Climate Change & Mastodon

Mark Niemann-Ross

2024-05-10

Abstract

This research answers the following questions:

* When discussing climate change, what additional topics are likely to appear in conversations?
* What terms, phrases, or hashtags are likely to appear in discussions about climate change?
* Which five Mastodon servers are most likely to host discussions about climate change?
* Which five individuals are most likely to discuss climate change?
* Are discussions about climate change positive or negative?

Table of contents

#| echo: false  
#| message: false  
#| warning: false  
#| error: false  
library(rtoot)  
library(tibble)  
library(htm2txt)  
library(quanteda)

Package version: 3.3.0  
Unicode version: 14.0  
ICU version: 71.1

Parallel computing: 8 of 8 threads used.

See https://quanteda.io for tutorials and examples.

findThisHashtag <- params$findThisHashtag  
instanceSampleSize <- params$instanceSampleSize  
  
# load in the list of instances ---------------  
masto\_instances <- readRDS(file = "~/Documents/NLP/nlpClimateChange/mastodon\_instances.RDS")  
  
# grab a sub-sample, nrows = instanceSampleSize  
masto\_instances\_sampled <- masto\_instances[sample(nrow(masto\_instances), instanceSampleSize), ]  
  
masto\_toots <- tibble()  
  
# build a tibble containing toots of #climatechange  
for (anInstance in masto\_instances\_sampled$name) {  
 try(silent = TRUE,   
 expr = {  
 masto\_anInstance <- get\_timeline\_hashtag(instance = anInstance,hashtag = "climatechange")  
 masto\_toots <- rbind(masto\_toots,masto\_anInstance)  
 },  
 )  
}  
  
# convert masto\_toots$content from html to text  
masto\_toots$content <- htm2txt(masto\_toots$content)  
  
# extract interesting parts of URL  
url\_parts <- strsplit(x = masto\_toots$"url",  
 split = "/")  
masto\_toots$instance <- sapply(url\_parts, "[[", 3)  
masto\_toots$contributor <- sapply(url\_parts, "[[", 4)  
  
# trim out the unwanted columns  
masto\_col\_names <- c("id","content","created\_at",  
"reblogs\_count","favourites\_count",  
"replies\_count","url","instance",  
"contributor")  
masto\_toots <- masto\_toots[ , masto\_col\_names]

# Introduction

Research on climate change helps us understand the complexities of the Earth’s climate system and the impacts of human activities on the environment. Research can develop models to predict future changes, assess risks, and formulate strategies to mitigate its effects.

# Methods

In this research, I pulled information from a random sample of Mastodon servers. The sample size was 20 from a list of 9995

# Results

## Related Topics

*When discussing climate change, what additional topics are likely to appear in conversations?*

In this research, #climateChange is an obvious starting point to discover obscured relationships. To find topics related to #climateChange, I followed this procedure:

1. Create a random list of mastodon instances
2. Ask each mastodon instance for any toots (messages) which include the #climateChange hashtag
3. From the list of toots, generate a list of tokens
4. Remove any tokens which do not begin with the hashtag symbol (#)
5. Create a document-feature matrix of the hashtags
6. Sort and print a list of the most frequent hashtags

# masto\_tootss is a tibble created at the top of this document  
# prep column names for quanteda  
masto\_col\_names[1] <- "doc\_id"  
masto\_col\_names[2] <- "text"  
names(masto\_toots) <- masto\_col\_names  
  
climateChange\_hashtags <- corpus(masto\_toots,   
 docid\_field = "doc\_id",  
 text\_field = "text",  
 unique\_docnames = FALSE) |>  
 tokens() |>  
 dfm() |>  
 dfm\_keep(pattern = "#\*")

This list of hashtags was created from a list of toots.

Total number of hashtags in this data set = 523

Total number of toots in this data set = 228

topfeatures(climateChange\_hashtags, n = 10)

#climatechange #climate #climatecrisis #environment   
 229 59 55 42   
 #science #degrowth #globalwarming #climateemergency   
 35 18 17 14   
 #nature #climateaction   
 13 11

## Related Terms and Phrases

*What terms and phrases are likely to appear in discussions about climate change?*

There are three items of interest identified with discussions on climate change: hashtags, single words (aka tokens), and phrases (aka n-grams). The above exercise identified hashtags; this section identifies words and phrases.

Identification of related phrases is a similar process to identifying related hashtags. Instead of removing all tokens *except* hashtags, finding tokens requires removing all hashtags, but leaving tokens intact.

# related terms (features)  
climateChange\_tokens <- corpus(masto\_toots,   
 docid\_field = "doc\_id",  
 text\_field = "text",  
 unique\_docnames = FALSE) |>  
 tokens(remove\_punct = TRUE,   
 remove\_numbers = TRUE,  
 remove\_symbols = TRUE) |>  
 tokens\_remove(pattern = stopwords("en")) |>  
 tokens\_remove(pattern = "#\*") |>  
 dfm()

### Terms related to #climatechange

climateChange\_tokens is filled with individual words that appear in climate change discussions. Here are the top-ten most frequent words.

topfeatures(climateChange\_tokens, n = 10)

degrowth climate can change people article now movement   
 90 89 70 70 56 47 44 44   
 global new   
 42 41

### Phrases related to #climatechange

Phrases can be any combination of words, a minimum of two but possibly more. This is accomplished through the use of n-grams, although there are more sophisticated methods for identifying phrases. Here is code to identify two-word phrases.

# related n-grams  
climateChange\_tok\_ngram <- corpus(masto\_toots,   
 docid\_field = "doc\_id",  
 text\_field = "text",  
 unique\_docnames = FALSE) |>  
 tokens(remove\_punct = TRUE,   
 remove\_numbers = TRUE,  
 remove\_symbols = TRUE) |>  
 tokens\_remove(pattern = stopwords("en")) |>  
 tokens\_remove(pattern = "#\*") |>  
 tokens\_ngrams() |>  
 dfm()

Here is the list of the ten most-frequent two-word phrases:

topfeatures(climateChange\_tok\_ngram, n = 10)

climate\_change   
 44   
 degrowth\_movement   
 25   
 full\_article   
 19   
 article\_--   
 19   
 friendly\_critique   
 17   
 critique\_degrowth   
 17   
 follow\_soon   
 17   
--\_https://medium.com/postgrowth/a-friendly-critique-of-the-degrowth-movement-f0bd2297072d   
 17   
 part\_excerpts   
 14   
 excerpts\_article   
 14

## Related Servers

*Which five Mastodon servers are most likely to host discussions about climate change?*

This research is built around a sampled list of servers 20 from a list of 9995 servers. Given patience and available CPU, it is possible to open up research to the entire fediverse of possible servers.

#| message: false  
#| warning: false  
#| error: false  
library("quanteda.textstats", quietly = TRUE)  
masto\_freqInstance <- corpus(masto\_toots,   
 docid\_field = "doc\_id",  
 text\_field = "text",  
 unique\_docnames = FALSE) |>  
 tokens() |>  
 dfm() |>  
 dfm\_keep(pattern = "#\*") |>  
 dfm\_group(instance) |>  
 textstat\_frequency(groups = instance) |>  
 (\(x) {table(x$group)})() |>  
 sort(decreasing = TRUE) |>  
 head(n=5)

Note that this listing fails to account for overall popularity of each server. For example, mastodon.social has approximately 15% of all mastodon accounts. The next largest has 12% and the numbers drop rapidly in subsequent servers.

masto\_freqInstance

mastodon.social beige.party climatejustice.social   
 104 70 63   
 mstdn.social flipboard.social   
 56 45

## Active Individuals

*Which five individuals are most likely to discuss climate change?*

It is important to recognize the impact of bots which are simply running to amplify a particular viewpoint, rather than individuals engaged in production conversation. By identifying a list of most frequent contributors, it is possible to ascertain if an account should be trusted.

masto\_freqIndividuals <- corpus(masto\_toots,   
 docid\_field = "doc\_id",  
 text\_field = "text",  
 unique\_docnames = FALSE) |>  
 tokens() |>  
 dfm() |>  
 dfm\_keep(pattern = "#\*") |>  
 dfm\_group(contributor) |>  
 textstat\_frequency(groups = contributor) |>  
 (\(x) {table(x$group)})() |>  
 sort(decreasing = TRUE) |>  
 head()

This produces a list of top-ten contributors from the random sample of mastodon instances:

masto\_freqIndividuals

@LibertyForward1 @ariadne @bradweitz objects   
 70 53 37 32   
 @nasa @ernandy   
 31 29

## Climate Change Sentiment

*Are discussions about climate change positive or negative?*

This research uses the simplest approach possible for calculating sentiment. Quanteda provides textstat\_polarity for quick sentiment analysis, producing a positive result if the conversation is “positive” or a negative result if the conversation is “negative.”

#| message: false  
#| warning: false  
#| error: false  
library("quanteda.sentiment", quietly = TRUE)

Attaching package: 'quanteda.sentiment'

The following object is masked from 'package:quanteda':  
  
 data\_dictionary\_LSD2015

masto\_sentiment <- textstat\_polarity(climateChange\_tokens,  
 dictionary = data\_dictionary\_LSD2015) |>  
 (\(x) {sum(x$sentiment)})()

The sentiment of conversations related to #climateChange is 18.7736448 . This is simple polarity, with negative values indicating a negative attitude and positive values indicating generally positive feelings. In this case, the sentiment value indicates generally positive attitudes about #climateChange.

# Discussion

Researching the discussion surrounding climate change is vital for several reasons:

1. **Understanding the Issue**: Climate change is a complex phenomenon with far-reaching implications. Research helps us grasp the various factors contributing to it, such as greenhouse gas emissions, deforestation, and industrial practices.
2. **Informing Policy Decisions**: Policymakers rely on scientific research to formulate effective policies addressing climate change. Understanding the latest findings helps in crafting legislation and regulations aimed at mitigating its impacts and transitioning to more sustainable practices.
3. **Mitigation and Adaptation Strategies**: Research provides insights into mitigation strategies to reduce greenhouse gas emissions and adapt to the changes already underway. This includes developing renewable energy sources, implementing sustainable agricultural practices, and planning for sea-level rise and extreme weather events.
4. **Public Awareness and Education**: Research findings help raise public awareness about the urgency of addressing climate change. Educating people about the causes, impacts, and potential solutions empowers them to make informed decisions in their daily lives and advocate for policy changes.
5. **International Cooperation**: Climate change is a global challenge that requires coordinated efforts from countries worldwide. Research serves as a common ground for international collaboration, facilitating data sharing, technology transfer, and joint initiatives to combat climate change collectively.
6. **Economic Implications**: Climate change can have significant economic consequences, affecting industries, employment, and overall economic stability. Research helps assess the economic risks and opportunities associated with climate change, informing investment decisions and business strategies.
7. **Environmental Justice**: Climate change disproportionately affects vulnerable communities, including low-income populations and indigenous peoples. Research helps identify these disparities and advocate for equitable solutions that prioritize environmental justice and social equity.