Social Network Analysis and Mining (2021) 11:87
https://doi.org/10.1007/s13278-021-00792-6

REVIEW PAPER

Heated conversations in a warming world: affective polarization in online climate change discourse follows real-world climate anomalies

Aman Tyagi[1] · **Joshua Uyheng**[2] · **Kathleen M. Carley**[2]

Received: 1 March 2021 / Revised: 13 August 2021 / Accepted: 17 August 2021 / Published online: 28 September 2021

Abstract

Climate change research describes online discourse as sharply polarized, echoing real-world divides in society. Yet while polarization in online climate change discourse has been extensively studied in terms of isolated communities and echo chambers, less is known about the extent of *affective polarization* that characterizes the hostile nature of intergroup interactions. Utilizing a combination of machine learning and network science tools, we design a methodological pipeline that quantifies the extent to which stance groups interact with more negative sentiment toward out-group members than in-group members. We apply this framework to 100 weeks of Twitter discourse about climate change. We find that deniers of climate change (Disbelievers) are more hostile towards people who believe (Believers) in the anthropogenic cause of climate change than vice versa. We also observe that Disbelievers use more words and hashtags related to natural disasters during more affectively polarized weeks as compared to Believers. Finally, using vector autoregression analysis, we find that climate anomalies in terms of both severe temperature and storms predict asymmetric shifts in online climate change discourse: Disbelievers grow more hostile toward out-groups, while Believers become less affectively polarized. These findings resonate with prior work on the asymmetric nature of polarization in contentious discourse, both around climate change and beyond. Our work also extends existing findings around temporal associations between climate anomalies, divided media representations, and real-world conflicts. We conclude with implications for science communication and studying affective polarization in online discourse, especially concerning the subject of climate change.

Keywords Climate change · Affective polarization · Stance detection · Online social networks

1 Introduction

Online social networks represent a powerful space for public discourse. Through large-scale, interconnected platforms like social media, diverse communities may potentially participate in open exchanges of views and information about a vast range of issues. However, research has increasingly demonstrated the dangers of *polarization* in online communication (Barberá et al. 2015; Tyagi et al. 2020; Tyagi and Carley 2020). Attributed to various psychological, social, and technological factors, intergroup communication in cyberspace has displayed tendencies to feature pathological dynamics especially concerning contentious issues (Geschke et al. 2019; Yarchi et al. 2020). For example, opposed groups may communicate in a highly balkanized fashion, such that members of an in-group are only minimally exposed to out-group members and their beliefs (Garimella et al. 2018; Karlsen et al. 2017). This phenomenon has been termed

This work was supported in part by the Knight Foundation and the Office of Naval Research Grants N000141812106 and N000141812108. Additional support was provided by the Center for Computational Analysis of Social and Organizational Systems (CASOS), the Center for Informed Democracy and Social Cybersecurity (IDeaS), and the Department of Engineering and Public Policy of Carnegie Mellon University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, Office of Naval Research or the US government.

✉ Aman Tyagi
amantyagi@cmu.edu

[1] CASOS Center, Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA

[2] CASOS Center, Institute for Software Research, Carnegie Mellon University, Pittsburgh, PA 15213, USA

interactional polarization. Polarization can also pertain to highly negative sentiments toward out-groups in the form of *affective polarization* (Anderson and Huntington 2017; Druckman and Levendusky 2019). Social scientific research examines how these phenomena are interconnected across a variety of contexts, such that online groups that disagree on a given topic are also more likely to be hostile toward each other (Uyheng and Carley 2020; Yarchi et al. 2020).

This paper focuses on quantifying affective polarization between two groups with opposing beliefs using Twitter discourse on an important social issue. One such issue which has received heated attention in online public discourse is climate change (Dunlap et al. 2016; Fisher et al. 2013; Tyagi et al. 2020). We focus on those who cognitively accept anthropogenic causes of climate change (*Believers*) and those who reject the same (*Disbelievers*). Previous work demonstrates sharp divergences in climate change beliefs and the emergence of interactionally polarized groups (Hamilton et al. 2015; Milfont et al. 2017; Tyagi et al. 2020). In other words, online discussions about climate change are *interactionally polarized*, implying the persistence of echo chambers between Believers and Disbelievers (Jang and Hart 2015; Eck et al. 2020; Williams et al. 2015).

Much less work, however, engages the question of *affective polarization* in online climate change discourse. A crucial limitation in prior work lies in the methodological options available to past researchers. Relying largely on manually annotated corpora and datasets of limited size, existing scholarship has faced barriers to measuring the emotional component of climate change discussions in a generalizable fashion (Anderson and Huntington 2017; Jang and Hart 2015; Eck et al. 2020). Drawing on recent advances in computational stance detection, targeted sentiment analysis, and network science measures, we design and deploy an integrated methodological pipeline for addressing this gap in the literature (Huang et al. 2018; Krackhardt and Stern 1988; Kumar 2020).

In this work, we specifically leverage computational methods to generate (a) automated stance labels for climate change Believers and Disbelievers, (b) individual measurements of the interaction valence between in-group and out-group members, and (c) broader assessments of group-level affective polarization. We demonstrate the utility of our framework by applying our methodology to a large-scale dataset of 100 weeks of online climate change discussion on Twitter. To illustrate the utility of our findings, we further show how our proposed affective polarization score relates to real-world climate anomalies. First, we show how changes in specific discourse around natural disasters relate to fluctuations in the level of asymmetric polarization between Believer and Disbeliever groups. Second, through vector autoregression analysis, we show that actual climate anomalies temporally predict asymmetric shifts in affective polarization for Believers and Disbelievers. Taken together, these results speak to the importance of understanding the emotional dimensions of intergroup conflicts online, particularly with respect to the pressing issue of climate change.

In sum, we probe the following research questions:

1. How can affective polarization be measured on a large-scale online conversation about climate change?
2. What levels of affective polarization are featured by climate change Believers or Disbelievers over time?
3. What is the relationship of affective polarization with the use of natural disaster related words?
4. How do real-world climate anomalies associate with levels of affective polarization among Believers and Disbelievers?

The subsequent sections of this paper are organized as follows. First, we provide an overview of related work, illustrating computational analysis of polarization in general terms and then in the case of climate change specifically in Sect. 2. We zero in on the dearth of principled empirical work on affective polarization, particularly in relation to online climate change discourse. Second, Sect. 3 presents our proposed methodological pipeline which integrates machine learning models and network science techniques to quantify affective polarization on online social networks. Third, we share our findings on our large-scale, long-term Twitter dataset in Sect. 4. Finally, we conclude in Sect. 5 with the implications of this work for understanding the state of climate change discourse on digital platforms as well as related empirical investigation of affective polarization on online social networks.

2 Related work

2.1 Computational analysis of polarization

Recognizing the ubiquity of online conflicts, extensive scholarship in the computational and social sciences has tackled the problem of polarization. More traditional approaches in offline settings have relied on survey measures to empirically assess divergence in beliefs between groups (Banks et al. 2020; Druckman and Levendusky 2019). But with burgeoning developments in computational methods—especially with respect to natural language processing and machine learning—automated methods have also arisen to leverage the vast digital traces linked to online activity (Huang et al. 2018; Kumar 2020).

General approaches to studying polarization infer individual attitudes from user information, such as the texts associated with an account on social media (e.g., Facebook comments, tweets). Group membership as well as group communication are similarly incorporated into analyses of polarization, by examining the beliefs of individuals in conjunction with their traceable patterns of digital interaction with other individuals.

Given various conceptualizations of polarization, different frameworks have been developed to quantify pathological patterns of communication across groups holding similar or opposed stances on a given issue (Darwish 2019; Demszky et al. 2019; Morales et al. 2015; Weber et al. 2013). Representing online conversations as graphical structures, social network approaches typically measure polarization in terms of a function of homophily in local community structures (Matakos et al. 2017; Stewart et al. 2018). In other words, the extent to which the likelihood that those holding similar views interact with each other—in contrast to those with whom they disagree. For example, one may quantify the probability of a random walk starting from a node belonging to a given stance group ending up in a node belonging to the same or a different stance group (Garimella et al. 2018; Tyagi et al. 2020; Yarchi et al. 2020).

More recent scholarship, however, emphasizes the importance of examining not just pathologically *isolated* communication, or interactional polarization; but also pathologically *hostile* communication, or affective polarization. Burgeoning evidence suggests that echo chambers represent an incomplete picture of polarization (Barberá et al. 2015). People holding opposed views, in fact, do interact with each other - but this does not necessarily mitigate polarization (Karlsen et al. 2017). Instead, research finds that intergroup exposures trigger further incivility (Anderson and Huntington 2017). Hence, reliable measures for affective polarization are needed, although the computational literature in this area remains in its nascent stages (Yarchi et al. 2020).

2.2 Climate change and polarized discourse

In the specific case of climate change discourse, analysis of polarized discourse has also represented a major research topic. Numerous studies link polarized beliefs about climate change to partisan divides, with more conservative individuals less likely to cognitively accept anthropogenic climate change than liberals (Dunlap et al. 2016; Hamilton et al. 2015). Past work specifically demonstrates that although higher levels of education and information access may increase the likelihood of climate change belief, these effects remain much lower among conservatives (Hamilton et al. 2015; McCright and Dunlap 2011). Such effects have been explained from the lens of elite signalling—whereby followers emulate the beliefs of their preferred political leaders—uneven exposure to information based on partisan media, as well as a generalized dislike for the members of the opposed ideological group (Bolsen and Shapiro 2018; Carmichael and Brulle 2017; Uyheng et al. 2021; Van Boven et al. 2018).

However, with time, scholars have also noted general trends toward increasing climate change beliefs overall (Milfont et al. 2017). Even if these do not necessarily translate into concrete support for policy (Fisher et al. 2013), the long-term instability of skepticism points to valuable ways forward for science communication (Jenkins-Smith et al. 2020). Collectively, these findings suggest the importance of accounting for the psychological processes surrounding climate change belief and disbelief, going beyond the transmission of information (Kahan et al. 2015).

These issues take on specific forms in cyberspace, where information flows are inextricably entangled with community dynamics. On social media, studies suggest employing social network analysis have uncovered robust evidence that online climate change discussions tend to exhibit echo chamber-like interactions (Williams et al. 2015; Tyagi et al. 2020). Qualitative analysis further showed that in rare instances of intergroup communication, more negative frames prevailed, featuring dismissal of climate change as a hoax, identity-based derailment of conversations, as well as overall higher levels of incivility (Anderson and Huntington 2017; Eck et al. 2020). Not withstanding the valuable idiographic insights derived from these studies, existing studies, however, rely on a minuscule fraction of the larger conversation to facilitate in-depth content analysis. Hence, larger-scale and more generalizable findings on the affective dynamics of online climate change discourse are notably lacking in the literature.

Finally, past studies suggest that these processes do not take place in a vacuum, but correlate with real-world climate anomalies. Various disciplines caution that clear causal chains are challenging to derive between climate change and the outbreak of conflicts in the real world (Evans 2019; Koubi 2019). Yet meta-analyses and reviews have suggested a convergence in the literature that climatological changes predict real-world conflicts across a range of spatial and temporal scales, linked to psychological, political, and economic conditions (Hsiang and Burke 2014). That said, other scholars have also shown that in certain cases, climate anomalies may drive further intergroup solidarity, pointing to a complex range of relationships between the climate and conflicts, and the need for further empirical work (Buhaug et al. 2014). On the level of discourse, one study has shown that top-down media representations of climate change appear to grow more polarized following anomalous fluctuations in global temperature (Pianta and Sisco 2020). However, to our knowledge, no work has systematically examined how these dynamics might drive conflicts in bottom-up discourse among the general public.

2.3 Contributions of this work

Motivated by the foregoing insights, this work seeks to contribute to the literature by offering an empirical analysis of affective polarization in online climate change discourse, using a computational methodology that integrates machine learning and network science techniques (Uyheng et al. 2019). As the succeeding sections demonstrate, this interoperable pipeline applies not solely to the topic of climate change, but may be deployed in a scalable and generalizable fashion across contentious issues writ large. For climate change research in particular, our framework overcomes methodological barriers present in prior work, including their common reliance on expensive survey or experimental measures, or manually annotated datasets in the context of social media research on climate change discourse (Hamilton et al. 2015; Milfont et al. 2017; Williams et al. 2015).

From a theoretical standpoint, we additionally contribute a nuanced operationalization of affective polarization as located on a group level. We unpack how group-level metrics valuably produce asymmetrical views of hostile behavior, thereby facilitating more fine-grained analysis of how different stance groups engage in varied levels of affectively polarized interactions. This conceptually aligns with the asymmetry of psychological factors characterizing climate change Believers and Disbelievers, especially over time (Dunlap et al. 2016; Jenkins-Smith et al. 2020; Van Boven et al. 2018).

Moreover, while established findings paint a picture of consistent echo chambers between climate change Believers and Disbelievers, we provide evidence for the flipside of these dynamics. We specifically quantify, over a larger-scale and longer-term dataset than previously examined in prior work, the extent to which intergroup interactions not only take place, but also systematically feature hostile emotional dynamics. Lastly, by showing their links to real-world climate anomalies, we also augment the existing literature on climate change and conflicts to include clashes in online public discourse (Pianta and Sisco 2020). This may inform possible data-driven interventions for policy-making beyond more prevalent frames of intergroup contact and science communication (Kahan et al. 2015).

3 Data and methods

3.1 Data collection

We collected tweets using Twitter's standard API[1] with keywords "Climate Change", "#ActOnClimate", "#ClimateChange". Our dataset was collected between August 26th, 2017 to September 14th, 2019. Due to server errors, the collection was paused from April 7th, 2018 to May 21st, 2018, and again from May 12th, 2019 to May 16th, 2019. We ignore these periods in our analysis. After deduplicating tweets, our dataset consisted of 38M unique tweets and retweets from 7M unique users. For our analysis, we aggregate tweets from each user for each seven-day period (1 week) to get a total of 100 weeks.

3.2 Stance labels

We use a state-of-the-art stance mining method (Kumar 2020) to label each user as a climate change Disbeliever or Believer. We use a weak supervision-based machine learning model to label the users in our dataset. The model uses a co-training approach with label propagation and text-classification. The model requires a set of seed hashtags essentially being used by Believers and Disbelievers. The model then labels seed users based on the hashtags used at the end of the tweet. Using the seed users, the model trains a text classifier and uses a combined user-retweet and user-hashtag network to propagate labels. In an iterative process, the model then labels users who are assigned a label by both methods with high confidence.

We set *ClimateChangeIsReal* and *SavetheEarth* as Believers seed hashtags and *ClimateHoax* and *Qanon* as Disbelievers seed hashtags. These hashtags have been shown to be used mostly by the respective groups (Tyagi et al. 2020). The algorithm labels 3.9M as Believers and 3.1M as Disbelievers[2]

3.3 Affective polarization metrics

We measure affective polarization in this work by combining outputs from an aspect-level sentiment model, a classic network science measure known as the E/I index (Krackhardt and Stern 1988) and Earth Mover's Distance (EMD) (Hitchcock 1941). Outputs are combined in the five steps which follow to produce dynamic group-level measurements of affective polarization.

[1] https://developer.Twitter.com/en/docs/tweets/search/overview/standard.

[2] We randomly sampled 1000 users from each group to manually validate the results. We label a user as Disbeliever if we find any Tweet akin to someone who does not believe in climate change or anthropogenic cause of climate change. Otherwise, we label the user as Believer. We observe that the average precision from manual validation of 2000 users is 81.2%. We use the parameter values as defined in Kumar (2020) as $k = 5000, p = 5000, \theta(I) = 0.1, \theta(U) = 0.0, \theta(T) = 0.7$.

3.3.1 Aspect-level sentiment

Aspect-level sentiment refers to the emotional valence of a given utterance toward one of the concepts it mentions. Sentiments toward specific entities are vital to consider in polarized discussions such as those we consider here. For instance, climate change Disbelievers might express negative feelings toward notions of greenhouse gases, while in agreement with fellow Disbelievers with whom they are interacting.

We utilize Netmapper to extract entities from each tweet, and predict the aspect-level sentiment of each tweet toward each entity (Carley et al. 2018). Netmapper is reasonably accurate for sentiment analysis tasks (Uyheng and Carley 2020; Uyheng et al. 2019). Word-level sentiment is computed based on the average of known valences for surrounding words within a sliding window. For the purposes of this work, each tweet by a certain agent i which mentions or replies to agent j is assigned an aspect-level sentiment score from -1 (very negative) to $+1$ (very positive) directed toward the concept "@[agent j]". This allowed us to compute affective dimensions to the communication between groups of the same or opposed stance groups.

3.3.2 Affective networks

Let $G^+ = (A, E^+)$ denote a positive interaction network where the set of vertices A contains all Twitter accounts in our dataset and the set of directed edges E^+ contains all positive-valenced mentions and replies between agents in A. Similarly, let $G^- = (A, E^-)$ denote a negative interaction network over the same set of agents A and the set of directed edges E^- representing their negative-valenced mentions and replies. Let S_{ij} denote the set of all aspect-level sentiments in tweets by agent i toward agent j, where $i, j \in A$. Then the weight w^+_{ij} of edge $e^+_{ij} \in E^+$ from i to j is given by $\sum_{x \in S_{ij}} \min(0, x)$. Conversely, the weight w^-_{ij} of edge $e^-_{ij} \in E^-$ from i to j is given by $\sum_{x \in S_{ij}} \min(0, -x)$.

3.3.3 E/I indices

We assess group-level differences in positive and negative interactions using Krackhardt's E/I index (Krackhardt and Stern 1988). For a given affective network, the E/I index intuitively captures the extent to which each stance group k engages in correspondingly valenced interactions with members of the out-group relative to their in-group (Uyheng and Carley 2020; Uyheng et al. 2021). Hence, for instance, high values of the E/I index for the negative interaction network would indicate that the given stance group interacts in a more negative way to their opponents relative to those who share their beliefs. To compute the E/I indices, let $A_k \subseteq A$ denote the set of agents belonging to stance k and $A_{k'}$ those who do not hold stance k. The E/I index of stance group k on the positive interaction network is therefore computed as follows:

$$P^+_k = \frac{E^+_k - I^+_k}{E^+_k + I^+_k} \tag{1}$$

where $E^+_k = \sum_{i \in A_k, j \in A_{k'}} w^+_{ij}$ and $I^+_k = \sum_{i,j \in A_k} w^+_{ij}$. On the other hand, the E/I index of stance group k on the negative interaction network is similarly computed thus:

$$P^-_k = \frac{E^-_k - I^-_k}{E^-_k + I^-_k} \tag{2}$$

where $E^-_k = \sum_{i \in A_k, j \in A_{k'}} w^-_{ij}$ and $I^-_k = \sum_{i,j \in A_k} w^-_{ij}$. Given the construction of P^+_k and P^-_k, we note that both values are bounded between -1 and $+1$.

3.3.4 Polarization valence

We find whether the interactions have negative valence or positive valence by defining polarization P_k as expressed below:

$$P_k = \frac{P^-_k - P^+_k}{2}. \tag{3}$$

In this work, we operationalize our view of affective polarization in terms of high E/I indices on the negative interaction network, and low values on the positive interaction network. P_k assigns positive values for groups that display disproportionately hostile or negative interactions toward the out-group relative to their in-group. Values close to 0, on the other hand, indicate relatively even levels of positive and negative interactions. Finally, negative values indicate that those holding stance k are more negative to their in-group but positive to their out-group. Values of P_k are usefully bounded between -1 and $+1$ for ease of interpretability and comparison over different networks and time periods.

3.3.5 Polarization magnitude

To find the magnitude of affective polarization we use Earth Mover's Distance (EMD) on the distribution of weighted edges for outgroup and ingroup interactions. This is similar to computing first Wasserstein distance between two 1D distributions (Ramdas et al. 2017). Similar to affective networks, we define $G = (A, E)$ as interaction network where the set of vertices A contains all Twitter accounts in our dataset and the set of directed edges E contains all valenced (positive or negative) mentions and replies between agents in A. In this case, we do not separate negative and positive valence graphs and treat weight w_{ij} of edge $e_{ij} \in E$ from i to

j as given by $\sum_{x \in S_{ij}} x$. Let u_k be distribution of w_{ij}, where $i \in A_k, j \in A_{k'}$, that is, for i and j belonging to different stance groups k and k'. Similarly, let v_k be distribution of w_{ij}, where $i, j \in A_k$, that is, for all i and j belonging to the same stance group k. For a group holding stance k, we define our proposed affective polarization metric as:

$$l_k = \begin{cases} -\int_{-\infty}^{+\infty} |U_k - V_k| & : P_k < 0 \\ \int_{-\infty}^{+\infty} |U_k - V_k| & : P_k \geq 0 \end{cases} \tag{4}$$

where U_k and V_k are the respective CDFs of u_k and v_k. Here, EMD is proportional to the minimum amount of work required to covert one distribution to another.[3] We use P_k to assign positive or negative valence to the EMD. Other techniques exist to find the difference in distribution such as KS-Test (Massey 1951). However, during our experiments, we found that EMD is able to capture more nuanced differences in distributions. This may be because the EMD can capture differences in heavy-tailed distributions better and it does not make any parametric assumptions (Ramdas et al. 2017).

Our affective polarization metric l_k is positive when $P_k > 0$. As noted in Sect. 3.3.4, a positive value would mean more hostility or negative sentiment in intergroup communication compared to intragroup communication. On the other hand, a negative value of l_k is when $P_k < 0$, meaning more positive sentiment in intergroup communication compared to intragroup communication.

3.4 Temporal links to real-world climate anomalies

To examine links between real-world climate anomalies and the affective polarization score introduced in this work, we utilize vector autoregression. By combining real-world records of climatological shifts with our assessment of online discourse, we can systematically probe their temporal associations.

3.4.1 Real-world climate anomalies

We take global records of real-world climate anomalies by referring to the databases of the National Climatic Data Center (NCDC).[4] We specifically use two records to denote real-world climate anomalies. The first measure we use is the NCDC time series on *global surface temperature anomalies*, which refer to average global deviations in surface temperature relative to expected reference values.[5] Higher values of this index indicate greater increases in surface temperature

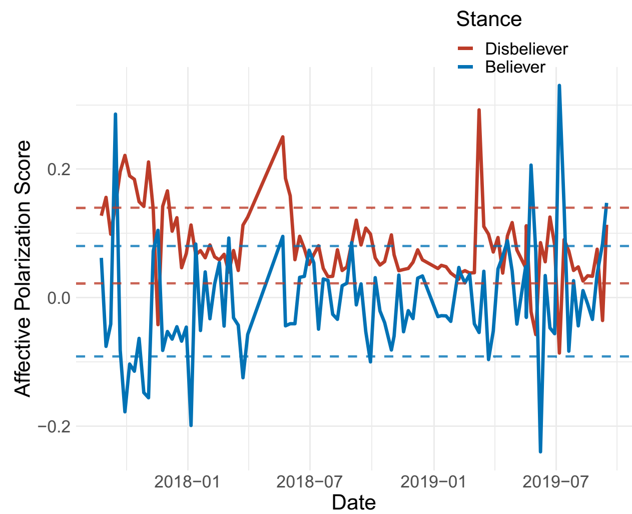

Fig. 1 Affective polarization metric (l_k) for Believers and Disbelievers of climate change. Higher positive values denote more hostility towards the other group. The dotted lines represent mean ±1 standard deviation, which for Believers is −0.091 and 0.080 and Disbelievers is −0.117 and 0.106. The analysis was done on data collected from 26th August 2017 to 14th September 2019 as described in Sect. 3.1

than expected. The second measure we use is the NCDC record for *extreme storm events*.[6] We specifically take the sum of all records of hurricanes, tropical storms, typhoons, and tropical depressions. For this latter measure, we utilize the logarithm of storm counts to account for differences in scale.

3.4.2 Vector autoregression

We then utilize vector autoregression to model temporal associations between affective polarization scores and real-world climate anomalies. Vector autoregression determines the relationship between past values of a set of variables on their later values (Toda and Phillips 1994). Vector autoregression models have been used widely across the social sciences, with relevant examples linking presidential rhetoric and economic outcomes (Wood et al. 2005), determining agenda-setting affects between political elites and their followers (Barberá et al. 2019), or the formation of echo chambers in response to automated bot activity (Uyheng and Carley 2020).

Here, we specifically regress levels of Believer and Disbeliever affective polarization scores on each other's levels of affective polarization, and include as covariates the climate anomalies described above. Theoretically, this tests whether Believer and Disbeliever levels of hostility predict later

[3] http://infolab.stanford.edu/pub/cstr/reports/cs/tr/99/1620/CS-TR-99-1620.ch4.pdf.

[4] https://www.ncdc.noaa.gov/cdo-web/.

[5] https://www.ncdc.noaa.gov/monitoring-references/faq/anomalies.php.

[6] https://www.ncdc.noaa.gov/stormevents/ftp.jsp.

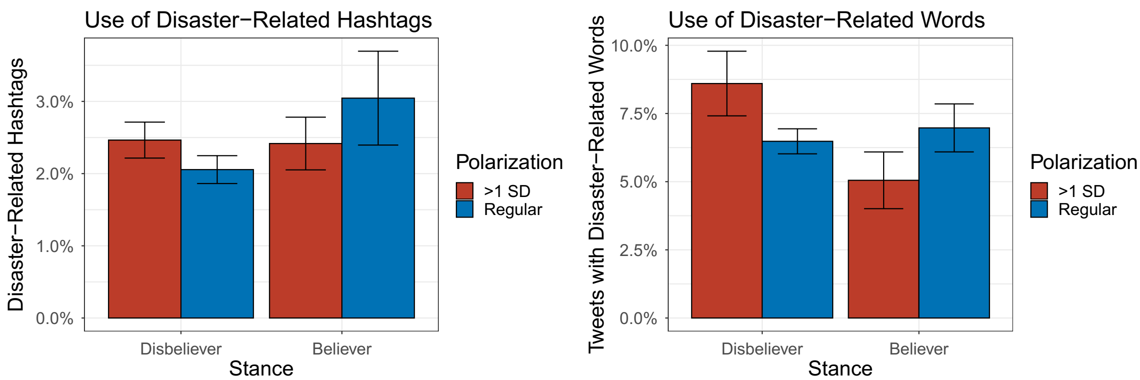

Fig. 2 Comparison of relative salience of disaster-related talk when the affective polarization metric is greater than 1 standard deviation or otherwise. Error bars represent ±1 standard errors. Left: Percentage of the top 100 most frequent hashtags containing natural disaster-related words. Right: Percentage of tweets with at least one natural disaster-related word

levels of intergroup hostility, while also examining whether fluctuations in affective polarization systematically follow from real-world climate anomalies. Our models include constant and trend terms, and determine the optimal number of lags using the Akaike Information Criterion (AIC).

4 Results

Using the metric defined in Equation 4, in this section, we explore how affective polarization between Believers and Disbelievers evolves over the 100 weeks under consideration. Then we explore how hostile periods are related to natural disaster-related words, and later, to real-world climate anomalies.

4.1 Long-term trends in affective polarization

We first look at how the affective polarization metric is changing over time in Fig. 1. Overall, our analysis found that climate change Disbelievers tended to exhibit high levels of hostility toward climate change Believers. This finding was relatively consistent throughout the 100-week period under observation, as the time series for climate change Disbelievers only very rarely goes below the threshold of 0, which would have been periods exhibiting similarly valenced interactions toward in-group and out-group members. Some weeks displayed exceptionally high levels of hostility toward climate change Believers, greater than one standard deviation from the mean. The standard deviation of l_k is lower for Disbelievers than for Believers, indicating that Disbelievers act in a much more consistent manner over the 100 weeks than Believers.

Climate change Believers, on the other hand, were not generally hostile toward Disbelievers, as the time series for climate change Believers tends to fluctuate over and under the threshold of 0. This indicates that climate change Believers communicate with in-group and out-group members with relatively similar emotional valence. However, on certain weeks, climate change Believers did also feature exceptionally high hostility scores. This suggests that climate change Believers may also behave in a hostile manner toward climate change Disbelievers, even if not consistently over the long term.

Manual investigation of the highly hostile weeks suggests some insights. These are the peaks represented in Fig. 1. During these weeks, Believers were either targeting former United States President Trump for his comments or policies on climate change, Pope's remarks on oil companies to act on climate, or Hurricane Maria.[7] On the other hand, Disbelievers are tweeting about natural disasters (winter storms, hurricanes, and wildfires), the Canadian government's decision on gas pipelines, or surprisingly about the 2017 Las Vegas shooting.[8]

4.2 Asymmetric hostilities in discussions of natural disasters

Next, to further investigate instances where hostility between Believers and Disbelievers is high we compare those weeks with weeks where hostility is low. We define hostile weeks as those data points where l_k is more than their respective group means plus 1 standard deviation. On Fig. 1, this corresponds to all the weeks where for Believers $l_k > 0.080$ and for Disbelievers $l_k > 0.140$. The number of such weeks for Disbelievers where $l_k > 0.140$ is 20 and for Believers where $l_k > 0.080$ is 12. We look further into these weeks as examples of exceptionally hostile weeks.

[7] https://en.wikipedia.org/wiki/Hurricane_Maria.

[8] https://en.wikipedia.org/wiki/2017_Las_Vegas_shooting.

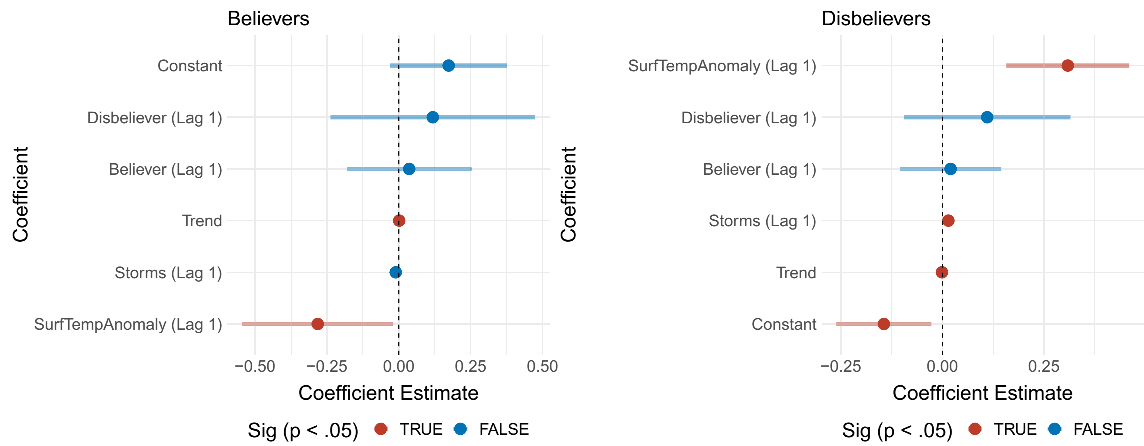

Fig. 3 Results of vector autoregression for believers (Left) and disbelievers (Right). Coefficients are presented with 95% confidence intervals. Statistically significant effects are depicted in red. The trend components are close to zero, but statistically significant

We use natural disaster-related hashtags and words to examine the salience of these topics in periods of sharpened hostility between the two groups. In the first plot of Fig. 2 we look at the top 100 most frequent hashtags used within those groups to find the percentage of hashtags related to natural disasters. On regular weeks within one standard deviation of the mean, Believers use more natural disaster-related hashtags than Disbelievers. However, during the exceptionally hostile weeks, Believers use these hashtags less.

Conversely, for Disbelievers, we observe the opposite behavior. Disbelievers use more natural disaster-related hashtags when they are more hostile towards Believers. We provide further evidence of this finding in the second plot of Fig. 2. Here, we look at the percentage of Tweets with at least one natural disaster-related word. We find similar patterns as mentioned above. Moreover, we find that a greater percentage of tweets from Disbelievers mention natural disaster-related words compared to Believers. This indicates that Disbelievers discuss natural disasters more when they are exceptionally hostile towards Believers compared to other weeks.

4.3 Asymmetric shifts in affective polarization following climate anomalies

Finally, we explore temporal associations between online moments of affective polarization and real-world climate anomalies. Figure 3 depicts the coefficients of the vector autoregression models for Believers and Disbelievers. For both groups, the best lag value was found to be one week in the past.

For Believers, statistically significant effects may be seen for the time series's trend component and the lagged values of surface temperature anomalies. This indicates that the overall trend of Believers' affective polarization appears to be increasing, even if the level of temporal increase is relatively small over time ($p < .05$). Furthermore, the statistically significant, negative coefficient for surface temperature anomalies suggests that Believers turn *less* affectively polarized following periods of greater surface temperature anomalies ($p < .01$). This indicates that, during these periods, Believers express less negative emotion toward Disbelievers. Alternatively, Believers may also express higher levels of in-group negative emotion following periods of higher surface temperature anomalies. Interestingly, these real-world fluctuations have greater predictive value for future levels of affective polarization than past levels of affective polarization. Disbelievers' hostility also does not predict Believers' future hostility. That said, we note that while for Believers, the overall model is statistically significant ($F(5, 93) = 2.642, p < .05$), the variance explained is only $R^2 = 12.44\%$, indicating many other factors may explain long-term affective polarization among climate change Believers.

Among Disbelievers, we detect negative constant ($p < .05$) and trend components ($p < .001$). This indicates that, although Disbelievers do attempt to be much more affectively polarized, their levels of hostility are decreasing (albeit very gradually) over time. Most interestingly, however, we see that real-world climate anomalies have the reverse predictive relationship with Disbeliever hostility compared to Believers. Here, we observe statistically significant and positive effects for *both* surface temperature anomalies ($p < .001$) and storms ($p < .01$). This suggests that in these periods of greater climate anomalies and extreme weather events, Disbelievers grow *more* hostile toward Believers. These effects are moreover more predictive than past levels of either Believer or Disbeliever affective polarization. The overall model is also statistically significant ($F(5, 93) = 12.31, p < .001$), and explains over three times as much variance than in the Believer case ($R^2 = 39.82\%$). While this still points to other factors which explain this variability, Disbeliever hostilities appear to be much more closely linked to real-world climate fluctuations than those of Believers.

5 Conclusions and future work

Taken together, our findings suggest the importance of considering affective polarization in online discourse, particularly concerning the subject of climate change. Whereas past studies had shed light on the echo chamber dynamics which characterized intergroup communication surrounding climate change (Tyagi et al. 2020; Williams et al. 2015), we show how this polarization also extends to the realm of emotion in the form of affective polarization (Yarchi et al. 2020). We extend existing studies which highlight the role of incivility and personalized framing in encounters between climate change Believers and Disbelievers by designing and deploying a scalable and interoperable technique for analyzing relative intra- and intergroup interaction valence (Anderson and Huntington 2017; Eck et al. 2020). This allowed us to quantify the extent of hostile communications between the two groups over a large-scale, long-term dataset, and explore underlying patterns in their fluctuations over time.

In particular, our findings empirically affirm the value of viewing polarization from an asymmetrical perspective. Related scholarship in political psychology underscores how ideological asymmetries underpin conflict dynamics across a variety of social issues (Jost 2017). In other words, the participation of two groups within polarized discourse does not necessarily mean that both groups engage in conflict in the same way. Prior work illustrates that these findings translate robustly to the digital sphere—political elites or opinion leaders who share moralized content online behave in distinct ways depending on their ideological orientations and group membership (Brady et al. 2017; Carmichael and Brulle 2017). Our proposed methodology offers an interoperable framework to examine these phenomena empirically and is broadly applicable to a wide range of contentious issues (Uyheng et al. 2019, 2021).

From an empirical standpoint, the present work contributes to the growing literature on asymmetric polarization by showing how these dynamics unfold both affectively and in the context of climate change conversations. Indeed, higher levels of hostility from Disbelievers present a specifically notable finding for social scientific scholarship on discourse about climate change. Longitudinal analysis in prior work suggests that generalized climate change beliefs over time are increasing (Hamilton et al. 2015; Milfont et al. 2017), and climate change Disbelievers, in particular, are more susceptible to potential belief change (Jenkins-Smith et al. 2020). However, significant psychological barriers remain for fuller acceptance of anthropogenic causes for climate change, and the corresponding urgency for responsive policy changes (Dunlap et al. 2016; Fisher et al. 2013). For instance, as previously pointed out, one psychological factor which impedes climate change Beliefs is not related to the climate at all, but anchors primarily on the feelings of dislike felt by one group towards the other (Van Boven et al. 2018). Higher levels of hostility among climate change Disbelievers toward climate change Believers thus constitutes one major obstacle for further dialogue between the two groups, as well as for science communication more broadly.

These insights are especially important to consider in light of our secondary set of findings. Our analysis suggests that further asymmetries arise between Believers and Disbelievers engagement with disaster words in relation to their levels of affective polarization, as well as following real-world climate anomalies. As we have noted, mixed results are present in the literature with respect to the linkages between climate change and various forms of conflict (Buhaug et al. 2014; Evans 2019; Hsiang and Burke 2014). Here, we similarly make no causal assumptions regarding our findings (Koubi 2019). However, we do point out that our findings appear to converge regarding the associative relationships between climate anomalies and affective polarization for both Believers and Disbelievers.

In particular, we note that comparable levels of natural disaster talk are seen when both groups are within average levels of affective polarization. Meanwhile, moments of increased affective polarization correlate with opposite behaviors for Believers and Disbelievers. Believers appear to shift to other areas of contention, whereas Disbelievers' increased invocation of disaster terms points to the more aggressive discussion of these catastrophes, albeit positioned in resistance to explanations related to anthropogenic climate change. With respect to real-world events, we also see that Believers featured *reduced* affective polarization following climate anomalies, whereas Disbelievers experienced *heightened* hostilities. Taken together, these findings suggest another layer of intractable conflict in beliefs, as major climate events do not appear to invite susceptibility of belief change for Disbelievers. Instead, they potentially incite more vigorous psychological resistance.

Collectively, this paper indicates significant benefits to studying affective polarization in online climate change discourse. Although social media discourse does not necessarily constitute a representative sample of a particular global population (Morstatter et al. 2013), digital platforms like Twitter nonetheless constitute a vital space for public conversations about important issues like climate change. It offers a large-scale snapshot of how the broader public may understand and interpret contentious issues like climate change, complementing existing work on top-down media representations (Pianta and Sisco 2020). Hence, these findings paint a useful picture of public discourse as situated specifically in cyberspace, which may also bear implications for how digitally mediated science communication and public policy may also be designed and implemented (Bolsen and Shapiro 2018; Kahan et al. 2015).

Besides the issue of demographic representativeness for online data, other limitations attend the present analysis. First, although we utilize a large number of tweets to characterize general affective behavior, this does not encompass those interactions that do not include our collection keywords. Second, the task of getting an aspect-level sentiment of each tweet towards other entities is a non-trivial task. We use Netmapper, which has been used with reasonable accuracy for multiple sentiment level tasks (Uyheng and Carley 2020; Uyheng et al. 2019). This paper focuses on designing a framework to get affective polarization scores between two competing groups, and we do not make an effort to improve aspect-level sentiment scores. Third, in our analysis, we use a list of natural disaster-related words. Communication about the natural disasters could also happen using specific names related to these disasters, for example, using "Dorian" instead of "Hurricane Dorian". Such analysis would require a more comprehensive list of natural disasters occurring worldwide during the 100 weeks. This is out of scope for the current work.

Recognizing the foregoing limitations, we also consider avenues for future work in this area. On a conceptual level, researchers may wish to expand the binary system of climate change beliefs assumed here. Affectively polarized dynamics between multiple groups may be a more challenging yet also potentially informative line of inquiry to explore given the diversity of positions held with respect to this complex issue. Acknowledging the non-neutrality of cyberspace, it would also be essential to consider whether disinformation maneuvers may also be involved in shaping the broader climate change discussion. Inauthentic bot-like accounts and trolls may unduly influence different groups by manipulating the flow of information or amplifying intergroup aggressions; such factors have been seen in relation to other contentious issues and may potentially be present here as well (Uyheng and Carley 2020). Methodologically, a computational analysis may extend our findings by performing a more fine-grained characterization of the types of hostility expressed by both groups. Natural language processing (e.g., topic models) may offer one way forward in this regard. We also acknowledge that other methods have previously been introduced for mining signed networks in quantifying online conflicts. The structural balance theory[9] popularized the use of sentiment networks to predict the nature of relationships. Numerous works use structural balance theory to partition groups into hostile or friendly groups. In this work, we use weighted sentiment networks to quantify affective polarization in the context of extensive social networks. Using stance labels with the partitioning of sentiment networks in the context of structural balance theory to quantify polarization opens up interesting scholarship questions. We leave that work for future research.

Finally, further studies may examine several hypotheses opened up by our results. For instance, social scientists may investigate real-world levels of experienced hostility by climate change Believers and Disbelievers toward opposed groups. These evidence bases would be valuable to accumulate in cross-cultural settings and over time—especially in connection with concurrent political shifts (Hamilton et al. 2015; Milfont et al. 2017), as well as real-world climate-related developments to track causal relationships, which this work suggests but does not empirically verify.

Appendix

List of natural disaster related words used in the analysis: avalanche, blizzard, bushfire, cataclysm, cloud, cumulonimbus, cyclone, disaster, drought, duststorm, earthquake, erosion, fire, flood, forestfire, gale, gust, hail, hailstorm, heatwave, high-pressure, hurricane, lava, lightning, low-pressure, magma, naturaldisasters, nimbus, permafrost, rainstorm, sandstorm, seismic, snowstorm, storm, thunderstorm, tornado, tremor, tsunami, twister, violentstorm, volcano, whirlpool whirlwind, windstorm.

Regression tables

See Tables 1 and 2.

Table 1 Estimation results for believers as discussed in Fig. 3 (left)

	Estimate	SE	*t* value	Pr(|*t*|)
Believer Lag 1	0.0361841	0.110574	0.327	0.74422
Disbeliever Lag 1	0.1181611	0.1816169	0.651	0.5169
SurfTempAnomaly Lag 1	−0.2820251	0.1341301	−2.103	0.0382
Storms Lag 1	−0.0106965	0.0094395	−1.133	0.26006
Constant	0.1733606	0.1037136	1.672	0.09798
Trend	0.0012666	0.0004204	3.013	3.33e−03

Table 2 Estimation results for disbelievers as discussed in Fig. 3 (right)

	Estimate	SE	*t* value	Pr(|*t*|)
Believer Lag 1	0.0200535	0.063672	0.315	0.753505
Disbeliever Lag 1	0.110212	0.1045808	1.054	0.294684
SurfTempAnomaly Lag 1	0.3086637	0.0772364	3.996	0.000129
Storms Lag 1	0.0146449	0.0054355	2.694	0.00837
Constant	−0.1442111	0.0597215	−2.415	0.017704
Trend	−0.0012278	0.0002421	−5.072	2.00e−06

[9] For more information, please refer to (Heider 2013).

References

Anderson AA, Huntington HE (2017) Social media, science, and attack discourse: how Twitter discussions of climate change use sarcasm and incivility. Sci Commun 39(5):598–620

Banks A, Calvo E, Karol D, Telhami S (2020) #Polarizedfeeds: three experiments on polarization, framing, and social media. The International Journal of Press/Politics, p 1940161220940964

Barberá P, Casas A, Nagler J, Egan PJ, Bonneau R, Jost JT, Tucker JA (2019) Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. Am Polit Sci Rev 113(4):883–901

Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R (2015) Tweeting from left to right: is online political communication more than an echo chamber? Psychol Sci 26(10):1531–1542

Bolsen T, Shapiro MA (2018) The US news media, polarization on climate change, and pathways to effective communication. Environ Commun 12(2):149–163

Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ (2017) Emotion shapes the diffusion of moralized content in social networks. Proc Natl Acad Sci 114(28):7313–7318

Buhaug H, Nordkvelle J, Bernauer T, Böhmelt T, Brzoska M, Busby JW, Ciccone A, Fjelde H, Gartzke E, Gleditsch NP et al (2014) One effect to rule them all? A comment on climate and conflict. Clim Change 127(3):391–397

Carley LR, Reminga J, Carley KM (2018) ORA and netmapper. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, New York

Carmichael JT, Brulle RJ (2017) Elite cues, media coverage, and public concern: an integrated path analysis of public opinion on climate change, 2001–2013. Environ Polit 26(2):232–252

Darwish K (2019) Quantifying polarization on Twitter: the Kavanaugh nomination. International conference on social informatics. Springer, New York, pp 188–201

Demszky D, Garg N, Voigt R, Zou J, Shapiro J, Gentzkow M, Jurafsky D (2019) Analyzing polarization in social media: method and application to tweets on 21 mass shootings. In: Proceedings of NAACL. Association for computational linguistics, Minneapolis, Minnesota, pp 2970–3005

Druckman JN, Levendusky MS (2019) What do we measure when we measure affective polarization? Public Opin Q 83(1):114–122

Dunlap RE, McCright AM, Yarosh JH (2016) The political divide on climate change: partisan polarization widens in the US. Environ Sci Policy Sustain Develop 58(5):4–23

van Eck CW, Mulder BC, Dewulf A (2020) Online climate change polarization: interactional framing analysis of climate change blog comments. Sci Commun p 1075547020942228

Evans GW (2019) Projected behavioral impacts of global climate change. Annu Rev Psychol 70:449–474

Fisher DR, Waggle J, Leifeld P (2013) Where does political polarization come from? Locating polarization within the US climate change debate. Am Behav Sci 57(1):70–92

Garimella K, Morales GDF, Gionis A, Mathioudakis M (2018) Quantifying controversy on social media. ACM Trans Soc Comput 1(1):1–27

Geschke D, Lorenz J, Holtz P (2019) The triple-filter bubble: using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. Br J Soc Psychol 58(1):129–149

Hamilton LC, Hartter J, Lemcke-Stampone M, Moore DW, Safford TG (2015) Tracking public beliefs about anthropogenic climate change. PLoS ONE 10(9):e0138208

Heider F (2013) The psychology of interpersonal relations. Psychology Press, Hove

Hitchcock FL (1941) The distribution of a product from several sources to numerous localities. J Math Phys 20(1–4):224–230

Hsiang SM, Burke M (2014) Climate, conflict, and social stability: what does the evidence say? Clim Change 123(1):39–55

Huang B, Ou Y, Carley KM (2018) Aspect level sentiment classification with attention-over-attention neural networks. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, New York, pp 197–206

Jang SM, Hart PS (2015) Polarized frames on climate change and global warming across countries and states: evidence from Twitter big data. Glob Environ Chang 32:11–17

Jenkins-Smith HC, Ripberger JT, Silva CL, Carlson DE, Gupta K, Carlson N, Ter-Mkrtchyan A, Dunlap RE (2020) Partisan asymmetry in temporal stability of climate change beliefs. Nat Clim Chang 10(4):322–328

Jost JT (2017) Ideological asymmetries and the essence of political psychology. Polit Psychol 38(2):167–208

Kahan DM, Jenkins-Smith H, Tarantola T, Silva CL, Braman D (2015) Geoengineering and climate change polarization: testing a two-channel model of science communication. Ann Am Acad Pol Soc Sci 658(1):192–222

Karlsen R, Steen-Johnsen K, Wollebæk D, Enjolras B (2017) Echo chamber and trench warfare dynamics in online debates. Eur J Commun 32(3):257–273

Koubi V (2019) Climate change and conflict. Annu Rev Polit Sci 22:343–360

Krackhardt D, Stern RN (1988) Informal networks and organizational crises: an experimental simulation. Soc Psychol Quart 51:123–140

Kumar S (2020) Social media analytics for stance mining a multimodal approach with weak supervision. Ph.D. thesis, Carnegie Mellon University

Massey FJ Jr (1951) The Kolmogorov–Smirnov test for goodness of fit. J Am Stat Assoc 46(253):68–78

Matakos A, Terzi E, Tsaparas P (2017) Measuring and moderating opinion polarization in social networks. Data Min Knowl Disc 31(5):1480–1505

McCright AM, Dunlap RE (2011) The politicization of climate change and polarization in the American publics views of global warming, 2001–2010. Sociol Q 52(2):155–194

Milfont TL, Wilson MS, Sibley CG (2017) The publics belief in climate change and its human cause are increasing over time. PLoS ONE 12(3):e0174246

Morales A, Borondo J, Losada JC, Benito RM (2015) Measuring political polarization: Twitter shows the two sides of Venezuela. Chaos Interdiscip J Nonlinear Sci 25(3):033114

Morstatter F, Pfeffer J, Liu H, Carley KM (2013) Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In: Proceedings of the 7th international AAAI conference on weblogs and social media, ICWSM 2013. AAAI press, pp 400–408

Pianta S, Sisco MR (2020) A hot topic in hot times: how media coverage of climate change is affected by temperature abnormalities. Environ Res Lett 15(11):114038

Ramdas A, Trillos NG, Cuturi M (2017) On Wasserstein two-sample testing and related families of nonparametric tests. Entropy 19(2):47

Stewart LG, Arif A, Starbird K (2018) Examining trolls and polarization with a retweet network. In: Proceedings of the ACM WSDM, workshop on misinformation and misbehavior mining on the web

Toda HY, Phillips PC (1994) Vector autoregression and causality: a theoretical overview and simulation study. Economet Rev 13(2):259–285

Tyagi A, Babcock M, Carley KM, Sicker DC (2020) Polarizing tweets on climate change. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, New York, pp 107–117

Tyagi A, Carley KM (2020) Divide in vaccine belief in Covid-19 conversations: implications for immunization plans. medRxiv

Tyagi A, Field A, Lathwal P, Tsvetkov Y, Carley KM (2020) A computational analysis of polarization on Indian and Pakistani social media. Social informatics. Springer, New York

Uyheng J, Carley KM (2020) Bot impacts on public sentiment and community structures: comparative analysis of three elections in the Asia-Pacific. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, New York, pp 12–22

Uyheng J, Magelinski T, Villa-Cox R, Sowa C, Carley KM (2019) Interoperable pipelines for social cyber-security: assessing Twitter information operations during NATO Trident Juncture 2018. Comput Math Organ Theo 26:1–19

Uyheng J, Tyagi A, Carley KM (2021) Mainstream consensus and the expansive fringe: characterizing the polarized information ecosystems of online climate change discourse. In: Proceedings of the 13th ACM web science conference 2021, pp 196–204

Van Boven L, Ehret PJ, Sherman DK (2018) Psychological barriers to bipartisan public support for climate policy. Perspect Psychol Sci 13(4):492–507

Weber I, Garimella VRK, Batayneh A (2013) Secular vs. islamist polarization in egypt on twitter. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM ’13. Association for Computing Machinery, New York, NY, USA, pp 290–297

Williams HT, McMurray JR, Kurz T, Lambert FH (2015) Network analysis reveals open forums and echo chambers in social media discussions of climate change. Glob Environ Chang 32:126–138

Wood BD, Owens CT, Durham BM (2005) Presidential rhetoric and the economy. J Polit 67(3):627–645

Yarchi M, Baden C, Kligler-Vilenchik N (2020) Political polarization on the digital sphere: a cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. Polit Commun 38:1–42