Evolution of the U.S. Television News Narrative on Climate Change

Adam Block, Helena Caswell, Dylan Lewis, Disha Trivedi

Massachusetts Institute of Technology

Research Question

Print and televised media reporting on climate change influences the public perception of climate change, which in turn affects support for systemic policies to reduce greenhouse gas emissions and for individual actions to mitigate climate change. Over two thirds of Americans get their news often or sometimes from television. In this analysis, we look at ten years of data from three television stations: CNN, Fox News, and MSNBC to address 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?

Data Description & Preprocessing

The data used for this analysis includes television transcript snippets related to climate change coverage across CNN, MSNBC, and Fox News from July 2009-January 2020, from the GDELT Project. The features of the data points include time of day and date of the mention, the TV news network, the show, and a snippet of the transcribed audio. This dataset provides the ability to compare the TV networks over time on the subject of climate change in order to answer the research question posed. We follow standard NLP text preprocessing by removing punctuation and numbers, converting all letters to lower-case, lemmatizing, removing standard English stopwords & corpus-specific stopwords, and tokenizing the data into words.

Frequency Analysis

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. Table 1 shows the correlation of monthly frequency of climate change mentions between networks, which appears to track with the partisan lean of the networks, with MSNBC (furthest left) least correlated with Fox News (furthest right).

	CNN	Fox News	MSNBC
CNN	1.000000	0.759091	0.760837
Fox News	0.759091	1.000000	0.712548
MSNBC	0.760837	0.712548	1.000000

Tab. 1: Network Climate Change Mention Frequency Correlation

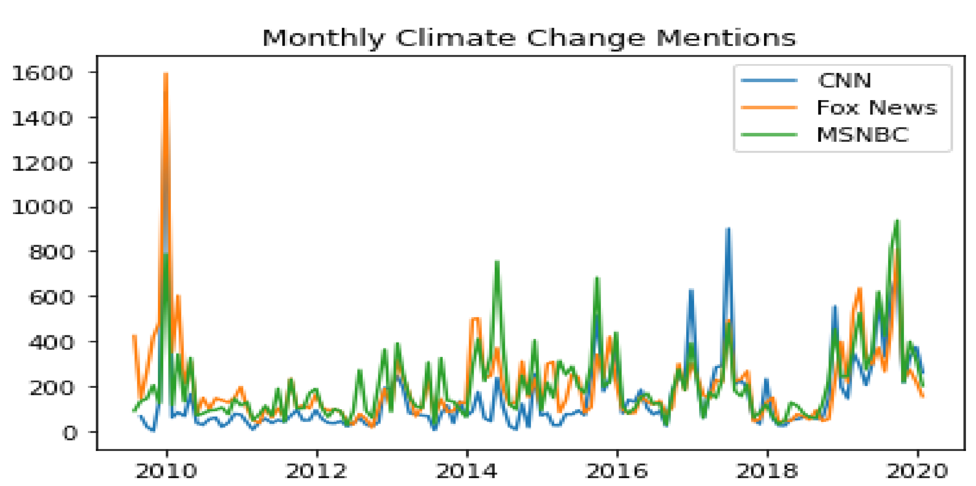

Fig. 1: Frequency of Climate Change Mentions by Month and Network

To see if natural disaster declarations were at all correlated with climate change mentions, we looked at US Natural Disaster Declarations from the Federal Emergency Management Agency (FEMA), which had daily data by county for incidents caused by extreme weather events. The county specific declarations had the useful effect of magnifying larger disasters that affect multiple counties, which should be more likely to be covered by the media. As shown in the table 2, the correlation between disaster declarations and climate change media mentions is overall very low over the time series, but is higher in more liberal TV media.

However, as shown in the figure below, total mentions and correlations with disaster declarations increased remarkably starting in late 2018, coinciding with the release of the IPCC's Special Report on Global Warming of 1.5 ºC in October 2018, and the rise of youth activists such as Greta Thunberg and the Sunrise movement.

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

Tab. 2: Correlation between Climate Change Mention Frequency and Natural Disaster Declarations

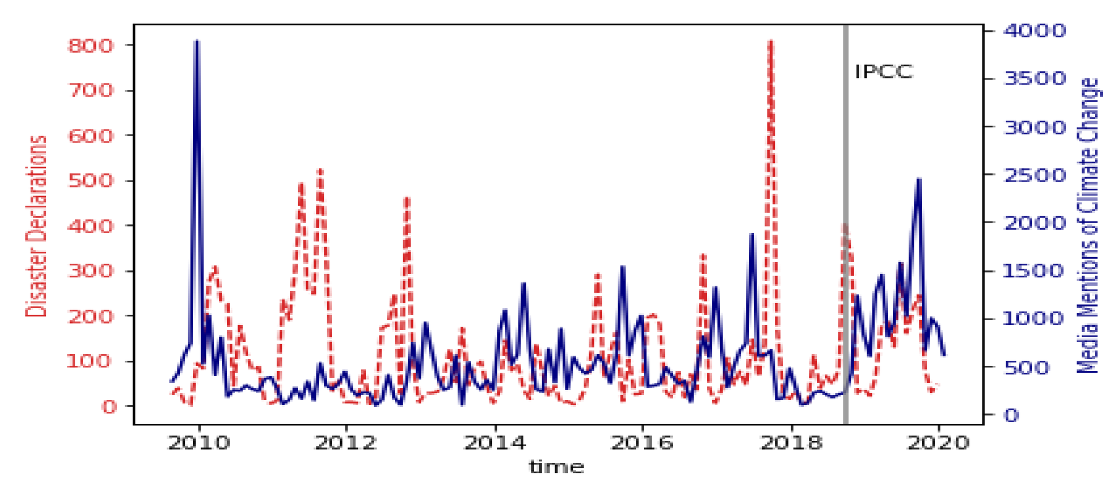

Fig. 2: Frequency of Climate Change Mentions and Disasters Declarations

Content Analysis: TF-IDF & Cosine Similarity

To undergo our content analysis, we need to featurize the news snippets, or collections of snippets, which form the documents in our corpus. We form a document for each year in our dataset, i.e. all snippets in 2009, and transform the documents into a $l2$-normalized unit vector TF-IDF embedding. From this featurization, we form a document-term matrix, where the rows correspond to the TF-IDF embedding of a document and the column represents a unique word in the corpus ($\sim$34,000 words). Thus, entry i, j corresponds to the normalized TF-IDF score of word j in document i (word j's relative importance for the i^{th} document). We also construct a document-term matrix for documents created for networks in specific years, i.e. all CNN snippets in 2015. We calculate the cosine similarity between the document embeddings to yield a measure of similarity between the documents:

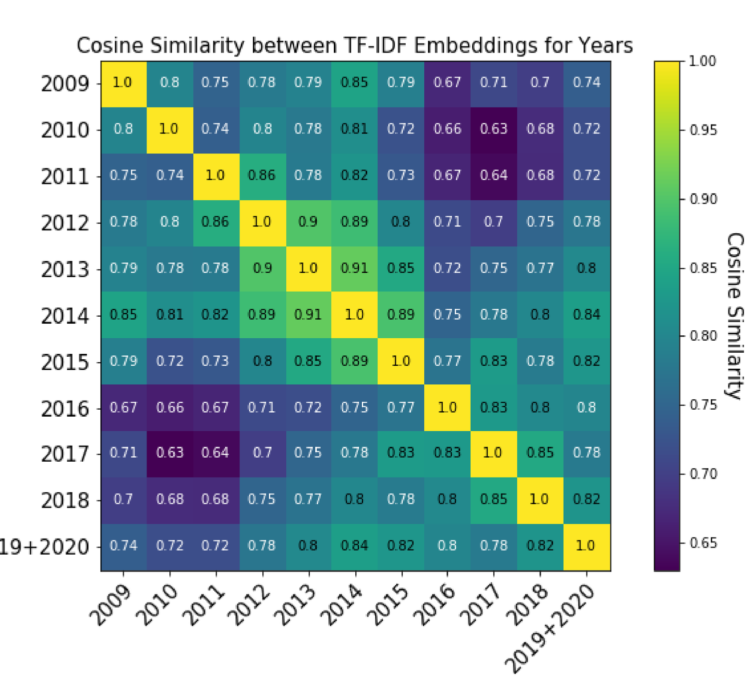

Fig. 3: Cosine Similarity between Year TF-IDF Embeddings

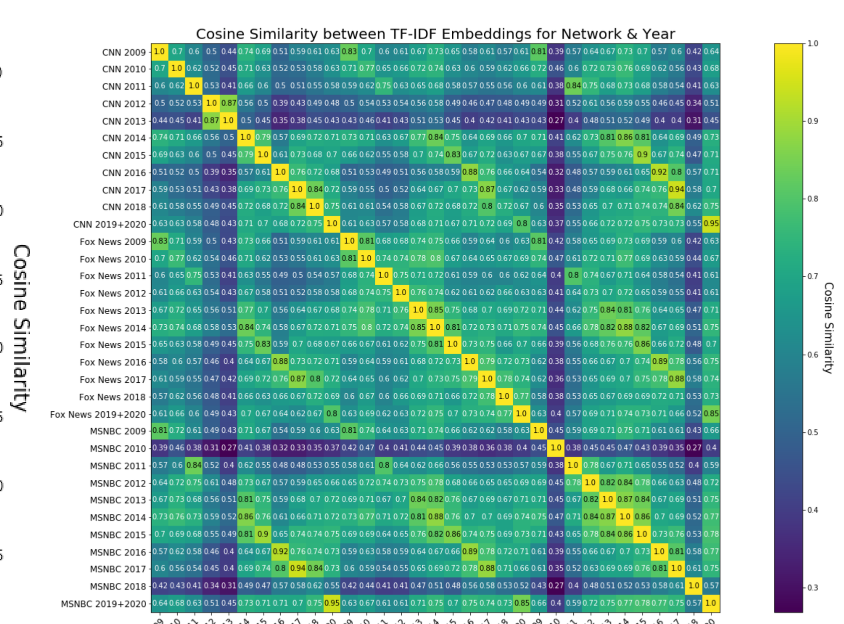

Fig. 4: Cosine Similarity between Network & Year TF-IDF Embeddings

We observe in Figure 3, the embeddings of the years at the beginning of the dataset, 2009-2012, are most dissimilar to the embeddings of the years at the end of the dataset, 2016-2019+2020. This implies that the words most important in the content of climate change mentions were most different between these years than with any other years in the dataset. Finally, the cosine similarity heatmap in Figure 4 reveals documents which are largely dissimilar not only within the network over the years but also with the other networks over the years. These documents include: CNN 2012, CNN 2013, MSNBC 2010 and MSNBC 2018. Excluding these documents, we observe that there are relatively high levels of similarity (≥ 0.75) between networks within the same year for all years. Finally, excluding the documents mentioned above, we observe the similarity between the embeddings for CNN & MSNBC increase with subsequent years from 0.81 in 2009 to as high as 0.95 in 2019+2020.

Content Analysis: Topic Modeling

In an effort to better quantify some of the previous analysis, we turn to topic modelling. Using the technique of Nonnegative Matrix Factorization (NMF) on TF-IDF normalized document-term matrices and cross-validation, we find that 15 topics has the highest coherence score.

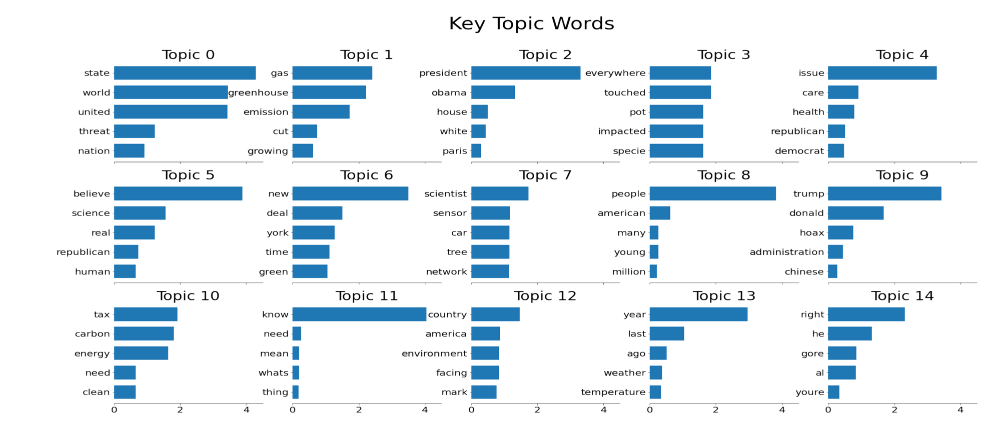

Fig. 5: Top 5 most relevant words by topic.

We also count the monthly topic mentions for each channel and each topic. We see several anomalous topics such as topic 12. We also see some topics being mentioned much more frequently after the Trump administration began:

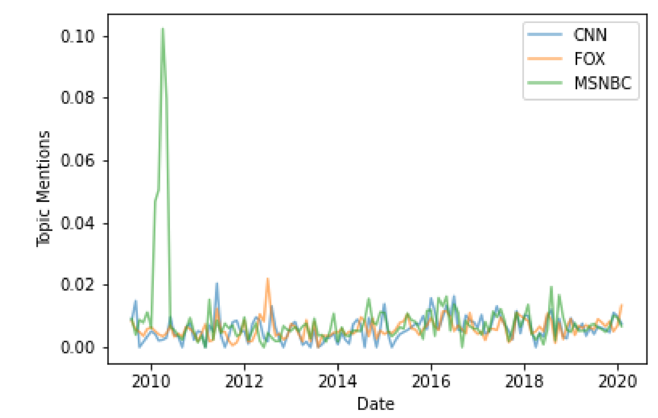

Fig. 6: Frequency by month of mentions of Topic 12

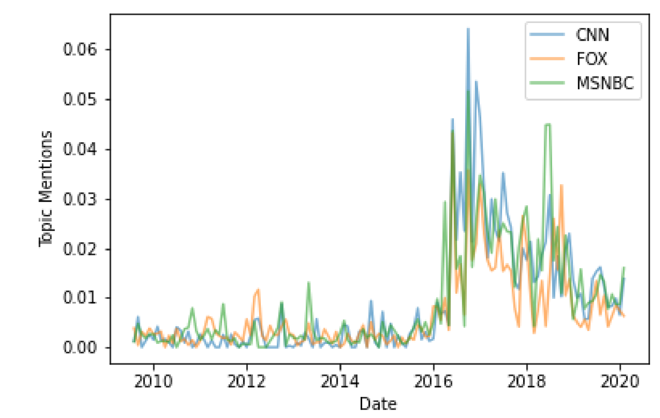

Fig. 7: Frequency by month of mentions of Topic 9

We tested to see whether this trend was statistically significant under a simple model of independent Gaussian noise added to a different mean for each topic-channel-era triple, with two eras broken at January 2017. After using Holm-Bonferroni to correct for multiple hypotheses, we have the following significant hypotheses:

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

Interpretation of Results

Climate change TV 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. This is 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 content similarity. Topic analysis also finds that the majority of 15 topics found in topic analysis had significant changes in mean on some of the topics at the time of Donald Trump's inauguration.