2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Affective Polarization in Online Climate Change Discourse on Twitter

Aman Tyagi
CASOS, Engineering and Public Policy
Carnegie Mellon University
Pittsburgh PA, USA
amantyagi@cmu.edu

Joshua Uyheng
CASOS, Institute for Software Research
Carnegie Mellon University
Pittsburgh PA, USA
juyheng@cs.cmu.edu

Kathleen M. Carley
CASOS, Institute for Software Research
Carnegie Mellon University
Pittsburgh PA, USA
kathleen.carley@cs.cmu.edu

***Abstract*—Online social media has become an important platform to organize around different socio-cultural and political topics. An extensive scholarship has discussed how people are divided into echo-chamber-like groups. However, there is a lack of work related to quantifying hostile communication or *affective polarization* between two competing groups. This paper proposes a systematic, network-based methodology for examining affective polarization in online conversations. Further, we apply our 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. Moreover, Disbelievers use more words and hashtags related to natural disasters during more hostile weeks as compared to Believers. These findings bear implications for studying affective polarization in online discourse, especially concerning the subject of climate change. Lastly, we discuss our findings in the context of increasingly important climate change communication research.**

***Index Terms*—climate change, affective polarization, stance detection, online social networks**

I. Introduction

Online social networks represent a powerful space for public discourse. However, research has increasingly demonstrated the dangers of *polarization* in online communication [1]–[3]. 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 [4], [5]. This phenomenon has been termed *interactional polarization*. Polarization can also pertain to highly negative sentiments toward out-groups in the form of *affective polarization* [6], [7]. In this paper, we focus on quantifying affective polarization between two groups with opposing beliefs using Twitter discourse on a significant social issue.

One significant issue which has received heated attention in online public discourse is climate change [8]–[10]. We focus on those who cognitively accept anthropogenic causes of climate change (*Believers*) and those who reject the same (*Disbelievers*). Previous work demonstrates not only sharp divergences in climate change beliefs but also the emergence interactionally polarized groups [10]–[12]. Much less work, however, engages the question of *affective polarization* in online climate change discourse. Relying consistently 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 [6], [13], [14].

This work leverages 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. Furthermore, we link our findings to natural disasters words to explain important climate change belief constructs.

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) Do climate change Believers or Disbelievers feature greater levels of affective polarization?
3) What is the relationship of affective polarization with use of natural disaster related words[1]

II. Related Work

A. Computational analysis of polarization

Social network approaches typically measure polarization in terms of 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 [2], [4], [15]. More recent scholarship, however, suggests that echo chambers represent an incomplete picture

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 U.S. government.

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

[1]We provide the list of natural disaster related words in our project repository: https://github.com/amantyag/affective_climate_change

of polarization [1]. People holding opposed views, in fact, do interact with each other - but this does not necessarily mitigate polarization [5]. Instead, research finds that intergroup exposures trigger further incivility [6]. Hence, reliable measures for affective polarization are needed, although the computational literature in this area remains in its nascent stages [15].

B. Climate change and polarization

Numerous studies link polarized beliefs about climate change to partisan divides [8], [11]. However, with time, scholars have also noted general trends toward increasing climate change beliefs overall [12]. Even if these do not necessarily translate into concrete support for policy [9], the long-term instability of skepticism suggests the importance of accounting for the psychological processes surrounding climate change belief and disbelief [16].

On social media, studies suggest that online climate change discussions tend to exhibit echo chamber-like interactions [10], [17]. 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 [6], [14]. 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.

C. Contributions of this work

Motivated by the foregoing insights, our framework combines machine learning and network science methods in a novel, scalable, and generalizable fashion for ready application in a variety of contentious issues. This 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 [11].

From a theoretical standpoint, we additionally contribute a nuanced operationalization of affective polarization as located on a group level. This conceptually aligns with the asymmetry of psychological factors characterizing climate change Believers and Disbelievers, especially over time [8], [18], [19]. Finally, on an empirical level, our work also extends prevailing scholarship on polarized climate change discourse. We specifically quantify, over a larger-scale and longer-term dataset than previously examined in prior work, the extent to which intergroup interactions systematically feature hostility. This may inform possible data-driven interventions for policymaking beyond more prevalent frames of intergroup contact and science communication [16].

III. Data and Methods

A. Data collection

We collected tweets using Twitter's standard API[2] 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 seven day period (1 week) to get a total of 100 weeks.

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

B. Stance labels

We use a state-of-the-art stance mining method [20] 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 [10]. The algorithm labels 3.9M as Believers and 3.1M as Disbelievers. We provide details of manual validation of stance results and the parameters in our project repository https://github.com/amantyag/affective_climate_change.

C. 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 [21] and Earth Mover's Distance (EMD) [22].

1) Aspect-level sentiment: Aspect-level sentiment refers to the emotional valence of a given utterance toward one of the concepts it mentions. We utilize Netmapper to extract entities from each tweet, and predict the aspect-level sentiment of each tweet toward each entity [23]. 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]".

2) Affective networks: Let $G^+ = (V, E^+)$ denote a positive interaction network where the set of vertices V contains all Twitter accounts in our dataset and the set of directed edges E^+ contains all positive-valenced mentions and replies between agents in V. Similarly, let $G^- = (V, E^-)$ denote a negative interaction network over the same set of agents V 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 $i, j \in V$. 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) E/I indices: We assess group-level differences in positive and negative interactions using Krackhardt's E/I index [21]. 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 [24]. 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 $V_k \subseteq V$ denote the set of agents belonging to stance k and $V_{k'}$ 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 $E_k^{+} = \sum_{i \in V_k, j \in V_{k'}} w_{ij}^{+}$ and $I_k^{+} = \sum_{i,j \in V_k} w_{ij}^{+}$. 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 $E_k^{-} = \sum_{i \in V_k, j \in V_{k'}} w_{ij}^{-}$ and $I_k^{-} = \sum_{i,j \in V_k} w_{ij}^{-}$. Given the construction of P_k^{+} and P_k^{-}, we note that both values are bounded between -1 and $+1$.

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

$$P_k = P_k^{-} - P_k^{+}. \tag{3}$$

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.

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. Similar to affective networks, we define $G = (V, E)$ as interaction network where the set of vertices V 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 V. 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 $i \in V_k, j \in V_{k'}$ and let v_k be distribution of w_{ij}, where $i \in V_k, j \in V_k$. For a group holding stance k, we define our novel 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.

Our novel affective polarization metric l_k is positive when $P_k > 0$. As noted in §III-C4, 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.

IV. Results

We first look at how the affective polarization metric is changing over time in figure 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 indicates 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 much more organized manner over the 100 weeks than Beleivers. 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.

To 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 mean plus 1 standard deviation, i.e. from figure 1, 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 exceptional hostilie weeks.

Next, we use natural disaster-related words as a proxy to determine how natural disasters play a role in hostility between the two groups. In the first plot of figure 2 we look at the top 100 most frequent hashtags used within those groups to find the percentage of hashtags related to natural disasters. As expected, Believers use more natural disaster-related hashtags than Disbelievers. However, during the exceptional hostile weeks Believers use less of these hashtags. Interestingly, Disbelievers show the exact 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 econd plot of figure 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,

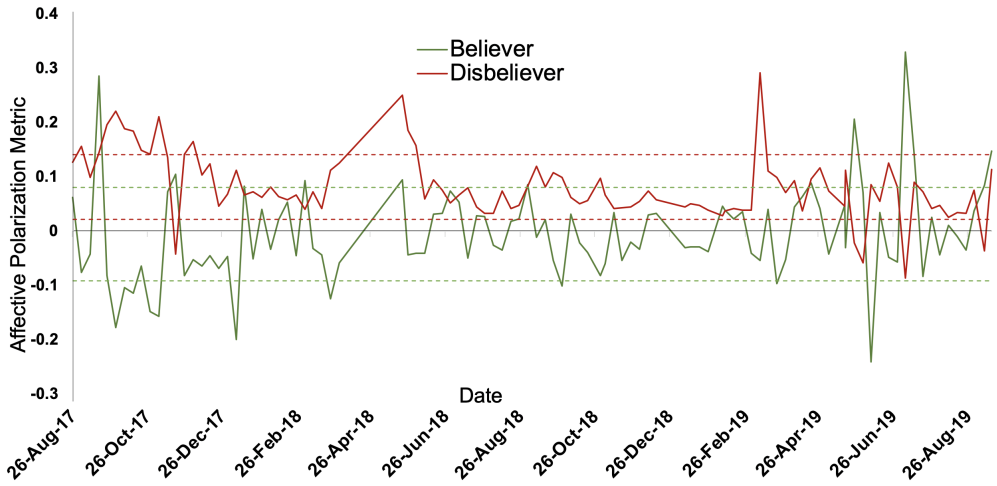

Fig. 1. Affective polarization metric (l_k) for Believers and Disbelievers of climate change. Higher positive values denote more hostility towards the other group. The dotted lines represent mean ±1 standard deviation, which for Believers is -0.091 and 0.080 and disbelievers is -0.117 and 0.106. The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

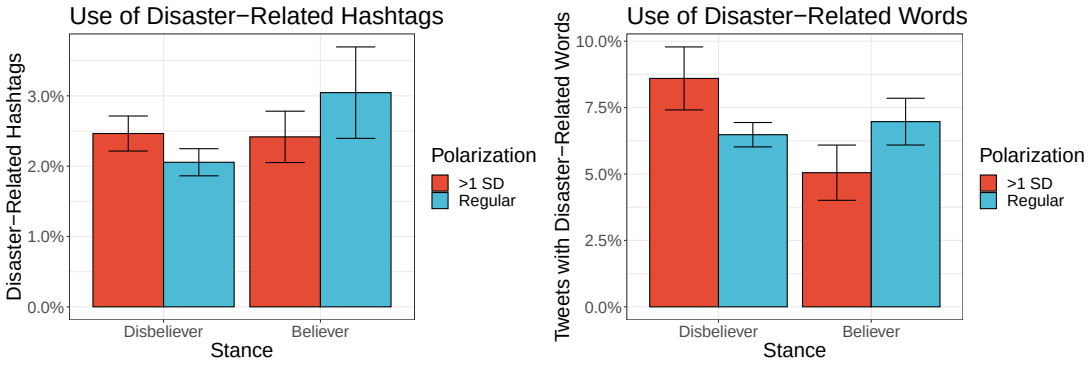

Fig. 2. Comparison of relative salience of disaster-related talk when the affective polarization metric is greater than 1 standard deviation or otherwise. Error bars represent ±1 standard errors. **Left:** Percentage of the top 100 most frequent hashtags containing natural disaster-related words. **Right:**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

we find that a greater percentage of Tweets from Disbelievers mention natural disaster-related words compared to Believers. This indicates that Disbelievers are calling out natural disasters more when they are exceptionally hostile towards Believers compared to other weeks.

V. Discussion 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 [17], we show how this polarization extends also to the realm of emotion in the form of affective polarization. We extend existing studies which highlight the role of incivility and personalized framing in encounters between climate change Believers and Disbelievers [6], [14] by introducing a scalable technique for analyzing relative intra- and intergroup interaction valence. This allowed us to quantify the extent of hostile communications between the two groups over a large-scale, long-term dataset - thereby validating existing findings in a generalizable manner as well as showing their relative stability over time.

Furthermore, we highlight the value of viewing polarization from an asymmetrical perspective. Indeed, higher levels of hostility from Disbelievers present a specifically notable finding for social scientific scholarship on climate change discourse. Longitudinal analysis in prior work suggests that generalized climate change beliefs over time are increasing [11], [12], and climate change Disbelievers in particular are more susceptible to potential belief change [18]. But significant cognitive barriers remain for fuller acceptance of anthropogenic causes for climate change and the corresponding urgency for responsive policy changes [8], [9]. Higher levels of hostility among climate change Disbelievers toward climate change Believers constitutes one such obstacle for further dialogue between the two groups. As past studies suggest, 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 [19].

These insights are especially important to consider given 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. Although comparable levels are seen when both groups are within average levels of our metric, moments of increased affective polarization correlate with opposite behaviors for Believers and Disbelievers. Believers appear to shift to other areas of contention, such that their aggression is characterized by non-disaster topics. In contrast, Disbelievers' increased invocation of disaster terms points to more aggressive discussion of these catastrophes, albeit positioned in resistance to explanations related to anthropogenic climate change. This introduces 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.

Besides the issue of demographic representativeness for online data, other limitations attend the present analysis. First, although we have a large number of tweets to characterize general affective behavior, however, it does not encompass those interactions which 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 [24], [25]. The focus of this paper is on designing a framework to get affective polarization score between two competing groups and we do not make an effort to improve aspect-level sentiment scores.

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 important to consider whether disinformation maneuvers may also be involved in shaping the wider 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 [24].

REFERENCES

[1] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, "Tweeting from left to right: Is online political communication more than an echo chamber?" *Psychological Science*, vol. 26, no. 10, pp. 1531–1542, 2015.

[2] A. Tyagi, A. Field, P. Lathwal, Y. Tsvetkov, and K. M. Carley, "A computational analysis of polarization on Indian and Pakistani social media," in *Social Informatics*. Springer, 2020.

[3] A. Tyagi and K. M. Carley, "Divide in vaccine belief in covid-19 conversations: Implications for immunization plans," *medRxiv*, 2020.

[4] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis, "Quantifying controversy on social media," *ACM Transactions on Social Computing*, vol. 1, no. 1, pp. 1–27, 2018.

[5] R. Karlsen, K. Steen-Johnsen, D. Wollebæk, and B. Enjolras, "Echo chamber and trench warfare dynamics in online debates," *European Journal of Communication*, vol. 32, no. 3, pp. 257–273, 2017.

[6] A. A. Anderson and H. E. Huntington, "Social media, science, and attack discourse: How Twitter discussions of climate change use sarcasm and incivility," *Science Communication*, vol. 39, no. 5, pp. 598–620, 2017.

[7] J. N. Druckman and M. S. Levendusky, "What do we measure when we measure affective polarization?" *Public Opinion Quarterly*, vol. 83, no. 1, pp. 114–122, 2019.

[8] R. E. Dunlap, A. M. McCright, and J. H. Yarosh, "The political divide on climate change: Partisan polarization widens in the US," *Environment: Science and Policy for Sustainable Development*, vol. 58, no. 5, pp. 4–23, 2016.

[9] D. R. Fisher, J. Waggle, and P. Leifeld, "Where does political polarization come from? Locating polarization within the US climate change debate," *American Behavioral Scientist*, vol. 57, no. 1, pp. 70–92, 2013.

[10] A. Tyagi, M. Babcock, K. M. Carley, and D. C. Sicker, "Polarizing tweets on climate change," in *To appear International Conference SBP-BRiMS*, A. H. Halil Bisgin, C. Dancy, , and R. Thomson, Eds. Springer, 2020.

[11] L. C. Hamilton, J. Hartter, M. Lemcke-Stampone, D. W. Moore, and T. G. Safford, "Tracking public beliefs about anthropogenic climate change," *PloS One*, vol. 10, no. 9, p. e0138208, 2015.

[12] T. L. Milfont, M. S. Wilson, and C. G. Sibley, "The public's belief in climate change and its human cause are increasing over time," *PloS one*, vol. 12, no. 3, p. e0174246, 2017.

[13] S. M. Jang and P. S. Hart, "Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data," *Global Environmental Change*, vol. 32, pp. 11–17, 2015.

[14] C. W. van Eck, B. C. Mulder, and A. Dewulf, "Online climate change polarization: Interactional framing analysis of climate change blog comments," *Science Communication*, p. 1075547020942228, 2020.

[15] M. Yarchi, C. Baden, and N. Kligler-Vilenchik, "Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media," *Political Communication*, pp. 1–42, 2020.

[16] D. M. Kahan, H. Jenkins-Smith, T. Tarantola, C. L. Silva, and D. Braman, "Geoengineering and climate change polarization: Testing a two-channel model of science communication," *The ANNALS of the American Academy of Political and Social Science*, vol. 658, no. 1, pp. 192–222, 2015.

[17] H. T. Williams, J. R. McMurray, T. Kurz, and F. H. Lambert, "Network analysis reveals open forums and echo chambers in social media discussions of climate change," *Global Environmental Change*, vol. 32, pp. 126–138, 2015.

[18] H. C. Jenkins-Smith, J. T. Ripberger, C. L. Silva, D. E. Carlson, K. Gupta, N. Carlson, A. Ter-Mkrtchyan, and R. E. Dunlap, "Partisan asymmetry in temporal stability of climate change beliefs," *Nature Climate Change*, vol. 10, no. 4, pp. 322–328, 2020.

[19] L. Van Boven, P. J. Ehret, and D. K. Sherman, "Psychological barriers to bipartisan public support for climate policy," *Perspectives on Psychological Science*, vol. 13, no. 4, pp. 492–507, 2018.

[20] S. Kumar, "Social media analytics for stance mining a multi-modal approach with weak supervision," Ph.D. dissertation, Carnegie Mellon University, 2020.

[21] D. Krackhardt and R. N. Stern, "Informal networks and organizational crises: An experimental simulation," *Social Psychology Quarterly*, pp. 123–140, 1988.

[22] F. L. Hitchcock, "The distribution of a product from several sources to numerous localities," *Journal of Mathematics and Physics*, vol. 20, no. 1-4, pp. 224–230, 1941.

[23] L. R. Carley, J. Reminga, and K. M. Carley, "ORA & Netmapper," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 2018.

[24] J. Uyheng and K. M. Carley, "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, 2020, pp. 12–22.

[25] J. Uyheng, T. Magelinski, R. Villa-Cox, C. Sowa, and K. M. Carley, "Interoperable pipelines for social cyber-security: Assessing Twitter information operations during NATO Trident Juncture 2018," *Computational and Mathematical Organization Theory*, pp. 1–19, 2019.