2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Approaches for Quantifying Video Prominence, Narratives, & Discussion: Engagement on COVID-19 Related YouTube Videos

Jennifer Jin
Accenture Applied Intelligence
Arlington, Virginia, United States
jennifer.jin@accenture.com

Sophia Lam
Accenture Applied Intelligence
Arlington, Virginia, United States
sophia.s.lam@accenture.com

Onur Savas, Ph.D
Accenture Applied Intelligence
Arlington, Virginia, United States
orcid.org/0000-0001-5540-9880

Ian McCulloh, Ph.D
Accenture Applied Intelligence
Arlington, Virginia, United States
ian.mcculloh@accenture.com

***Abstract*—Initial scientific studies suggest the spread of extreme content, "the COVID-19 infodemic," likely plays a crucial role in news spread about the "the COVID-19 pandemic." In this paper, we quantify the evolution of polarization and engagement in YouTube social networks about public-health interventions for COVID-19. Although YouTube is a major information and news source with high engagement with younger populations, the platform is not widely researched in social network analysis. Discussions about coronavirus on social media can influence how individuals interpret news about the disease and affect their compliance with various non-pharmaceutical interventions. We compare coronavirus video content by identifying three subgroups of public-health intervention-related videos: individual interventions, government interventions, and medical interventions, as well as seven video title narratives. The polarization index measures the level of agreement with the video content using votes: likes and dislikes. The engagement index measures the level of user interaction by comparing views, votes, and number of comments. We observe that over time, engagement for the intervention video subgroups has increased whereas the diffusion for other non-intervention videos has decreased, which suggests that information about COVID-19 interventions has become more popular as the pandemic develops. Additionally, YouTube's search ranking algorithm seems to strongly take into account video polarization as videos that remain prominent in the search rankings have polarization scores 37% lower than videos that are removed from the top ten results. Topicality of video content may also play a role as medical treatment-related videos are the least promoted in the search results amongst the video subgroups despite having low video polarization. Engagement is lowest overall on medical intervention videos, which may be due to vaccine and treatment development as a topic being downgraded quickly from YouTube's search results. We recommend further research into YouTube's search result ranking model to better understand YouTube's role in the spread of news and information about coronavirus and other topics. Despite focusing on the COVID-19 pandemic, the methods for analyzing YouTube videos may be applied to other events or crises.**

***Index Terms*—social networks, social media, narratives, polarization, engagement, YouTube, videos, comments, COVID-19, coronavirus**

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00

I. INTRODUCTION

A. Background

Since December 2019, the COVID-19 pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to over 44 million cases and 1.17 million deaths, as of this writing. The outbreak first began in Wuhan, China and has been reported in more than 188 countries and territories. The virus is spread through people in close contact via respiratory droplets from coughing, sneezing, and talking. Therefore, public health officials and organizations have recommended interventions such as social distancing, mask wearing, and handwashing as methods for decreasing spread. Various media sources and politicians have polarized the implementation of interventions [1]. Through this research, we seek to understand the effects of polarization in discussion and YouTube video content around COVID-19 by posing the following research questions (RQs):

RQ1. Do the narratives around COVID-19 video content highlight public intervention methods?

RQ2. Does polarization affect which videos are promoted on YouTube?

RQ3. Do YouTube videos about COVID-19 interventions have greater engagement and polarization than other COVID-19 content?

B. COVID-19 and Polarization

Fear, anxiety, and uncertainty in the society and economy has given way for misinformation to spread as quickly as the virus itself. Misinformation was widespread during the development of other epidemics like HIV or the bubonic plague. However, mass media and social networks have allowed rapid spread of false information about the disease. Even famous figures and state governments have perpetuated rumors. Due to the spread of fake news, World Health Organization declared an "infodemic." Misleading information potentially hinders public health efforts used to halt disease transmission [2].

Misleading information may also be a contributor for polarization and politicization of the pandemic. Common topics for misinformation include conspiracy theories, misreporting, transmission, prevention, and treatment. Some theories claim the origins of the virus are in biological warfare while other stories involve xenophobic attacks. In the United States, anti-mask wearing and anti-lockdown narratives are prevalent and politicized, even though masks are effective for reducing disease transmission. Many drugs, such as hydroxychloroquine, or at-home remedies have been promoted as treatments despite having little scientific evidence. Research focusing on newspaper article analysis found that traditional news coverage early in the pandemic was highly politicized with polarization levels comparable to debates around climate change. The misinformation and polarization surrounding the disease have led to a resistance against interventions and best practices to reduce transmission and fatalities [3].

C. Platform Choice

Although Twitter is a popular platform for research, other sites like Instagram, Facebook, Reddit, or YouTube have more engagement and reach with United States households across many age groups. Specifically, Pew Research Center found that 73% and 68% of adults use YouTube and Facebook respectively whereas only 22% of adults use Twitter [4]. The video platform also has high engagement with young people, with over 90% of millennials and Generation Z respondents using YouTube [4]. As a result, news content on YouTube is of particular interest to understand younger audiences' response to events. A Politico survey in September 2020 found that 39% of respondents aged 18-34 used YouTube daily as a U.S. elections news source, the highest of all the platforms asked, which included Twitter, newspapers, and TV news [5].

YouTube is a major information and news source, with over 2 billion active users worldwide. Over a quarter of users in a January 2020 Pew Research survey reported receiving their news through YouTube [6]. Additionally, almost two-thirds of respondents to a March 2020 survey of social media users believed they would increase their consumption of YouTube content if asked to stay at home. Discussion about the disease on social media can influence how individuals receive news about the disease and affect their compliance with stay-at-home orders and other interventions. Overall, despite widespread usage across, YouTube is an understudied social media platform. Social media platforms beyond Twitter with greater usage and frequency of engagement should be investigated to better understand online conversations.

D. YouTube and Polarization

Like other social media platforms, YouTube may also be a driver for greater political polarization and radicalization. However, the platform has not been widely researched for its potential for increasing divisiveness debates.

A study focusing on the effect of YouTube videos related to science and conspiracy theories shared on Facebook found that content had a polarizing effect on users [7]. Users focus on specific content types and aggregate the information into separate groups or echo chambers. Other research combines Twitter and YouTube data and observes a strong correlation between shares on Twitter and the views of the video, indicating that popular content on YouTube is often promoted on other platforms [8]. Past research does not discuss the sentiments of YouTube users themselves and their engagement with the publisher of the content or other viewers via comments or likes.

Additionally, other research on YouTube promoted videos has focused on the recommended videos playlist, where videos "auto-play" after another. This paper will focus instead on the start of the video watching process, when the viewer is initially seeking information and finding video content through the search bar. Viewers are more likely to actively pay attention to videos at the start of their watching experience and gaining insight into what videos are promoted by YouTube's search ranking model can help video publishers achieve higher prominence in search results.

II. Methodology

A. YouTube Video and Comment Data Collection

To investigate and understand the conversation surrounding interventions for COVID-19 prevention and treatment, we focus on YouTube as a platform. In particular, we investigate the videos promoted by YouTube through its search algorithms and ranking. The top ten video results from a set of coronavirus-related search queries on YouTube were tracked every four hours, starting from January 28 to July 4, 2020. Video metadata and comments were collected on 6,668 unique videos with over 750,000 unique comments and replies. Videos may appear multiple times in the dataset if they remained in the top 10 search result for later scraper runs. Data collection was performed on a scheduled basis through UiPath, a robotic process automation (RPA) tool. Additionally, a Python-based web scraper relying on the Selenium and BeautifulSoup4 libraries was used to supplement data collection on-demand [9]. We collect the attributes listed in Table I.

The data scraper collects the metadata and comments of YouTube's top 10 videos results from a set of coronavirus-related search queries every 4 hours. Some search queries include "virus epidemic," "covid19 quarantine," "covid testing," "china outbreak," and "coronavirus cure."

By collecting the top 10 video results, we develop a better understanding of the kind of content that YouTube promotes. The average viewer who is interested in learning about the pandemic will first be exposed to this content and is less likely to scroll past the first page of search results to find additional content. Therefore, scraping was restricted to the first 10 results to serve as a snapshot of what is currently trending on YouTube. Similarly, comments were extracted from the content visible after two full scroll-throughs of the page, representing the top comments as an average user would view them for that video.

B. Classifying YouTube Video Content

Using key terms related to social distancing and other disease prevention practices, videos were divided based on title into three subgroups of interventions: individual interventions, which focus on the steps private individuals take to

TABLE I
YOUTUBE VIDEO ATTRIBUTES

Type	Attribute	Description
Video Metadata	title	Video title
	url	Video URL
	owner	Video publisher display name
	owner_channel	Video publisher channel URL
	search_query	Search query
	scrape_time	Video metadata scrape time
	transcript	Video transcript
	description	Video description
	category	Video category[1]
	datetime	Video post time
	views	Video view count
	type	Video type
	likes	Like vote count
	dislikes	Dislike vote count
	owner_subscribers	Video publisher subscriber count
	duration	Video duration
	num_comments	Video total comment count
Comment	author	Commenter display name
	author_channel	Commenter channel URL
	text	Comment text
	reply_ct	Comment reply count
	comment_datetime	Comment post time
	comment_url	Comment URL
	scrape_time	Comment scrape time
	comment_votes	Comment net like & dislike count

[1] As designated by video uploader at publication time

prevent coronavirus spread; government interventions, which are society-level prevention strategies, like travel bans and business closures; and medical interventions, which are medical approaches for treating and preventing COVID-19. The title keyword matches for each intervention category are shown in Table II. A video with a title that matched multiple intervention groups, such as "Coronavirus Updates: Fauci to Self-Quarantine After White House Exposure, Johnson Eases UK Lockdown" would be categorized as containing content regarding individual as well as government interventions.

To further understand the videos being analyzed, especially the 70.3% of videos categorized as "non-intervention," a total of seven narratives were identified from the video title corpus: (1) self-imposed & mandatory quarantines, (2) the beginnings of the COVID-19 outbreak, (3) the spread of COVID-19 within the USA, (4) general politics, in particular President Trump's response, (5) the rising coronavirus-related death toll in the USA, (6) the global economic impact, and (7) the development of treatments & vaccines. As shown in Table IV, video title narratives generally align with intervention subgroups. For example, over 70% of videos in the medical interventions subgroup also focused on the coronavirus vaccines and cures narrative.

C. Intervention Categorization

We categorized COVID-19 interventions into multiple subgroups based on video titles. Individual interventions focus on methods individuals can take to prevent coronavirus spread such as mask wearing or proper hand washing. Government interventions analyze society-level prevention strategies, like travel bans and business closures. Medical interventions investigate pharmaceutical approaches for treating and preventing COVID-19. Table II includes the key search terms for each subgroup. The 4,907 videos that did not fall into any of the three intervention subgroups outlined in Table II were categorized as "other" and were considered to contain non-intervention-related content. Note that these intervention subgroups are not mutually-exclusive, and a video could be considered to contain content about both medical and individual interventions, for example.

TABLE II
INTERVENTION KEY TERMS

Subgroup	Key Terms	Volume	Volume %
Individual	Social distanc* Self* Isolat* Mask* Quarantine	343	4.9%
Government	Travel ban* Travel restriction* Lockdown* Shutdown* Shelter in place Stay at home Clos* (closure) Contact trac* Quarantine* Evacuat* Reopen*	796	11.4%
Medical	Vaccin* Trial* *medicine* Disinfect* Treatment* Drug* Pharm* Hydroxychloroquine Remdesivir Hospital* Cure* Pharm*	930	13.3%
Other	N/A	4,907	70.3%

[1] The asterisk depicts a wildcard to denote any number of characters. For example, "Telemedicine" and "medicines" would match "*medicine*" but "medical" would not.
[2] Key terms are not case-sensitive.

D. Narratives

To develop further understanding of the videos being analyzed, a topic modelling analysis using gensim's Latent Dirichlet Allocation (LDA) module was performed on trigrams based upon collected video titles. Hyperlinks and numbers were removed from the title corpus, tokenized with the ntlk library's word_tokenize method, and then lemmatized using parts-of-speech tagging through the Stanford Core NLP pipeline [10]. Table III contains the seven narratives that were identified from the video title corpus.

The narratives generally align with the intervention subgroups in Table IV. For example, over 70% of videos in the

TABLE III
VIDEO BREAKDOWN BY VIDEO NARRATIVE

No.	Description	Key Terms	Volume	Volume %
1	Self-imposed & mandatory quarantines	china inside quarantine epidemic mandatory	402	6.0%
2	Initial COVID-19 outbreak	outbreak china wuhan italy	389	5.8%
3	Spread of COVID-19 in USA	outbreak usa spread case new york	1314	19.7%
4	Politics & President Trump	trump president china protest	440	6.6%
5	USA COVID-19 Death Toll	death toll usa rise outbreak	1554	23.4%
6	Global Economic Impact	global economy pandemic impact	613	9.2%
7	Medical Treatments	vaccine pandemic treatment cure	1952	29.3%

TABLE IV
VIDEO BREAKDOWN OF INTERVENTION VS. NARRATIVE

		Intervention			
No.	**Description**	***Indi.***	***Govt.***	***Med.***	***Other***
1	Self-imposed & mandatory quarantines	39.9%	19.6%	1.8%	4.5%
2	Initial COVID-19 outbreak	10.2%	16.0%	1.9%	5.0%
3	Spread of COVID-19 in USA	14.6%	14.2%	10.3%	22.3%
4	Politics & President Trump	8.5%	6.8%	4.9%	6.7%
5	USA COVID-19 Death Toll	6.7%	15.5%	6.5%	27.9%
6	Global Economic Impact	5.0%	7.9%	3.1%	10.5%
7	Medical Treatments	15.2%	20.1%	71.4%	23.0%

medical interventions subgroup also fell into the coronavirus vaccine and treatment development narrative. The primary narrative on videos about individual interventions, which were defined as behaviors such as mask-wearing, self-isolating after exposure, and maintaining social distancing, was about quarantine measures. Coronavirus-related videos that did not fall into an intervention subgroup were more likely to have narratives around the spread of coronavirus and the increasing death count in the US.

E. Quantifying Engagement and Discussion on YouTube

We developed two indices to describe the discussion and interactions on a video: engagement and polarization. Engagement is roughly how "popular" a video is and describes the degree of viewer interaction on the video. Polarization concerns how "controversial" the video content is by measuring the contrast between viewer opinions on the video. Both indices are scaled to be between 0 and 100.

We quantify polarization and engagement on YouTube videos by computing their corresponding indices. For each video, these indices will change over time as the contributing factors and the events around the topic change. Both indices range from 0 to 100.

Polarization measures the contrast between viewer opinions on the video. Some platforms, such as Reddit, compare like and dislike ratios as a measure of controversy for their post sort options. We use a similar method to extract the contrast between viewer opinion with the polarization index. Equation (1) calculates the polarization index, namely , using the aggregated likes and dislikes on the video. Higher polarization is strongly correlated with ratio of likes (l) to dislikes (d) being equal to 1.

$$\pi = 100 - (l - d)/(l + d) \tag{1}$$

Engagement measures the degree of viewer interaction on the video. Users can directly interact and leave feedback through viewing the content itself, writing comments and replies, and adding likes or dislikes on the video. We apply a scaling factor for the volume of video views (v), comments (m), likes (l), and dislikes (d) and use a logarithmic transform to calculate the engagement index in (2).

$$\sigma = \log(c_1 * v + c_2 * (d + l) + c_3 * m)) \tag{2}$$

for some scaling factors c_1, c_2, c_3. We use scaling factors to accommodate varying magnitudes in the number of views, likes, and dislikes on a given video. In our data, we observe that Empirical results showed $c_2/c_3 = 10$ and $c_1/c_2 = 100$. The logarithmic transform reduces skew from the distribution of video view count.

F. Search Result Ranking

To better understand which videos are promoted by YouTube's search ranking algorithm, we tracked videos that appeared in the top ten search results for longer periods of time. 1,239 (18.6%) of the collected videos appeared in the top 10 search results for coronavirus-related search terms for more than 3 days. Videos that remained in the top ten search results over multiple scrape attempts are represented in the video data set multiple times, with resulting scores calculated as of each time information had been extracted.

III. Results

A. Index Comparison

Engagement and polarization differ between the coronavirus intervention video subgroups. Fig. 1 and Table V display the distribution of index scores.

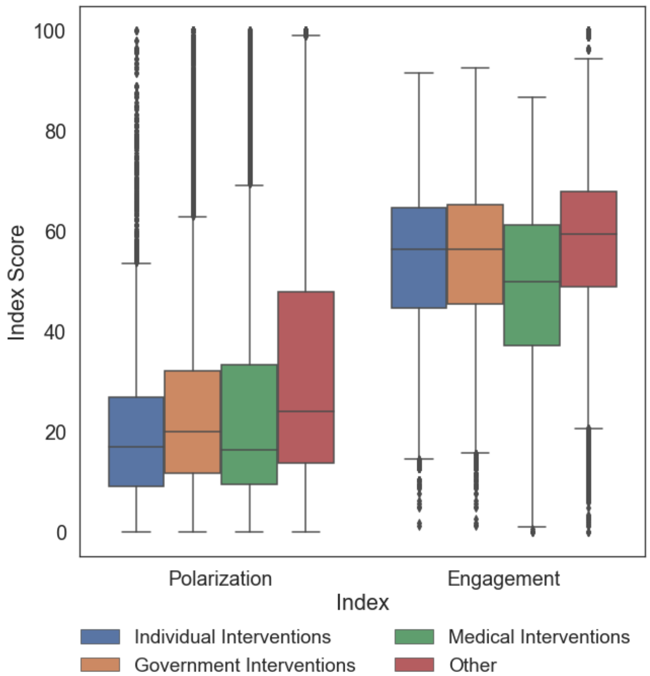

Fig. 1. Boxplot of metrics on videos per intervention subgroup.

TABLE V
Metrics For Video by Intervention Subgroup

		Intervention			
Index	**Metric**	***Indi.***	***Govt.***	***Med.***	***Other***
Polarization	Median	17.74	21.05	16.99	24.99
	Average	23.15	27.95	26.33	33.30
	Std. Dev.	18.18	22.08	24.30	24.62
Engagement	Median	56.44	56.41	48.87	59.39
	Average	54.06	54.31	49.05	57.75
	Std. Dev.	16.72	15.77	15.85	14.46

Coronavirus-related videos not about interventions, such as general news reporting, had the highest polarization score of the subsets with a median polarization index of 24.99. While both medical and individual intervention videos were less polarized, with scores of 16.99 and 17.74, respectively, the standard deviation of polarization for medical intervention videos is similar to the standard deviation on non-intervention videos. While generally the content of medical-intervention videos is not particularly polarizing, there are some videos that have controversial content for their viewers. By comparison, videos about individual interventions had both the lowest polarization score (17.74) and distribution (18.18). Individual interventions, like voluntarily self-quarantining or wearing a mask, are not inherently associated with political matters. Additionally, the narratives less associated with individual interventions are the USA coronavirus death toll and global economy, which are observed to be more polarizing narratives.

Videos about medical interventions are the least popular subset, with a median engagement score of 48.87, compared to over 56.4 for both individual and government intervention-related videos, as well as 59.39 for non-intervention coronavirus videos.

Polarization scores increase across time as shown in Fig. 2. Medical intervention videos have the greatest increase with a slope of 0.21 when applying a linear regression model. For comparison, the slope for government intervention videos was 0.11, for non-intervention ("other") videos 0.09, and approximately zero for individual intervention videos.

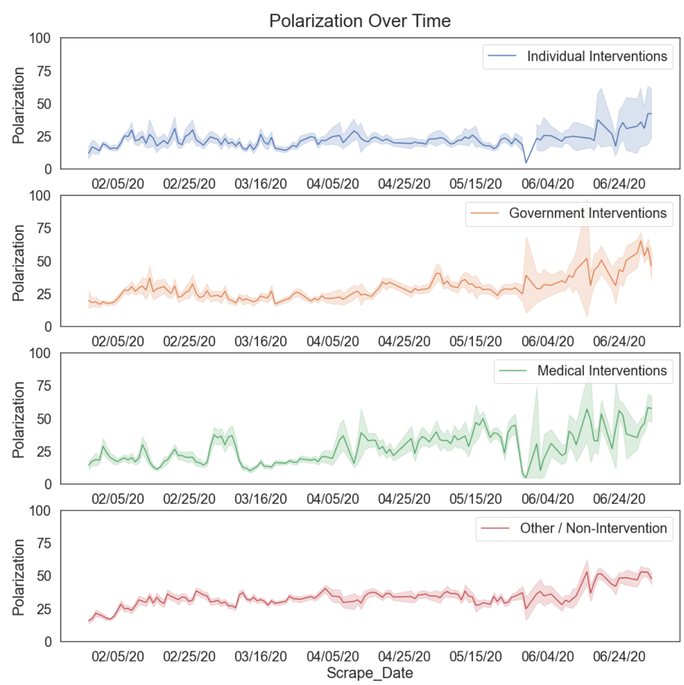

Fig. 2. Polarization on videos per intervention subgroup over time.

Engagement increases the most for videos about individual interventions as depicted in Fig. 3, with a slope of 0.10, as opposed to 0.06 for government, 0.04 for medical, and -0.02 for non-intervention videos. While the net increase is small for medical intervention videos with a linear regression slope of 0.04, we observe that engagement increased in the months of April and May 2020 but tapered off in June 2020

B. Narrative Comparison

The two metrics for audience engagement were calculated on the video data subset by narratives expressed in the video title and are depicted in Fig. 4 and Table VI. The highest degree of polarization is observed on videos with narratives around the coronavirus-related death counts in the United States. While videos with political narratives on average are

Fig. 3. Engagement on videos per intervention subgroup over time.

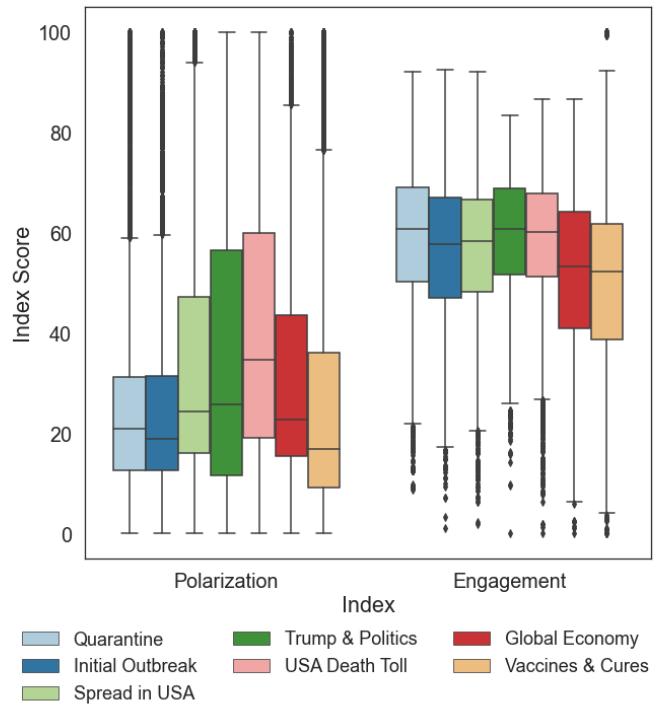

Fig. 4. Boxplot of metrics on videos by video title narrative.

not particularly polarized, with a median score of 29.27, their spread is the highest of all the narratives at a standard deviation of 27.05.

Videos with narratives around the novel coronavirus's effect on the global economy, as well as with narratives around medical treatments and vaccines for coronavirus receive the least engagement, with scores of 53.21 and 52.29, respectively. In comparison, videos about quarantines and about politics, in particular U.S. President Trump, have higher engagement scores of above 60.7.

TABLE VI
METRICS FOR VIDEO BY NARRATIVE

#	Description	Polarization *Med.*	Polarization *Avg.*	Polarization *Std. Dev.*	Engagement *Med.*	Engagement *Avg.*	Engagement *Std. Dev.*
1	Self-imposed & mandatory quarantines	22.26	26.78	19.96	60.74	59.51	14.47
2	Initial COVID-19 outbreak	19.30	26.60	21.14	57.67	55.92	14.06
3	Spread of COVID-19 in USA	26.84	34.80	23.43	58.34	56.53	13.34
4	Politics & President Trump	29.27	36.49	27.05	60.76	59.72	13.74
5	USA COVID-19 Death Toll	35.80	41.24	24.90	60.10	58.73	13.10
6	Global Economic Impact	22.20	32.05	24.84	53.21	52.04	15.16
7	Medical Treatments	18.18	27.34	24.35	52.29	50.27	16.29
Grand Total		22.90	31.29	24.32	57.41	55.63	15.31

C. Video Prominence in Video Search Results and Index Development over Time

1,239 (18.6%) of the collected videos appeared in the top 10 search results for coronavirus-related search terms for more than 3 days. These videos were prominent in search rankings for a median of 6 days. Characteristics of videos that persist in search results are shown in Table VII. The narratives for repeated videos tend to be more around quarantines, the initial spread of COVID-19 and less around the USA COVID-19 death toll or the global economic impact of COVID-19.

Tracking the videos that remained in search results for over 3 days showed that both calculated indices increased over time. When comparing the indices from the time the video was first scraped to the indices on August 2, the median polarization index went from 23.2 to 24.9 and median engagement increased from 47.2 to 72.7. The growth of the polarization index suggests that the longer videos are available for public consumption, the less valuable additional engagement is on the video. The videos promoted by the YouTube search results page for longer periods of time have 9.9% higher engagement and 37.0% lower polarization scores than videos that do not remain in the first 10 search results. As YouTube's interest is in keeping users on the platform longer, this combination suggests that YouTube prefers less controversial videos in its video ranking.

As seen in Fig. 5, the subset of videos that drop off of the first search results page the fastest is non-intervention coronavirus videos. Of the intervention subgroups, however, medical

TABLE VII
METRICS BY PROMINENCE IN SEARCH RESULTS

Metric	Videos ranked for $<$ 3 days	Videos ranking for 3+ days	All Videos
Volume	5,429	1,239	6,668
% Livestream	5.23%	3.03%	3.78%
Views[1]	62,052	178,625	123,449
Duration[12]	250.5	240	246
Transcript Available	65.45%	72.00%	69.76%
Likes[1]	668	1,600	1,200
Dislikes[1]	130	201	173
Comment Volume	471	956	751
Polarization Index[1]	31.34	19.75	22.90
Engagement Index[1]	53.87	59.22	57.41

[1] Median value displayed.
[2] Measured in seconds (s).

intervention videos have the steepest decline in continued search ranking, which suggests that YouTube is promoting new different content relating to medical interventions. Individual intervention videos are the most persistent, with 40.5% of videos still returning in the top ten search results three days after first appearance.

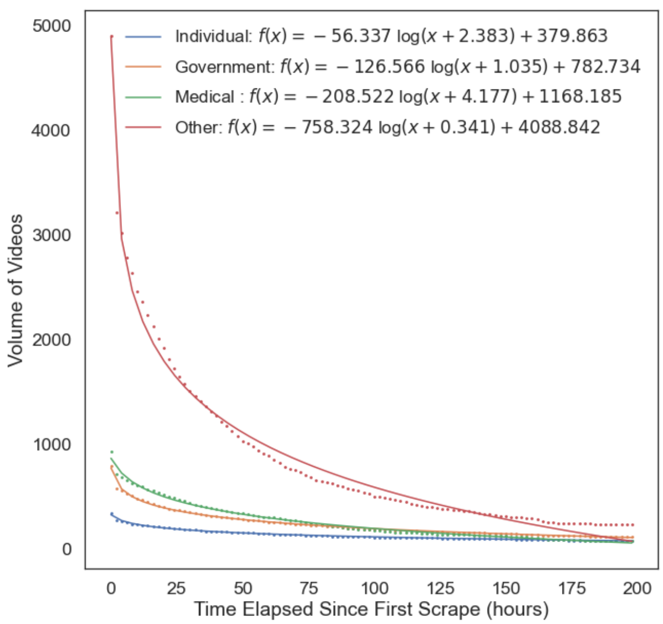

Fig. 5. Volume of intervention videos by length of time in top 10 search results, shown with fitted logarithmic growth functions.

The video narrative that had the sharpest decline in video prominence was the development of coronavirus vaccines and treatments, as shown in Fig. 6. These videos fell out of the top ten search results over 25% faster than videos with the highest polarization narrative, which is around the USA coronavirus death toll. This observation implies that the video ranking model that powers YouTube's search results considers factors beyond how controversial viewers find the video content.

Videos about the global economic impact of coronavirus also have low engagement scores, though they maintain their status in YouTube's top ten search results once present. This observation implies that viewers may simply not be interested in watching economic news about coronavirus.

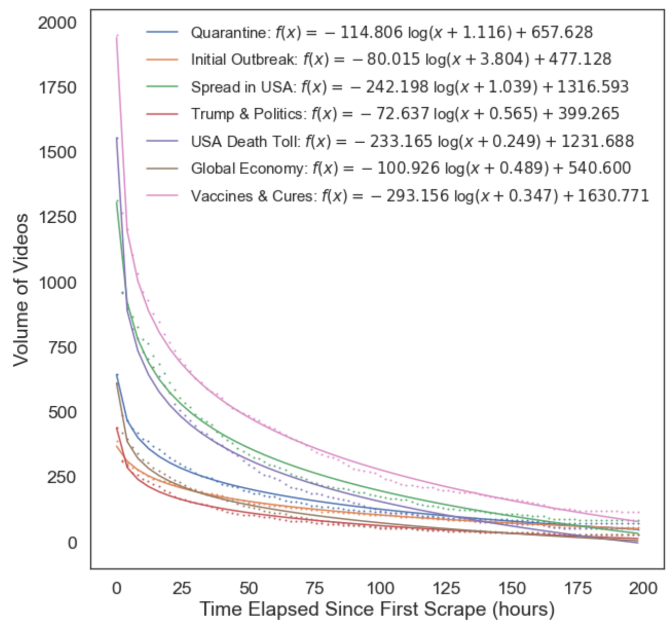

Fig. 6. Volume of videos by length of time in top 10 search results, shown with fitted logarithmic growth functions.

IV. DISCUSSION

We set out to quantify the discussion and engagement on coronavirus-related YouTube videos. The quality of viewer interactions with the video, as well as the way YouTube promotes certain videos in its search result pages, was monitored by tracking the top ten search results for various coronavirus-related search queries. In addition to comparing metrics across intervention subgroups and video narratives, we also investigated how YouTube as a platform affects which videos are shown to viewers by observing the videos that were promoted by YouTube's search result ranking algorithm and remained in the first page of video results for longer periods of time. Through our research, we are able to answer the following research questions:

RQ1. Do the narratives around COVID-19 video content highlight public intervention methods?
RQ2. Does polarization affect which videos are promoted on YouTube?
RQ3. Do YouTube videos about COVID-19 interventions have greater engagement and polarization than other COVID-19 content?

In general, we found two of the seven narratives involve public health interventions: 1) self-imposed & mandatory quarantines, which corresponds to the "individual interventions" subgroup, and 2) medical treatments, which has overlap with the "medical interventions" subgroup. Medical treatment related content is the most commonly observed narrative in our collected YouTube video dataset. Other narratives served to inform viewers about the origin, spread, economics, and politicization of the disease. Videos about coronavirus deaths in the USA were not only the most popular narrative but also the most polarized, even in comparison to videos about political topics and the U.S. President. Updates on the number of coronavirus-related deaths may be watched by bipartisan viewers with a wider variety of viewpoints, as opposed to political videos, which might appeal more to a narrower audience. These highly polarized videos about the death toll decline in search results on YouTube quickly, which aligns with our observation that polarized videos are less likely to be promoted in search results over time.

In general, it appears that YouTube puts preference to less polarizing videos when returning videos for a search query. Videos promoted in search results for over three days were 37.0% less polarized than videos that do not remain in the first page of search results. YouTube may reduce the ranking for controversial content that receives more down votes and negative activity. Videos about individual interventions have the lowest polarization scores and spread, as well as the least growth in polarization over time. This may contribute to why YouTube's ranking algorithm promotes individual intervention videos for longer periods of time than other videos.

Comparing videos from when they first appeared in search results to later scrape times shows that, as expected, engagement accumulates over time for each individual video. However, polarization also increased, which implies that older videos tends to attract watchers who disagree more with the content. When looking at the entire video data set over time, engagement decreased for non-intervention videos but increased for intervention-related videos, suggesting there is more interest in combating coronavirus through learning more about public health interventions than at the start of the pandemic. In particular, the viewership and engagement on videos about individual interventions to prevent coronavirus spread increased the most over the observed time period. While the net increase in engagement is small for medical intervention videos, we observe that relative engagement increased in the months of April and May but tapered off in June 2020. Despite this increase, videos about medical interventions are the ultimately least popular subset of the observed videos.

Videos with narratives around COVID-19 vaccine and treatment development share significant overlap with medical intervention videos. Both groups of videos have the least viewer engagement, which could stem from a lack of promotion by YouTube's search ranking model. This finding is in conflict with observations of how video polarization inversely correlates with YouTube promotion as medical-related videos have the lowest polarization metrics across the entire COVID-19 video data set. For example, medical intervention videos decline in search result rankings at a 64.8% faster rate than government intervention videos. Videos about vaccines and treatments development dropped out of the top ten results 25% faster than videos about the U.S. coronavirus death toll, which have the highest observed video polarization scores. Medical videos may have particularly high turnover in the search results due to constantly updating new developments. We note that political-narrative videos and economic-related videos do not seem to be adversely affected in search result rankings despite being topical as well.

For future research, we plan to refine engagement and polarization measures. Engagement does not account for how long a video has been available, so this index can be improved to factor in time as an element to identify videos that are more "viral" and attract attention in a short time span as opposed to videos that steadily receive comments and views. In addition to user interactions within the YouTube platform, integrating data from other social media sites can provide us insights into users and social networks. Our polarization index is currently calculated at the video-level, so we plan to expand the polarization index by considering the type and polarity of discussion in the comments section through sentiment analysis.

Two approaches were used for classifying YouTube videos by content—one by whether the video is related to public health interventions, and the other by identifying narratives in the video title. Both methods rely on the video title accurately describing the video content, which means that some videos may be miscategorized due to vague, unspecific, or unrelated titles. We also plan to improve video categorization by incorporating video descriptions and transcripts if available to better account for variances in video titles.

Acknowledgment

This research was funded by Accenture Applied Intelligence to support clients responding to SARS-CoV2, better estimate future impacts, and understand discussions on under-utilized social media platforms.

References

[1] C. Funk and A. Tyson, "Partisan differences over the pandemic response are growing," June 2020.
[2] WHO, *Working together to tackle the "infodemic"*. WHO, Jun 2020.
[3] P. S. Hart, S. Chinn, and S. Soroka, "Politicization and polarization in covid-19 news coverage," *Science Communication*, vol. 42, no. 5, pp. 679–697, 2020.
[4] A. Perrin and M. Anderson, "Share of u.s. adults using social media, including facebook, is mostly unchanged since 2018," Apr 2019.
[5] R. Dugyala and K. Rahman, "6 things to know about gen z, politics and 2020," Oct 2020.
[6] G. Stocking, P. van Kessel, M. Barthel, K. E. Matsa, and M. Khuzam, "Many americans get news on youtube, where news organizations and independent producers thrive side by side," Sep 2020.
[7] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi, "Users polarization on facebook and youtube," *PLOS ONE*, vol. 11, pp. 1–24, 08 2016.
[8] A. Abisheva, V. R. K. Garimella, D. Garcia, and I. Weber, "Who watches (and shares) what on youtube? and when? using twitter to understand youtube viewership," in *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, (New York, NY, USA), p. 593–602, Association for Computing Machinery, 2014.
[9] S. Lam, J. Jin, S. Alawi, O. Savas, and I. McCulloh, "Evolution of misinformation, polarization, and engagement in youtube social networks: A case study of public health interventions for covid-19." (Poster Presentation at NETSCI2020 Conference), September 2020.
[10] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60, 2014.