Analysis of evolution of meme trends on 4chan.org's /pol/ board via image clustering

J. Jin, E. Williams, S. Lam, O. Savas, E. Hohman, M. Bosch-Ruiz, P. Rodrigues
Accenture Applied Intelligence
Arlington, VA 22203

November 5, 2020

1 Introduction

Internet memes rapidly accelerate the spread of ideas on social media platforms. Although memes are primarily consumed as entertainment, they often communicate messages that influence viewers' political opinions [4]. With the advent of new multimodal classification techniques, we can garner new insights about the content and culture of a platform. [2] analyzes memes using image, text, and face encoding and applies this to a large online conversation around a political event, and [5] finds that 30% of the image-with-text memes in their sample have identifiable themes that are politically relevant and shared more often by Democrats than Republicans.

Often new memes are created from "templates" based on key elements and features of already popular memes. These templates can be exploited by "meme factories," which are coordinated networks of creators, for social or monetary gain. Employing meme templates help users achieve widespread dissemination, urge call-to-actions, affect discourse, and spread misinformation [1]. The emergence of the novel coronavirus, COVID-19, introduced an influx of Internet memes reflecting on the infection rate, origin, and location of cases. Given 4chan.org's /pol/ imageboard's reputation as a vector for misinformation [7], we conducted a study of meme templates and their use in knowledge spread throughout the COVID-19 pandemic to answer:

1. What computer vision methods can we use to identify and cluster meme templates?
2. What and how are memes used in COVID-19 related discussions on 4chan.org?

2 Methodology

2.1 Data Collection

We collect our image and text data from 4chan.org's "politically incorrect" board, /pol/. We focus on this board as users are prolific producers and consumers of memes that spread misinformation. /pol/ frequently originates alternative news and conspiracy theories that cross over to mainstream social networks like Twitter, Facebook [8], and YouTube [6]. From January 1 to September 30, 2020, we collected 894,346 unique threads from /pol/. We subset to 69,722 coronavirus-related threads containing 860,051 images by searching the first post for the regex string "virus|covid|rona |corona". Due to limitations in time and compute power, we randomly sampled 10,000 images from each month for analysis.

2.2 Image Clustering

Memes are challenging to cluster using computational methods because of the countless ways in which meme templates can be edited and manipulated. Among these many manipulations, individuals who post memes often edit text, add panels, remove panels, or superimpose other images on the meme template. We found dbscan clustering with perceptual hashing was too brittle for the task and failed to group many edited memes together. We adapted an unpublished open-source approach to meme clustering that relies on feature extraction [3]. We experimented with several backbone models, including ResNet50, VGG16, and VGG19, but found EfficientNetB6 to be the most robust to noise. Using feature extraction, even without fine-tuning, resulted in substantial improvements in the performance of our clustering algorithms.

Following feature extraction with an EfficientNetB6 backbone, we clustered images based on their encodings with dbscan, hdbscan, umap + hdbscan, and flat hierarchical clustering.

We found, qualitatively, that flat hierarchical clustering, with a cophenic distance 40% of the maximum distance in the distance matrix, resulted in sufficiently robust meme clusters.

3 Results

Throughout the pandemic, we observe a shift in the types of meme templates and images posted by 4chan users in threads about coronavirus. Although /pol/ users in COVID-19 related discussions began with more humorous hazmat suit memes in January, there are more references to scientific information on charts and graphs over time. This suggests that discussion shifts towards making claims around COVID-19 case counts and projections. Politically charged memes and image content also became more popular on the platform. The largest clusters for each month are listed in Table I and Figure I.

Each month, the largest cluster consisted of over 1,000 screenshots of news and social media posts. We also see representation from specific meme templates across the data, including variants of Pepe the Frog, That Feel Bro, and references to the military and weaponry.

Table 1: Image Volume per Cluster & Month

Month	***Jan***	***Feb***	***Mar***	***Apr***	***May***	***Jun***	***Jul***	***Aug***	***Sep***
news/media	1205	1447	1627	1591	1622	1572	1686	1774	1235
pepe	204	123	129	110	72	154	103	105	158
feels guy	28	34	39	79	45	54	30	35	23
anime	90	99	128	358	309	486	160	143	186
cartoons	236	356	292	133	91	140	251	307	309
military/weapons	45	50	36	58	34	37	30	43	31
people	168	242	310	318	194	318	287	261	380
hazmat	516	64	28	30	6	0	4	7	8
gas mask	4	43	26	20	15	5	7	7	7
store	0	20	22	4	0	0	0	0	0
graphs	158	215	175	184	155	133	191	114	81
tables	17	5	23	28	48	27	91	42	12
maps	16	42	26	9	23	24	20	49	33
us maps	0	4	6	0	8	17	10	34	19
maga momiji	13	13	18	21	27	26	37	41	30
total	4483	4437	4391	4537	4216	4861	4370	4394	4147
% labelled	67	67	69	70	71	70	70	72	72

The Hazmat suit meme, depicting figures in a yellow hazardous materials coverall, initially dominated conversations with 516 images identified in January 2020. This template quickly drops in frequency, with image volume declining by 88% in February. From May onward, we see that the Hazmat suit meme is no longer popular. Similarly, image clusters depicting gas masks become more prominent in February and March 2020 but are rarely seen after May. Other ephemeral image clusters are pictures of grocery store shelves and toilet paper, which are briefly popular in February and March.

In contrast, the volume of images with data tables and graphs increases about 35% from 158 to 215 respectively in February. There is sustained interest in tables and graphs, with a peak of over 280 graphs and tables identified in July.

In general, political content increases as the United States approaches the 2020 Elections in November. While map clusters are observed over the entire year, starting in June maps of the United States, particularly in reference to the United States electoral college, start to appear more often. The MAGA Momiji meme, which shows support for President Donald Trump, becomes more prominent after June 2020.

4 Conclusions

Although memes are a popular vehicle for spreading ideas and misinformation, the medium is not heavily explored. In this research, we provide a new method of clustering memes. Grouping memes by template allow us to understand the evolution of the template over time and popularity of the template on a social media platform.

Understanding and perspectives around coronavirus changed as more information was discovered and reported. We notice a similar evolution in use of meme templates on

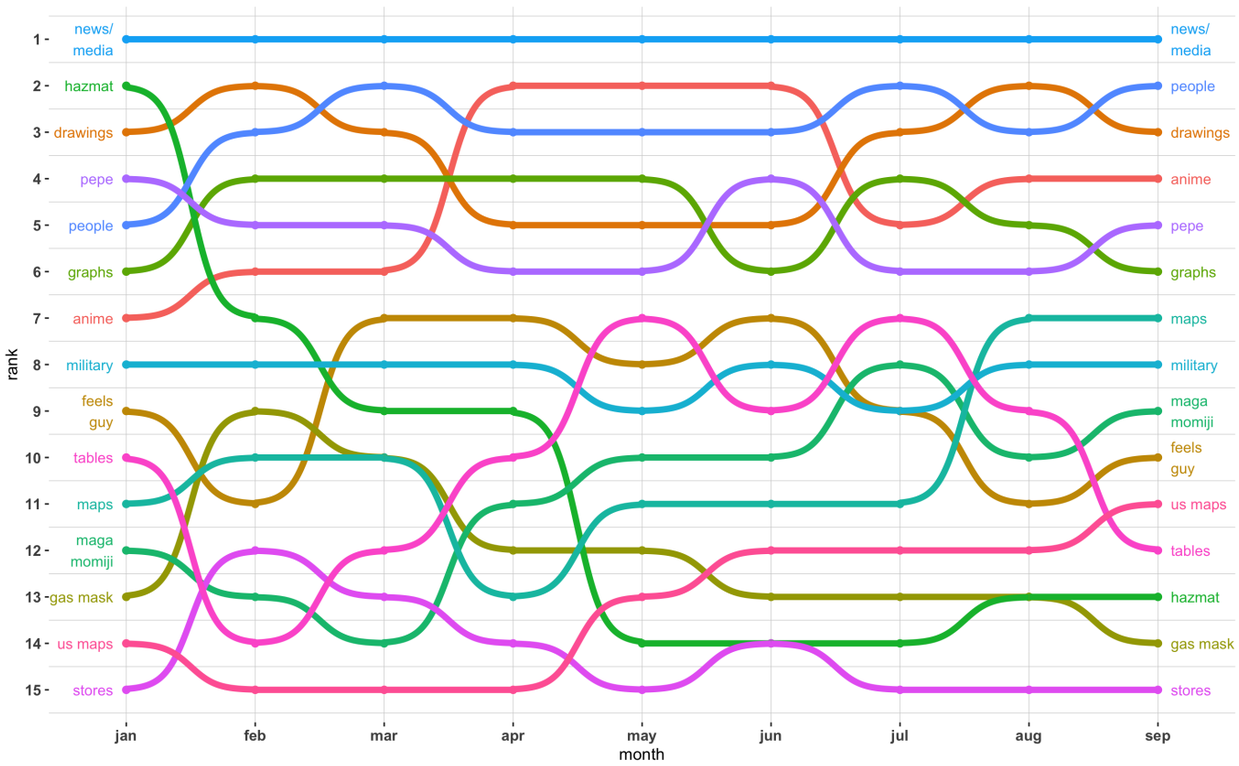

Figure 1: Change in Cluster Rank by Image Volume per Month.

4chan.org. Initially, posters on the /pol/ imageboard used Hazmat suit and gas mask templates for coronavirus-related threads. Over time, users integrate more images of graphs, tables, and other scientific information. Political meme templates and images also became more popular as the discussion approached the U.S. presidential election in November.

4.1 Future Work

/pol/ threads expire and are removed from 4chan.org's server at a rapid pace. We observed users starting new threads with the same text and images multiple times over a period of time, leading to duplicate images. Studying these "re-posted" threads will help understand what images and discussion are being actively promoted by 4chan users to reach a larger audience.

While this clustering approach worked well for the majority of meme templates, the approach struggled with some animation. Some memes, while still grouped with similar memes, were incorrectly split between multiple clusters. In future work, we plan to formally quantify clustering errors and fine-tune our backbone model on meme templates to improve multidimensional encoding representation.

References

[1] Crystal Abidin. Meme factory cultures and content pivoting in singapore and malaysia during covid-19. *The Harvard Kennedy School Misinformation Review*, 1, 2020.

[2] David M. Beskow, Sumeet Kumar, and Kathleen M. Carley. The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning. *Information Processing and Management*, 57(2):102170, 2020.

[3] Cakan and Caglar. deep_meme_clustering. https://github.com/caglorithm/deep_meme_clustering, 2018.

[4] Joan Donovan. How memes got weaponized: A short history. 2019. Last accessed 4 November 2020.

[5] Yuhao Du, Muhammad Aamir Masood, and Kenneth Joseph. Understanding visual memes: An empirical analysis of text superimposed on memes shared on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):153–164, May 2020.

[6] Gabriel Hine, Jeremiah Onaolapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. Kek, cucks, and god emperor trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. May 2017.

[7] Lee Knuttila. User unknown: 4chan, anonymity and contingency. *First Monday*, 16(10), 2011.

[8] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. "the web centipede: Understanding how web communities influence each other through the lens of mainstream and alternative news sources". 17th ACM Internet Measurement Conference (IMC 2017), 2017.