ in 2020). For his opponent candidate, the correlations between these two media pairings are weaker ($r = 0.891$ in 2016 and 0.845 in 2020 across credibility types; $r = 0.902$ in 2016 and 0.869 in 2020 across ideology types). Interestingly, in 2020, the coverage of Trump achieves a higher level of alignment than that of Biden at the topic level ($\bar{r} = 0.951$ per media pairing for Trump and $\bar{r} = 0.857$ for Biden), but not at the keyword level ($\bar{r} = 0.538$ per media pairing for Trump and $\bar{r} = 0.602$ for Biden). This reveals the level at which media diverge for a given candidate. Different types of media organize similar priorities for Trump-related topics, but the specific keywords used in their discussions are poorly coordinated. Whereas for Biden, although the topic-level agendas are not as well aligned as his Republican counterparts, the keywords used
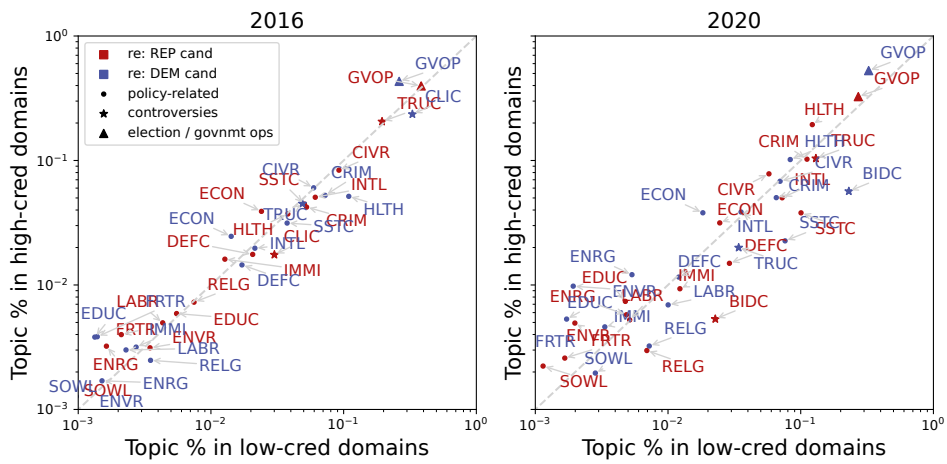
259

Figure 3: Comparisons of topic proportion between low- and high-credibility domains in 2016 (left) and 2020 (right). Please refer to SM Table 3 for the list of topics the acronyms in the figure stand for.
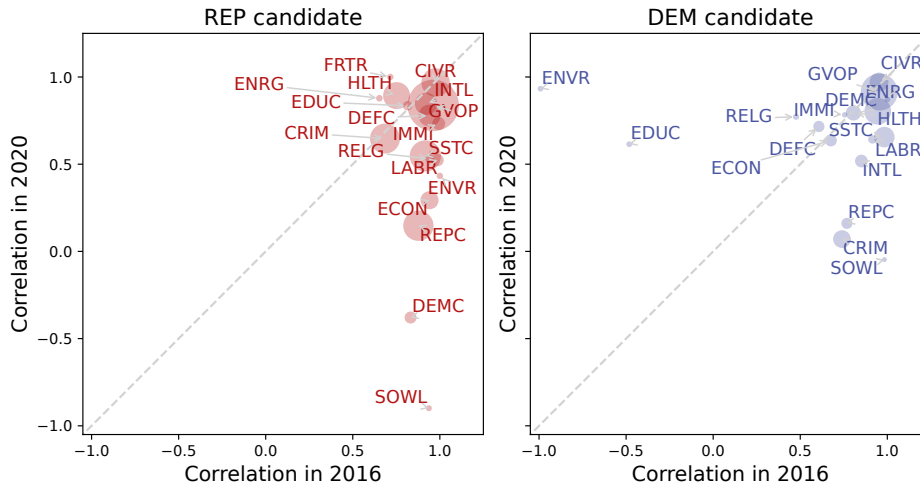
.

Figure 4: By-topic alignment of top 500 keywords across years between low- and high-credibility news media in 2016 (left) and 2020 (right). Dot size is a function of the overall frequency of the corresponding topic.

in different media types have relatively greater overlaps.

Candidate Controversies: Key Attributes as Divergence Drivers

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

Candidate-wise, we find that higher proportions of at-

tention on "DEM candidate controversies" (topic related to Clinton controversies in 2016 and Biden controversies in 2020) from low-credibility and right-leaning media are the main source of salient agenda divergence for the Democratic candidate. Focusing on blue dots in Figure 3, we see that the two largest topics that deviate significantly from the diagonal line are "DEM candidate controversies" and "government operations". Low-credibility and right-leaning media highlight "DEM candidate controversies" more than their counterparts, limiting the attention devoted to "government operations". Such deviations have become more salient in 2020. We again observe similar patterns for the credibility and ideology divide. The plot summarizing the ideology results is omitted here for brevity (see SM Figure 9).
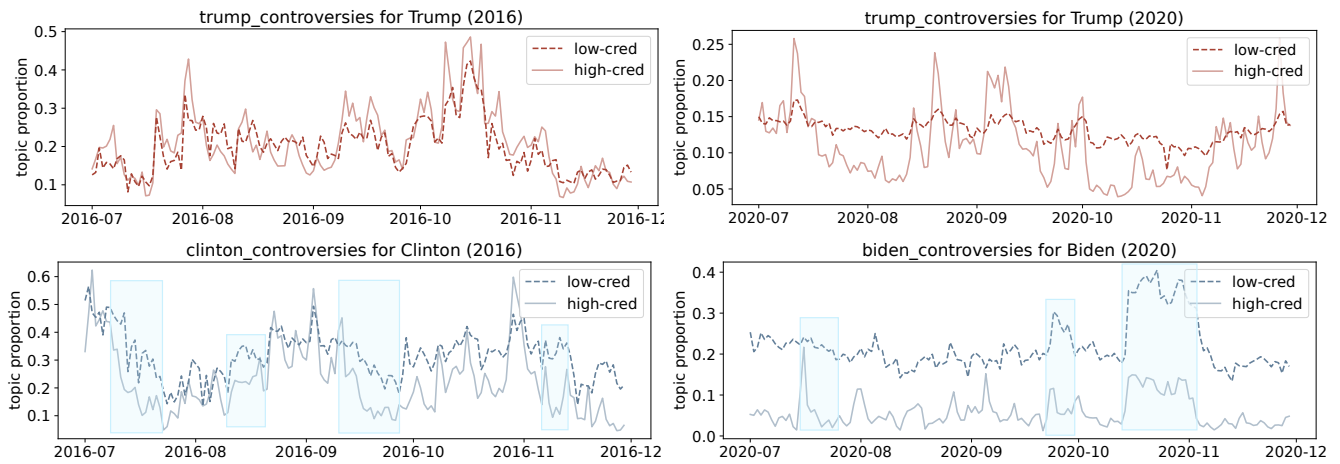
Figure 5: Time series of the topic "REP/DEM candidate controversies" from low- and high-credibility media in 2016 (left) and 2020 (right). Blue rectangle boxes in the second row highlight a few periods with drastic drops in overall topic alignment, when low- and high-credibility media have different spotlighting intensity or attention spans on "DEM candidate constroversies".

The attention on "DEM candidate controversies" not only diverges at the aggregated level but also signals the specific point in time when different media types will diverge. To illustrate this, we examine the temporal dependence between the attention disparity on "DEM candidate controversies" and the temporal fluctuations in overall topic alignment. Specifically, we apply a uni-variate ordinary least squares (OLS) model for topic k, using the time series of temporal alignment as the dependent variable Y, and the time series of the temporal *difference* in the proportional attention on topic k between a given media pairing as the independent variable X. We find that regressing the difference in "DEM candidate controversies" can explain more of the variance (i.e., achieve the highest R-squared values) than any other topic, especially for Biden in 2020 between left- and right-leaning media (R-squared = 0.7565). We report the full results in SM Table 2. Comparing the time series of "DEM candidate controversies" between low- and high-credibility media (see Figure 5) with that of temporal topic alignment between low- and high-credibility media (blue lines in SM Figure 11 A1 and A2), we see that agenda divergence is brought forward by the misaligned attention spans or the different spotlighting intensity on "DEM candidate controversies". For Clinton's case in 2016, for instance, we can link some dramatic drops in temporal alignment to the time periods when low-credibility media discussed "Clinton controversies" much longer than high-credibility media after breaking events such as Bill Clinton and Loretta Lynch's meeting in early July and Hillary Clinton fainting in mid-September.

Furthermore, at the keyword level, we notice that the drop in keyword alignment for "REP candidate controversies" contributes greatly to the drop in overall keyword alignment from 2016 to 2020. As shown in Figure 4, "REP candidate controversies" is among the noteworthy topics that have a dramatic drop in keyword alignment from 2016 to 2020 and that occur frequently enough to have a sizable impact on the overall alignment. Looking closer at specific aspects (i.e., keywords) addressed for "REP candidate controversies", we see that high-credibility media dedicate more attention to Trump's family members; and that low-credibility media put the spotlight on the deep-state conspiracy, and push stories co-mentioning Trump with figures such as Jeffrey Epstein, Adam Schiff, and Roger Stone. While there is a lack of consensus on central aspects of "REP candidate controversies" across different media types in 2020, the divergence of aspects for "DEM candidate controversies" is much weaker[9]. Conditioned on the topic "DEM candidate controversies", we notice that controversies centered around Hunter Biden are heavily debated on both sides of media with high occurrences of keywords "Hunter Biden", "laptop" and "Ukraine". Here, we focused on the alignment across credibility groups. The patterns observed for ideology are, again, similar and omitted for brevity (see SM Figure 10 for ideology results).

Temporal Dynamics: Who Leads and Who Follows?

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

We assess the IAS power based on (i) the number of attributes one media type leads for the other, as well as (ii) the

---

[9] Among the top 500 frequent keywords for Trump-related texts in 2020, we correlate keywords that belong to "Trump controversies". Pearson's R = 0.1446 between low- and high-credibility media; Pearson's R = 0.1139 between left- and right-leaning media. In Biden-related texts, the corresponding Pearson's Rs for keywords belonging to "Biden controversies" are 0.8095 between low- and high-credibility media, and 0.7597 between left- and right-leaning media.
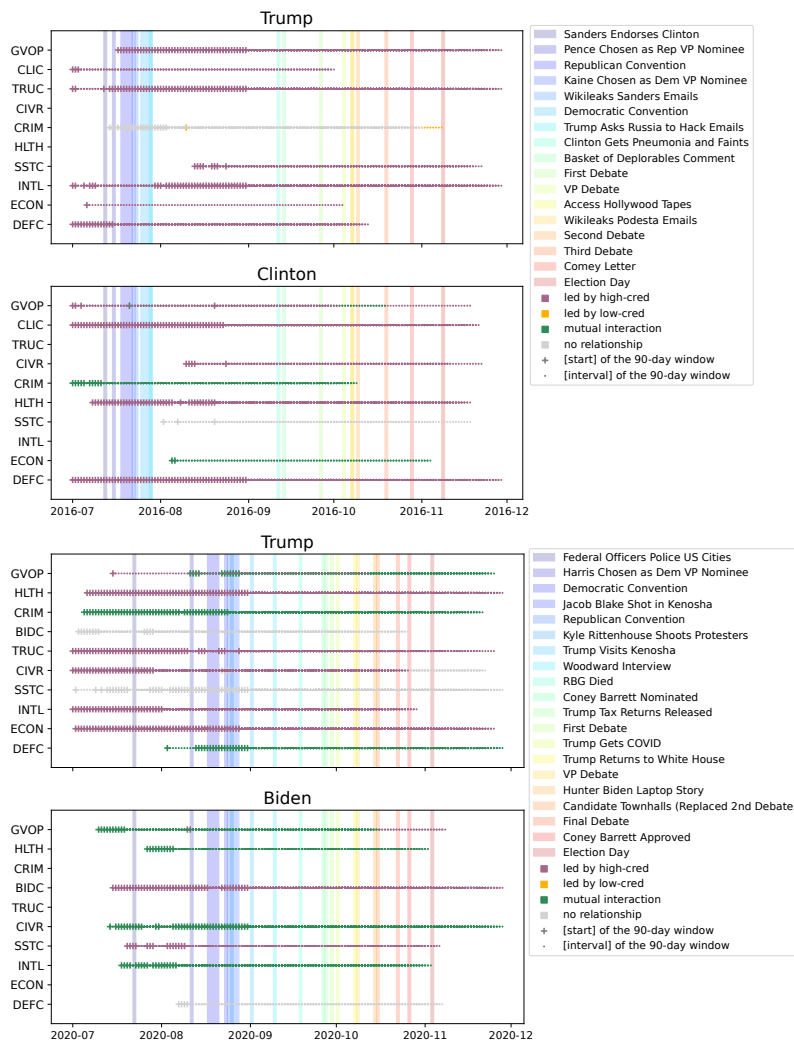
Figure 6: Granger causality results between low- and high-credibility media for 2016 (the first and second figures) and 2020 (the third and fourth figures). We test Granger causalities in a sliding window of 90 days and display robust results that appear in more than 95% of the bootstrapping runs. Each plus sign marks the starting point of the 90-day sliding window with a significant result. We include results for the top 10 topics that show frequently in all news headlines for a given election year.

length of time period during which such IAS power can persist. We summarize this information in the sliding-window plots displayed in Figure 6, where the starting points of all 90-day windows with a significant and robust Granger causality result are marked with plus signs ($+$). Each plus sign is followed by 90 dots ($\cdot$) colored the same as the plus sign to visually demonstrate the full length of sliding windows. For example, for Trump 2016 there is only one time window with significant and robust results on the attribute "economy (ECON)", spanning from early July to early October.

Overall, we see that high-credibility media serve as the dominant actor in IAS, setting the agenda for more candidate attributes than low-credibility media. Out of the top 10 attributes that appear frequently in a given year, high-credibility media lead the agenda of 5.5 attributes for the Re-

publican candidate and 3 attributes for the Democratic candidates on average, with varying window lengths[10]. Meanwhile, we do not see low-credibility media persistently leading the agenda on any attribute in either election season. Despite the encouraging results, we observe that high-credibility media's IAS power has declined from 2016 to 2020 for the Democratic candidate, with a decrease in terms of the number of attributes it leads (from 4 to 2), and the total number of windows it leads (from 165 to 60). Furthermore, agendas between high- and low-credibility media appear to be more intertwined, mutually interacting with each other on 3 attributes for Trump and 4 for Biden in 2020, but only 1 for Clinton in 2016.

Notably, while "candidate controversies" acts as a cru-

_____

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

cial attribute that drives the divergence of the media agenda, discussions of "REP (DEM) candidate controversies" in the coverage of the Republican (Democratic) candidate are always led by high-credibility media. Based on Figure 6, the longest consecutive high-credibility-leading windows (i.e., consecutive time windows with a significant and robust result of high-credibility media taking the lead) lasts for 135 days (45 windows) for the Republican candidate and 134 days (44 windows) for the Democratic candidate on average[11]. Moreover, the election-wise comparison re-iterates the diminishing IAS power of high-credibility media specifically on "DEM/REP candidate controversies", as the length of the longest consecutive high-credibility-leading windows shrinks from 141.5 days (51.5 windows) per candidate in 2016 to 127.5 days (37.5 windows) per candidate in 2020. Such shrinking in window length happens more severely for the Democratic candidate (ΔL = -20 days) than the Republican candidate (ΔL = -8 days).

To sum up, high-credibility media is more powerful in IAS compared to low-credibility media, as it leads the agenda for more attributes and consistently for longer periods of time; however, the IAS power of high-credibility media has declined from 2016 to 2020, together with a few more attributes seeing mutually interacting agendas in 2020 (e.g., "crime" for Trump, "government operation" and "healthcare" for Biden). Contrasting these patterns with the IAS dynamics between left- and right-leaning media (see SM Figure 13), we see shared patterns between high-credibility and left-leaning media in terms of their dominant role in IAS in general, as well as their weakening leader advantage from 2016 to 2020, especially for the Democratic candidate. While some level of symmetry does exist between credibility and ideology, we see the value of separately addressing IAS along these two dimensions. Right-leaning media clearly take a more active role than low-credibility media in 2020, persistently setting agenda for a few attributes of the Republican candidate (e.g., leading "civil rights" and "international affairs" for Trump in 2020).

Conclusions and Limitations

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

Overall, we observe a high level of agenda alignment in candidate coverage between low- and high-credibility media. The agenda convergence indicates that low- and high-credibility media still share a common ground for candidate-related discussions on broad issues; however, the initiator of such assimilation remains unclear. Low-credibility media could be borrowing stories from traditional players with higher credibility levels, due to their limited resources to independently produce impactful news stories in the fierce attention battleground. Alternatively, past work also shows that traditional media can spread misinformation, especially

---

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

by indexing political elite talking points (Muddiman et al. 2022). High-credibility media might be loosening their journalistic standards in order to attract and retain their audience, generating stories of disputable issues that are easier to be re-packaged into low-credibility clickbaits, as we see in our results the significant proportions of attention devoted to "candidate controversies" rather than policy topics in high-credibility media.

In addition, our study adds to the growing body of literature that highlights partisan asymmetries in the news ecosystem (e.g., Budak 2019; Guess, Nyhan, and Reifler 2020) by demonstrating the stronger alignment in agendas for the Republican candidate compared to the Democratic candidates. Past work shows that media exhibit their bias largely through negative depictions of the opposing side, as opposed to positive depictions of the preferred side (Soroka 2014; Budak, Goel, and Rao 2016). Here, the diverging agendas for the Democratic candidate provide evidence that such bias may extend to selective coverage of topics.

We also observe meaningful shifts in IAS when comparing 2016 to 2020, which underscore two valuable insights. First, it is a caution against over-generalization from studies focused on a single case study. Second, it shows that the U.S. news ecosystem is still in flux, with IAS powers of different media types shifting. Our results thus motivate future researchers to carry out large-scale empirical studies to examine well-established communication theories using contemporary datasets.

There are various limitations to this work. First and foremost, although we use terms such as "influence" and "Granger causation" when assessing IAS, the IAS process captured in our study is based on correlational analysis. While we follow the terminology used in scholarship and theorize about setting the agenda, we caution the reader that the associations found here are not sufficient evidence of a causal relationship. Secondly, our dictionary-based model utilizes context-specific topics and keywords related to a certain issue (e.g., presidential candidates), which limits its generalizability. While the dictionaries themselves are not generalizable, the pipeline we introduced for constructing and validating topic dictionaries is. Our modeling approach allows interpretability and cross-year comparisons. Finally, we use a set of existing source lists for determining the credibility of different websites, where sizable disagreements exist across lists constructed by different fact-checkers and scholars (Bozarth, Saraf, and Budak 2020). Furthermore, the limited coverage of ideology labels has restricted our scope of analysis when comparing left- and right-leaning media. We encourage future studies to explore more source lists of domain credibility and ideology, and incorporate a better-labeled dataset for such parallel analysis.

Our findings also identify new directions for future work of IAS. First, it is worth following up with more recent datasets to examine if the IAS trends identified in 2016 and 2020 continue in future elections. Second, we notice a few signals for the insufficient explanatory power of the current models in capturing temporal "causation" between left- and right-leaning media. For instance, in SM Figure 13, we see the absence of significant and robust IAS results in most 90-

day sliding windows, particularly for the Democratic candidate. This may be a substantive finding: left- and right-wing media fail to set each other's agendas. Or, this may be a result of linear regression models failing to capture the increasingly complicated agenda interactions. Such obscurity invites future explorations of different methodologies to validate or extend our findings. Third, our parallel analysis points out that the divisions along ideology and credibility share some structural features but are not entirely overlapping. Future work could look into the interplay between these two dimensions.

Ethics Statement and Broader Impact

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

Apart from cautiously deriving the implication, we have incorporated the following ethical considerations: (1) once collected and preprocessed, the headline dataset is stored on the server with restricted access; (2) we remove personally identifiable information from the MTurk output; (3) we actively communicate with MTurk workers who raise questions or concerns, and make sure that those who attentively work on the labeling tasks are fairly compensated (even if they fail the screening); and (4) we release the dictionary and the model source code in a GitHub repository[12].

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

Acknowledgements

References

Agiesta, J. 2020. Read: The methodology behind new polling project The Breakthrough. *CNN*.

Allcott, H.; Gentzkow, M.; and Yu, C. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2): 2053168019848554.

Baum, M. A.; and Groeling, T. 2008. New media and the polarization of American political discourse. *Political Communication*, 25(4): 345–365.

Bode, L.; Budak, C.; Ladd, J. M.; Newport, F.; Pasek, J.; Singh, L. O.; Soroka, S. N.; and Traugott, M. W. 2019. *Words that matter: How the news and social media shaped the 2016 Presidential campaign*. Brookings Institution Press.

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

Bozarth, L.; Saraf, A.; and Budak, C. 2020. Higher ground? How groundtruth labeling impacts our understanding of fake news about the 2016 US presidential nominees. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 48–59.

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

Brosius, H.-B.; and Kepplinger, H. M. 1990. The agenda-setting function of television news: Static and dynamic views. *Communication research*, 17(2): 183–211.

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

Budak, C.; Bozarth, L.; Bond, R. M.; Margolin, D.; Jones, J. J.; and Garrett, R. K. 2023a. Bursts of contemporaneous publication among high-and low-credibility online information providers. *New Media & Society*, 14614448231183617.

Budak, C.; Goel, S.; and Rao, J. M. 2016. Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1): 250–271.

Budak, C.; Jomini Stroud, N.; Muddiman, A.; Murray, C. C.; and Kim, Y. 2023b. The Stability of Cable and Broadcast News Intermedia Agenda Setting Across the COVID-19 Issue Attention Cycle. *Political Communication*, 1–21.

Chaffee, S. H.; and Metzger, M. J. 2001. The end of mass communication? *Mass communication & society*, 4(4): 365–379.

Churchill, R.; Singh, L.; Ryan, R.; and Davis-Kean, P. 2022. A Guided Topic-Noise Model for Short Texts. In *Proceedings of the ACM Web Conference 2022*, 2870–2878.

Couts, A.; and Wyrich, A. 2016. Here are all the 'fake news' sites to watch out for on Facebook.

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

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/. Accessed: 2024-03-31.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gentzkow, M.; Shapiro, J. M.; and Stone, D. F. 2015. Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1, 623–645. Elsevier.

Gillin, J. 2017. PolitiFact's guide to fake news websites and what they peddle.

Golan, G.; and Wanta, W. 2001. Second-level agenda setting in the New Hampshire primary: A comparison of coverage in three newspapers and public perceptions of candidates.

---

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

Journalism & Mass Communication Quarterly, 78(2): 247–259.

Groshek, J.; and Clough Groshek, M. 2013. Agenda trending: Reciprocity and the predictive capacity of social network sites in intermedia agenda setting across issues over time. *Available at SSRN 2199144*.

Guess, A. M.; Nyhan, B.; and Reifler, J. 2020. Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour*, 4(5): 472–480.

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

Guo, L.; and Vargo, C. 2020. "Fake news" and emerging online media ecosystem: An integrated intermedia agenda-setting analysis of the 2016 US presidential election. *Communication Research*, 47(2): 178–200.

Hearings. 2017. The Policy Agendas Project at the University of Texas at Austin. www.comparativeagendas.net, Accessed September 26, 2017.

Hermann, P.; Svrluga, S.; and Miller, M. E. 2016. Alleged gunman tells police he wanted to rescue children at DC pizza shop after hearing fictional internet accounts. *The Washington Post*, 5.

Kim, J.; and Min, Y. 2015. An issue attention cycle analysis of the network agenda setting model. *The power of information networks: New directions for agenda setting*, 132.

Kiousis, S.; Bantimaroudis, P.; and Ban, H. 1999. Candidate image attributes: Experiments on the substantive dimension of second level agenda setting. *Communication Research*, 26(4): 414–428.

Kiousis, S.; Mitrook, M.; Wu, X.; and Seltzer, T. 2006. First-and second-level agenda-building and agenda-setting effects: Exploring the linkages among candidate news releases, media coverage, and public opinion during the 2002 Florida gubernatorial election. *Journal of Public Relations Research*, 18(3): 265–285.

Krippendorff, K. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Lee, J. K. 2007. The effect of the Internet on homogeneity of the media agenda: A test of the fragmentation thesis. *Journalism & Mass Communication Quarterly*, 84(4): 745–760.

Levendusky, M. 2013. *How partisan media polarize America*. University of Chicago Press.

Lind, F.; Gruber, M.; and Boomgaarden, H. G. 2017. Content analysis by the crowd: Assessing the usability of crowdsourcing for coding latent constructs. *Communication methods and measures*, 11(3): 191–209.

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

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

McCombs, M.; Llamas, J. P.; Lopez-Escobar, E.; and Rey, F. 1997. Candidate images in Spanish elections: Second-level agenda-setting effects. *Journalism & Mass Communication Quarterly*, 74(4): 703–717.

McCombs, M.; and Reynolds, A. 2009. How the news shapes our civic agenda. In *Media effects*, 17–32. Routledge.

McCombs, M.; and Valenzuela, S. 2020. *Setting the agenda: Mass media and public opinion*. John Wiley & Sons.

McCombs, M. E.; and Shaw, D. L. 1972. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2): 176–187.

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

Muddiman, A.; Budak, C.; Murray, C.; Kim, Y.; and Stroud, N. J. 2022. Indexing theory during an emerging health crisis: how US TV news indexed elite perspectives and amplified COVID-19 misinformation. *Annals of the International Communication Association*, 46(3): 174–204.

Muddiman, A.; Stroud, N. J.; and McCombs, M. 2014. Media fragmentation, attribute agenda setting, and political opinions about Iraq. *Journal of Broadcasting & Electronic Media*, 58(2): 215–233.

Nicholson, S. P. 2021. *Voting the agenda: Candidates, elections, and ballot propositions*. Princeton University Press.

Reese, S. D.; and Danielian, L. H. 2012. Intermedia influence and the drug issue: Converging on cocaine. In *Communication campaigns about drugs*, 29–45. Routledge.

Shoemaker, P. J.; and Reese, S. D. 2013. *Mediating the message in the 21st century: A media sociology perspective*. Routledge.

Soroka, S. N. 2014. *Negativity in democratic politics: Causes and consequences*. Cambridge University Press.

Stroud, N. J. 2011. *Niche news: The politics of news choice*. Oxford University Press.

Sweetser, K. D.; Golan, G. J.; and Wanta, W. 2008. Intermedia agenda setting in television, advertising, and blogs during the 2004 election. *Mass Communication & Society*, 11(2): 197–216.

Van Zandt, D. 2015. Media Bias / Fact Check.

Vargo, C. J.; Basilaia, E.; and Shaw, D. L. 2015. Event versus issue: Twitter reflections of major news, a case study. In *Communication and information technologies annual*, volume 9, 215–239. Emerald Group Publishing Limited.

Vargo, C. J.; and Guo, L. 2017. Networks, big data, and intermedia agenda setting: An analysis of traditional, partisan, and emerging online US news. *Journalism & Mass Communication Quarterly*, 94(4): 1031–1055.

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

Zimdars, M. 2016. My "fake news list" went viral. But made-up stories are only part of the problem. *Washington Post*.

Ethics Checklist

1. For most authors...

   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes. We're interested in addressing the divide across different news consumption niches by empirically measuring agenda alignment, and we do not anticipate any of the aforementioned harms.

   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes. We cautiously use the term "Granger causation" to avoid overstating the temporal relationships we infer.

   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes. We do that by referencing widely tested approaches as well as carefully validating the output generated by our topic model.

   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes. We discuss the de-biasing method implemented to minimize the potential bias introduced by varying snapshotting frequencies of Wayback machine in Section *Dataset and Preprocessing*.

   (e) Did you describe the limitations of your work? Yes, please refer to Section *Limitations and Future Work.*

   (f) Did you discuss any potential negative societal impacts of your work? Yes, please refer to Section *Ethics Statement and Broader Impact.*

   (g) Did you discuss any potential misuse of your work? No. We do not anticipate any direct misuse of our work since the model and the results are only applicable in a specified context. We try to minimize the possibility of our results being misinterpreted by carefully framing the conclusions and supplementing clarifications when necessary (e.g., clarifying high-credibility media's role in agenda divergence).

   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, please refer to Section *Ethics Statement and Broader Impact.*

   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.

2. Additionally, if your study involves hypotheses testing...

   (a) Did you clearly state the assumptions underlying all theoretical results? Not applicable. Our work is driven by an open question instead of a theoretically grounded hypothesis.

   (b) Have you provided justifications for all theoretical results? NA

   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

   (e) Did you address potential biases or limitations in your theoretical framework? NA

   (f) Have you related your theoretical results to the existing literature in social science? Yes. For instance, the lower degree of keyword-level alignment (compared to topic-level alignment) is in line with previous work.

   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes. We include these in the Section *Ethics Statement and Broader Impact.*

3. Additionally, if you are including theoretical proofs...

   (a) Did you state the full set of assumptions of all theoretical results? NA

   (b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? No. We will publish the code upon acceptance.

   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes. For the hyperparameters, we did some sensitivity analysis and chose the ones that had the best results. For data splits, we randomly split them.

   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? No, because it is of little relevance to our paper.

   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, because the model does not require much computational resource. Both the dictionary-based topic model and the guided topic model were run on a single CPU.

   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, please refer to Methods – Dictionary-based Topic Modeling – Validating Output.

   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No. When comparing model labels with human labels, we noticed that human labelers are more likely to label a text as controversies-related more than the model, which means the downstream counts of the topic "candidate controversies" would be rather conservative. However, after internal inspections of these text examples, we do not think the conservative perspective would significantly mislead our findings because (1) the output time series of "candidate controversies" is able to capture major events

of candidate scandals and (2) we would caution readers to take human labels as *the* groundtruth, given the suboptimal level of human-human agreement overall.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? Yes, we utilize lists of online news labels from existing studies and referenced them in Section *Dataset and Preprocessing*; we also leveraged Guided Topic-Noise Model created by Churchill et al. (2022) and cited the work in Section *Methods – Dictionary-based Topic Modeling – Constructing topic dictionaries*.

(b) Did you mention the license of the assets? No. There is no existing license for the assets we utilize in this paper.

(c) Did you include any new assets in the supplemental material or as a URL? No. We release the current version of topic dictionaries in a GitHub repository and will update the repository if any future changes occur. The headline dataset is available upon request.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating NA

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. While the headline dataset does not contain personally identifiable information, it may contain low-credibility information. When releasing a sample on MTurk for the topic labeling task, we explained the context of the data and the task (e.g., the data source, and the purpose of our research) to minimize the risk of misleading workers.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? Yes, please refer to Section *Ethics Statement and Broader Impact*. More details regarding how we conform to these guidelines will be included in the Datasheet released with the topic dictionary.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Yes. We release the topic dictionary with the Datasheet.

6. Additionally, if you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots? Yes, please refer to SM Figure 7.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes, please refer to Section *Ethics Statement and Broader Impact*.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Yes, we included it in our discussion of the MTurk task. Please refer to Methods – Dictionary-based Topic Labeling – Validating Output.

(d) Did you discuss how data is stored, shared, and deidentified? Yes, please refer to Section *Ethics Statement and Broader Impact*.

Supplementary Materials

Collecting Human Labels from MTurk   Our quality-control pipeline involves (a) selecting workers who have acquired a *Masters* qualification and reside in the U.S., and (b) blocking workers who fail to correctly label any of the two screening texts in a single HIT (see the full pipeline in SM Figure 8). We include detailed instructions on the top of the task page, describing the data sources, the purpose of our study, as well as the quality assurance steps we take to decide whether to accept a submission (see the task interface in SM Figure 7). Workers are compensated 1.0 USD per HIT, achieving an hourly rate of 15 USD at a relaxing speed of 4 minutes per HIT.

The entire task costs 407 USD, providing labels for 240 news headlines, 240 survey responses, and 240 tweets sampled from our data[13], with each text being read by three MTurk workers. In total, 40 workers are involved in our study, all of whom have contributed at least one qualified assignment. 19 out of these workers were blocked from further submitting for failing the screening questions. Out of 284 total submitted assignments, we use 263 qualified ones (93.36%) to generate labels for 718 text inputs. We assess the reliability of the workers by computing inter-rater reliability (Krippendorff's alpha $= 0.4385$ for 2016 and 0.4251 for 2020) for the primary topic, a commonly used measure in the literature (Krippendorff 2018). The reliability, while low by traditional content analysis standards, is significantly higher than accepted levels for crowd-sourced approaches (Lind, Gruber, and Boomgaarden 2017).

Figure 7: Screenshots of labeling task interface we launched on Amazon Mturk.

---

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

Pre-task screening
- has completed at least 1000 HITs
- acceptance rate ≥ 98%
- resides in US
- has *Masters* qualification

qualified MTurkers

Fail to correctly label any screening text

Correctly label at least one screening texts

if they spend less than 30 seconds per HIT

if they spend at least 30 seconds per HIT

Accept

Reject without paying

Reject and pay
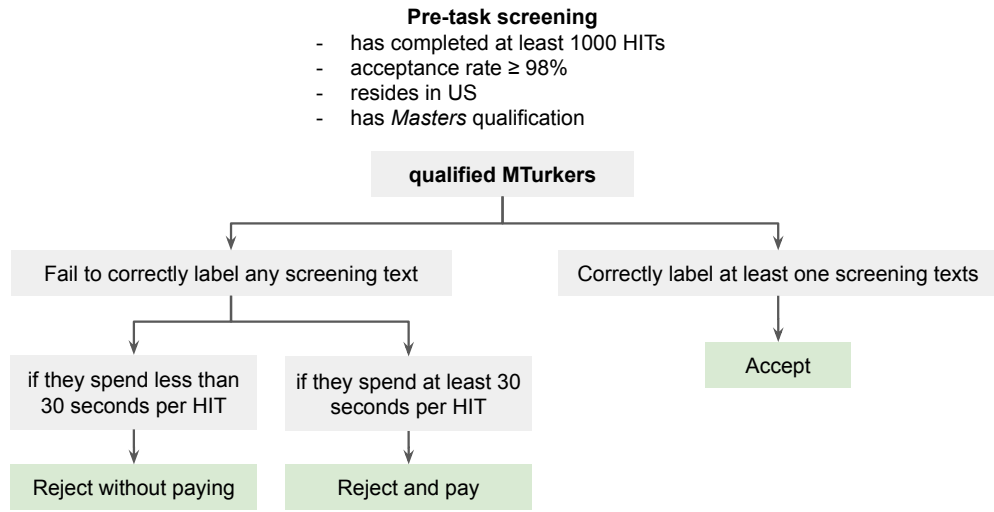
Figure 8: Diagram of our quality-control pipeline on MTurk.
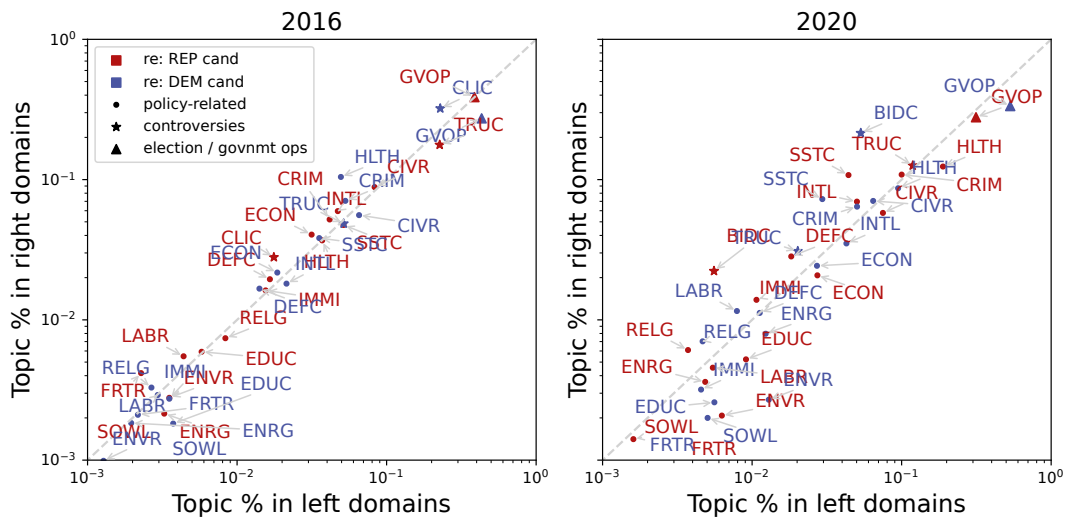
Figure 9: Comparisons of topic proportion between left- and right-leaning domains in 2016 (left) and 2020 (right).

269
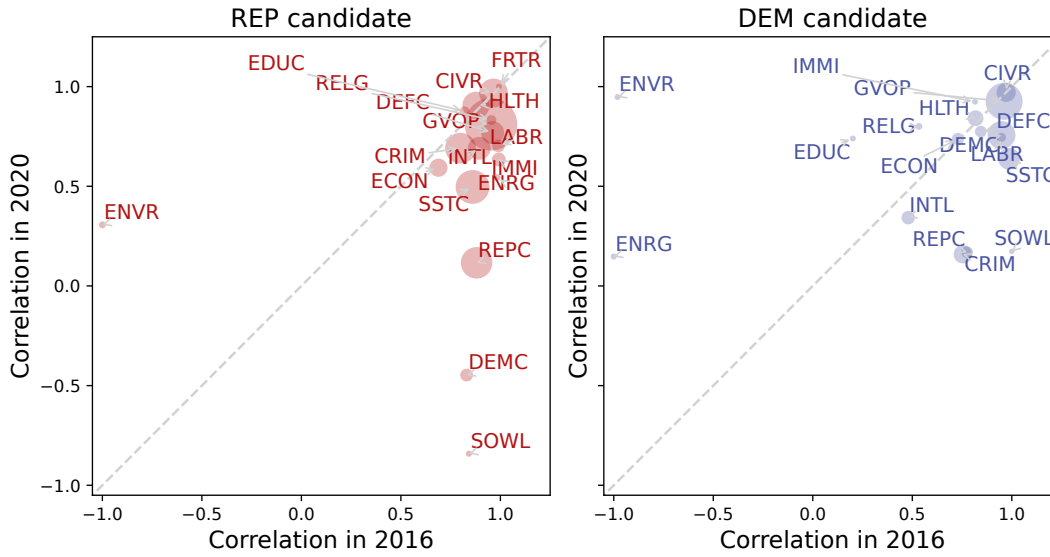
Figure 10: By-topic alignment of top 500 keywords across years between left- and right-leaning news media in 2016 (left) and 2020 (right). Dot size is a function of overall frequency of the corresponding topic.
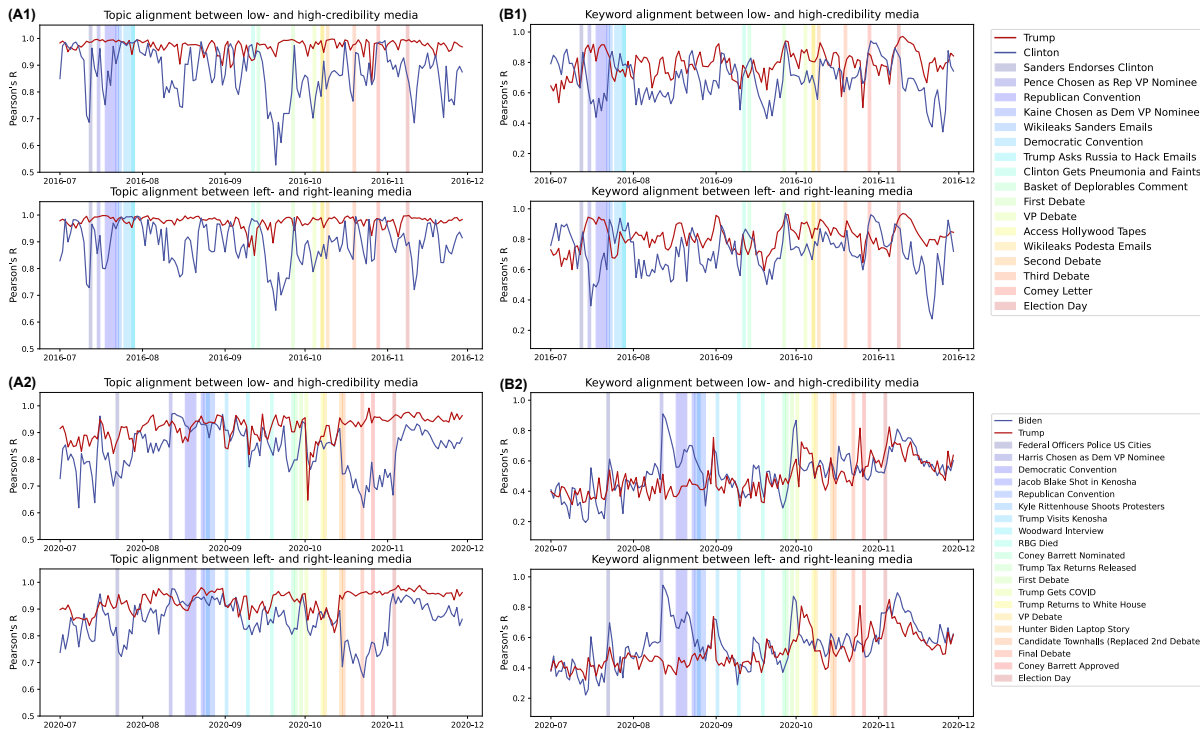
Figure 11: (A) Topic and (B) keyword alignment over time during in the 2016 (A1 and B1) and the 2020 (A2 and B2) U.S. Presidential Election campaign. Two pairings of media have been applied: (i) between low- and high-credibility news media, and (ii) between left- and right-leaning news media.

Between low- and high-credibility media

	Trump 2016					Clinton 2016			
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
---	---	---	---	---	---	---	---	---	---	---	---
1	clinton_controversies	0.0818	-0.3985	0.9783	0.0004	1	clinton_controversies	0.3771	-0.8581	0.9519	0.0000
2	healthcare	0.0463	0.3867	0.9734	0.0078	2	government_ops	0.3415	0.7204	0.9767	0.0000
3	government_ops	0.0348	-0.0975	0.9739	0.0215	3	healthcare	0.3292	-1.0742	0.9188	0.0000
4	religion	0.0347	0.5750	0.9738	0.0215	4	civil_rights	0.1625	1.0886	0.8809	0.0000
5	energy	0.0187	0.8224	0.9741	0.0925	5	defence	0.0485	1.3085	0.8693	0.0064
6	immigration	0.0136	-0.2392	0.9739	0.1532	6	foreign_trade	0.0439	4.3758	0.8706	0.0096
7	environment	0.0130	-0.5692	0.9731	0.1618	7	environment	0.0377	4.9955	0.8741	0.0165
8	sstc	0.0126	0.1448	0.9739	0.1694	8	trump_controversies	0.0297	0.6729	0.8750	0.0336
9	intl_affairs	0.0115	0.1368	0.9724	0.1891	9	economy	0.0177	0.6136	0.8706	0.1022
10	foreign_trade	0.0108	0.5925	0.9723	0.2035	10	energy	0.0077	-2.2671	0.8713	0.2826
11	defence	0.0027	0.1100	0.9731	0.5243	11	labour	0.0063	0.8403	0.8731	0.3298
12	trump_controversies	0.0026	0.0285	0.9746	0.5341	12	sstc	0.0038	-0.2669	0.8725	0.4481
13	economy	0.0025	0.0698	0.9729	0.5430	13	intl_affairs	0.0030	0.3211	0.8734	0.5031
14	education	0.0021	0.2084	0.9734	0.5721	14	religion	0.0028	0.6867	0.8712	0.5182
15	civil_rights	0.0007	-0.0266	0.9734	0.7446	15	education	0.0010	-0.5998	0.8709	0.6997
16	labour	0.0003	0.0644	0.9733	0.8464	16	crime	0.0010	0.0950	0.8699	0.6969
17	social_welfare	0.0001	-0.0875	0.9734	0.9063	17	social_welfare	0.0006	-0.8843	0.8714	0.7598
18	crime	0.0000	0.0007	0.9733	0.9945	18	immigration	0.0002	0.2195	0.8722	0.8507

|	Trump 2020					Biden 2020			
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
---	---	---	---	---	---	---	---	---	---	---	---
1	healthcare	0.4583	0.4737	0.9459	0.0000	1	biden_controversies	0.4331	-1.3262	1.0412	0.0000
2	government_ops	0.2056	-0.4172	0.9039	0.0000	2	healthcare	0.1132	0.5269	0.8452	0.0000
3	sstc	0.1400	-0.8924	0.9735	0.0000	3	sstc	0.1115	-1.5748	0.9071	0.0000
4	intl_affairs	0.0631	-0.8651	0.9339	0.0018	4	economy	0.0608	1.0088	0.8434	0.0022
5	labour	0.0593	-2.5242	0.9155	0.0025	5	social_welfare	0.0288	3.3612	0.8529	0.0367
6	education	0.0584	1.5050	0.9247	0.0027	6	defence	0.0245	1.3161	0.8400	0.0541
7	energy	0.0389	-2.5416	0.9152	0.0149	7	immigration	0.0152	-1.7457	0.8364	0.1303
8	social_welfare	0.0379	4.8195	0.9195	0.0162	8	energy	0.0120	0.7488	0.8432	0.1798
9	trump_controversies	0.0332	-0.2389	0.9193	0.0246	9	foreign_trade	0.0107	1.3465	0.8413	0.2040
10	immigration	0.0315	1.7172	0.9131	0.0286	10	civil_rights	0.0092	0.2771	0.8371	0.2405
11	environment	0.0266	-1.2478	0.9129	0.0446	11	environment	0.0072	-0.5412	0.8334	0.2992
12	biden_controversies	0.0221	-1.6594	0.9462	0.0676	12	crime	0.0067	-0.2425	0.8412	0.3143
13	civil_rights	0.0211	-0.2589	0.9133	0.0742	13	intl_affairs	0.0034	0.2468	0.8402	0.4745
14	foreign_trade	0.0201	-3.0306	0.9143	0.0818	14	trump_controversies	0.0022	0.1638	0.8371	0.5654
15	defence	0.0107	-0.4614	0.9224	0.2055	15	education	0.0012	0.4482	0.8399	0.6753
16	economy	0.0019	0.1600	0.9190	0.5917	16	government_ops	0.0011	0.0406	0.8460	0.6850
17	crime	0.0013	0.0634	0.9177	0.6551	17	labour	0.0005	-0.2058	0.8399	0.7870
18	religion	0.0000	-0.1049	0.9183	0.9344	18	religion	0.0002	-0.1412	0.8390	0.8626

Between left- and right-leaning media

	Trump 2016					Clinton 2016			
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
---	---	---	---	---	---	---	---	---	---	---	---
1	trump_controversies	0.3421	0.2847	0.9884	0.0000	1	clinton_controversies	0.5109	-0.8366	0.9702	0.0000
2	government_ops	0.2855	-0.2427	0.9773	0.0000	2	government_ops	0.4927	0.7247	0.9977	0.0000
3	sstc	0.0922	0.3410	0.9772	0.0001	3	healthcare	0.2963	-0.8535	0.9295	0.0000
4	clinton_controversies	0.0615	-0.3001	0.9797	0.0021	4	civil_rights	0.1551	0.8905	0.8996	0.0000
5	healthcare	0.0538	0.3623	0.9760	0.0040	5	defence	0.0303	0.8668	0.8905	0.0319
6	environment	0.0267	-0.7085	0.9757	0.0443	6	trump_controversies	0.0237	0.5026	0.8945	0.0585
7	immigration	0.0231	-0.2711	0.9766	0.0618	7	environment	0.0229	3.2566	0.8937	0.0630
8	religion	0.0224	0.4008	0.9763	0.0660	8	foreign_trade	0.0093	1.6879	0.8918	0.2368
9	economy	0.0209	-0.1762	0.9772	0.0756	9	economy	0.0072	0.3279	0.8916	0.2983
10	education	0.0074	-0.3383	0.9759	0.2905	10	intl_affairs	0.0021	0.2250	0.8933	0.5754
11	foreign_trade	0.0047	0.3403	0.9754	0.4012	11	education	0.0012	0.5600	0.8933	0.6672
12	labour	0.0033	0.2042	0.9758	0.4789	12	religion	0.0010	0.3518	0.8919	0.6928
13	energy	0.0031	0.2888	0.9763	0.4982	13	social_welfare	0.0009	0.8794	0.8928	0.7166
14	civil_rights	0.0022	-0.0409	0.9761	0.5634	14	immigration	0.0008	-0.3322	0.8919	0.7338
15	crime	0.0005	-0.0260	0.9762	0.7785	15	crime	0.0002	-0.0345	0.8930	0.8660
16	defence	0.0005	-0.0400	0.9761	0.7898	16	labour	0.0001	0.0661	0.8924	0.9272
17	social_welfare	0.0002	0.1137	0.9759	0.8604	17	energy	0.0000	0.1051	0.8923	0.9526
18	intl_affairs	0.0001	-0.0114	0.9761	0.9003	18	sstc	0.0000	-0.0246	0.8923	0.9336

	Trump 2020					Biden 2020			
rank	topic	r squared	coefficient	constant	p-value	rank	topic	r squared	coefficient	constant	p-value
---	---	---	---	---	---	---	---	---	---	---	---
1	healthcare	0.4520	0.3817	0.9539	0.0000	1	biden_controversies	0.7565	-1.5721	1.0982	0.0000
2	government_ops	0.2913	-0.4029	0.9178	0.0000	2	sstc	0.1202	-1.4664	0.9219	0.0000
3	sstc	0.2464	-0.9607	0.9911	0.0000	3	healthcare	0.1155	0.4772	0.8642	0.0000
4	education	0.1291	1.8162	0.9395	0.0000	4	foreign_trade	0.0656	2.9863	0.8640	0.0014
5	social_welfare	0.0806	5.7017	0.9332	0.0004	5	social_welfare	0.0512	4.0195	0.8752	0.0051
6	labour	0.0713	-2.2454	0.9293	0.0009	6	government_ops	0.0315	0.1947	0.8931	0.0286
7	immigration	0.0548	1.8365	0.9261	0.0037	7	economy	0.0224	0.5495	0.8608	0.0656
8	foreign_trade	0.0403	-3.4842	0.9272	0.0132	8	defence	0.0201	1.0680	0.8593	0.0817
9	energy	0.0353	-1.9664	0.9293	0.0204	9	immigration	0.0156	-1.5855	0.8561	0.1255
10	intl_affairs	0.0338	-0.5139	0.9409	0.0233	10	crime	0.0079	-0.2360	0.8606	0.2747
11	environment	0.0244	-0.9693	0.9275	0.0547	11	energy	0.0074	0.5273	0.8614	0.2928
12	civil_rights	0.0232	-0.2202	0.9275	0.0611	12	civil_rights	0.0071	0.2192	0.8570	0.3008
13	economy	0.0227	0.4459	0.9340	0.0640	13	religion	0.0039	0.5572	0.8571	0.4460
14	biden_controversies	0.0168	-1.1736	0.9514	0.1116	14	environment	0.0034	-0.3347	0.8549	0.4745
15	trump_controversies	0.0092	-0.1019	0.9320	0.2405	15	labour	0.0025	0.4178	0.8558	0.5407
16	crime	0.0003	-0.0262	0.9317	0.8198	16	intl_affairs	0.0016	0.1527	0.8592	0.6218
17	religion	0.0002	0.1607	0.9311	0.8765	17	trump_controversies	0.0006	0.0763	0.8575	0.7653
18	defence	0.0001	0.0442	0.9311	0.8814	18	education	0.0001	0.1030	0.8585	0.9146

Table 2: Summary table of OLS results to model temporal alignment between low- and high-credibility media (top), and between left- and right-leaning media (bottom). The dependent variable is the time series of temporal alignment (Pearson's R) in the overall topic distribution between low- and high-credibility media, or between left- and right-leaning media; the independent variable is the time series of the temporal difference in the proportional attention devoted to a certain topic (high-credibility subtracted by low-credibility, or right-leaning subtracted by left-leaning).
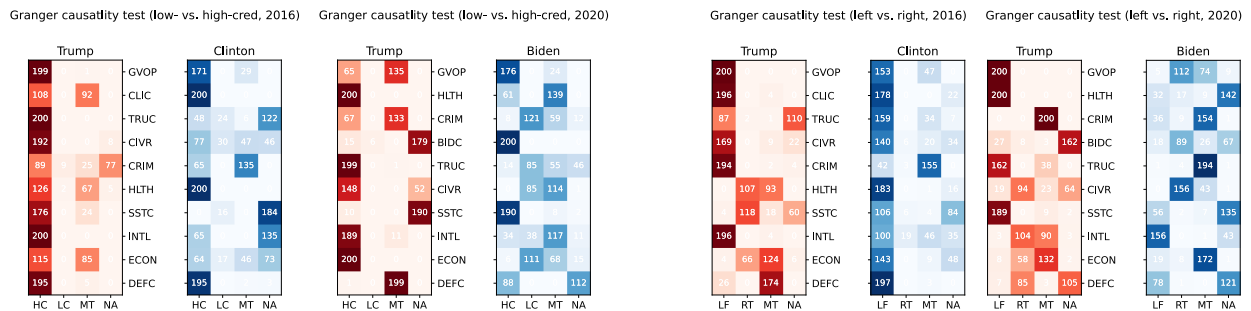
Figure 12: Granger causality between low- and high-credibility media (the left four columns) and between left- and right-leaning media (the right four columns). The cell values are the number of times a given type of IAS relationship appears significant out of 200 bootstrapping runs (sampling 80% of the data). Types of IAS relationship (along the X-axis) are displayed in abbreviations: HC means led by high-credibility media; LC means led by low-credibility media; LF means led by left-leaning media; RT means led by right-leaning media; MT means we found significant results (i.e., mutual interaction) in both directions; NA means we found no significant results in either direction. Significance threshold for p-value is 0.05. We include results for the top 10 topics (in descending order) that show frequently in all news headlines for a given year.
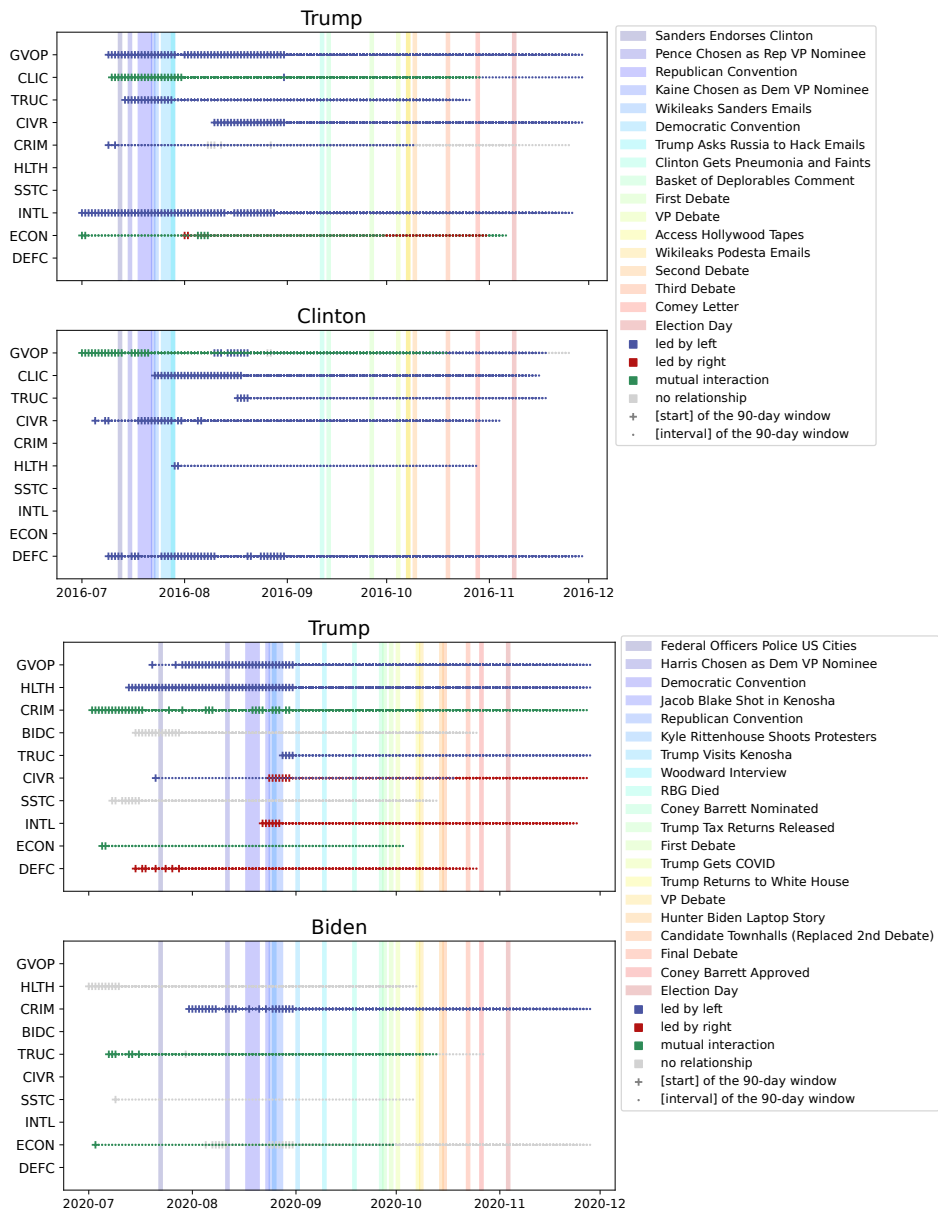
Figure 13: Granger causality results between low- and high-credibility media for 2016 (the first and second figures) and 2020 (the third and fourth figures). We test Granger causalities in a sliding window of 90 days and display robust results that appear in more than 95% of the bootstrapping runs. Each plus sign marks the starting point of the 90-day sliding window with a significant result. We include results for the top 10 topics that show frequently in all news headlines for a given year.

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

Table 3: List of topics in our dictionary. "Candidate controversies" shows up in their corresponding election year (e.g., Biden controversies only show up in 2020).

	Model variant		Headline			Survey			Tweets			Average across three sources		
year	keywords filtering/weighting	# of topics per text	model-human	human-human	diff.	model-human	human-human	diff.	model-human	human-human	diff.	model-human	human-human	diff.
2016	**strongly only**	**primary only**	**0.4881**	**0.5663**	**-0.0771**	**0.5661**	**0.6311**	**-0.0650**	**0.3564**	**0.4591**	**-0.1028**	**0.4702**	**0.5519**	**-0.0816**
2016	strongly + weakly (equal weights)	primary only	0.4566	0.5689	-0.1123	0.5537	0.6287	-0.0751	0.3068	0.4677	-0.1609	0.4390	0.5551	-0.1161
2016	strongly + weakly (different weights)	primary only	0.4876	0.5680	-0.0804	0.5419	0.6269	-0.0850	0.3118	0.4697	-0.1579	0.4471	0.5549	-0.1077
2016	strongly only	primary + secondary	0.2116	0.5993	-0.3877	0.3056	0.6337	-0.3282	0.2197	0.5026	-0.2829	0.2456	0.5785	-0.3329
2016	strongly + weakly (equal weights)	primary + secondary	0.1956	0.6015	-0.4059	0.2768	0.6331	-0.3563	0.1627	0.5021	-0.3394	0.2117	0.5789	-0.3672
2016	strongly + weakly (different weights)	primary + secondary	0.1989	0.6002	-0.4014	0.2739	0.6346	-0.3607	0.1646	0.5019	-0.3374	0.2124	0.5789	-0.3665
2016	strongly only	all included	0.2018	0.6619	-0.4601	0.3096	0.7023	-0.3927	0.2239	0.5914	-0.3675	0.2451	0.6519	-0.4068
2016	strongly + weakly (equal weights)	all included	0.1834	0.6646	-0.4813	0.2751	0.7036	-0.4285	0.1683	0.5896	-0.4213	0.2089	0.6526	-0.4437
2016	strongly + weakly (different weights)	all included	0.1846	0.6641	-0.4794	0.2748	0.7032	-0.4283	0.1679	0.5914	-0.4234	0.2091	0.6529	-0.4437
2020	**strongly only**	**primary only**	**0.4570**	**0.5149**	**-0.0580**	**0.5426**	**0.6021**	**-0.0595**	**0.3725**	**0.4700**	**-0.0975**	**0.4573**	**0.5290**	**-0.0717**
2020	strongly + weakly (equal weights)	primary only	0.4533	0.5116	-0.0583	0.5467	0.6015	-0.0549	0.3628	0.4704	-0.1077	0.4542	0.5279	-0.0736
2020	strongly + weakly (different weights)	primary only	0.4649	0.5133	-0.0484	0.5572	0.6039	-0.0467	0.3652	0.4716	-0.1065	0.4624	0.5296	-0.0672
2020	strongly only	primary + secondary	0.2252	0.5232	-0.2980	0.2650	0.6308	-0.3658	0.2226	0.4965	-0.2739	0.2376	0.5502	-0.3126
2020	strongly + weakly (equal weights)	primary + secondary	0.2105	0.5252	-0.3147	0.2457	0.6267	-0.3811	0.2024	0.4928	-0.2904	0.2195	0.5482	-0.3287
2020	strongly + weakly (different weights)	primary + secondary	0.2174	0.5272	-0.3099	0.2472	0.6263	-0.3791	0.1996	0.4944	-0.2948	0.2214	0.5493	-0.3279
2020	strongly only	all included	0.2272	0.6337	-0.4065	0.2657	0.6933	-0.4276	0.2235	0.5878	-0.3643	0.2388	0.6383	-0.3995
2020	strongly + weakly (equal weights)	all included	0.2198	0.6357	-0.4159	0.2407	0.6948	-0.4541	0.1899	0.5898	-0.3999	0.2168	0.6401	-0.4233
2020	strongly + weakly (different weights)	all included	0.2202	0.6339	-0.4137	0.2391	0.6944	-0.4552	0.1905	0.5891	-0.3986	0.2166	0.6391	-0.4225

Table 4: Complete agreement scores for model-human and human-human comparisons.