The Evolution of the News Narrative on Climate Change

Adam Block Helena Caswell Dylan Lewis Disha Trivedi

May 17, 2021

1 Abstract

Print and televised media reporting on climate change influences the public perception of climate change, which in turn affects support for individual actions and systemic policies to mitigate climate change. Over two thirds of Americans get their news often or sometimes from television. In this analysis, we examine ten years of data from three highly-viewed television stations whose viewership is on the rise: CNN, Fox News, and MSNBC. We find that frequency of climate change mentions follows similar patterns across networks over time, yet the content of the climate change mentions greatly differ across networks in certain years and are generally similar in others. The mentions of climate change in the media follow trends that appear to influenced by political and environmental factors such as election cycles, international diplomacy, and natural disaster events.[1]

2 Introduction

Print and televised media reporting on climate change influences the public perception of climate change [Antilla, 2008], which in turn affects support for systemic policies and individual actions to mitigate climate change, such as regional emissions targets and purchases of hybrid vehicles respectively [Chen and Zhao, 2019]. Even online news and social media's rise in the past decade, approximately 68% of Americans get their news often or sometimes from television [Center, 2020]. At the same time, US public opinion on climate change has shifted significantly in the past decade, rising from only 49% of the population believing that global warming would harm the U.S in 2008 to 62% believing the same in 2018 [YPC, 2020, Ballew, 2020].

Within the context of this decade-long rise in public opinion that climate change will harm the U.S., we seek to investigate the simultaneous changes in climate change TV news coverage to ask the following research question:

How has the frequency and content of top American English-speaking news media coverage of climate change evolved in the past ten years—and what environmental and political factors have influenced the trends?

If identified, these factors serve an important role in shaping news coverage, which correspondingly shapes all-important public opinion about climate change's risk and policy and personal choice to offset it. Conclusive documentation exists that the three networks on which this project focuses reach a key cross-section of U.S. viewership. As major television networks—among the top 7 most watched television networks overall in 2020 and the top 25 most watched networks consistently between 2009 and 2020—these networks attract anywhere from half a million to over 2 million daily viewers on any average evening over the last decade. Primetime news viewership of Fox News, CNN, and MSNBC is on the rise, which further motivates for our study of these networks' climate change mentions over time[Reports, 2019]. Figure 17, drawn from media summaries of Nielsen yearly television ratings data, demonstrates this rise.

[1]The code for this project can be found here: https://github.com/dyllew/ids.131-fp

3 Related Work

There is ample evidence that the environmental and political factors do influence news media coverage and corresponding public opinion on climate change. In a review of related work, public perception of climate risk has been shown to be strongly influenced by concurrent extreme weather events [Kim and Marlon, 2020]. We also found work from Ruz et al. 2020 that indicates that natural disasters are strongly correlated with negative sentiment associated with environmental discourse on social media. In the same work, Ruz et al. 2020 found that social movements, particularly those around climate and environment, were associated with heterogeneous sentiment changes on Spanish-speaking Twitter media. Sentiment analysis and Bayesian classifier use on such data also enabled network mapping between related patterns of environmental and natural disaster-associated words [Ruz, 2020].

Political affiliations and media partisanship also influence political coverage and events. A study by the Feldman et al. on 2007-2008 survey data of climate change coverage on Fox News, CNN, and MSNBC found a negative association between Fox News viewership and acceptance of global warming and a positive relationship between CNN and MSNBC viewership and acceptance of global warming [Feldman and Leiserowitz, 2011]. The authors also found that Fox News takes a more dismissive tone towards climate change than CNN or MSNBC and features interviews with a greater ratio of climate change deniers than believers. This partisan news divide maps to beliefs about climate change, as 95% of liberal Democrats, compared to 41% of conservative Republicans, believe in climate change's existence.

Dahal et al. 2019 conducted computational analysis on U.S.-based users' Twitter data to gauge public opinion on climate change. Their sentiment analysis demonstrated a strong negative correlation between Twitter discourse on climate change in response to extreme weather or political events [Dahal and Li, 2019]. Similar work has been conducted on British and Spanish Twitter data, where sentiment analysis found a more negative association with climate change in Britain than in Spain and determined positive sentiments attached to discussion of renewable energy on both countries [Loureiro and Alló, 2020].

Much like the above studies, Jost et al. 2019 conducted a sentiment analysis, albeit non-computational, of Canadian print and televised media to determine the positive or negative associations attached to the 'change' portion of 'climate change' [Jost, 2019]. Their findings indicated that sentiments attached to 'change' varied with political factors that influence, block, or drive policy change.

Such findings demonstrate that environmental and political factors influence climate change discourse in text-based media. These findings correspondingly dovetail with and motivate our investigation of environmental and political factors' influence on American English-speaking television media.

4 Data Description, Preprocessing, and Exploration

For this analysis we have decided to use a dataset of TV news transcript snippets related to climate change coverage across CNN, MSNBC, and Fox News from July 2009-January 2020, from the GDELT Project.[2] These data were gathered using a TV News Archive[3] and searching for snippets with the following keywords: "climate change", "global warming", "climate crisis", "greenhouse gas", "greenhouse gases", and "carbon tax". This dataset includes the UTC time and date of the mention, the station, the show, and a text snippet corresponding to a 15 second clip of the transcribed audio from the show regarding the climate change mention. Although the dataset includes similar data from the BBC, due to the relative paucity of the BBC data—the BBC snippets extends only from 2017-January 2020—and our primary focus on the American TV media landscape, we exclude these data from consideration. The resulting American media-only dataset has 19,304 snippets from CNN, 25,865 snippets from Fox News, and 26,429 snippets from MSNBC for a total of 71,598 snippets of climate change coverage. The features of the data discussed below as well as the abundance of data points among the individual networks over time provides us the ability to compare the frequency & content of climate change mentions between the TV networks over time in order to answer the research question posed.

[2] Link to Climate TV News Dataset
[3] Link to TV News Archive

4.1 Text Preprocessing Pipeline for the Climate TV News Corpus

Before we can analyze the data, we need to clean it. The first point to note is that there is missing data from CNN from the month of October 2009, so we remove this data from consideration. While our data have ancillary features (including URLs to sources, the name of the show, and some identifying information), we are primarily concerned with the date of airing for frequency analysis and the text snippet for content analysis. We follow standard Natural Language Processing (NLP) practice by removing punctuation and numbers, converting all letters to lower-case, and tokenizing the data. Finally, we lemmatize each word, i.e. we convert each word to its root, in order to better match similar semantics (e.g. we might hope that *go, going*, and *went* are all treated as the same word). As we are working with relatively short snippets, we do consider the structure of the sentence to be of lesser importance and thus tokenize by single-word chunks as opposed to more complicated n-grams. Finally, we use the NLTK package in Python [Bird et al., 2009] to remove standard English stopwords and identify corpus-specific stopwords—words that are common to many of the snippets which we deem too ubiquitous and meaningless to contribute to any signal in the differentiating between snippets—in order to better distill each snippet's signal. In our analysis, we treat each snippet or at other times, a group of related snippets, as what is typically referred to as a *document* in NLP literature. We use the terms interchangeably, but we do make clear the distinction when using a single snippet vs. a group of snippets. In addition to removing the English-language stopwords provided in the NLTK package, we found and removed stopwords specific to the Climate TV News Corpus before conducting our analysis.

4.2 Identifying and Removing Corpus-Specific Stopwords

We remove corpus-specific stopwords by first removing the standard stopwords provided in the NLTK package. We then use a max document frequency threshold of 20% (i.e. terms which appear in greater than 20% of the snippets in the entire dataset) and removed words found to be above this threshold. Finally, we also removed any other words that do not contribute to understanding the snippet's semantics or differentiation of the signal in later analysis (i.e. commonly modifiers, pronouns, and adverbs like *some, back, theyre, he, look, really*). The full list of corpus-specific stopwords we found is listed in Appendix Section B. After removing stopwords, we find that the Climate TV News corpus vocabulary consists of 33,755 unique words.

4.3 Identifying Most Frequent Named Entities by Network and Year

In exploring this dataset, we sought to gain a course understanding of who, what, and where each network discussed, and in what corresponding years, so we applied a pre-trained Named Entity Recognition (NER) model provided by the python spaCy package[4] as a well as a pre-trained true-caser model[5] as our dataset was originally all lower case text in order to capitalize potential named entities. We break the corpus of snippets up into documents or collections of snippets, where each document is the concatenation of all snippets for a specific network in a specific year (i.e. all Fox News snippets in 2010). We then apply the true caser model to upper case potential named entities and apply the NER model on the transformed document to extract the named entities. We present the 50 most frequent named entities for each network and year document below in Figure 18, where the size of the entity is directly correlated to its frequency in the document.

From these word clouds, we observe that the most frequent named entities across the years and the networks are typically politicians, specifically presidents, presidential, and vice-presidential candidates such as Barack Obama, Donald Trump, Al Gore, Sarah Palin, Mitt Romney, Newt Gingrich, Hillary Clinton, etc. We also observe frequent mentions of countries, cities, or states, i.e. the US, China, India, Copenhagen, Paris, Alaska, New York, and California. Finally, we observe the frequent mentions of US governmental agencies, bodies, or political parties such as the White House, EPA, Senate, Congress, Republican(s), Democrat(s).

4.4 Determined Methods of Analysis

For the analysis presented herein, we note that we decided not to pursue sentiment analysis as was originally stated in the proposal. Although it would have added breadth to our analysis, it would have been at the

[4] Link to Pre-trained NER spaCy model
[5] Link to Pre-trained True Caser Model

expense of not pursuing deeper analysis in Time Series and TF-IDF streams of analysis. As a result, our analysis in these streams goes deeper than we originally planned for in the proposal as we added dynamic time warping to the Time-Series Analysis stream and added cosine similarity to the TF-IDF stream. We still leverage breadth in methods covering three distinct methods of analysis in depth, namely Time-series, TF-IDF & Cosine Similarity, and Topic Modeling to answer the research question posed.

5 How does the Frequency of Climate Change Mentions Vary Over Time?

First, we looked at how the frequency of climate change mentions vary over time for each network. We found that the frequency of climate change mentions varies significantly over time, peaking with well known events such as the 2009 United Nations Climate Change Conference (UNCCC), 2015 Paris Agreement (PA), 2018 IPCC Special Report, and Democratic primary debates throughout 2019.

Climate change mentions over time are strongly and positively correlated between networks. The correlations between networks' average daily and monthly climate change mentions are shown in Table 1 and 2 respectively.

	CNN	Fox News	MSNBC
CNN	1.00	0.55	0.57
Fox News	0.55	1.00	0.52
MSNBC	0.57	0.52	1.00

Table 1: Daily Correlation Between Network Climate Change Frequencies

	CNN	Fox News	MSNBC
CNN	1.00	0.76	0.76
Fox News	0.76	1.00	0.71
MSNBC	0.76	0.71	1.00

Table 2: Monthly Correlation Between Network Climate Change Frequencies

Based on these correlations, we determined that each network's climate change mentions are generally highly correlated with those of their peers. The greatest correlation exists between CNN and MSNBC's average monthly mentions of climate change. As both networks have a partisan lean to the left, such a relationship is to be expected. The overall correlations track with the partisan lean of the networks, with MSNBC being the most far left and thus the most different and least correlated with Fox News, which is the furthest right. This partisan gap explains MSNBC and Fox News' lowest correlations of 0.52 and 0.71 daily and monthly correlations, respectively. We have also visualized monthly and daily climate change mentions below, with the daily mentions visualizations collected in Figures 38 & 39.

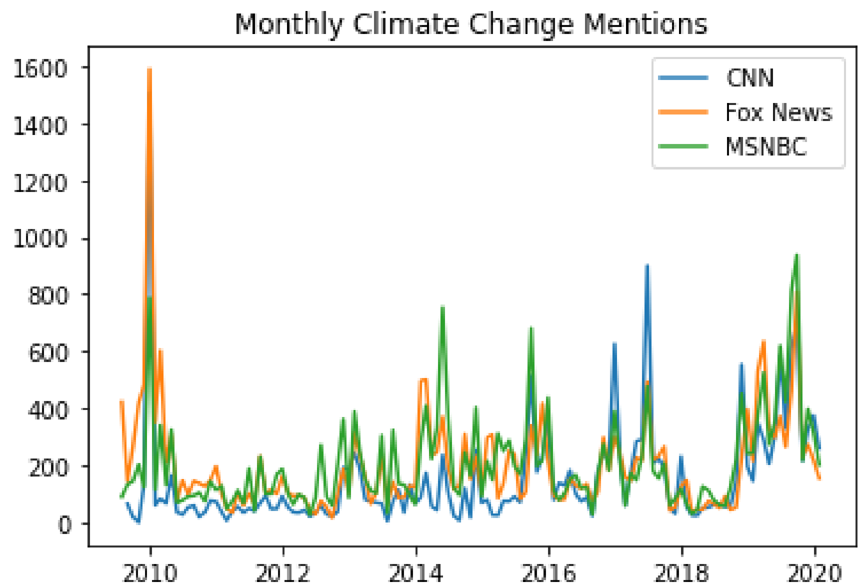


Figure 1: Frequency of Climate Change Mentions by Month and Network

Based on our research, we hypothesized that environmental factors, and specifically natural disasters might be correlated with frequency of climate change mentions. For this analysis, we looked at two different US natural disaster datasets. Although it is likely that some natural disasters outside the US make American television news and are discussed with connection to climate change, we focused on US natural disasters for simplicity. We looked at two different US natural disaster datasets. The first was on US billion dollar disasters from the National Oceanic and Atmospheric Administration (NOAA), by year [NOAA, 2020]. The figure below shows the frequency of climate change mentions by each network against disasters. It is clear that there is no significant correlation by year for number of billion dollar disasters. We also looked at correlations by disaster type, in case some disasters might be more tied to climate change than others, but did not find any disaster types that appeared to correlate annually with climate change mentions.

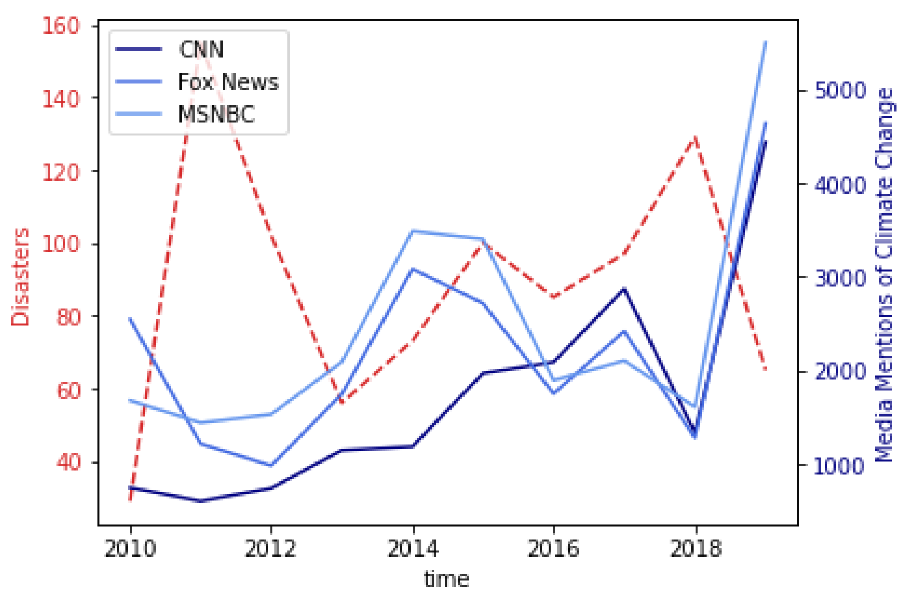


Figure 2: Frequency of Climate Change Mentions and Billion Dollar Disasters

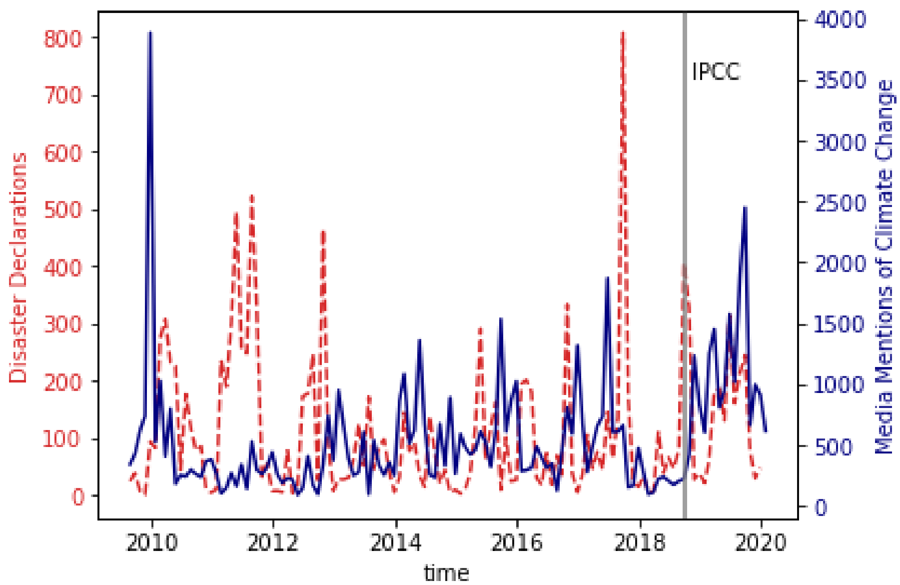


Figure 3: Frequency of Climate Change Mentions and Disasters Declarations

However, this is not entirely surprising, given that natural disasters may not dominate news cycles on climate change at an annual scale. We also wanted to look at a more time-granular dataset to see if there were correlations by day or month, so we looked at US Natural Disaster Declarations from the Federal Emergency Management Agency (FEMA), which had daily data [FEMA, 2021]. This data is by county, which has the effect of magnifying larger disasters that affect multiple counties. This feature is useful, since larger disasters should be more likely to be covered by the media, and amplification may allow us to better see that in our analysis. From this dataset we selected extreme weather incident types that could be affected by climate change (Fires, Severe Storms, Hurricanes, Floods, Coastal Storms, Tornados, Typhoons, Snow, Severe Ice Storms, Mud/Landslides, and Freezing) and excluded all others (Earthquakes, Volcanos, Toxic Substances, Terrorists, Chemical, Dam/Levee Break, Biological). For this dataset, we analyzed correlation by month since disasters often last more than one day.

Correlation between disaster declarations and media mentions of climate change was low over the time series, but varied between stations, with MSNBC having a correlation almost six times higher with extreme weather natural disaster declarations than CNN and eight times higher than Fox News. When plotting the total (CNN + MSNBC + Fox News) mentions by month against the monthly disaster declarations, we noticed that the two appeared to correlate much more after the release of the IPCC's Special Report on Global Warming of 1.5 ºC in October 2018. The IPCC report galvanized and coincided with youth activists such as Greta Thunberg and the Sunrise Movement, as well as leading many policymakers to consider more aggressive carbon mitigation goals. As shown in the figure below, it appears to have marked an overall increase in mentions of climate change. However, it also marks a remarkable increase in the correlation between climate change mentions and extreme weather declarations, with correlation for MSNBC increasing from 0.061 to 0.43. Post IPCC report, the trend of MSNBC correlating the most, then CNN, then Fox News still holds. This suggests that the mentioning of climate change in relation to natural disasters is more frequent in more liberal TV media, and was very infrequent across networks until recent years. More information on this is given in the content analysis.

	MSNBC	CNN	Fox News
Full Time Series Correlation	0.061	0.011	0.008
Post IPCC Report Correlation	0.43	0.30	0.15

Table 3: Correlation between Climate Change Mention Frequency and Natural Disaster Declarations

To compare the number of daily climate change mentions across networks, we use dynamic time warping, a method that allows comparison across time series data when the time indices between time series does not perfectly align (is non-Euclidean)[Zhang, 2019]. The indices in each network's daily climate change mentions data do not align because not every network recorded a daily climate change reference.

Dynamic time warping maps time indices from one series onto one or more indices in another series, outputting a measure of Euclidean distance between the two series. This distance measurement indicates the degree of separation (or difference) between the series[Zhang, 2020]. Table 4 includes the pairwise distance measurements between each of the three time series. While the distances are similar, MSNBC and CNN's largest distance from each other indicates that these two networks make mention of climate change at the most different daily interpolations.

Network	Distance
MSNBC with CNN	19,905
CNN with Fox News	18,525
MSNBC with Fox News	17,151

Table 4: Dynamic Time Warping Distance Between Daily Climate Change Mentions Data on 3 Networks

Because we wanted to detect anomalies within our data, as well as check for the effect of trend and seasonality, we conducted an STL decomposition into trend, season, and noise/residue on our total and individual network series[Statmodels, 2021b]. STL has two key drawbacks: (1) it assumes additivity across seasonal/trend/residual components and (2) it does not assume calendar year changes in periodicity, and requires user specification of such periodicity[Statmodels, 2021a]. To address the periodicity drawback, we ran the STL decomposition with periodicity 1, 2, and 4 years (in Python: 365, 730, and 1460 days respectively), corresponding to one year, one midterm election cycle, one presidential election cycle. These political cycles were chosen as our period because the literature review indicates strong linkage between political cycles and media coverage. However, "news cycles" of frequent topic-based coverage can in fact be quite short and variable—anywhere from a day (the typical newsroom 24 hour cycle) to a week (for one-time hot topics, like a political debate) to two months (for longer topics, like the Paris Agreement negotiations)[Lab, 2019]. These iterations did not show significant differences in the seasonal and trend components.

Based on our plots of daily and monthly individual network data (see Appendix C.5) as well as daily and monthly aggregates of all networks, we suspect our time series are fairly stationary. However, analyzing the standard deviation of the residual and determining if it crosses a threshold enables anomaly detection. We also used a "robust" and "non-robust" decomposition, with the robust model using a data-dependent weighting function to reweight the data, which reduces the impact of anomalies. Using the generalized extreme student deviate test for outliers across the residual component, a test which allows calculation for any number of outliers if a suspected number is not known, we detected one outlier at the year 2010 for the "non-robust" STL decomposition. We detected no outliers using the "robust" decomposition. Our STL decomposed aggregate number of daily climate mentions on all three networks appears to be fairly stationary. Below, we visualize the "non-robust" decomposition for one-year (12 month) periodicity of total monthly average climate change mentions across the 3 networks. We see seasonality when using this year-long variation, which intriguingly suggests that particular seasons feature spikes in climate change mentions in news overall. Perhaps winter weather coverage, which may include climate-induced weather event severity, influences the yearly spike toward the start of every year (corresponding to January). However, we see no upward trend in climate change coverage within this model, indicating that despite climate change's effects and news viewership both increasing, news coverage on the topic has not. Here, we're also able to visualize the one outlier in year 2010, which corresponds to the average monthly data for January 2010. During this period, Congressional midterm elections resulted in a newly Republican-controlled Congress and the 2010 IPCC Report was released, to brainstorm about possible political and climate-related confluences with this outlier.

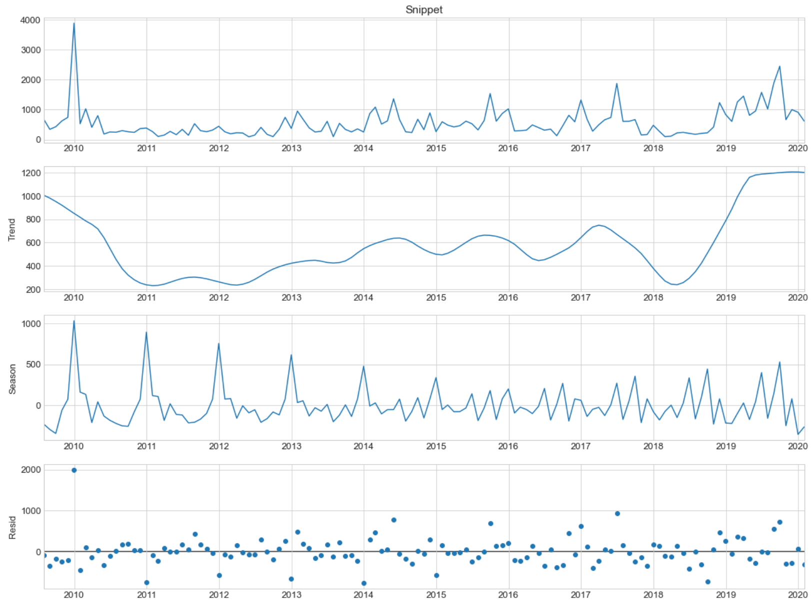


Figure 4: Frequency of Climate Change Mentions and Disasters Declarations

For STL Decomposition for single networks daily data with 1 week and 1 year periodicity, as well as our robust vs. non-robust total network visualizations, see Figure 41. Outlier detection on all three separate networks yielded no outliers for the robust nor the non-robust decomposition. For STL Decomp on single networks' monthly data with 1 year periodicity, see Figure 42. The monthly data plots show a yearly seasonality across networks, which corresponds to the seasonality we see in the plot above.

6 How Does the Content of Climate Change Mentions Vary by Network and over Time?

To gain insight into how the content of mentions of Climate Change varies by network and over time, we use two distinct methods of content analysis. We begin our content analysis by merging together related snippets on the basis of Network, Year, and Network & Year in order to form a document. We then use Term Frequency-Inverse Document Frequency (TF-IDF) to construct normalized numerical word embeddings for each document that allow us to extract the most important and relevant words to each document and we compute the pairwise cosine similarity between document embeddings to capture the similarity in their relative word importances as we assume this similarity measure to be a reasonable proxy for their content similarity. Additionally, we utilize Topic Analysis to unveil abstract topics defined by words in the corpus and study how the topics vary by network and time.

6.1 Identifying Important Words and Similarity by Network, by Year, and by Network & Year

6.1.1 Document TF-IDF Featurization

The tf-idf statistic is a useful tool for extracting important keywords or words relevant to a document in a corpus. A tf-idf score for a word in a specific document is an importance score based on the word's occurrences in the document (term-frequency) as well as the inverse of the presence of the word across all documents in the corpus (inverse document frequency) [Rajaraman and Ullman, 2011]. It follows from this calculation that tf-idf has an advantage over typical term frequency in a document for extracting important words as it weights words which appear in many documents in a corpus as less important and less informative than words which appear in few documents [Pedregosa et al., 2011]. The implications for using tf-idf for extracting important words for each document are further illustrated in the calculation shown in Appendix section A.1. We note that in our analysis, we use a variation of the IDF, which incorporates smoothing in order to not completely ignore words which appear in every document in the corpus (e.g. *president*).

In our analysis, we use the TF-IDF vectorizer provided in the scikit learn python package [Pedregosa et al., 2011] to construct a document-term matrix transforming our text data into a numerical representation for analysis. Each row of this matrix represents a vector (or embedding) corresponding to a document at index i. Each column corresponds to a unique word at index j in the corpus vocabulary, thus the number of columns (length of document vector) is equal to the number of unique words in the entire corpus. Each entry in this matrix corresponds to the normalized TF-IDF score for the word w_j in document d_i. Each row is unit norm, $L2$-normalised, such that the sum of the squared elements of the vector is equal to 1.

This matrix representation allows us to parse out which words were most important to each document by looking at the k top-ranked words by normalized TF-IDF score for each document and provides a numerical vectorization of our text data enabling us to compute content similarity between networks, years, and networks in specific years.

6.1.2 Cosine Similarity

We calculate a measure of similarity between documents by calculating the cosine similarity between document tf-idf vectors as follows, where doc_A and doc_B are the tf-idf vectors from the document-term matrix:

$$\text{Cosine Similarity}(doc_A, doc_B) = \cos(\theta) = \frac{doc_A \cdot doc_B}{\|doc_A\| \|doc_B\|}$$

Since the rows in our document-term matrix are unit norm, this formula reduces to become the dot product between the tf-idf vectors. We note that since TF-IDF scores cannot be negative, the range of the cosine similarity will range from $[0, 1]$, in which documents that are entirely dissimilar will have a cosine similarity score of 0 (No overlap in relative word importances between the documents) and document which are extremely similar (i.e. same exact normalized TF-IDF score for each word in the corpus) will have a cosine similarity score of 1.

6.1.3 Different Levels of Document Granularity

For our analysis, we split the corpus into documents at three different levels of granularity. For each of the following levels, we compute the tf-idf document-term matrix, extract and visualize the top 50 words with highest TF-IDF scores for each document, and compute the cosine similarity between the documents:

1. **By Network**: Each document is constructed from the concatenation of all snippets for a specific news network: CNN, Fox News, and MSNBC (3 documents in the corpus)

2. **By Year**: Each document is constructed from the concatenation of all snippets in a specific year, e.g. all snippets in 2009 (11 documents in the corpus).

   Note: We combine the snippets in 2019 and 2020 together as there was only one month's worth of data for 2020, January 2020.

3. **By Network and Year**: Each document is constructed from the concatenation of all snippets for a specific network in a specific year, e.g. all MSNBC snippets in 2015 (33 documents in the corpus).

   Note: Similar to the level above, we combine 2019 & 2020 snippets together for each network.

6.1.4 By Network Documents

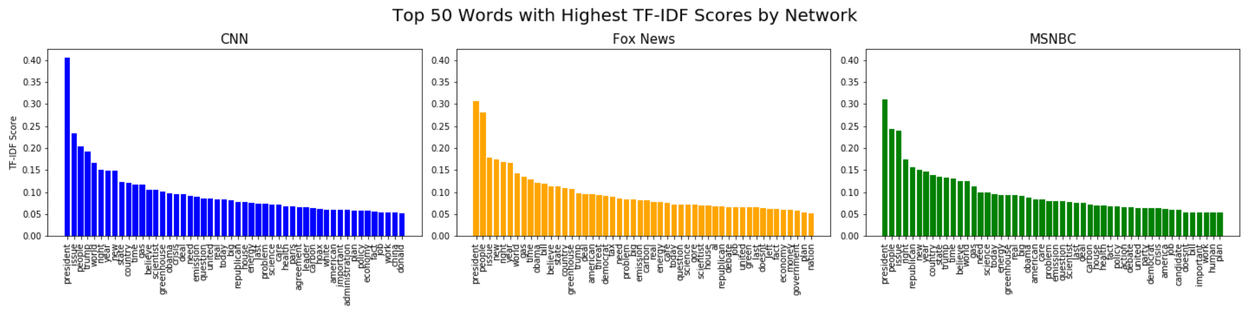


Figure 5: Top 50 words with Highest TF-IDF Scores by Network Bar Plots

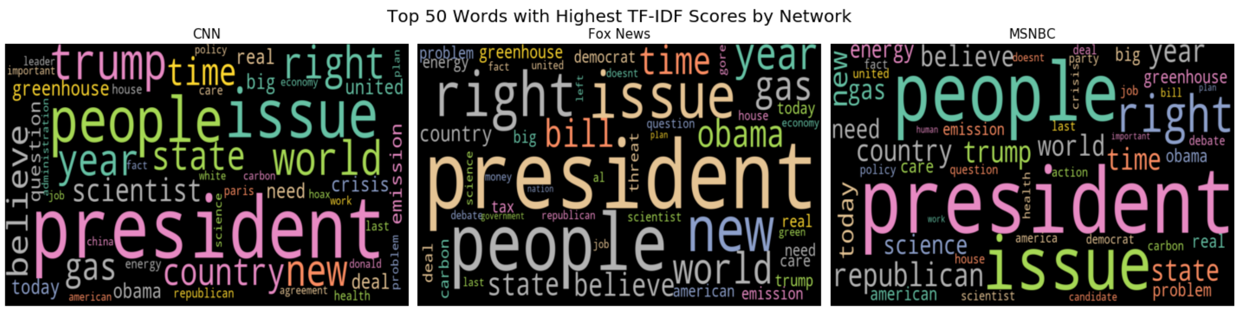


Figure 6: Top 50 words with Highest TF-IDF Scores by Network Word Clouds

At this level of document granularity, we see that many of the words in the top 50 words with highest TF-IDF for each network are the same although they do differ slightly in their actual TF-IDF scores as seen in Figures 5 & 6. This large overlap in the most important words is the result of using the smoothing variation of the IDF term in the TF-IDF calculation coupled with having only a few documents that use the same words most frequently. The smoothing variation has the implication that words that appear in all of the networks will not be completely ignored, but nonetheless have the lowest IDF weighting contribution of any words in the corpus. For words which are present in each of the networks and have a high TF-IDF score, such as *president, people, issue, trump, republican, world, right*, and *new*, differ in TF-IDF score only on the basis of their relative frequency (TF) for each of the networks respectively. For example, *president* is the word with highest TF-IDF score for each of the networks, although it has the highest score TF-IDF for CNN at ∼0.40, and ∼0.32 for Fox News and MSNBC. Since the IDF weighting for the word *president* is the same for each of the networks as it appears in all the networks, this implies that the word *president* occurs relatively more frequently for CNN than it does for the other networks, but its relative frequency in all of networks is so high that it makes TF-IDF high even with the lowest possible IDF. Although for many of these words, the IDF score is the same between the networks, the TF-IDF score still reveals how relatively important a word was for one network compared to the others. We observe a similar instance of this with the word *trump*, which has a TF-IDF score of ∼0.19 for CNN, ∼0.11 for Fox News, and ∼0.14 for MSNBC, although this difference is less dramatic than it was for president, it still highlights the relative importance of the word for CNN compared to the other networks.

The following words are overlapping between the networks in their top 50 most important words sorted in alphabetical order:

american, believe, big, carbon, care, country, deal, emission, energy, fact, gas, greenhouse, house, issue, job, last, need, new, obama, people, plan, president, problem, question, real, republican, right, science, scientist, state, time, today, trump, united, world, year

We also present the set of words which do not overlap between all of the networks top 50 most important words for each of the network (e.g. words which were in the top 50 for CNN but either not in the top 50 for both of the other networks or only present in the top 50 for one other network). These non-overlapping words are ordered by descending TF-IDF score for the network include:

CNN:
crisis, health (assumed to be from health care mentions), *paris, agreement, leader, hoax, white* (assumed to be from White House mentions), *important, adminstration, policy, economy, work, china, donald*

Fox News:
bill, threat, democrat, tax, gore, al, debate, green, doesnt, left, economy, money, government, nation

MSNBC:
health (assumed to be from health care mentions), *policy, action, debate, party, democrat, crisis, america, candidate, doesnt, bill, important, work, human*

We use the set of words created from the union of the top 50 most important words for each of the networks shown in Figure 5 and study these words using the TF-IDF scores found at the Network and Year document granularity in Section 6.1.6 in order to study and compare the evolution of the importance of these words over the years for each of the networks.

From the most important words shared between all of the networks, we see that content of climate change mentions regards a mix of words about: government officials and organizations such as *president, obama, trump*, and *house*; climate change topics and policies such as *carbon emissions, greenhouse gases, energy, science, scientist, plan*, and *deal*; the U.S. and the world such as *american, country, united states, world*; US political parties such as *republican/republicans*; and other words such as *believe, big, fact, issue, job/jobs, last, need, new, problem, question, real, right, time, today*, and *year*.

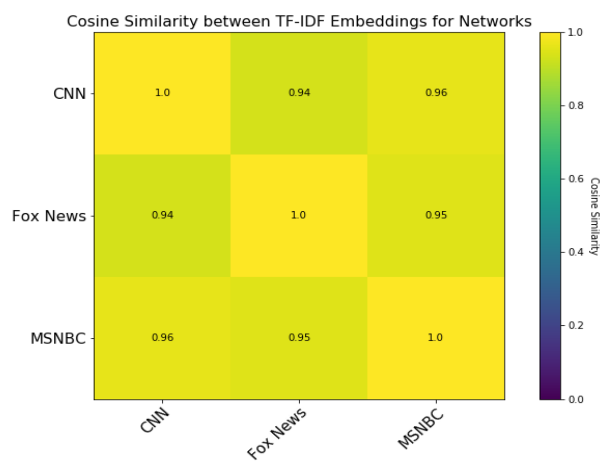


Figure 7: Cosine Similarity Heatmap between Network Documents TF-IDF Word Embeddings

We compute the cosine similarity between each of the Networks TF-IDF word embedding vectors as shown in Figure 7. The pairwise cosine similarities between each of the network embeddings is very high: 0.94 between CNN and Fox News, 0.95 between Fox News and MSNBC, and 0.96 between CNN and MSNBC. We observe that the cosine similarity shared between the liberal leaning networks is the highest, albeit marginally, and the cosine similarity shared between CNN and Fox News is the lowest although still very high.

Using finer document granularity at the level of years as well as networks and years shows larger variety in which terms are important to specific years and to networks in specific years as there is less overlap between top words between the documents than at the by network level since the idf will not be the same for many of the top words for these documents. These finer levels more clearly depict how the content of the climate change mentions varies over time and between networks over time.

6.1.5 By Year Documents

Since there are almost 4 times as many documents at this level, we notice far less overlap in the top 50 most important words to each year than at the level of networks. There are words which had high TF-IDF score and ubiquitous presence in the documents at the network level which also have similar importance and ubiquity at this level, namely *president* as it has the top TF-IDF score for many of the documents in this level. Some other words which remain important over the years as well include *people*, the president in that year (*obama* or *trump*), *greenhouse gases, carbon emissions, science, scientist*, and *republican/republicans*. However at this level, we observe that political figures other than Obama and Trump emerge in some of the years important words, tracking with major political events occurring in those years such as presidential elections. We see examples of this seen in Figures 8 & 9, where in the 2011 document, we see important terms for the year 2011 regarding major Republican party presidential candidates running in the 2012 presidential election such as Mitt Romney, Newt Gringrich, Rick Perry, and Jon Huntsman.

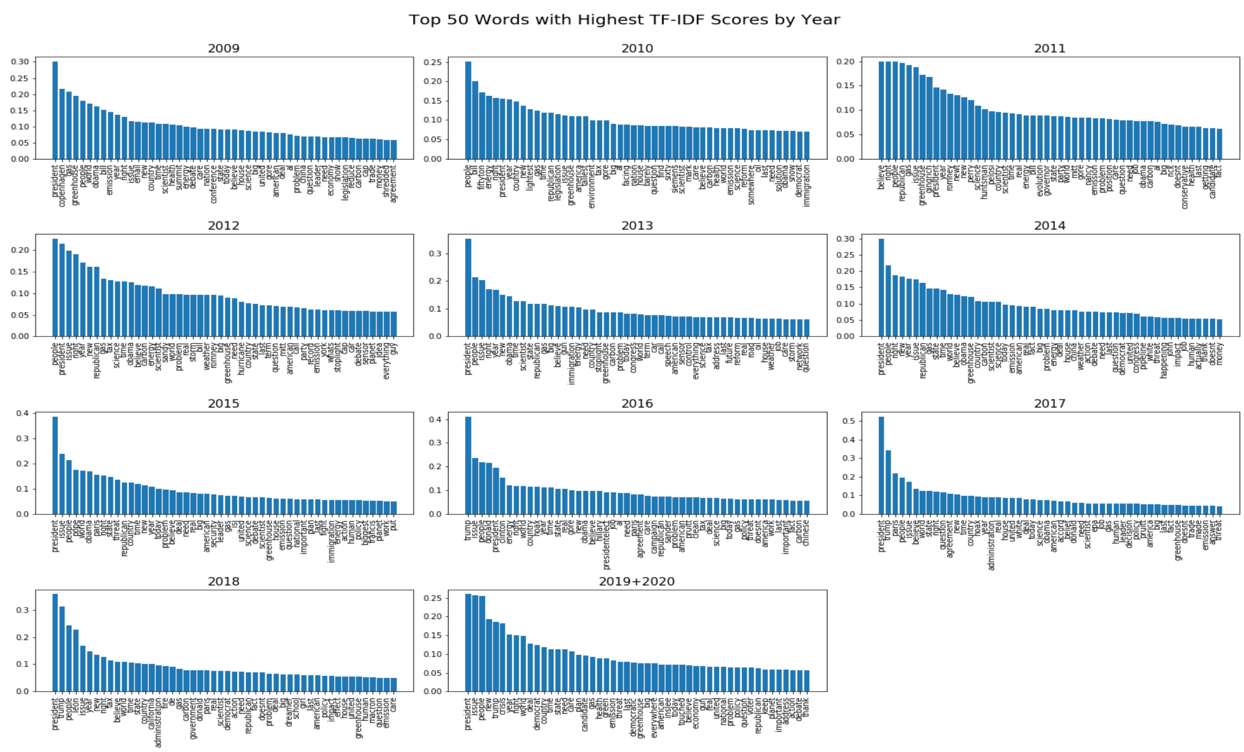


Figure 8: Top 50 Words with Highest TF-IDF Scores by Year Bar Plots

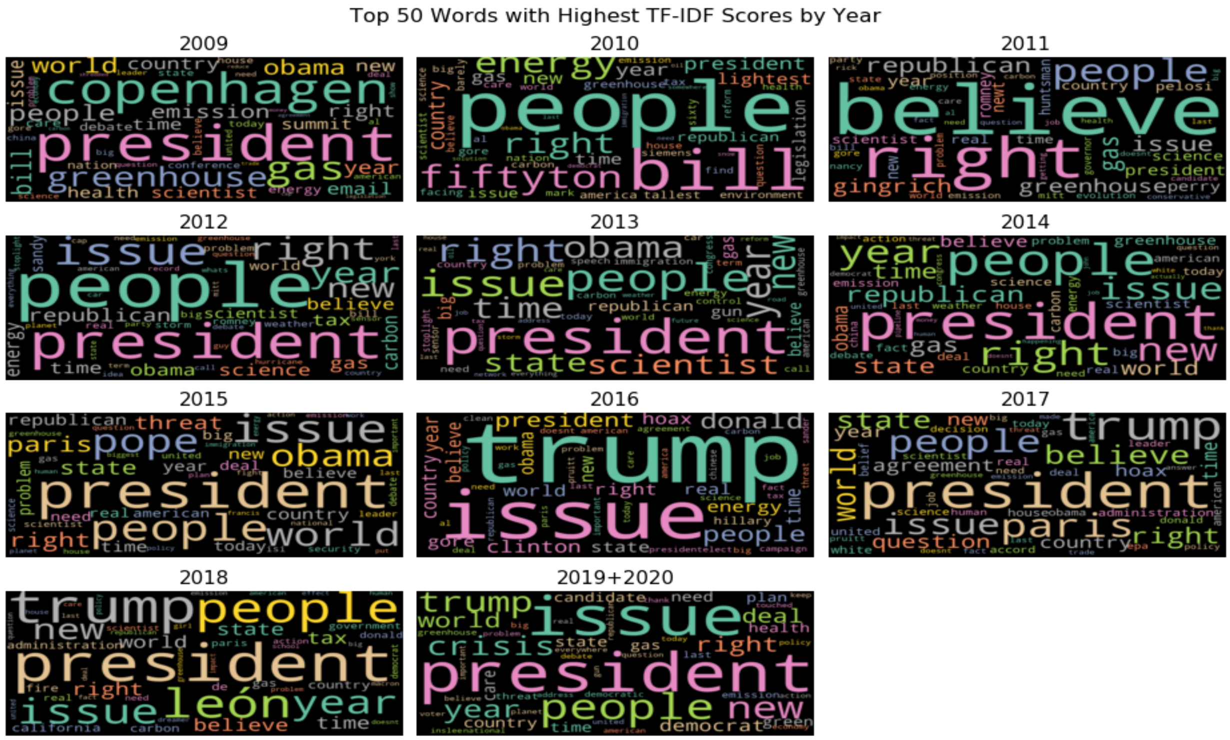


Figure 9: Top 50 Words with Highest TF-IDF Scores by Year Bar Plots

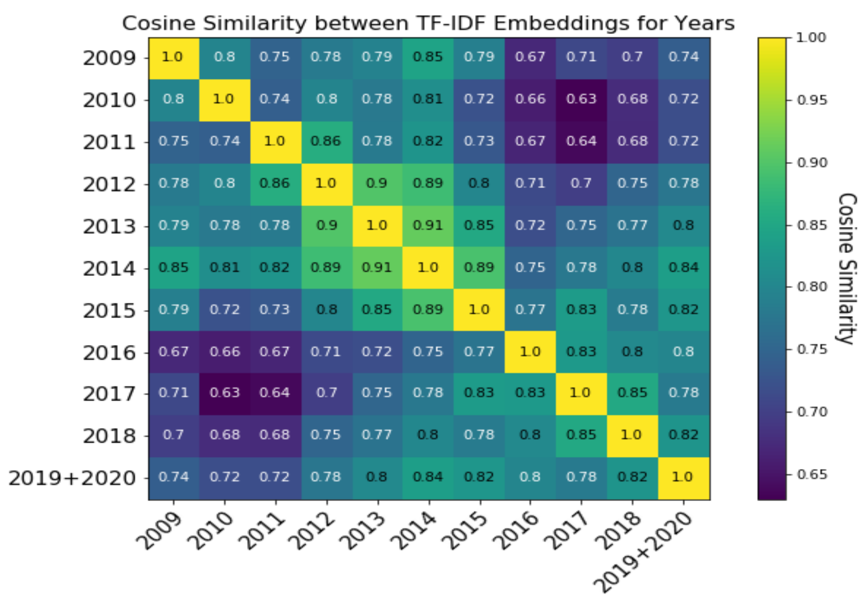


Figure 10: Cosine Similarity Heatmap between Year Documents TF-IDF Word Embeddings

Relatedly, for the 2016 presidential election, we observe the republican and democratic presidential candidates as important terms, namely, Donald Trump and Hillary Clinton for the 2016 document. Additionally, we see *presidentelect* as an important term for 2016 as Donald Trump was elected president in the 2016 election. Other political figures important to 2016 document are Bernie Sanders and Scott Pruitt in 2016 and 2017. Al Gore is important in 2009-2011 and 2016. Nancy Pelosi is important in 2011. Kevin De León and French President Emmanuel Macron are important in 2018.

Events or words regarding climate change which are not carbon emissions or greenhouse gases emerge at the year level, such as the Copenhagen Climate Change Conference (word: *Copenhagen*), the Climatic Research Unit email controversy (word: *email*), *reduce* for 2009; *fiftyton, environment, oil, tax, siemens* in 2010; *fact* in 2011; *real, record, planet, tax* in 2012; *real, fact* in 2014; Pope Francis declaring that the Catholic Church views climate change as a moral issue, the Paris Agreement, *planet, real* in 2015; *hoax, Paris Agreement, clean, tax* in 2016; *hoax, Paris Agreement, real, EPA* in 2017; *tax, impact, real, fact, effect* in 2018; *crisis, green new deal, real, planet* in 2019+2020.

Words regarding weather emerge as important at this level including *snow* in 2009 and 2010, *hurricane, sandy* (assumed to be referring to Hurricane Sandy which occured in 2012), *weather, storm* in 2012, *weather* and *storm* in 2013, *weather* in 2014, and *fire* in 2018.

From the computed cosine similarity in Figure 10, we observe relatively high cosine similarity between most of the years (i.e. cosine similarity of 0.75 or higher). However, the TF-IDF word embeddings of early years of the dataset, 2009-2012, are most dissimilar to the latter years of the dataset, 2016-2019+2020, having the lowest cosine similarity scores between years ranging from 0.63 between 2010 and 2017 and 0.78 between 2012 and 2019+2020. We observe the lowest cosine similarity scores for 2009 are with 2016 (0.67), 2017 (0.71), 2018 (0.7), and 2019+2020 (0.74). 2010, 2011, and 2012 also share their lowest similarity scores with the referred latter years. This implies that the content of climate change coverage in the early years (2009-2012) differed most from the content of climate change coverage in the latter years (2016-2019+2020) and vice-versa than with the years inbetween (2013-2015). This dissimilarity tracks with the Barack Obama's first presidential term that's emcompassed in the years in the dataset (2009-2012) and Donald Trump's presidential term encompassed by the years present in the dataset (2016-2020). These time-adjacent presidents held opposing stances on climate change validity and policies.

6.1.6 By Network and Year Documents

Similar to the year level, we observe in Figures 19 & 20 events and topics in specific years emerge as important words for all of the networks, e.g. the Copenhagen Conference in 2009, but more interestingly, we observe that within the same year the words that are important to each of the networks can be very different. This becomes even more clear when plotting the set of the union of words which were most important to each of the networks found in Section 6.1.4 using the TF-IDF Scores found at the Network & Year level.

We plot these words by network and over time using the TF-IDF embeddings found in the by network and year level in Figures 21 & 22. Some words have spikes in importance for all of the networks at certain points in time such as *paris, agreement, trump, gore, green, hoax, china, crisis, candidate,* and *believe*. There are some words which are mostly important to one network more so than the others over most years. For Fox News these words include: *bill, doesnt, gore, green, left, money*, and *people*. For MSNBC these words include: *action, candidate, human, party, republican*, and *science*. There are also words in which there is a single spike or multiple spikes in importance for a specific network in time where a word's importance for the network is much higher at the spikes than for the other networks at the spikes. For CNN these words include: *administration, big, china, crisis, debate, democrat, doesnt, emission, energy, hoax, important, issue, leader, need, paris, president, united, work, white*, and *world*. For Fox News these words include: *administration, american, believe, carbon, care, deal, democrat, government, job, last, new, obama, tax, threat*, and *time*. For MSNBC these words include: *action, america, believe, big, care, country, debate, fact, health, issue, last, need, people, policy, problem, question, real, state, time*, and *today*.

Where the networks do not overlap in their most important words include words which revolve around the dire nature of climate change, such as the words *crisis, important* for CNN and MSNBC; *health care* as another important policy topic that is highly polarized by partisanship for CNN & MSNBC; *threat* for Fox News but as we see in the Network and Year document level, this word is most important to Fox News in 2015 and appears to be in regards to terrorism rather than climate change as other important words for Fox News in 2015 include *terrorism, ISIS*, and *security*; climate policy and plans *Paris Agreement* for CNN and *policy* for both CNN & MSNBC, and *action* for MSNBC; the economy and climate-economic policies, *economy* for CNN & Fox News and *tax, money, green* for Fox News; regards to Donald Trump and climate change denial *leader, hoax, white, administration, china, donald* for CNN; references to partisanship and debates: *democrat/democrats, debate* for Fox News and MSNBC, *left* for Fox News, and *party, candidate* for MSNBC; *Al Gore*, a prominent politician who advocated for climate action throughout the years of this data for Fox News.

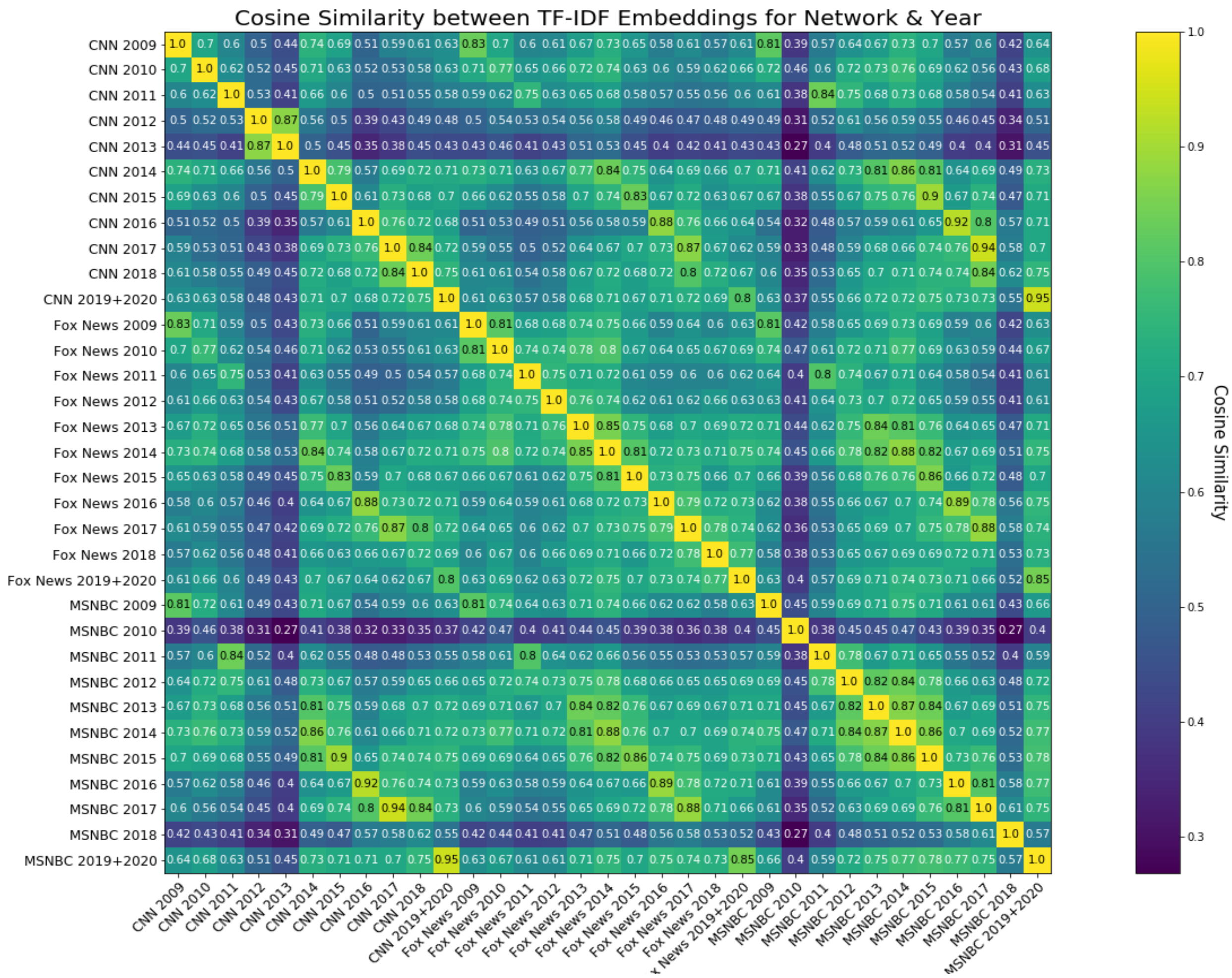


Figure 11: Cosine Similarity Heatmap between Network & Year Documents TF-IDF Word Embeddings

In the cosine similarity heatmap of the Network and Year in Figure 11 there are some documents that are striking dissimilar to most other documents (seen as dark violet lines). These documents include CNN 2012, CNN 2013, MSNBC 2010, and MSNBC 2018. The important words for the respective networks in these years appear to have correspondence to the top words which define the anomalous topics mentioned for each of these networks in 6.2, namely topic 7 for CNN in 2012 & 2013 and topic 12 for MSNBC in 2010.

We compare pairwise, the cosine similarites between the networks over time, i.e. pairs of networks in the same year. Excluding the documents mentioned above, generally networks in the same year have high cosine similarity (this is seen by the diagonal light green patterns in the heat map). Between CNN and Fox News, we observe their similarity values are highest in 2016 and 2017 at 0.88 and 0.87 respectively and they have the highest cosine similarities among the pairwise comparisons for the years 2009, 2010, and 2018. Between Fox News and MSNBC, we observe their similarity values are highest in 2014, 2016, and 2017 at 0.88, 0.89, and 0.88 respectively and they have the highest cosine similarities among the pairwise comparisons for the years 2012, 2013, 2014. When comparing the similarity between the liberal leaning networks, CNN and MSNBC, the cosine similarity between them increases with time with the exception of the dissimilar documents referenced in the paragraph above, having highest cosine similarity in 2019 + 2020 at 0.95. CNN and MSNBC have the highest cosine similarities among the pairwise comparisons for the years 2011, 2015, 2016, 2017, and 2019+2020.

We thus find that there were some years where a specific network was strikingly dissimilar with the networks in that year and with networks in most other years (including itself). We observe this occur twice for CNN (2012 & 2013) as well as MSNBC (2010 & 2018), but not at all for Fox News. Additionally, excluding those years of large dissimilarity, the networks are generally pairwise similar in content in the same year, and the liberal-leaning networks appear to be increasing in similarity in the content of their climate change mentions over time.

6.2 Topic Analysis

In an effort to better quantify the results discussed above, we turn to topic analysis. While an exploratory treatment of key words related to the questions of interest is certainly important, in this section we consider another way to identify important words: topic analysis. While we initially considered using Latent Dirichlet Allocation (LDA) [Blei et al., 2003] to conduct the topic modelling, we chose instead to use the simpler, faster Nonnegative Matrix Factorization (NMF) technique, introduced by [Lee and Seung, 1999] and applied to topic modelling in [Pauca et al., 2004]. The motivation for this change is primarily one of practicality given the faster optimization of the latter model and our limited computational resources.

The intuition behind NMF models is simple and extends beyond topic modeling. Given a matrix $M \in \mathbb{R}^{m \times n}$, and an integer $k \ll \min(m, n)$, we wish to find two matrices $W \in \mathbb{R}^{m \times k}$ and $H \in \mathbb{R}^{k \times n}$ such that $M = WH$; this process amounts, essentially, to low-rank projection onto a basis of k basis topics that would form the feature space if this model were considered as a simple linear auto-encoder. If M is composed of nonnegative entries only (i.e., M is a *nonnegative matrix*), such as in a document-term matrix, we can restrict W, H to be nonnegative as well. Thus we wish to approximate M as well as possible such that W, H are of the correct size and nonnegative. While the objective can be chosen to be a variant of KL divergence, for textual topic mining, we use the Frobenius norm to measure quality of approximation [Pauca et al., 2004]. The nonnegativity constraints both encourage sparsity and are meaningful: if M is a document-term matrix, it is unclear how to interpret a negative value in a topic-document or term-topic matrix. While it is theoretically possible to conduct a topic analysis using the original term-document matrix, it is more common to use the matrix normalized by the TF-IDF scores described previously.

With the methodology described, the only remaining task is to pick the number of features k. In order to do this, we turn to coherence scores, which essentially measure how internally consistent the topics are. More precisely, the coherence of a topic is intended to measure whether the topic is a "natural" class and so the success of a quantification thereof can be measured in its correlation with human perception. Given a coherence measure, we can apply cross-validation to estimate coherence scores for topic models with different numbers of topics and select the optimal k. While there are a number of competing coherence scores including UMASS, UCI, and NPMI (all described in [Röder et al., 2015]), empirically the measure C_v has been found to correlate best with human perception and so we use this [Röder et al., 2015]. Using GenSim [Řehůřek and Sojka, 2010], for $5 \leq k \leq 16$, we train an NMF topic model and evaluate the coherences, with the results in Figure 12:

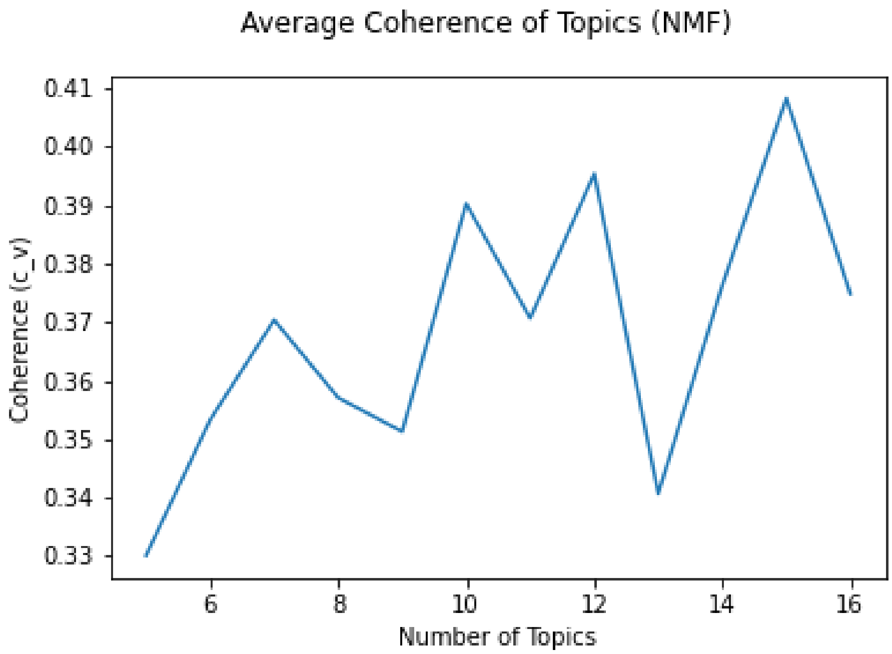


Figure 12: Plot of C_v against number of topics; evaluated on the corpus of aggregate snippets across all channels with each document corresponding to a month.

We see that the optimal number of topics is $k = 15$. While the coherence is not all that high, we consider this sufficient for two reasons. First, we suspect that the coherence is drawn down by the aggregation of unlike, short documents (note that the snippets are just that: cut sections of a more cohesive whole). Second, the topic analysis is primarily accomplished to draw out the relevant topic clusters for an analysis of how they change over time rather than further textual analysis and so low coherence is reasonable.

Having determined the correct number of topics, we train a topic model for each topic. The results can be found in Figure 13.

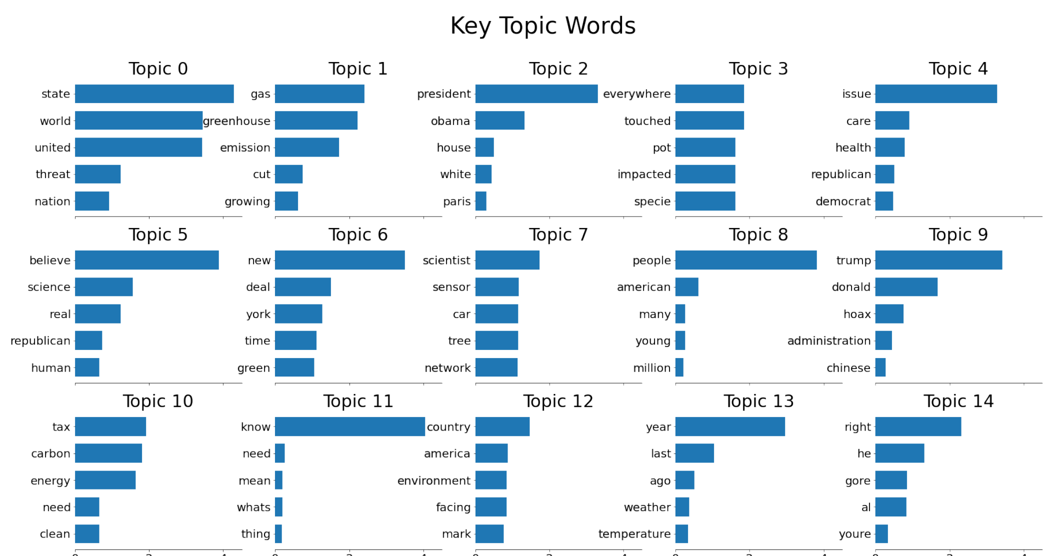


Figure 13: Top 5 words associated to each topic, as measured by their value in the topic basis.

We thus take these topics (and in particular their top five words) as relevant markers for the subjects of climate change discussion on these three networks. We are now ready to consider how mentions of these topics vary over time. As a crude proxy of mentions, we count the number of times any one of the top five words for each topic appears in a network's coverage for each month, then normalize by the total number of words in the document. To illustrate, we provide a plot of how frequently the 15 topics are mentioned each month on MSNBC in Figure 14.

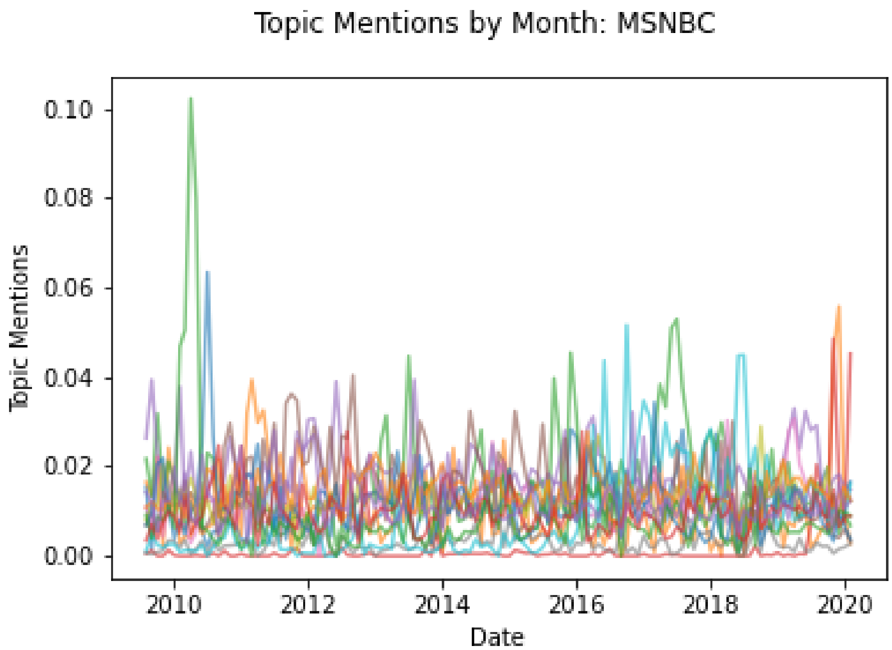
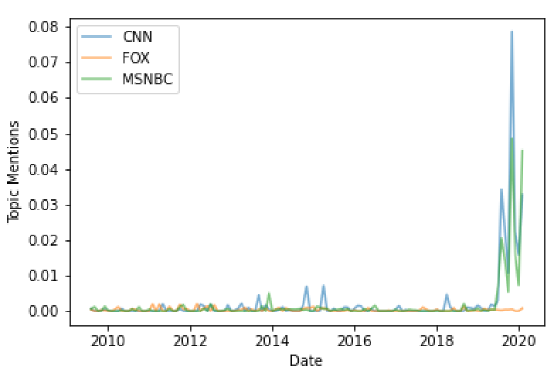
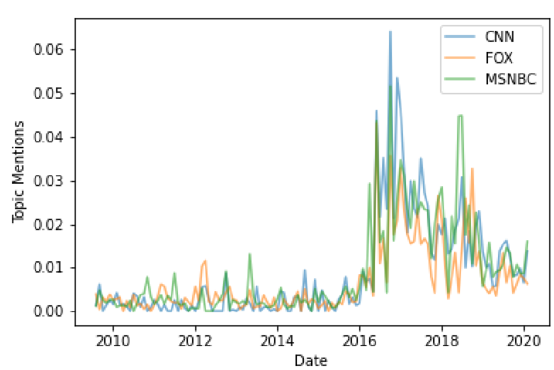


Figure 14: Monthly mentions of each topic on MSNBC, normalized by total word count. Each coloured line corresponds to a topic.

Similar plots for the other networks can be found in Appendix C.3. We are more interested, however, in how the topics evolve over time and in particular if each networks's coverage changes over time. To do this, we fix a topic and plot the frequency over time for each network. As an example, we do this for topics 3 and 9 in Figure 15.



(a) Topic 3



(b) Topic 9

Figure 15: Frequency of Monthly topic mentions for topics 3 and 9.

Plots for the other topics can be found in Appendix C.4. Topic 3 (characterized by the words "everywhere," "touched," "pot," "impacted," and "specie") and Topic 9 (characterized by the words "trump," "donald," "hoax," "administration," and "chinese") are interesting because they only appear to be commonly discussed sometime after the 2016 presidential election. In order to see if other topics satisfy this trend we employ a very simple model. In particular, we suppose that the frequency of mention for each topic in each epoch (before Trump and after Trump) is modeled by an independent Gaussian with mean μ_{Ti} where $0 \leq T \leq 14$ is the topic and $i \in \{0, 1\}$ is the epoch. We define the pre- verse post-Trump breakpoint to be January 2017, when he was inaugurated. Other reasonable choices include November 2016 when he was elected, or perhaps even earlier when he entered the "national conversation." Due to a desire to reduce researcher degrees of freedom, we considered the inauguration to be the clearest Schelling point and chose thus. While the independence is not fully justified due to autocorrelation, the plots make this assumption at least somewhat reasonable. Except for the anomalies of topic 7 on CNN and topic 12 on MSNBC (see Appendix C.4 for plots), the Gaussianity and independence assumptions appear *prima facie* reasonable. We thus conduct a t-test for change of mean for each of these channel-topic pairs and then apply the Holm-Bonferroni procedure [Holm, 1979] to correct for multiple hypotheses. At the $\alpha = .05$ level, we are left with the following nine significant changes of mean:

Channel	Topic	Adjusted p-Value
CNN	9	5.438×10^{-9}
MSNBC	9	3.579×10^{-5}
FOX	9	4.705×10^{-5}
FOX	4	.0093
CNN	2	.0097
FOX	8	.0134
CNN	3	.0159
MSNBC	3	.0210
MSNBC	5	.0270

Note that the next smallest Adjusted p-values after the ones listed above are .0690 and .1422, so there definitely appears to be a significance gap under this model between the significant hypotheses and the others.

Considering the plot above and the key words, it is unsurprising that Topic 9 is highly significant in mean change pre- and post-administration change. Topic 3 appears to be associated with the economy and so it is also unsurprising that such discussion would change in frequency between presidential administrations, although from the plot it appears that such a change is actually a more recent phenomenon. The other significant topics are 2, which is related to presidential policy and the Paris accords, 4, which is related to governmental policy, 5, which is related to "believing science," and 8, which appears to be centred on discussion of demographic trends in climate opinion. An overview of the top words associated to each topic is in Figure 13 and a more complete list can be found in the project repository on github. We conclude our discussion of hypothesis testing by emphasizing that these tests are meant to be purely exploratory. Due to the scarcity of data, a holdout set of reasonable size was impractical and thus the hypotheses are constructed as functions of the data (in particular, the NMF was trained on the same set on which the tests were evaluated). A better analysis would use techniques along the lines of [Jewell et al., 2019], except conditioning on the results of the NMF model. This optimization and analysis problem is significantly beyond the scope of the current work.

We conclude our topic analysis on an exploratory note. In particular, we have noted that the Trump administration change does not significantly affect coverage of all topic-channel pairs. We might hope to find some other times that do affect coverage. To accomplish this, we use the ruptures package in python [Truong et al., 2020] for computations. More specifically, we use the same model as above where each channel-topic-epoch triple has a fixed mean with subGaussian noise additive noise at each month. The goal is now to find the epoch break point, which we do by least squares. The objective, then, for each channel-topic pair, (T, c) to minimize

$$\min_{\substack{B_{T,c} \\ \mu^1_{T,c}, \mu^2_{T,c}}} \sum_{t < B_{T,c}} (x_t^{T,c} - \mu^1_{T,c})^2 + \sum_{t > B_{T,c}} (x_t^{T,c} - \mu^2_{T,c})^2$$

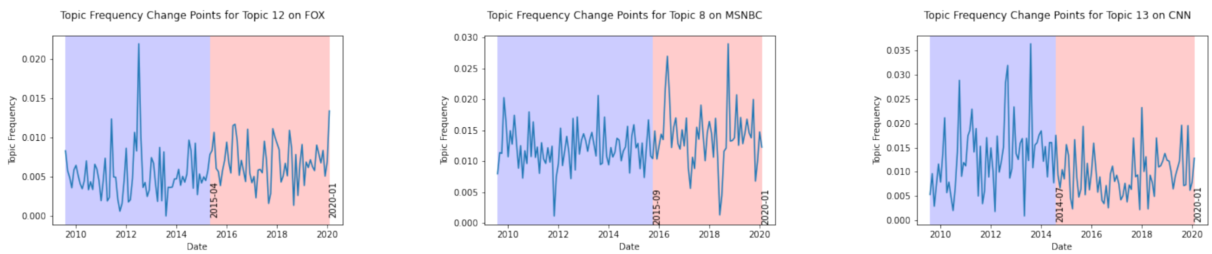


Figure 16: Detected breakpoints. We use blue and pink shading to denote the pre- and post-breakpoint epochs.

where $B_{T,c}$ is the breakpoint for the channel-topic pair and $\mu^i_{T,c}$ are the means for each epoch. We do not care about the means, but wish only to find the breakpoints. Again, we use the package ruptures to optimize this objective and plot three examples in Figure 16 Due to the large quantity, a complete list of the plots of the detected breakpoints for each topic-channel pair can be found in our online repository[6]. While space constraints prevent a full analysis of the changepoint detection, we note that many of the found changepoints occur in 2014-2015. Our hypothesis for this phenomenon is that a general change in tenor in climate-related coverage that was occuring then, possibly related to the buildup before the Paris Agreement toward the end of 2015.

7 Conclusion

Climate change television news media coverage frequency and content appears to be significantly driven by political events more so than environmental factors. The frequency of climate change mentions follow similar patterns by network, with clear influence of political events such as the 2009 UNCCC, 2015 Paris Agreement, and 2019 Democratic primary debates driving climate news coverage. We additionally see this reflected in the content of the climate mentions over time as words describing the political events occurring at the time tend to be the most important words for each of the networks in that specific year coupled with the tendency of different networks in the same year to have high similarity in their content. We see this reflected in the majority of 15 topics found in topic analysis, with significant changes in mean on some of the topics at the time of the Trump's inauguration.

In recent years, there appears to also be much more correlation with natural disaster declarations across all networks, perhaps indicating greater connection of climate change with extreme weather events, following the 2018 IPCC Report. This follows with public opinion surveys that show public opinion shifting to more Americans believing that climate change would harm the United States.

The content of climate mentions in the years in the dataset corresponding to the Trump presidency (2016-2020) are most dissimilar to years in the dataset corresponding to Obama's first presidential term (2009-2012) in which the political climate on climate change was vastly different in those respective eras. In the years 2010 & 2018, the content of MSNBC differed greatly from the other networks in those years and with other networks and itself in other years. This similarly occurred for CNN in 2012 & 2013. With the exception of these years, the similarity of content of climate mentions between CNN and MSNBC, the liberal-leaning networks, has been increasing. While there is variation between networks that tracks with the partisan alignment, overall the frequency and content between networks correlates quite strongly.

The fact that climate change is predominantly mentioned in the context of political events, and with reference to politicians and with other polarizing issues such as healthcare rather than with connection to weather or impacts of climate change seems significant in shaping Americans views of climate change as a politically polarizing issue rather than a shared obstacle for humanity. This follows other work showing that political actors have replaced scientific actors over time leading to further polarization. Further research investigating climate change coverage by popular news networks in other countries' could illustrate the ways in which climate change is framed in the news around the world and how that corresponds to political polarization of the issue.

[6] https://github.com/dyllew/ids.131-fp/tree/main/plots/ChangePointDetection

References

Climate Change in the American Mind: National Survey Data on Public Opinion (2008-2018) . *Yale Program on Climate Change Communication (YPCCC) George Mason University Center for Climate Change Communication (Mason 4C).*, 2020. doi: 10.17605/OSF.IO/JW79P.

Liisa Antilla. Self-censorship and science: a geographical review of media coverage of climate tipping points. *Public Understanding of Science*, 19(2):240–256, 2008. doi: 10.1177/0963662508094099.

Leiserowitz A. Roser-Renouf-C. Rosenthal S. A. Kotcher J. E. Marlon J. R. Lyon E. Goldberg M. H. Maibach E. W. Ballew, M. T. Climate change in the american mind: Data, tools, and trends. *Environment: Science and Policy for Sustainable Development*, 61(3):4–18, 2020. doi: 10.1080/00139157.2019.1589300.

Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

Pew Research Center. 86% of Americans Get News Online from Smartphone, Computer or Tablet, 2020. URL https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/.

Mrinal Ghosh Yong Liu Chen, Yubo and Liang Zhao. Media coverage of climate change and sustainable product consumption: Evidence from the hybrid vehicle market. *Journal of Marketing Research*, 56(6): 995–1011, 2019. doi: 10.1177/0022243719865898.

Kumar Sathish Dahal, Biraj and Zhenlong Li. Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, 9(1):1–20, 2019. doi: https://doi.org/10.1007/s13278-019-0568-8.

Edward W. Maibach Connie Roser-Renouf Feldman, Lauren and Anthony Leiserowitz. Climate on cable: The nature and impact of global warming coverage on fox news, cnn, and msnbc. *The International Journal of Press/Politics*, 17(1):3–31, 2011. doi: 10.1177/%2F1940161211425410.

FEMA. US Natural Disaster Declarations, 2021. URL https://www.kaggle.com/headsortails/us-natural-disaster-declarations.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

Sean Jewell, Paul Fearnhead, and Daniela Witten. Testing for a Change in Mean after Changepoint Detection. *arXiv preprint arXiv*, 2019. doi: 1910.04291.

Ann Schwebel Shoshana Jost, François Dale. fix. *Environmental Science and Policy*, 96(1):27–36, 2019. doi: https://doi.org/10.1016/j.envsci.2019.02.007.

Ballew M. Lacroix K. Leiserowitz-A. Kim, L. and J. Marlon. How does the american public perceive climate disasters? 2020.

Neiman Lab. A Typical Big News Story in 2018 Lasted about 7 Days (until We Moved on to the next Crisis), 2019. URL https://www.niemanlab.org/2019/01/a-typical-big-news-story-in-2018-lasted-about-7-days-until-we-moved-on-to-the-next-crisis.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

Maria L. Loureiro and Maria Alló. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the u.k. and spain. *Energy Policy*, 143(1):111–90, 2020. doi: https://doi.org/10.1016/j.enpol.2020.111490.

NOAA. Billion Dollar Weather and Climate Disasters, 2020. URL https://www.ncdc.noaa.gov/billions/time-series.

V Paul Pauca, Farial Shahnaz, Michael W Berry, and Robert J Plemmons. Text mining using non-negative matrix factorizations. In *Proceedings of the 2004 SIAM International Conference on Data Mining*, pages 452–456. SIAM, 2004.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Anand Rajaraman and Jeffrey David Ullman. *Data Mining*, page 1–17. Cambridge University Press, 2011. doi: 10.1017/CBO9781139058452.002.

Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. http://is.muni.cz/publication/884893/en.

Nielsen Reports. Top-Rated Channels: TV Network Winners Losers - Variety, 2019. URL https://variety.com/2020/tv/news/network-ratings-2020-top-channels-fox-news-cnn-msnbc-cbs-1234866801/.

Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408, 2015.

Henríqueza Pablo A. Mascareño-Aldo Ruz, Gonzalo A. Sentiment analysis of twitter data during critical events through bayesian networks classifiers. *Future Generation Computer Systems*, Volume 106(1):92–104, 2020. doi: https://doi.org/10.1016/j.enpol.2020.111490.

Statmodels. STL Decomposition — Forecasting: Principles and Practice, 2021a. URL https://Otexts.com/fpp2/.

Statmodels. Seasonal-Trend Decomposition Using LOESS (STL) — Statsmodels, 2021b. URL https://www.statsmodels.org/stable/examples/notebooks/generated/stl_decomposition.html.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

Jeremy Zhang. Understanding Dynamic Time Warping - The Databricks Blog, 2019. URL databricks.com/blog/2019/04/30/understanding-dynamic-time-warping.html.

Jeremy Zhang. Dynamic Time Warping, 2020. URL https://towardsdatascience.com/dynamic-time-warping-3933f25fcdd.

A Formulas

A.1 TF-IDF Calculation

$$\text{Term Frequency} = tf(w_i, d_j) = \frac{n_{w_i,d_j}}{\sum_k n_{w_k,d_j}}$$

$$\text{Inverse Document Frequency}^* = idf(w_i, N) = \log(\frac{N}{df_{w_i}})$$

$$\text{TF-IDF}(w_i, d_j, N) = tf(w_i, d_j) * idf(w_i, d_j, N)$$

w_i is a word in the corpus vocabulary.
d_j is a document in the corpus.
n_{w_i,d_j} is the count of the occurrence of w_i in document d_j.
$\sum_k n_{w_k,d_j}$ is the total number of words in document d_j.
N is the number of documents in the corpus.
df_{w_i} is the number of documents in the corpus in which word w_i is present.

* As noted in 6.1, in our analysis we use a variation of IDF:

$$idf(w_i, N) = \log(\frac{N+1}{df_{w_i}+1})$$

This variation incorporates smoothing by the addition of 1 to the numerator and denominator of the argument to the logarithm. This has the effect of adding a document to our corpus which has every term in the corpus vocabulary contained within it. We use this variation so that words which appear in every document in the corpus are not ignored [Pedregosa et al., 2011]. Though such words are frequent across the corpus, they do bring semantic meaning to our analysis and thus are not noise, which is why we aim to preserve them.

B Climate TV News Corpus-Specific Stopwords

The corpus-specific stopwords we found are:

climate, change, global, warming, im, say, he, thats, dont, thing, like, going, get, got, one, make, talk, would, said, well, could, back, want, lot, let, also, see, way, look, look, much, come, go, saying, come, some, say, take, something, there, theres, that, really, talking, think, theyre, reporter, know, part, report, many, even, next, whether, around, week, good, day, first, two, point, still, youre, coming, weve, may

C Other Plots

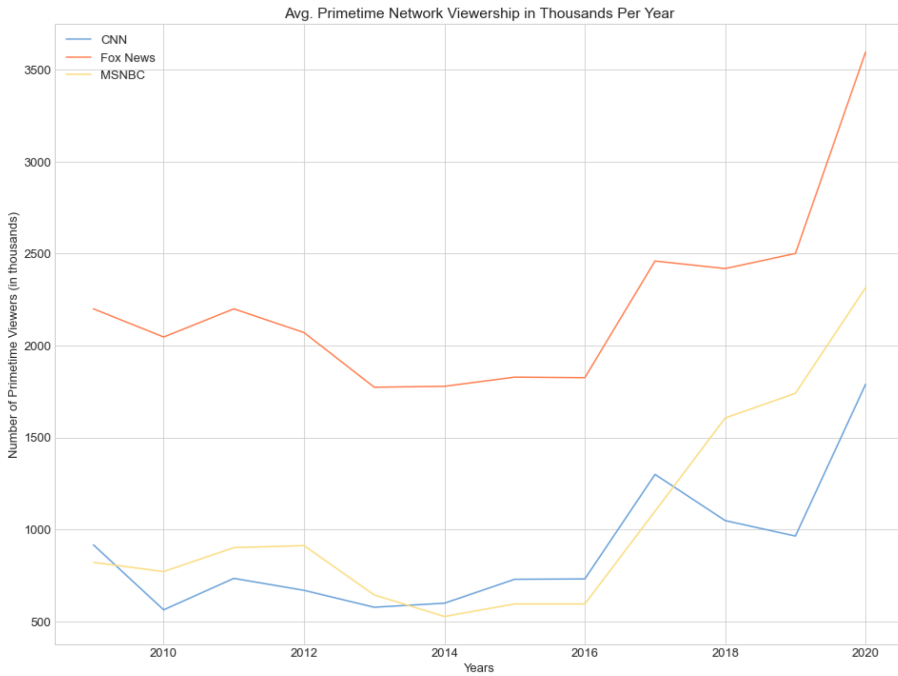


Figure 17: Yearly Average Primetime Network Viewership Over the Last Decade

Figure 18: 50 Most Frequent Named Entities by Network and Year

C.1 Top 50 Words with Highest TF-IDF Score by Network and Year

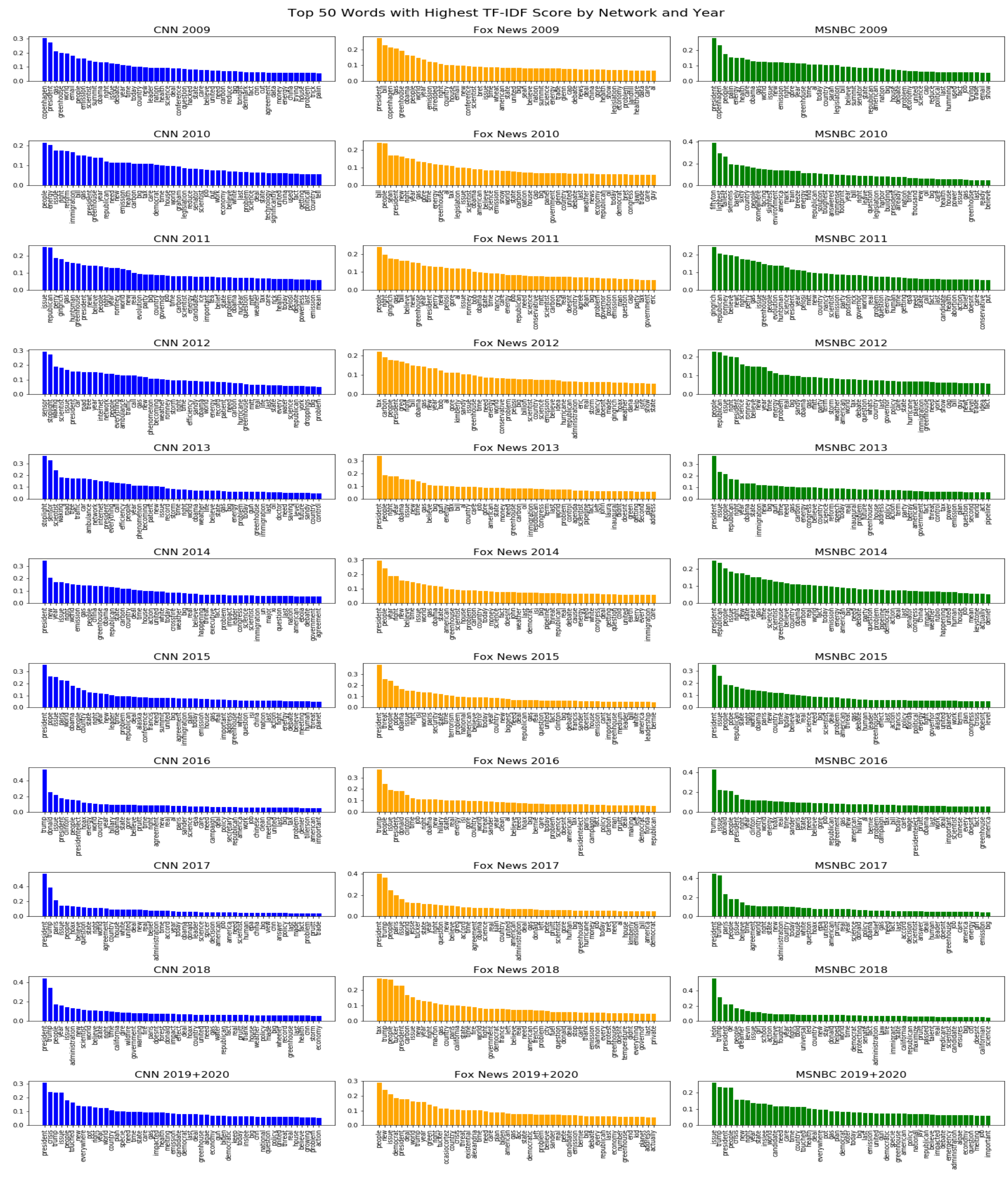


Figure 19: Top 50 Words with Highest TF-IDF Scores by Network & Year Bar Plots

Figure 20: Top 50 Words with Highest TF-IDF Scores by Network & Year Word Clouds

C.2 TF-IDF Scores by Network over Time of Words with Highest TF-IDF Scores among the Networks

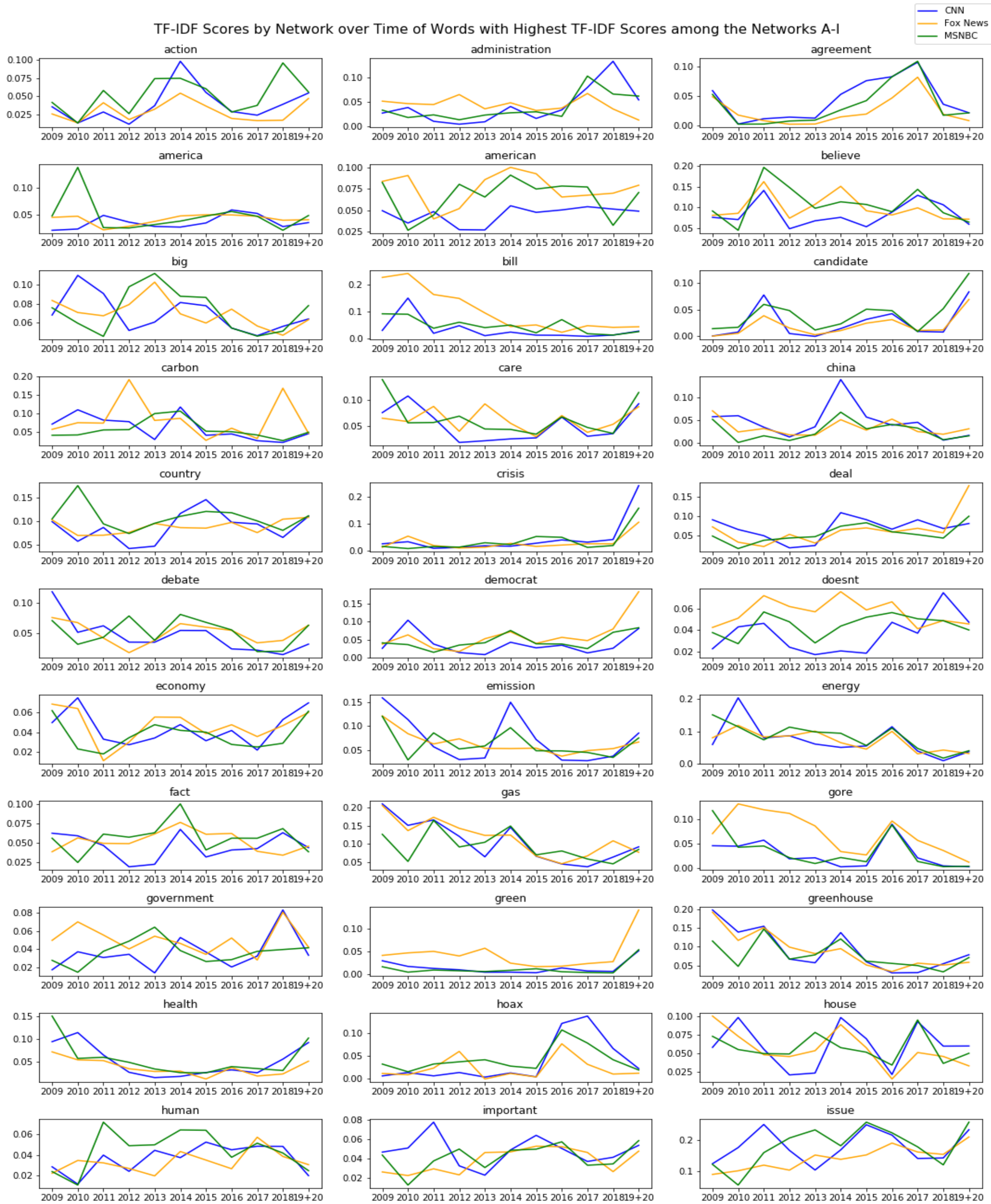


Figure 21: TF-IDF Scores by Network over Time of Words with Highest TF-IDF Scores among the Networks J-Z

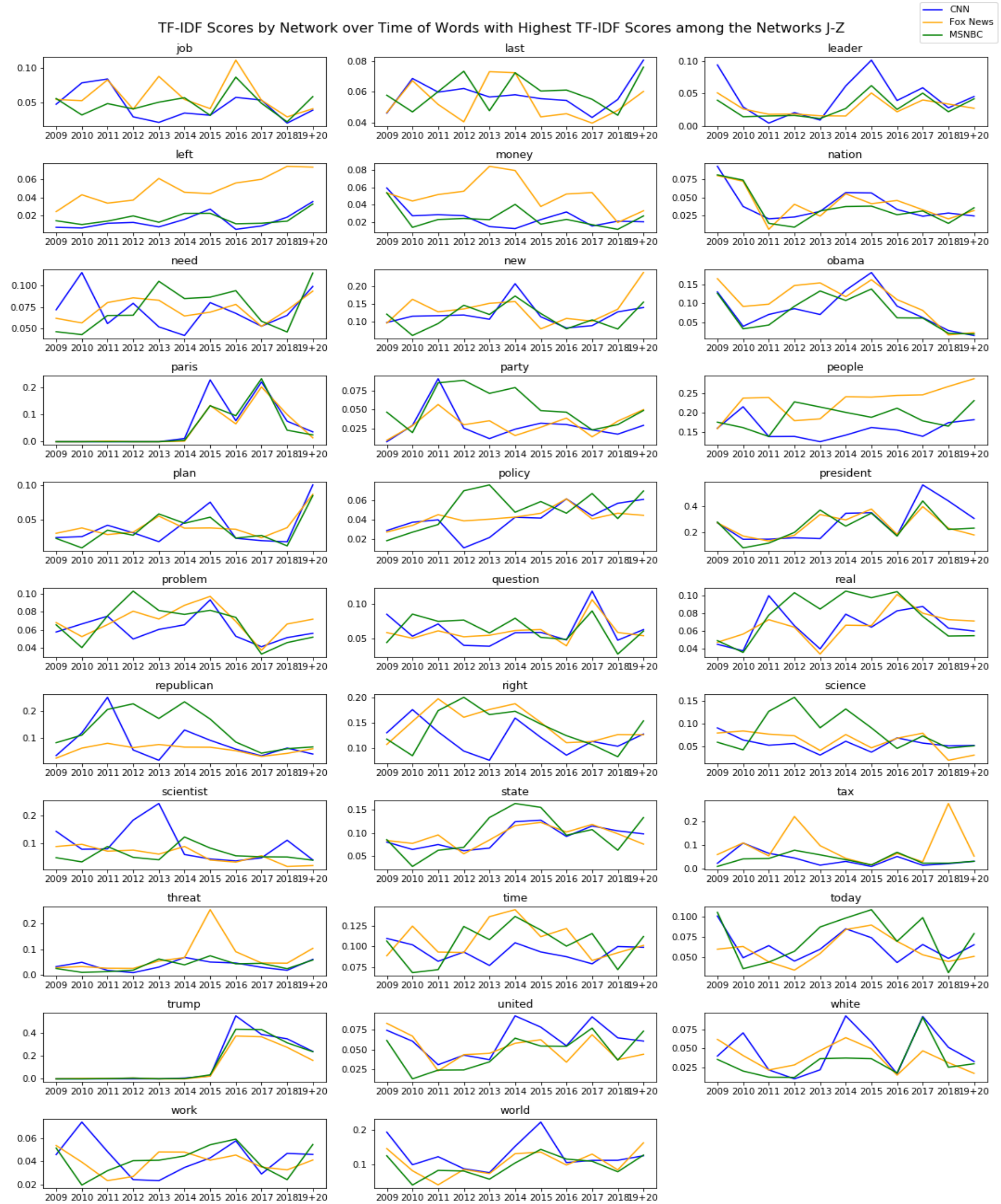


Figure 22: TF-IDF Scores by Network over Time of Words with Highest TF-IDF Scores among the Networks J-Z

C.3 Monthly Topic Mentions by Channel

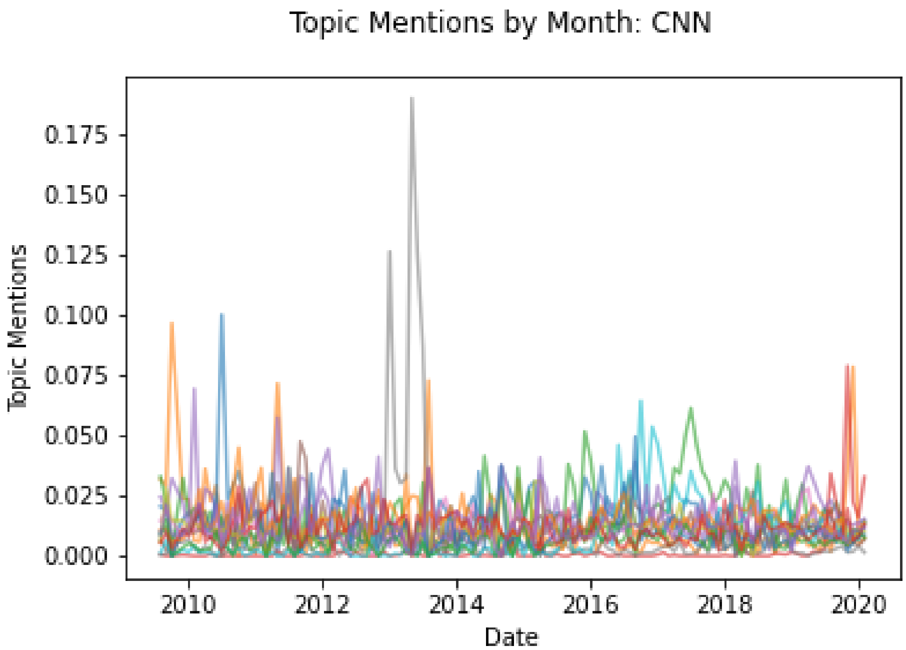


Figure 23: Frequency of topic mentions per month on CNN.

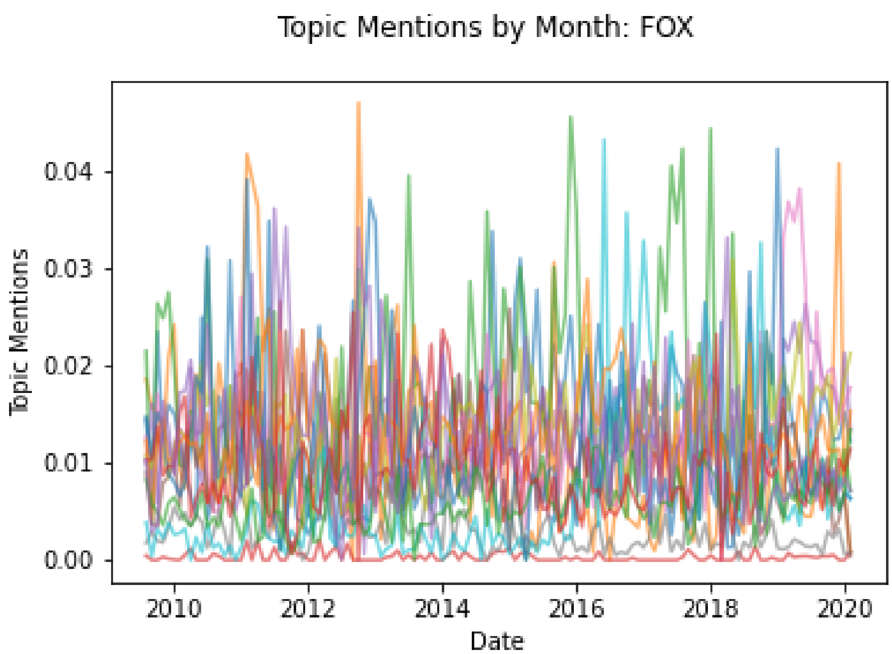


Figure 24: Frequency of topic mentions per month on FOX.

C.4 Monthly Topic Mentions by Topic

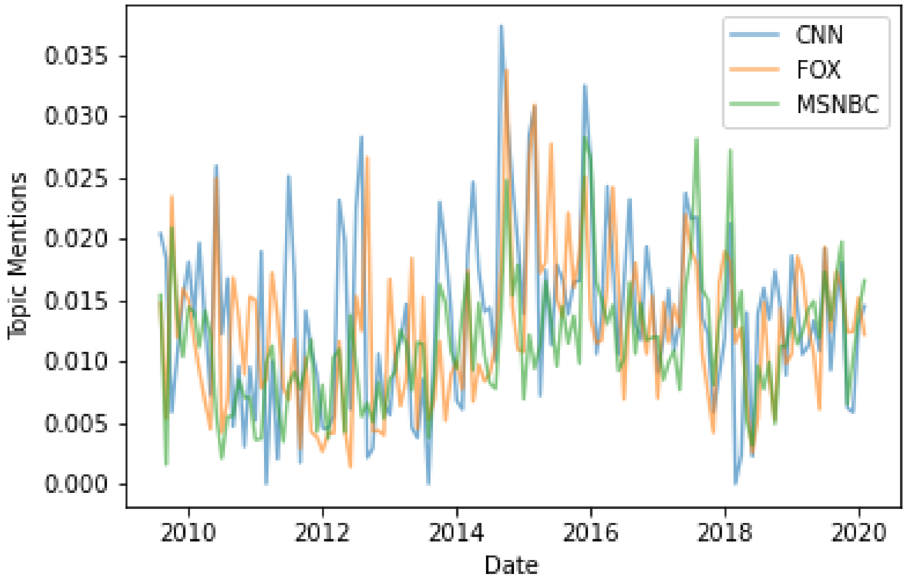


Figure 25: Frequency of topic mentions per month for topic 0.

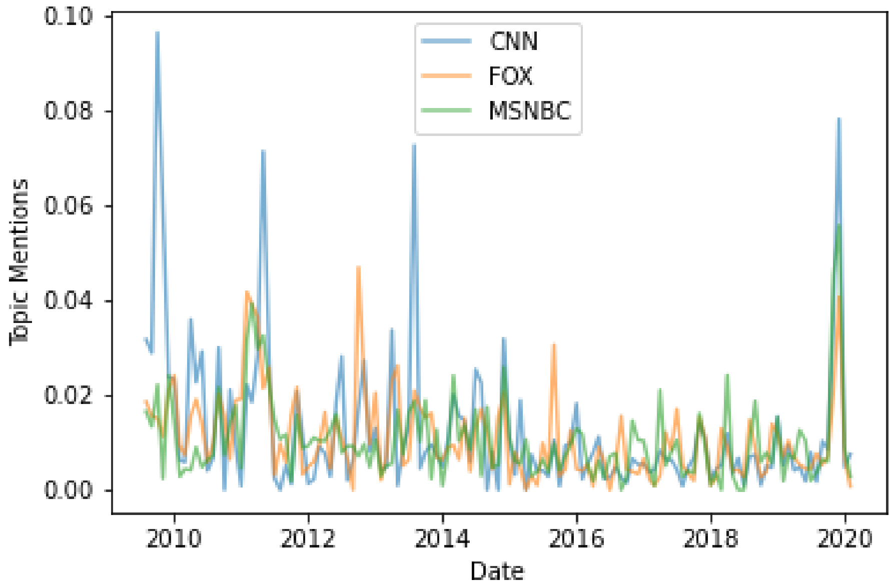


Figure 26: Frequency of topic mentions per month for topic 1.

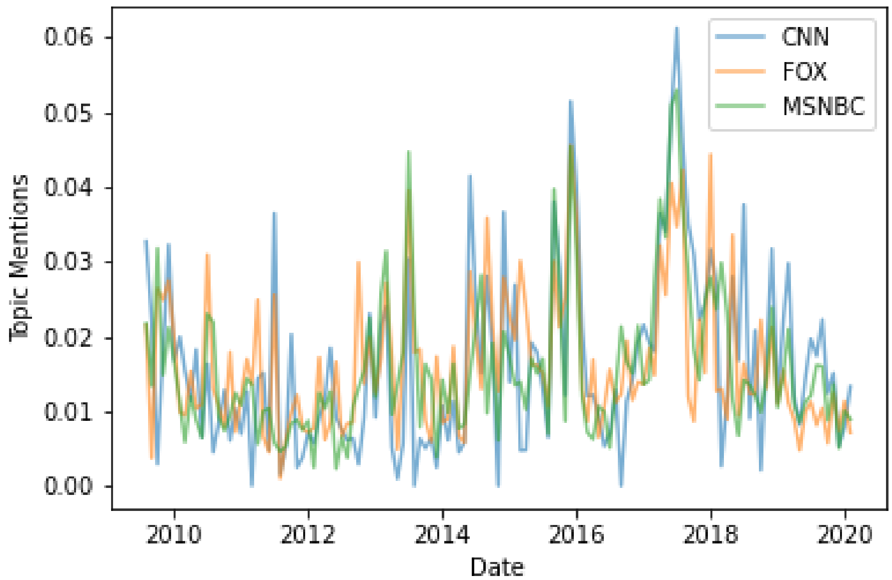


Figure 27: Frequency of topic mentions per month for topic 2.

Figure 28: Frequency of topic mentions per month for topic 4.

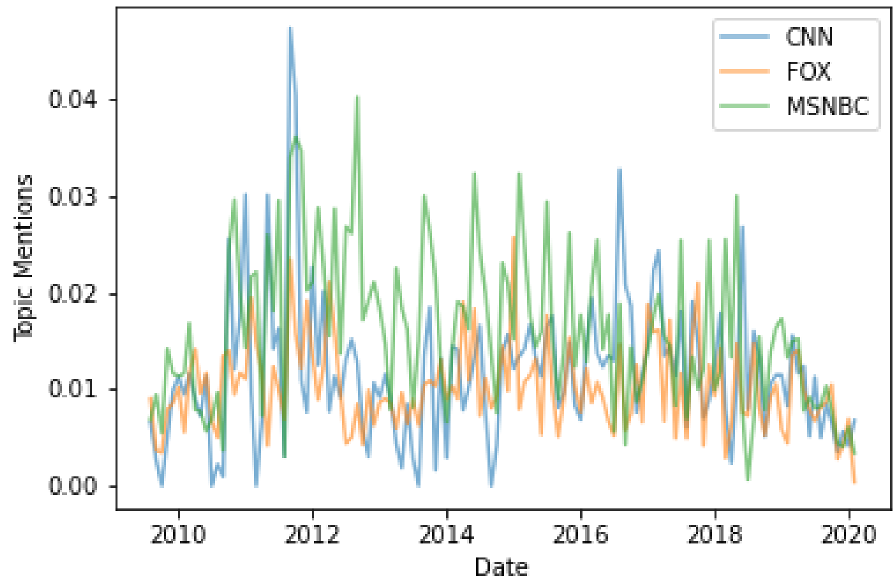


Figure 29: Frequency of topic mentions per month for topic 5.

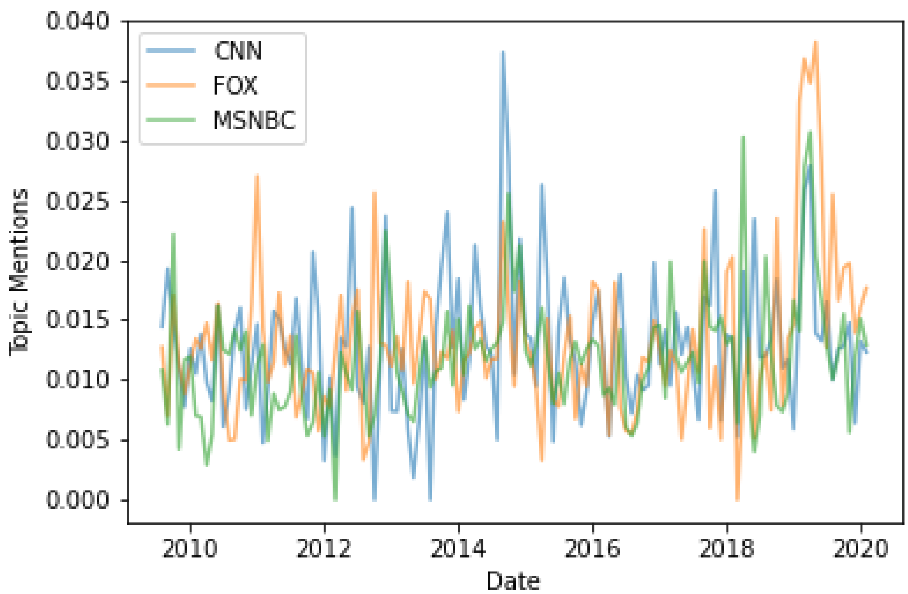


Figure 30: Frequency of topic mentions per month for topic 6.

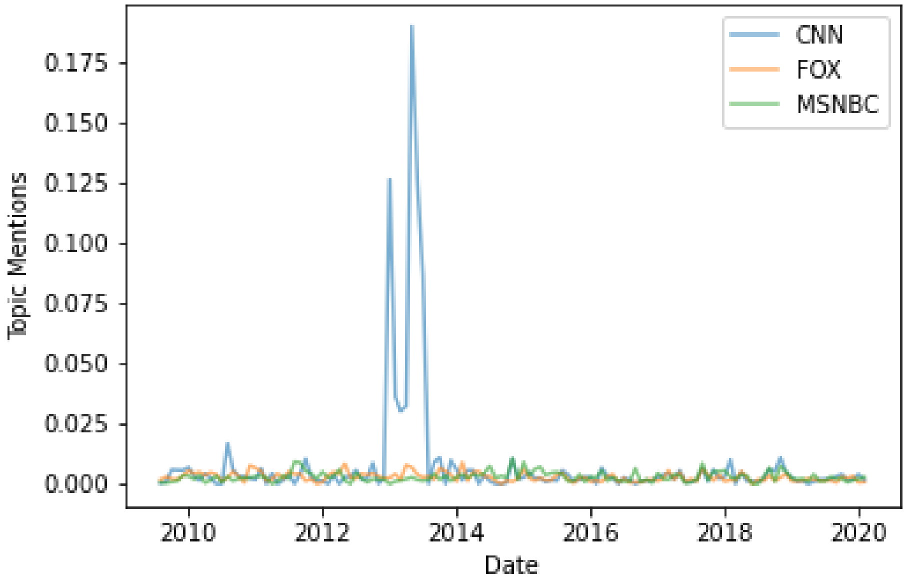


Figure 31: Frequency of topic mentions per month for topic 7.

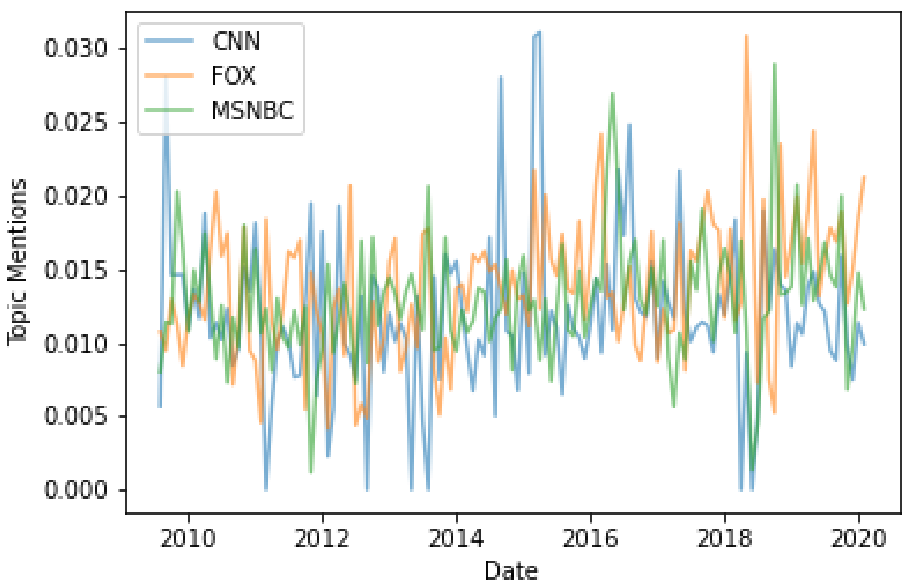


Figure 32: Frequency of topic mentions per month for topic 8.

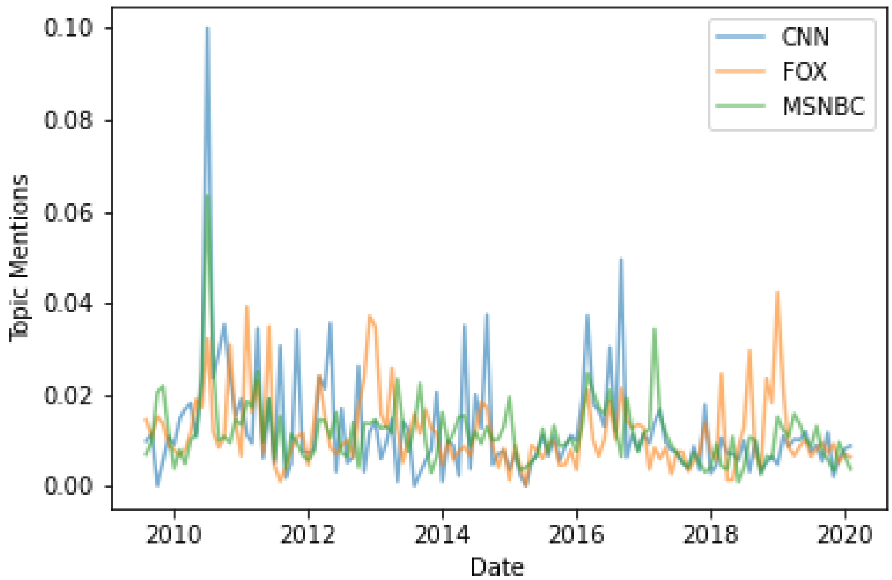


Figure 33: Frequency of topic mentions per month for topic 10.

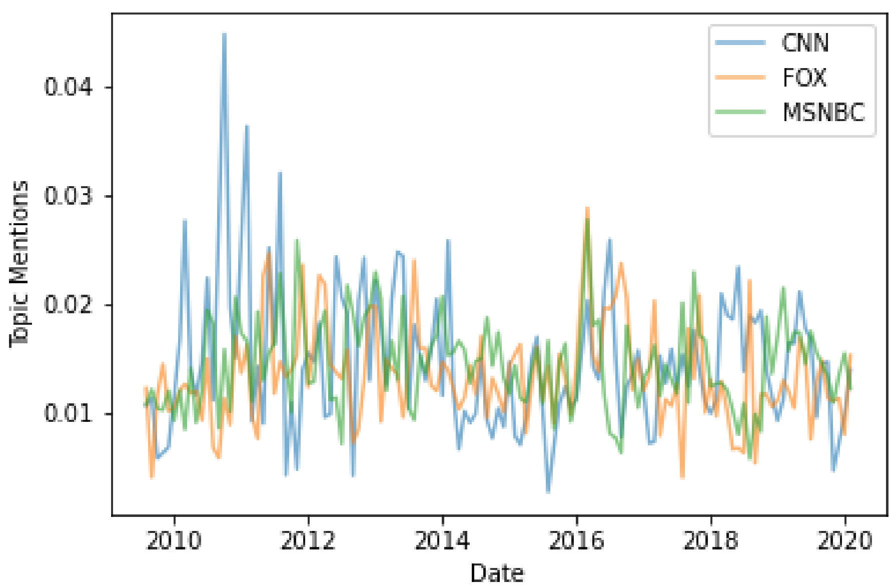


Figure 34: Frequency of topic mentions per month for topic 11.

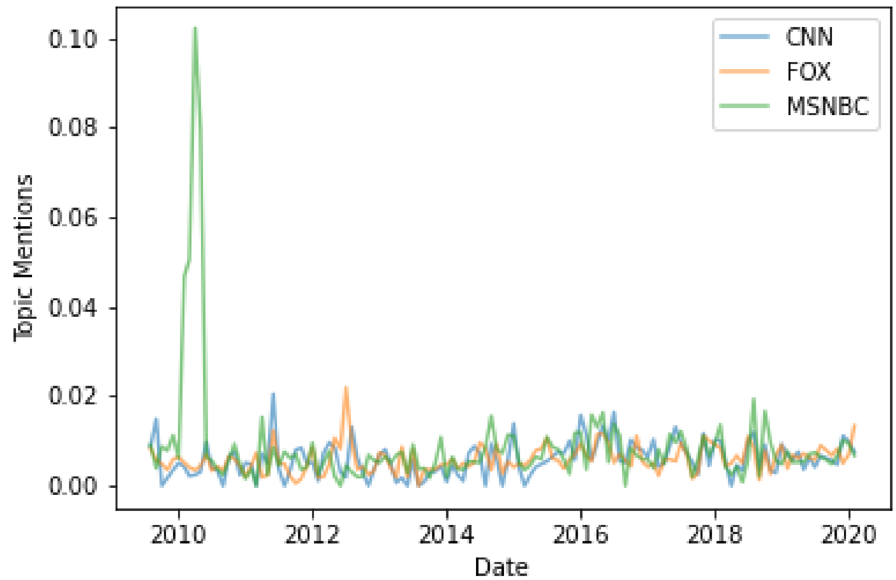


Figure 35: Frequency of topic mentions per month for topic 12.

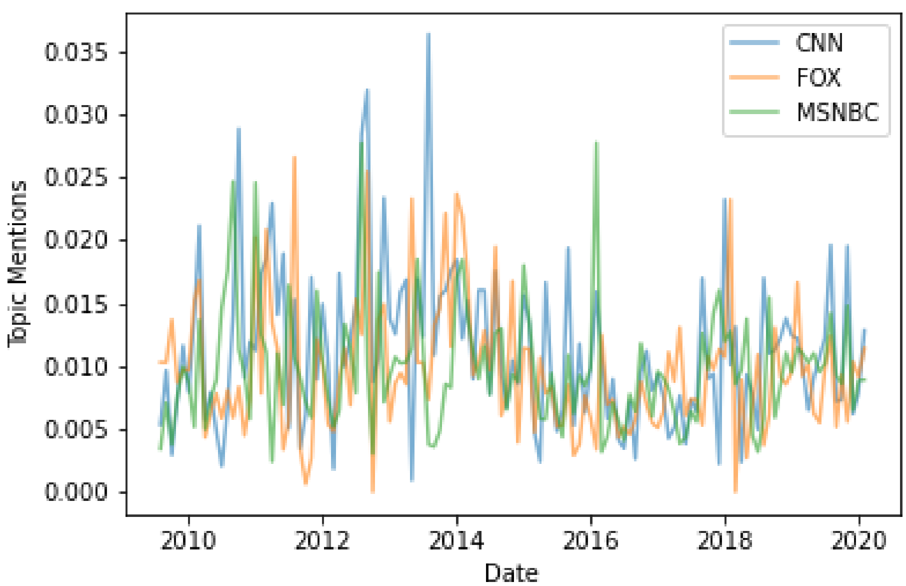


Figure 36: Frequency of topic mentions per month for topic 13.

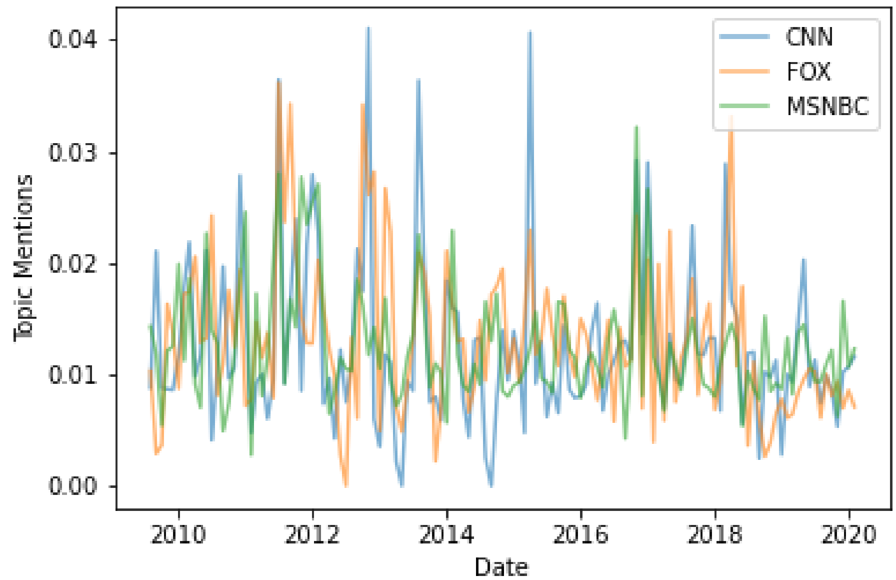


Figure 37: Frequency of topic mentions per month for topic 14.

C.5 Time Series Analysis Supplemental Figures: Daily News Network Mentions, Disaster Data by Disaster Type, and STL Decomposition

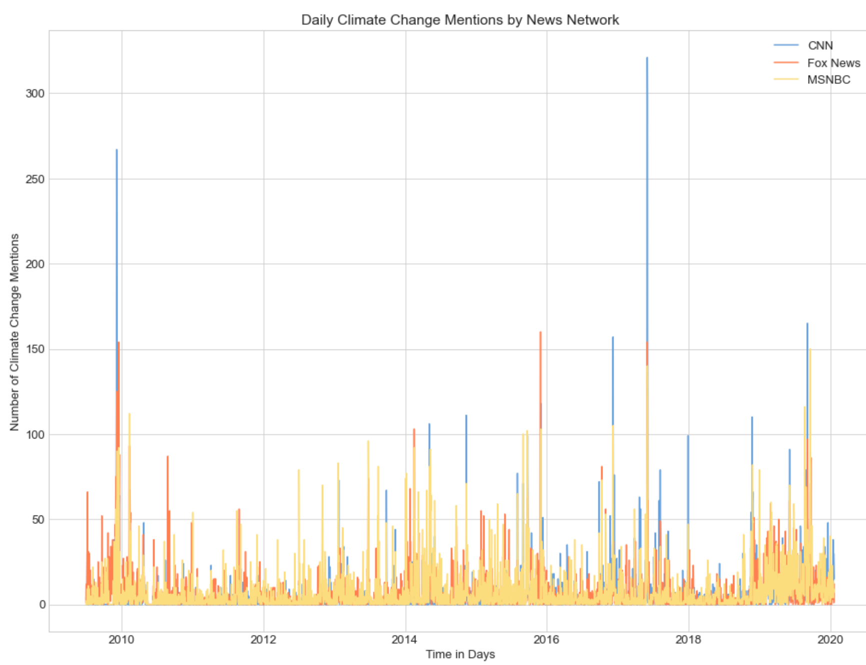


Figure 38: Climate Change Mentions By Day and Network

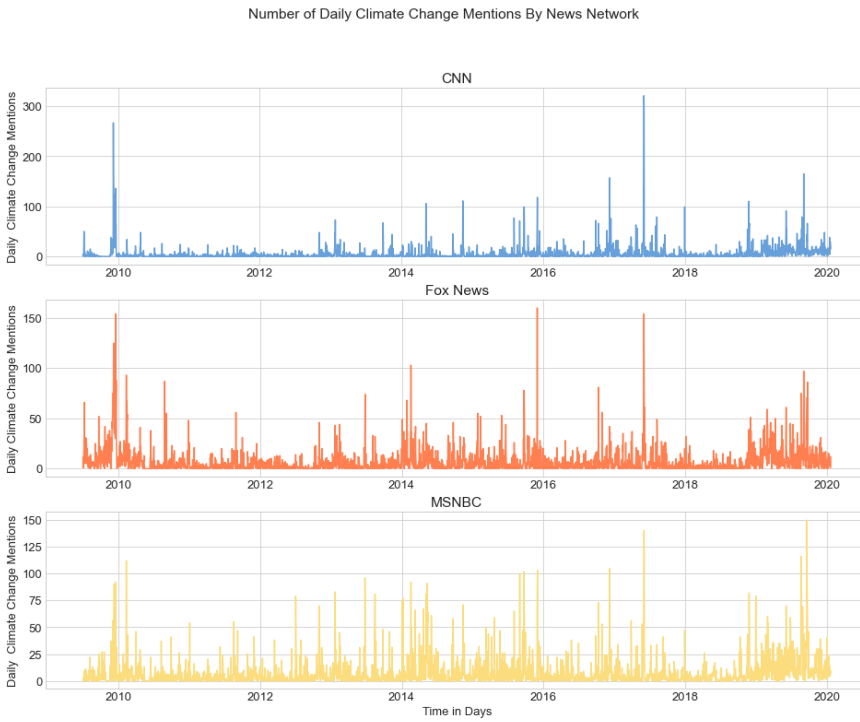


Figure 39: Disaggregated, Daily Climate Change Mentions By Network

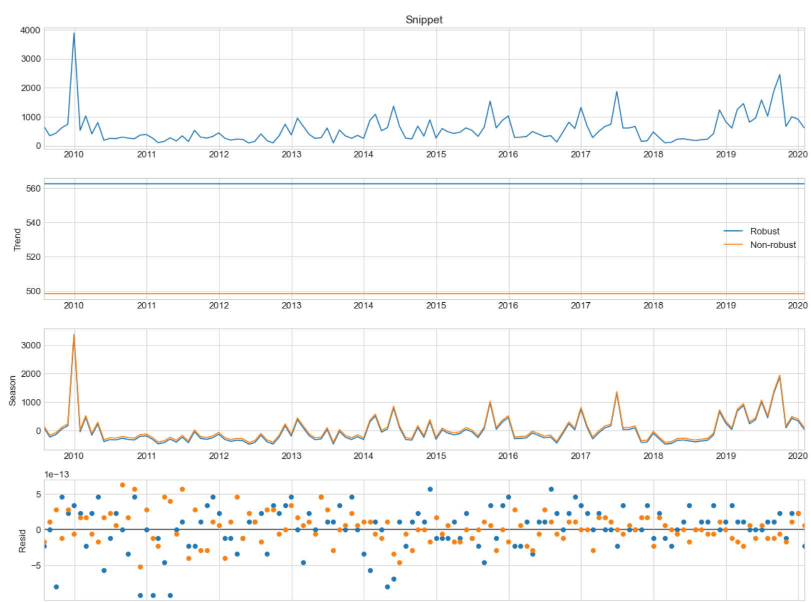
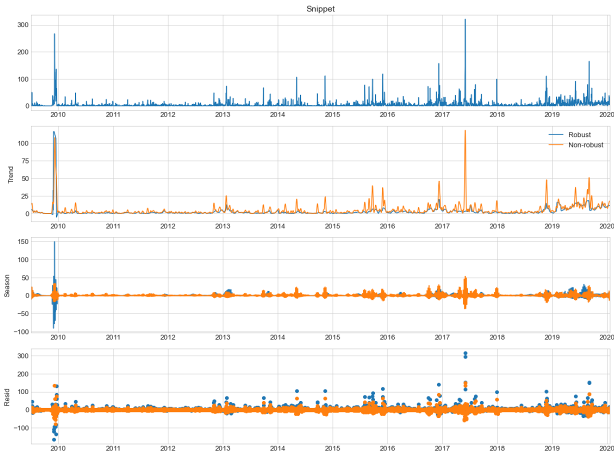
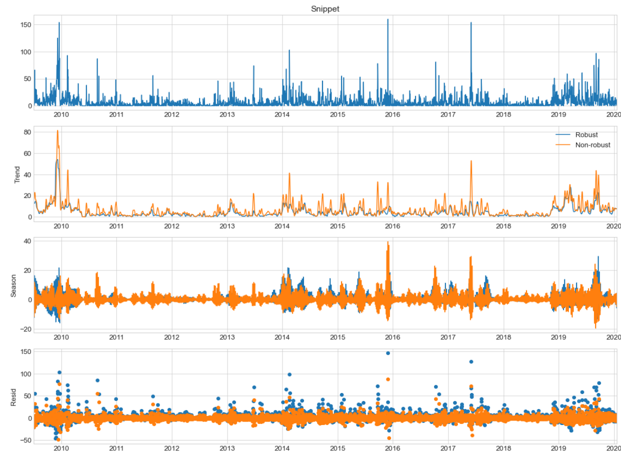


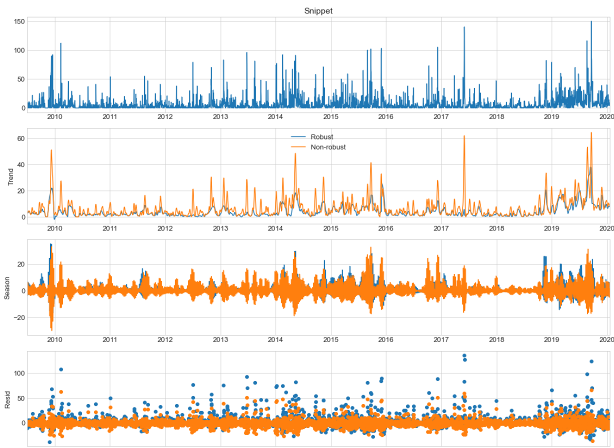
Figure 40: STL Decomposition, Robust and Non-Robust, for Total Climate Change Mentions with 1-Year Periodicity



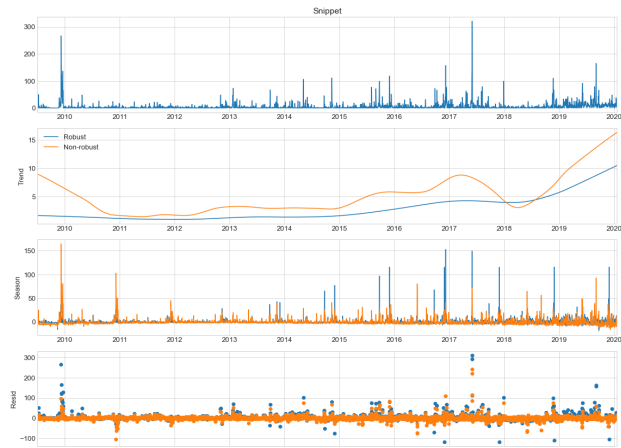
(a) CNN STL, 1 Week Periodicity



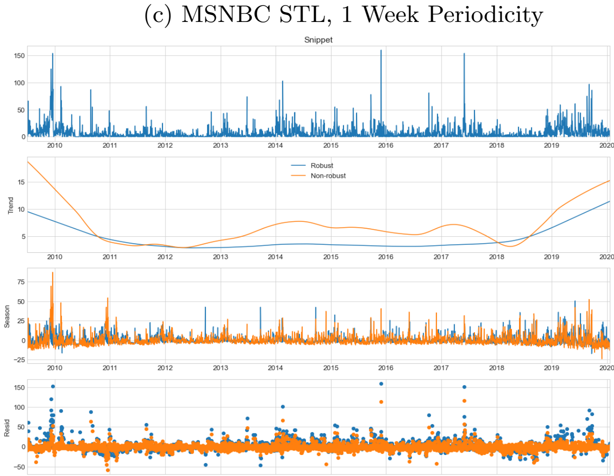
(b) Fox News STL, 1 Week Periodicity



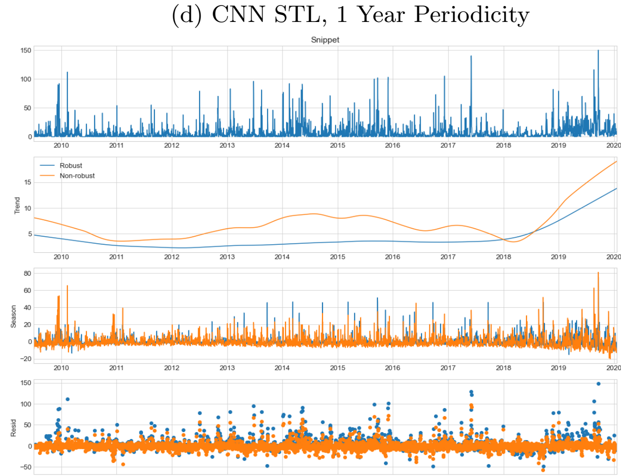
(c) MSNBC STL, 1 Week Periodicity



(d) CNN STL, 1 Year Periodicity

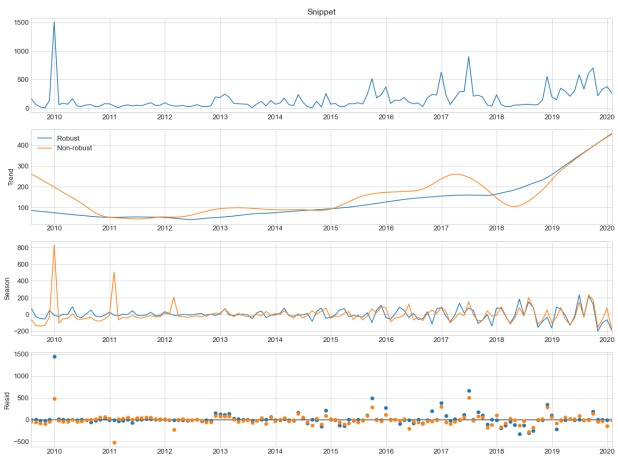


(e) Fox News STL, 1 Year Periodicity

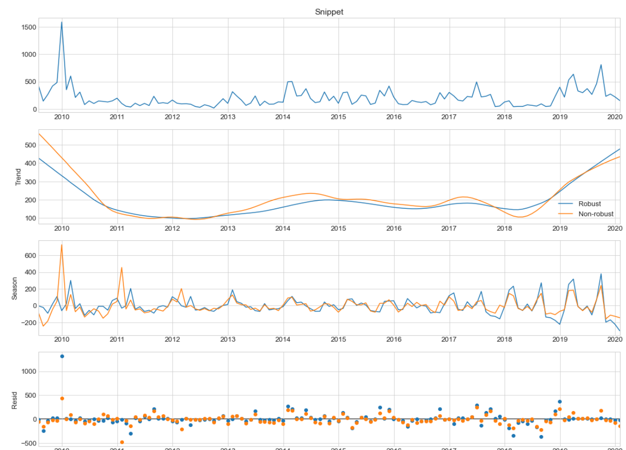


(f) MSNBC STL, 1 Year Periodicity

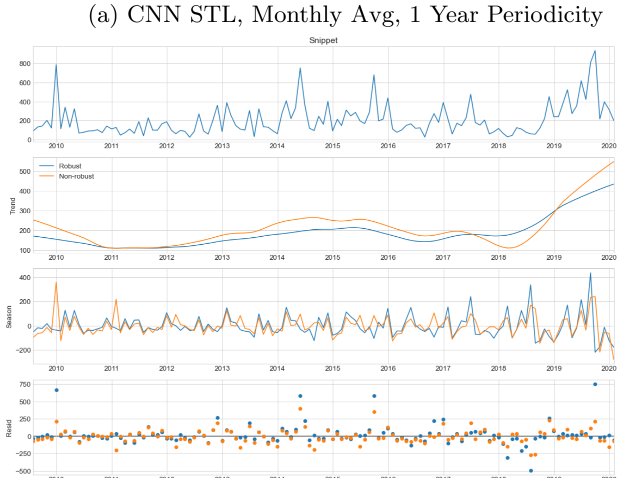
Figure 41: STL Decomposition, Robust and Non-Robust, Over Networks Daily Climate Change Mentions



(a) CNN STL, Monthly Avg, 1 Year Periodicity

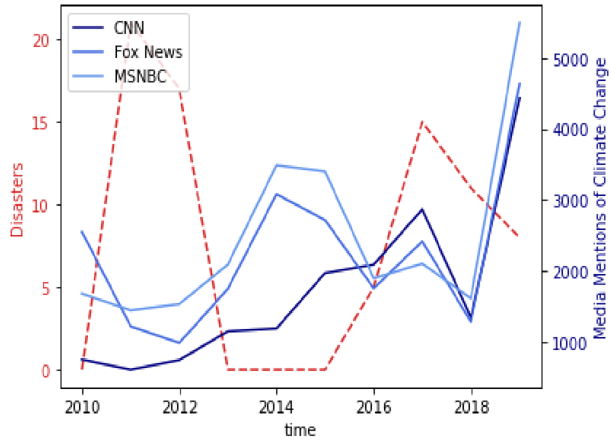


(b) Fox News STL, Monthly Avg, 1 Year Periodicity

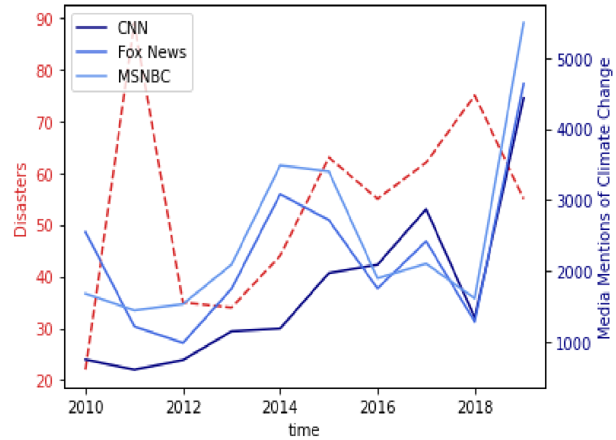


(c) MSNBC STL, Monthly Avg, 1 Year Periodicity

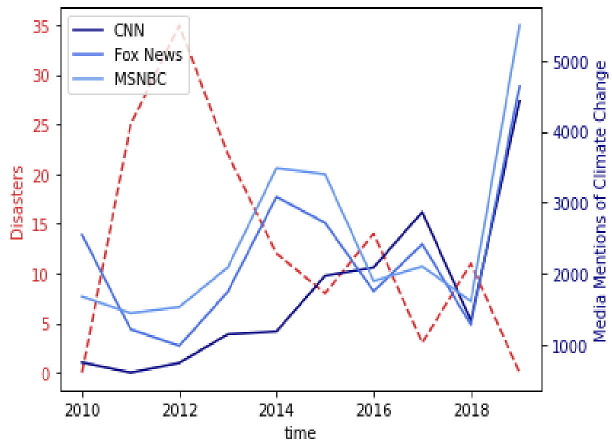
Figure 42: STL Decomposition, Robust and Non-Robust, Over Networks' Monthly Avg Climate Mentions



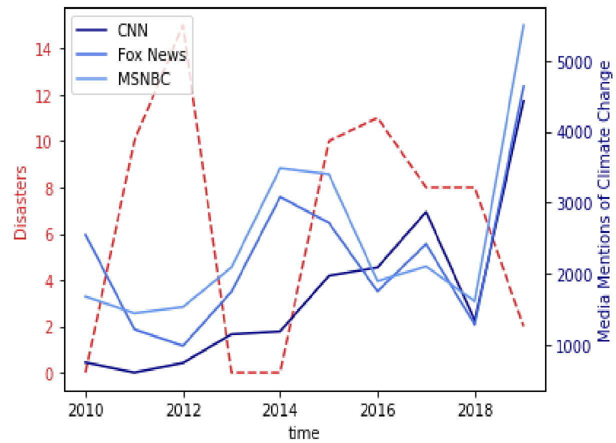
(a) Tropical Cyclones



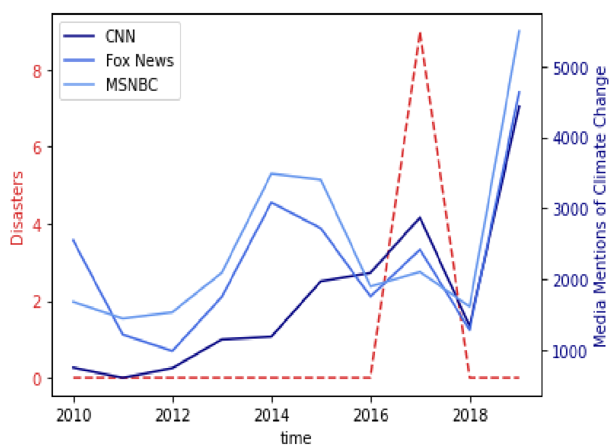
(b) Severe Storms



(c) Droughts



(d) Wildfires



(e) Freeze

Figure 43: Natural Disaster Declaration Data Against Networks Over Time, By Disaster Type