Analyzing Climate Change Discussions on Reddit

Conference Paper · December 2022

DOI: 10.1109/CSCI58124.2022.00150

6 authors, including:

Mohammad Parsa
University of California, Berkeley
9 PUBLICATIONS **22** CITATIONS

SEE PROFILE

Analyzing Climate Change Discussions on Reddit

Mohammad S. Parsa, Haoqi Shi, Yihao Xu, Aaron Yim, Yaolun Yin, and Lukasz Golab
University of Waterloo, Canada
Contact author: Lukasz Golab, lgolab@uwaterloo.ca

***Abstract*—Climate action is one of the United Nations Sustainable Development Goals. We contribute to this effort by analyzing climate change topics on the Reddit social curation platform, which contains over 100,000 discussion communities. We leverage the community-oriented structure of Reddit to answer two questions: 1) Which online communities discuss climate change? and 2) What do they discuss? In addition to the presence of groups of activists and skeptics, we find climate change discussions on question-answering and opinion-sharing communities, as well as in a community for teenagers. We then run a topic modeling algorithm on the five largest climate change communities. We find concerns about skeptics, political policies doubting the severity of climate change and the consequences of climate change, as well as discussions of climate actions that can be taken by individuals. Finally, an analysis of the URLs mentioned in discussions highlights influential content creators and underscores the need to avoid misinformation on these sources.**

***Index Terms*—social media, climate change, text mining**

I. Introduction

"Taking urgent action to combat climate change and its impacts" is one of the 17 Sustainable Development Goals established by the United Nations[1]. An important component of any climate action plan is an understanding of public opinion, which can be done through social media analysis.

A recent survey of research on social media and climate change [1] identified 35 relevant articles, of which 28 analyzed Twitter data, and the remaining few studied Facebook, Youtube and Weibo. Prior work on Twitter studied the spatio-temporal distribution of, and sentiment towards, climate-related content, usually identified by the presence of hashtags such as #climatechange and #globalwarming. There are also studies of the characteristics of the creators of this content as well as their followers and those who retweet their messages. Findings include correlations between climate-related tweeting activity and weather and political events, and the presence of online communities of activists and skeptics.

We apply text mining methods to the Reddit social curation platform. Reddit is divided into over 100,000 discussion communities referred to as *subreddits*, whose names start with "r/". Subreddit names indicate the topic, e.g., r/politics. A subreddit consists of posts, which are meant to start discussions, while each post is associated with zero or more comments.

While previous work on social media mining of climate change discussions focuses on Twitter, [1], we observe that Reddit is a valuable source of information on this topic. Reddit is organized into interest-based communities with descriptive names, allowing us to identify high-level topics discussed from a climate change standpoint, and to isolate communities focusing on climate change. We leverage this observation to answer the following research questions:

1) *Which* communities discuss climate change? To answer this question, we identify subreddits whose discussions frequently mention terms related to climate change.
2) *What* is being discussed? We answer this question by running a *topic modeling* algorithm on five subreddits we identified as focusing on climate change and its impacts: r/climate, r/environment, r/climatechange, r/climateskeptics, and r/climateOffensive. We also extract URLs that are frequently mentioned in discussions on these five subreddits to find information sources and discussion triggers.

II. Data and Methods

This analysis was implemented in Python 3.6 and is based on a publicly available copy of Reddit – all posts and comments from all subreddits – downloaded from the Google Big Query data warehouse[2], spanning from January 2016 to August 2019 inclusive.

A. Identifying Climate Change Terms

To identify terms suggesting climate change topics, we started with three online climate change glossaries, available at the U.S. Environmental Protection Agency (EPA)[3], the Global Change Research Program[4], and Wikipedia[5]. Altogether, these glossaries contain 422 unique terms. We removed terms with more than four words (most of which were highly specific, e.g., 'carbon capture and sequestration'), terms with other common uses (e.g., 'feedback'), and generalized the term 'renewable energy' to 'renewable' to capture different syntactic expressions of this concept. At this point, 303 terms remained.

Next, for each subreddit, we counted the number of *substring* matches against these 303 terms. Substring matching means that, for example, the term 'renewable' matches 'renewables' as well as 'renewable energy' and 'renewable sources'. We found that many terms occurred infrequently (under 500 times in all of Reddit, such as 'arctic shrinkage', 'carbon offset', and 'environmental justice'). Furthermore,

[1]https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[2]https://cloud.google.com/bigquery

[3]https://19january2017snapshot.epa.gov/climatechange/glossary-climate-change-terms_.html

[4]https://www.globalchange.gov/climate-change/glossary

[5]https://en.wikipedia.org/wiki/Glossary_of_climate_change

979-8-3503-2028-2/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00093

upon manual inspection, we found that many terms occurred in non-climate contexts. For example, 'climate' was frequently mentioned by gaming subreddits such as r/csgotrade, r/fo4, r/NoMansSkyTheGame, and r/Market76. After removing rare terms and those mainly used in other contexts, we were left with 11 terms mentioned in the context of climate change: *'carbon dioxide', 'carbon footprint', 'carbon tax', 'climate change', 'climate model', 'emission', 'fossil fuel', 'global warming', 'greenhouse gas', 'renewable', and 'sea level'*.

B. Identifying Climate Change Communities

Next, we identified subreddits whose posts and comments contain the most substring matches of the 11 climate terms. Among them we found five subreddits, which we call *climate subreddits*, whose names and accompanying descriptions suggest that they focus on climate change and its impacts: r/climate, r/environment, r/climatechange, r/climateskeptics, and r/climateOffensive.

(We also searched for subreddits whose names match our 11 climate terms, using Reddit's search interface at reddit.com/subreddits. We additionally found r/CarbonFootprint, r/CarbonTax, r/climate_discussion, r/ClimateActionPlan, r/climatescience, r/GlobalClimateChange, and r/GlobalWarming. However, these subreddits are smaller than our chosen five and they placed outside the top-50 subreddits overall, in terms of the number of matches to our 11 climate terms. Thus, they are not included in topic modeling.)

Table I summarizes the five climate subreddits in terms of: the number of posts, the number of non-empty posts (that have some text, not just an image or just a URL, and have not been removed by moderators), the number of comments, the percentage of posts without any comments, the number of users (who posted or commented at least once), the percentage of one-time users (those who posted or commented exactly once), the percentage of content (posts plus comments) created by the top ten most active users, the percentage of comments having a positive sentiment (using HuggingFace's Transformers sentiment analyzer [2]), and the percentage of comments identified as inflammatory (using the HateSonar hate speech classifier [3]).

The biggest, in terms of users, non-empty posts and comments, is r/environment; the smallest is r/climateOffensive, describing itself as focusing on solutions rather than discussions. There are many empty posts in r/climatechange and r/climateskeptics, corresponding mostly to posts with only images, with some deleted by moderators. Note that r/climateskeptics focuses on doubts and skepticism around climate change, allowing us to study climate skeptics in detail. All five subreddits have a high percentage of one-time users, as is common in social media, and all but r/environment have 20 or more percent of post and comments created by the ten most active users. Finally, we note the low percentage of positive comments (not surprisingly, lowest in r/climateskeptics and highest in r/climateOffensive), and the very low percentage of hate speech (likely due to the removal of inappropriate content by moderators).

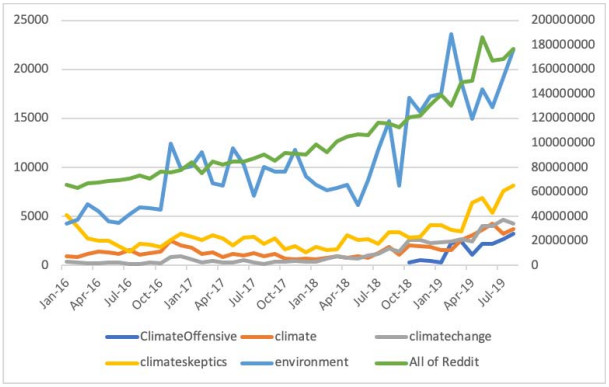

Fig. 1. Time series of the number of posts plus comments on climate subreddits (Y-axis on the left), and the number of posts plus comments on all of Reddit (Y-axis on the right).

Figure 1 shows the number of posts plus comments on the five climate subreddits, as well as the number of posts and comments on all of Reddit, aggregated by month. The *All of Reddit* curve uses the Y-axis on the right, ranging from 0 to 200 Million, whereas the other curves use the Y-axis on the left, ranging from 0 to 25,000. The content generated on the five climate subreddits is growing at about the same pace as Reddit as a whole. Note that the plot for r/climateOffensive starts in late 2018, which is when this subreddit was created. Other climate subreddits were created before 2016. Peaks in the time series correspond to climate-related events, such as the Paris Agreement to mitigate global warming in November 2016, the California wildfires in August 2018, and the Green New Deal, a proposed US program to address climate change and economic inequality, in February 2019.

C. Topic Modeling

Next, we performed topic modeling (separately) for each of the five climate subreddits. Since posts are meant to initiate discussions, they are good indicators of discussion topics. Thus, we only used posts for topic modeling, not comments.

Before topic modeling, we applied typical text processing steps such as tokenization, stopword removal, conversion to lower case, and lemmatization (i.e., grouping together the *inflected* forms of a word).

Next, we converted each post to a corresponding TF-IDF vector in a standard way. The ith coordinate of a TF-IDF vector for a given post corresponds to the frequency of the ith word in the post (TF) divided by the total number of posts in the given subreddit that mention the i word at least once. TF-IDF takes into account both the rarity of a word within the given subreddit and its relevance to the given post.

We then ran the Non-negative Matrix Factorization (NMF) topic modeling algorithm [4] with the vectorized posts as input. NMF partitions the input into clusters such that the vectors assigned to each cluster are similar (i.e., the corresponding

TABLE I
STATISTICS OF CLIMATE SUBREDDITS.

Subreddits	r/climatechange	r/climateskeptics	r/climate	r/ClimateOffensive	r/environment
Posts	7617	16340	1356	592	4727
Non-empty posts	1252	1278	1356	592	4727
Comments	55761	177625	67071	9497	448637
% Posts without comments	26.7	12.4	22.3	11.5	39.1
Users	6903	7756	11560	4372	101095
% One-time users	46	37.7	48.5	48.7	51.5
% Top 10 users content	20.9	29.2	17.5	19.2	3.8
% Positive Sentiment	26.7	22.7	27.2	34.8	25.4
% Hate Speech	4.2	3	2.7	1.9	2.4

posts use similar words and therefore they may discuss a similar topic). For each cluster, the algorithm outputs a list of the ten most representative words. Additionally, for each cluster, we identified the ten most frequent word bigrams and trigrams (i.e., sequences of two and three consecutive words, respectively) within the posts assigned to this cluster. Finally, to confirm the topics corresponding to the discovered clusters, we manually examined a 5% sample of posts with the most comments assigned to each cluster.

When running NMF, we must specify the number of clusters or topics. We executed NMF with the number of topics set from 5 to 50, and calculated the *coherence* [5] of the resulting topics. A higher coherence indicates that the representative words for each topic are related in a semantic sense. Coherence was highest at 5 topics for r/climate and r/climateOffensive, 10 topics for r/climateskeptics, 15 topics for r/climatechange, and 25 topics for r/environment.

D. URL Analysis

For a further understanding of discussion topics and information sources used when starting the discussions, we extracted the URLs mentioned in the posts (using the tldextract[6] parser), and identified the most frequently mentioned URL domains in each climate subreddit.

III. RESULTS

A. Which Communities Discuss Climate Change?

Figure 2 shows the number of matches per year of each of the 11 climate terms, over the *posts* within all subreddits. Recall that the dataset spans only up to August 2019, meaning that the 2019 numbers are not complete. We observe a year-over-year growth for most terms, especially 'climate change', and we note that general terms such as climate change are more popular than technical terms such as carbon tax. We obtained similar conclusions from the number of matches over all the *comments*.

Tables II and III show the top-30 subreddits in the number of climate term substring matches within the *posts* and *comments*, respectively. Both tables are sorted by descending match count for 2019 to focus on recent trends; the units in Table III are in thousands of comments. We highlight the climate subreddits in bold, and we discuss notable others below.

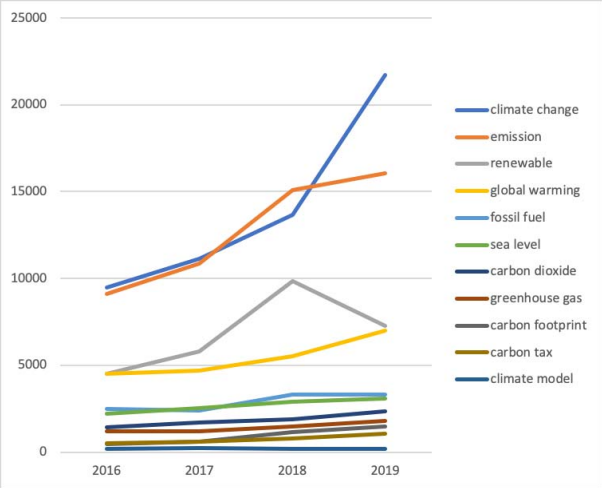

Fig. 2. Number of matches to climate change vocabulary by term and by year; posts only

Beginning with Table II, counting matches within posts, we note the presence of "automated" communities such as r/autotldr, which is a subreddit that automatically summarizes long posts from *all* subreddits. Not surprisingly, this subreddit contains the most matches to our climate terms. Other examples include r/removalbot, which logs comments removed by users or moderators, and r/TalkativePeople, which collects posts made by frequent posters on various subreddits.

Next, we find climate discussions in general question-answering and opinion-sharing communities such as r/unpopularopinion and r/NoStupidQuestions, and in specific populations such as r/teenagers. Furthermore, climate-related terms appear in r/vegan.

We also find political communities in Table II, such as r/DebateRightists (which appears to be a log of questions posted on r/askaconservative), the subreddit corresponding to the political podcast Chapo Trap House, and the subreddit for the former U.S. presidential candidate Andrew Yang, as well communities discussing conspiracies (r/conspiracy) and the potential collapse of civilization (r/collapse).

[6]https://github.com/john-kurkowski/tldextract

TABLE II
NUMBER OF MATCHES TO CLIMATE CHANGE VOCABULARY WITHIN POSTS, BY SUBREDDIT AND BY YEAR.

subreddit	2016	2017	2018	2019	total
autotldr	6384	7844	4871	4379	23478
unpopularopinion	34	110	843	2481	3468
DebateRightists	0	0	214	1842	2056
removalbot	122	455	1122	1732	3431
TalkativePeople	0	0	0	1332	1332
collapse	475	740	970	1050	3235
teenagers	24	49	92	1017	1182
NoStupidQuestions	198	329	601	889	2017
environment	360	338	731	773	2202
climatechange	79	122	414	607	1222
conspiracy	276	553	426	548	1803
ChapoTrapHouse	0	82	383	536	1001
YangForPresidentHQ	0	0	0	503	503
climate	227	222	309	463	1221
climateskeptics	219	195	200	402	1016
CrohnsDisease	313	362	479	398	1552
nullthworldproblems	0	0	0	366	366
ClimateOffensive	0	0	32	364	396
quzangle	0	0	0	349	349
changemyview	330	388	406	339	1463
civ	71	35	104	331	541
copypasta	75	101	178	317	671
MHoCCampaigning	0	60	256	275	591
AskScienceDiscussion	287	298	264	265	1114
MechanicAdvice	175	245	309	264	993
vegan	160	257	275	262	954
GlobalPowers	564	336	341	244	1485
rant	78	114	91	237	520
dirtypenpals	169	165	288	235	857
UlcerativeColitis	7	38	182	232	459

TABLE III
NUMBER OF MATCHES TO CLIMATE CHANGE VOCABULARY WITHIN COMMENTS, BY SUBREDDIT AND BY YEAR, IN UNITS OF THOUSANDS.

subreddit	2016	2017	2018	2019	total
worldnews	189.8	268.5	257.3	167.6	883.2
politics	252.1	204.4	193.7	143.5	793.8
AskReddit	199.7	216.8	226.3	136.2	778.9
Futurology	61	86.8	103.3	70.5	321.6
canada	28.4	34.3	53	41.6	157.2
unpopularopinion	0.5	2.4	18.6	34.6	56.1
The_Donald	45.4	72.5	47.3	32.8	198
collapse	24.6	32.7	54.3	31.5	143.1
environment	22.3	30	34.2	29.3	115.8
ukpolitics	12.7	19.9	25.7	28.4	86.7
news	58.6	93.4	65.8	27	244.8
europe	19.4	30.1	37.4	26.3	113.2
australia	23	20.5	27.7	24.8	95.9
todayilearned	40.9	42.6	41.7	24.6	149.8
science	51.9	55.1	42.2	20.9	170.1
pics	26.1	37.3	39	19.8	122.2
ChapoTrapHouse	0	7	26.1	19.5	52.6
CanadaPolitics	13.6	13.8	30.8	19.4	77.6
climateskeptics	18.8	17.5	18.7	17.1	72.1
NoStupidQuestions	6.3	10.2	18.7	16.2	51.3
changemyview	17.5	23.1	28.2	15.5	84.3
dataisbeautiful	17.8	20.8	12.6	15.2	66.3
neoliberal	0	17	24.9	15.1	57
space	27.7	38.1	35	14.8	115.6
conspiracy	15.3	37.5	25.1	14.8	92.7
Showerthoughts	11.4	8.2	20.5	14.6	54.8
PoliticalHumor	1.2	10.8	20.3	14.2	46.5
explainlikeimfive	41.9	33.8	23.1	13.8	112.6
videos	31.7	21.3	19.9	13.1	86
climatechange	3.3	3.9	15.1	12.4	34.7

r/nullthworldproblems appears to be humorous take on the notion of *first world problems*, discussing fictional problems faced by alternate universes.

Finally, we see subreddits dedicated to Crohn's disease and Ulcerative Colitis as a side effect of our substring matching methodology. Users report that their disease is in *remission*, which is a substring match to 'emission'.

Moving on to Table III, counting matches within comments, we again see general question-answering subreddits such as r/AskReddit and r/unpopularopinion, as well as political subreddits, including r/politics, r/ukpolitics, and r/TheDonald (referring to Donald Trump). Not seen before in Table II are general subreddits discussing news and region-specific communities for Canada, Europe, and Australia. These are very large subreddits (some of the largest on Reddit) with very long discussions on a variety of topics, of thousands of comments per post.

B. Which Topics Are Discussed?

First, we summarize the climate-oriented discussion topics on notable non-climate subreddits from Tables II and III, based on manual inspection of a sample of posts matching our climate terms. For each subreddit, the corresponding sample contains 100 such posts with the most comments.

In r/teenagers, these posts urged others to take actions to reduce climate change and discussed political policies proposed by Greta Thunberg and Donald Trump. In r/vegan, there were suggestions to not consume animal products in order to reduce carbon emissions.

In r/unpopularopinion, there were complaints about Greta Thunberg, climate change deniers, and organizations perceived to not be acting in a way that is constructive towards the climate change cause (e.g., Greenpeace campaigning against nuclear energy).

In r/NoStupidQuestions, common discussions included questions about what people can do about climate change, what people should do if earth becomes inhabitable, why climate skeptics exist, and questions about current events (e.g., attitudes towards Greta Thunberg and understanding criticism about philanthropists' commitments to fight climate change).

In r/changemyview, users were debating controversial climate change related topics such as nuclear energy and the influence of politics on climate change beliefs.

Finally, climate-oriented discussions on the political subreddits focused on climate change policies, discussions on news and regional communitites discussed climate-related news, r/conspiracy redditors expressed doubts about climate change, and r/collapse included discussions of the potential collapse of civilization due to climate change.

Next, we summarize the results of topic modeling of the five selected climate subreddits.

1) r/environment: This subreddit discusses a variety of topics, including concerns about skeptics, policies, influence of human activity on the environment, and potential solutions to mitigate climate change. 14% of the total posts represent concerns about climate change skeptics, including discussion about Donald Trump. Further 18% discuss environmental and political policies as well as anxiety about climate change and related policies. We also see discussions of various solutions (40% total) such as reducing meat consumption, protecting the environment, banning single-use plastic bags and bottles, reducing carbon emissions, recycling, eco-friendly products, planting trees, and renewable energy sources.

There are also topics related to how human activity influences the environment: ocean plastic waste, pollution, water quality, oil spills, and transportation (especially electric vehicles, and whether they actually reduce climate change).

Two topics are unique: discussions of careers that benefit the environment, and discussions of climate change content (videos/documentaries) found on social media (Youtube).

2) r/climatechange: Many topics debate the reality and severity of climate change (including recent scientific reports), and point out the difficulty of convincing skeptics (48% of the posts). A number of topics discuss the consequences of climate change such as ice melting and temperature rising (37% of the posts). There are also discussions of possible causes of climate change such as carbon emissions, possible solutions such as renewable sources of energy, and specific discussions about effectiveness of the Paris Agreement in solving climate change, the challenge it will face, and the recent U.S. withdrawal from the plan.

3) r/climateskeptics: Topics are related to skepticism: towards the scientific consensus, towards renewable energy sources and their impact on climate change, towards climate change being real or not as severe as claimed, towards ideologies and policies such as Green New Deal, towards the reliability of scientific studies on climate change issues, towards the legitimacy of climate change, towards loss of permafrost and ice, and towards sea level rise. There are also discussions on categorizing the different kinds of skeptics, and comments on the U.S. withdrawing from the Paris Agreement.

4) r/climate: This subreddit discusses news and potential climate change action, the effect of climate change on people's lives, the causes and effects of sea level rise, the potential impact of fossil fuels and renewable energies on CO2 emissions, temperature change, and climate change, and general aspects of global warming (scientific research, evidence, and impact on humans).

5) r/ClimateOffensive: In line with the stated purpose of this subreddit, its topics focus on specific initiatives: actions that people may take, fundraising for organizations such as the Rainforest Trust and the Climate Foundation, ways to contribute to organizations that plant trees such as using the Ecosia search engine (which uses its profits to plant trees), and discussions driven by the Meadow Restored project. Finally, one topic summarizes general discussions and project updates.

C. Which URLs Are Mentioned?

In the five climate subreddits, the most widely cited reference is Reddit itself, i.e., users are citing other posts in their posts. These citations come most often from r/autotldr, which, as mentioned earlier, is a subreddit that automatically summarizes long posts from all subreddits. Next are Wikipedia and Youtube. Additionally, ClimateOffensive discussions (and also some on r/environment) frequently reference citizensclimatelobby.org, which is a non-profit organization helping people take climate action. Notably, only r/climateskeptics discussion refer to skeptical blogs such as wattsupwiththat.com.

Next, we highlight frequently mentioned Wikipedia pages and Youtube videos.

1) Wikipedia Pages: Most of these pages describe climate change terms, laws, policies, and individuals working in this field. However, in r/climateskeptics, users expressed skepticism by referring to Wikipedia pages such as "Obsessive Compulsive Disorder" and "Medieval Warm Period".

2) YouTube Videos: In all but r/climateskeptics, frequently mentioned videos discuss climate change and climate action. Not surprisingly, the most frequent videos mentioned in r/climateskeptics express skepticism towards climate change. We also found that most videos refer to intellectuals, including university professors, geographers, geologists and climate activists, namely Nathaniel Stinnett, James Hansen, Kevin Anderson, Anthony Leiserowitz, Peter Wadhams, and Richard Alley. Moreover, there is a YouTube channel called "potholer54" in which journalist Peter Hadfield discusses fake news in different topics, including climate change. His videos have been cited frequently in climate subreddits.

IV. Discussion and Related Work

A. Which Communities Discuss Climate Change?

The descriptive nature of subreddit names allows us to identify high-level topics whose discussions have been influenced by climate change. First, we note that there are political communities in Table II and III, which aligns with previous work on the political polarization of climate change discussions on Twitter [1], [6], [7]. Furthermore, as was the case in previous Twitter analysis [8]–[11], we found communities of climate change activists and skeptics.

We also found climate change discussions on general question-answering forums and teenagers' online communities. This observation appears contrary to some of the previous work on Twitter analysis. For example, Williams et al. [8] observed that climate change related discussion on Twitter is "characterised by attitude-based homophily and wide-spread segregation of users" into like-minded communities. Our observations on Reddit, however, demonstrate that climate change related discussion in mixed-attitude communities such as r/NoStupidQuestions has grown to a level that is higher than within segregated attitude based communities such as r/climateskeptics. Whether the structure of Reddit or the nature

of its communities contribute to this shift and whether a similar trend can be observed in recent Twitter discussion could be areas of exploration for future work.

Finally, we note the presence of climate-related terms on r/vegan, and we point out recent work asking whether dietary strategies shared on Instagram and Twitter are effective in mitigating climate change [12].

B. Which Topics Are Discussed?

Our topic modeling results suggest four general categories of discussion within climate communities:

- Expressing skepticism about the reality, possible causes, or severity of climate change. Every topic on r/climateskeptics belongs to this category.
- Expressing concerns about skeptics and political policies doubting the reality and severity of climate change. Several large topics on r/environment belong to this category.
- Discussing the effects and consequences of climate change. Many topics on r/climatechange, and some topics on r/environment and r/climate belong to this category.
- Discussing potential solutions and actions that can be taken by individuals to reduce climate change and its impacts. Every topic on r/climateOffensive belongs to this category, as well as some topics on r/environment, r/climatechange and r/climate.

We draw two main insights from the topic modeling analysis. First, climate skeptics appear to be questioning a variety of issues, ranging from scientific studies and findings such as the loss of permafrost, to political policies. This is not surprising as climate skeptics view climate change as a natural process which cannot be altered by human activities.

Second, the other four climate subreddits all include some discussion of actions that can be taken by individuals to combat climate change. These discussions stand out as being more positive than those involving anxiety caused by skepticism and climate change predictions and consequences. In particular, discussions of potential solutions should be of interest to climate scientists and policymakers wishing to identify popular programs and initiatives (such as the tree planting initiatives mentioned in r/environment and r/climateOffensive), and seeking new ways to encourage the public to take action.

We also point out two noteworthy topics on r/environment. One includes discussions of the role of electric vehicles (EVs) in combating climate change. Upon manual inspection, we found that these discussions mention battery disposal issues and concerns that the electricity used to power EVs may not generated from renewable sources. Previous work on mining online EV ownership forums identified several barriers to their adoption such as their higher up-front cost [13]. Based on our findings, automakers and policymakers should note that potential EV buyers may have additional concerns.

Furthermore, another topic discusses of careers that benefit the environment. This observation should be of interest to the environmental science and engineering industry.

Next, we compare our topic modeling findings with prior work. As summarized by Pearce et al. [1], while prior work on climate discussion on Twitter focused on spatiotemporal distributions on content, sentiment, and the social network properties of users tweeting about climate change, there have been some topic modeling studies. For example, Haunschild et al. [14] reported that "the most tweeted topics regarding climate change research focus on the consequences of climate change for humans". Similarly, Veltri and Atanasova [15] found that topics focus on "links between climate change and its consequences". Finally, Dahal et al. [16] identified "a very diverse collection of topics that covers politics, belief, economics, and environment". While we also found these topics on Reddit, we additionally identified discussions of specific actions people can take to combat climate change. This underscores our earlier observation about the utility of Reddit mining in assessing public opinion about specific climate change programs and initiatives.

There has also been work on topic modeling of the climate change contents found on the websites of 19 conservative North American organizations [17]. The documents analyzed in that work focused on questioning the legitimacy of climate change research and the cause of climate change. Our analysis shows that similar skepticism exists in the climate skeptics community on Reddit. Futhermore, our analysis of r/climateskeptics points out specific topics of debate beyond scientific consensus, such as renewable energy sources, loss of permafrost and ice, and rising sea levels.

Inline with several prior studies [9], [18]–[22], we found that social media users tend to relate temperature anomalies or local weather to support or deny climate change facts. However, we found that in some discussions, temperature anomalies were used both by activists and skeptics as a proof to support their standpoint.

Olteanu et al. [23] compared the coverage of events related to climate change on social media (Twitter) and online news media (collected by the GDELT project[7]). They found that "mainstream news sources frequently feature extreme weather events framed as being a consequence of climate change, as well as high profile government publications and meetings. In contrast, actions by individuals, legal actions involving governments, and original investigative journalism, feature frequently as viral events in social media". We can make similar conclusions about Reddit given the presence of discussions about specific individuals' actions and government policies.

Finally, we note that mainly r/climateskeptics appears to attract skeptics, with the other four climate subreddits containing little skeptical content. Prior work on Twitter also noted that climate change discussions are dominated by activists rather than skeptics [1], [8].

C. Which URLs Are Mentioned?

An important result of our URL analysis is the popularity of references to Wikipedia and Youtube in all five climate subreddits. As a result, it is important to ensure that climate misinformation does not appear on these platforms. We also

[7] www.gdeltproject.org

observed that many URLs referenced by r/climateOffensive redittors point to non-profit organizations supporting climate change activism, such as citizensclimatelobby.org. This finding is inline with prior research stating that in developed countries, the intention to behave environmentally friendly is more likely to translate to actual behaviour [24]. Similarly, prior work on climate change links on Twitter also observed the popularity of the online newspaper The Guardian [9]. However, in contrast to previous Twitter analyses [15], [25], our results show frequent use of references outside mainstream media such as Wikipedia and Youtube.

The Wikipedia and YouTube links on r/climateskeptics indicate that skeptical users criticize climate change by referring to the medieval warm period [26] and accusing climate activists such as Greta Thunberg of having an Obsessive-Compulsive Disorder (OCD). Similarly, prior work [27], [28] found that climate change skeptics often label scientists and others as alarmists. Finally, climate subreddits often included links to James Hansen's biography and Youtube videos, making him an influential figure in online climate discussions.

V. Conclusions

We analyzed climate change content on the Reddit social media platform using substring matching and topic modeling. We leveraged the community-oriented structure of Reddit to identify communities, and therefore high-level topics, influenced by climate change. We also identified specific topics discussed by five climate change subreddits. We hope that our findings related to the nature of the communities discussing climate change, as well as the topics of the discussions themselves, can provide climate scientists and policymakers with insights into public perception of climate policies.

One limitation of this study is that the findings represent the views of a biased sample of the public, namely those who are active on Reddit. An interesting direction for future work is to compare climate change opinions and attitudes of other population groups and other countries not represented on Reddit.

References

[1] W. Pearce, S. Niederer, S. M. Özkula, and N. Sánchez Querubín, "The social media life of climate change: Platforms, publics, and future imaginaries," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 10, no. 2, p. e569, 2019.

[2] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, and J. Brew, "Huggingface's transformers: State-of-the-art natural language processing," *ArXiv*, vol. abs/1910.03771, 2019.

[3] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," *arXiv e-prints*, p. arXiv:1703.04009, Mar. 2017.

[4] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 267–273.

[5] D. O'callaghan, D. Greene, J. Carthy, and P. Cunningham, "An analysis of the coherence of descriptors in topic modeling," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5645–5657, 2015.

[6] 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.

[7] 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, Oct. 2017.

[8] 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.

[9] A. P. Kirilenko and S. O. Stepchenkova, "Public microblogging on climate change: One year of twitter worldwide," *Global environmental change*, vol. 26, pp. 171–182, 2014.

[10] P. Kitcher, "The climate change debates," *Science*, vol. 328, no. 5983, pp. 1230–1234, May 2010.

[11] G. Wong-Parodi and I. Feygina, "Understanding and countering the motivated roots of climate change denial," *Current Opinion in Environmental Sustainability*, vol. 42, pp. 60–64, Feb. 2020.

[12] M. M. Davies, "Are we out to lunch? Are the dietary strategies we share on social media effective in mitigating climate change?" Master's thesis, Dalhousie University, 2019.

[13] T. Carpenter, L. Golab, and S. J. Syed, "Is the grass greener?: mining electric vehicle opinions," in *The Fifth International Conference on Future Energy Systems, e-Energy '14*, 2014, pp. 241–252.

[14] R. Haunschild, L. Leydesdorff, L. Bornmann, I. Hellsten, and W. Marx, "Does the public discuss other topics on climate change than researchers? a comparison of explorative networks based on author keywords and hashtags," *Journal of Informetrics*, vol. 13, no. 2, pp. 695–707, 2019.

[15] G. A. Veltri and D. Atanasova, "Climate change on twitter: Content, media ecology and information sharing behaviour," *Public Understanding of Science*, vol. 26, no. 6, pp. 721–737, Nov. 2015.

[16] B. Dahal, S. A. P. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Social Network Analysis and Mining*, vol. 9, no. 1, Jun. 2019. [Online]. Available: https://doi.org/10.1007/s13278-019-0568-8

[17] C. Boussalis and T. G. Coan, "Text-mining the signals of climate change doubt," *Global Environmental Change*, vol. 36, pp. 89–100, 2016.

[18] L. Zaval, E. A. Keenan, E. J. Johnson, and E. U. Weber, "How warm days increase belief in global warming," *Nature Climate Change*, vol. 4, no. 2, pp. 143–147, Jan. 2014.

[19] J. Finnis, A. Sarkar, and M. C. Stoddart, "Bridging science and community knowledge? the complicating role of natural variability in perceptions of climate change," *Global Environmental Change*, vol. 32, pp. 1–10, May 2015.

[20] S. B. Capstick and N. F. Pidgeon, "Public perception of cold weather events as evidence for and against climate change," *Climatic Change*, vol. 122, no. 4, pp. 695–708, Jan. 2014.

[21] P. J. Egan and M. Mullin, "Turning personal experience into political attitudes: The effect of local weather on americans' perceptions about global warming," *The Journal of Politics*, vol. 74, no. 3, pp. 796–809, Jul. 2012.

[22] A. Spence, W. Poortinga, C. Butler, and N. F. Pidgeon, "Perceptions of climate change and willingness to save energy related to flood experience," *Nature Climate Change*, vol. 1, no. 1, pp. 46–49, Mar. 2011.

[23] A. Olteanu, C. Castillo, N. Diakopoulos, and K. Aberer, "Comparing events coverage in online news and social media: The case of climate change," in *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015*, 2015, pp. 288–297.

[24] M. Morren and A. Grinstein, "Explaining environmental behavior across borders: A meta-analysis," *Journal of Environmental Psychology*, vol. 47, pp. 91–106, 2016.

[25] T. P. Newman, "Tracking the release of ipcc ar5 on twitter: Users, comments, and sources following the release of the working group i summary for policymakers," *Public Understanding of Science*, vol. 26, no. 7, pp. 815–825, 2017.

[26] S. K. Ritter, "Global warming and climate change," *Chemical and Engineering News*, vol. 87, pp. 11–21, 2009.

[27] S. Niederer, "'global warming is not a crisis!': Studying climate change skepticism on the web," *NECSUS. European Journal of Media Studies*, vol. 2, no. 1, pp. 83–112, Jan. 2013.

[28] B. Forchtner, A. Kroneder, and D. Wetzel, "Being skeptical? exploring far-right climate-change communication in germany," *Environmental Communication*, vol. 12, no. 5, pp. 589–604, Jun. 2018.