

Final Report

on

**Identifying New Messengers and
Misinformation/Disinformation Actors Using Social Media
Posts**

*Submitting as a part of **DACSS604 - Advanced Data Driven Storytelling** for the partial
fulfillment of the requirement for the final course/degree.*

Submitted by

Krishna Chaitanya Rao Kathala & Sathvik Thogaru
MS Data Analytics and Computational Social Sciences,
University of Massachusetts Amherst.

To

Course Instructor

Asch Harwood
Advanced Data Driven Storytelling

Client

Purpose - A Global Social Impact Consultancy

EXECUTIVE SUMMARY

Online misinformation and disinformation about the coronavirus pose a serious risk to the public's health, the former involves deliberate dissemination of false or misleading information, while the latter involves unintentional dissemination of the same. Social media platforms have made it easier for people to get information, which has hampered the public health response, causing widespread confusion, and increased the number of deaths during the pandemic. Looking ahead, the Center for American Progress [1] expects disinformation and misinformation about the coronavirus to shift and worsen. As public health conditions vary more widely across the United States, geographic variation is an ideal vector for malicious actors to exploit. With lack of robust local media ecosystems, it will be especially difficult for social media platforms to moderate place-based disinformation and misinformation. Purpose is a global social impact consultancy that works to build and support movements to advance the fight for an open, just, and habitable world. We use public mobilization and storytelling to help the leading organizations, activists, businesses, and philanthropies engaged in this fight, and we create campaigning labs and new initiatives that can shift policies and change public narratives when it matters most. In March 2020, Purpose and the United Nations launched Verified, a global communications infrastructure to address the most pressing impact of mis- and disinformation on COVID-19 and promote accurate, science-backed information. The objective of this study is to:

- Identify additional social media profiles/activities of current messengers,
- Identify new potential messengers using general social listening and network analysis, and
- Utilize social media content posts to identify any potential bad actors who we may wish to exclude from further engagement activities.

BACKGROUND

Misinformation about the coronavirus has plagued effective public health responses from the start. Early on, the World Health Organization (WHO) warned of infodemic of online misinformation [2] As public health officials urged the public to stay home and flatten the curve, experts raised the alarm that their messages were competing with a tide of misinformation and disinformation online [3][4][5][6] The uncertainty surrounding the coronavirus, paired with intense global demand for information, created a perfect storm of speculation, conspiracy, and sharing of false or even harmful information. Complicating the situation, prominent public figures—including celebrities and politicians—were among the primary drivers of engagement around COVID-19 misinformation in early 2020 [7] For average users and influencers alike, coronavirus misinformation is so prevalent on social media that it has become increasingly difficult to avoid participating in spreading false or misleading narratives. Compounding the problem in the United States, disinformation producers seized on the COVID-19 pandemic as a way to advance their goals and agendas by accelerating chaos online. Disinformation producers—and far-right and white supremacist ecosystems in particular [8] —have long sought to exploit crisis moments to amplify false, conspiratorial, and hateful narratives. Conspiracy theories have generally thrived in crisis [9] but the modern social media environment and the sudden forced movement of attention online during stay-at-home orders have been a gift to malicious actors. Leveraging prevailing uncertainty, demand for information, and an audience stuck online, these groups have effectively deployed disinformation strategies to pervert perceptions of public opinion and warp public discourse for their own gain. The coronavirus crisis was merely the most recent topic weaponized by the far-right disinformation ecosystem to spread falsified information, racist narratives, undermine democratic institutions, and cause chaos [10]. We first go over various terminologies pertaining to misinformation that differ from one another. After the 2016 US presidential election, "fake news" received a lot of media and academic attention. The phrase is

used interchangeably with different types of false information, including misinformation and disinformation, both of which disseminate ideas, rumors, and hypotheses quickly through personal connections or internet forums. They differ mostly in terms of goal and method of dissemination. While disinformation entails intentionally creating and disseminating false information with the purpose to harm, misinformation refers to inaccurate information that is accidentally generated and transmitted (Wardle and Derakhshan 2017). Lazer et al(2018) .'s definition of "fake news" as made-up information that closely resembles news media content does not adequately convey the complexity of the phenomenon, which can also include satire and information purposefully created to mislead in order to further a political or other objective (Wardle, 2017). Here, we'll concentrate on false information about health and wellbeing. Nandita Krishnan and colleagues examined 12 prominent social media and messaging platforms' (SMPs') responses to COVID-19 rumors by analyzing community standards, official press releases, and blog entries. Although the majority of platforms claimed to prohibit COVID-19 misinformation, several platforms' reactions lacked openness and clarity. Facebook, Instagram, YouTube, and Twitter responded in a relatively uniform manner, although other platforms differed in terms of material kinds that were forbidden, standards for how to respond, and strategies for dealing with false information. Only Twitter and YouTube explained how they go about using different treatments. These variations underline the necessity of establishing uniform standards across platforms to more effectively combat COVID-19 disinformation.

Purpose is actively monitoring the social media posts of messengers to the extent that they have provided social handles on a range of platforms including Facebook, Instagram, Twitter, Tiktok, and Youtube. In this project we are working to identify additional social media profiles/activities of current messengers, also identify new potential messengers using general social listening and network analysis, and utilize social media content posts to identify any potential bad actors who we may wish to exclude from further engagement activities.

PROBLEM STATEMENT

Purpose is a global social impact consultancy that works to build and support movements to advance the fight for an open, just, and habitable world. We use public mobilization and storytelling to help the leading organizations, activists, businesses, and philanthropies engaged in this fight, and we create campaigning labs and new initiatives that can shift policies and change public narratives when it matters most. In March 2020, Purpose and the United Nations launched Verified, a global communications infrastructure to address the most pressing impact of misinformation and disinformation on COVID-19 and promote accurate, science-backed information. Purpose has built a network of over 2000 "messaging" and over 80000 "information volunteers" to help disseminate Verified campaign content. These messengers have been engaged through a range of mechanisms, including digital recruitment, in country programs, and partnerships. We are actively monitoring the social media posts of messengers to the extent that they have provided social handles on a range of platforms including Facebook, Instagram, Twitter, Tiktok, and Youtube. The pressing challenges that were addressed in this study are as follows: identify new potential messengers using general social listening and network analysis, and utilize social media content posts to identify any potential bad actors who we may wish to exclude from further engagement activities.

PROPOSED SOLUTION

In this study we devised a rule engine to define a Good influencer and a Bad influencer based on the descriptive statistics, the rule engine is defined with an aim to identify new potential messengers using general social listening and network analysis, and later on make use of the social media content posts to identify any potential bad actors who we may wish to exclude from further engagement activities.

We defined an Influencer to be a potential messenger/good influencer if the posts sentiment is positive and has a certain number for Reach and on the value of AVE (Advertising Value Equivalency)

Good actor / Good influencer / Potential Influencer:

Condition: {Sentiment = positive, Reach > X, source, AVE > Z},)

We defined an Influencer to be a Good actor for the Positive sentiment of posts; and threshold of Reach and AVE values as defined by the employer.

Bad actor/bad influencer:

Condition: {Sentiment = negative, Reach < X, source, AVE < Z}

We defined an Influencer to be a Bad actor for the Negative sentiment of posts; and threshold of Reach and AVE values as defined by the employer.

Reach is defined as the total monthly visitors to the source or social followers to the Influencer.

AVE (Advertising Value Equivalency) is used to estimate the amount of revenue attributed to an article.

TECHNOLOGICAL SLACK

<u>Miro</u>	Brainstorming, Ideation
<u>JIRA/Agile</u>	Defined the weekly tasks, goals, deliverables, action items.
<u>Meltwater API</u>	Used to bring data from the Meltwater platform into your tools and solutions by using the Export Console to download data (JSON, CSV)
Python Programming	Creating a Meltwater saved search in Python using REST API
RStudio	Data Cleaning & Data Analysis
Library/Packages	R: Tidyverse, Lubridate Python: json, Requests
Figma	Wireframing, Prototyping
Google Data Studio	Building dashboards and visualization
GitHub	Code hosting platform for version control and collaboration.
Platforms explored	PowerBI, Tableau, Rapid Miner

Table 1: Technology Slack

METHODOLOGY

We had approached with a well defined methodological framework which includes right from the acquiring data through the Meltwater API, Data Consolidation to Data Cleaning, Data Wrangling and Normalization of the parameters to reduce the Reach and AVE (Advertising Value Equivalency) scale and further computing the statistical analysis which led an input to the Data Analytics and Visualization using the Google Data Studio for the Dashboarding

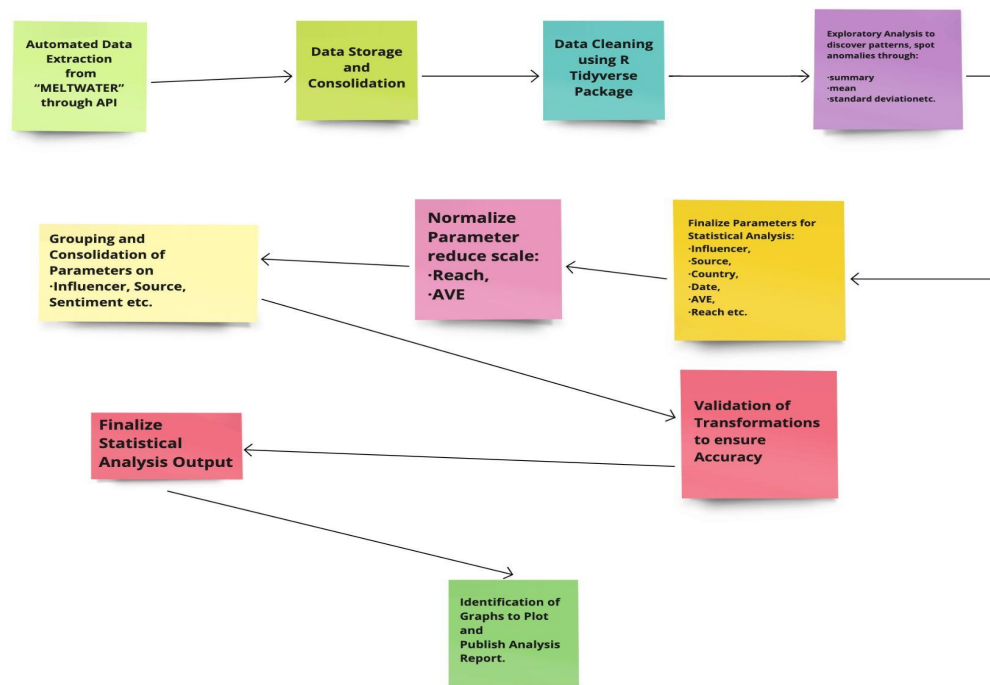


Figure: Methodological Framework for identifying potential influencers

We made use of the Meltwater API Export Console to create a one-time export of the social media data based upon our Meltwater Search Criteria. The data is available to download in JSON and CSV formats. To make a Meltwater Search we first need the Meltwater API credentials. The Meltwater API uses API

tokens to authenticate requests. For API calls authentication is performed by providing the API Token as a header named api key on the API request. All API requests must be made over HTTPS. Calls made over plain HTTP will fail. API requests without authentication will also fail.

The below mentioned code is used to create a new search in the Meltwater application and later export it using the Meltwater Exports console. We made use of the Visual studio code to run the below mentioned Python code.

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
# Install the requests library if it is not already installed -- pip3 install requests
import json
import requests
# Set up variables.
BASE_URL = "https://api.meltwater.com/v3/searches"
YOUR_MELTWATER_TOKEN = "*****"
SEARCH = ("vaccine" OR "covid" OR "Coronavirus" OR "Quarantine" or "Pandemic")
SEARCH_NAME = "DACSS604_Project"
headers = {"Content-Type": "application/json", "apikey": YOUR_MELTWATER_TOKEN}
# Meat of the API push.
r = requests.post(
    f'{BASE_URL}',
    headers=headers,
    data=json.dumps(
        {
            "search": {
                "type": "social",
                "category": "explore",
                "name": SEARCH_NAME,
                "query": {
                    "case_sensitivity": "no",
                    "not_keywords": ["Promotion", "offer", "marketing"],
                    "boolean": SEARCH,
                    "type": "boolean",
                },
            },
        }
    )
)
```

Data Types

There are two types of Meltwater Exports:

- One-time exports - these exports are run once, the data will not be refreshed. One-time exports are used when we just want to get an export of data for a specific time period.
- Recurring exports - these exports give you the ability to set up an export once and get results updated regularly. Recurring exports are used when we need to have the latest data for the predefined time span e.g. “last week's data refreshed every week”.

For our project we used the one-time export which is compatible with the resources that we have and the time given for the project. Below is the snapshot of the Meltwater Export console and the results.

Export Console

Search exports

Create ExportRefreshDelete Selected

Searches / Tags	Details	Status	Data
<input type="checkbox"/> DACSS604_Project {19742665}	ONE-TIME Fri, 2022-07-01 13:00 UTC to Sat, 2022-07-02 01:00 UTC	FINISHED	JSON CSV
<input type="checkbox"/> DACSS[Mis-Information] {19642117}	ONE-TIME Sat, 2022-08-13 04:00 UTC to Sun, 2022-08-14 04:00 UTC	FINISHED	JSON CSV
<input type="checkbox"/> DACSS[Mis-Information] {19642117}	ONE-TIME Mon, 2022-08-08 04:00 UTC to Tue, 2022-08-09 04:00 UTC	FINISHED	JSON CSV
<input type="checkbox"/> DACSS[Mis-Information] {19642117}	ONE-TIME Sun, 2022-08-07 14:00 UTC to Sun, 2022-08-07 22:00 UTC	FINISHED	JSON CSV

<<<1>>>

Figure: Snapshot of the Meltwater Export console and the results

The CSV format is ideal for importing into a spreadsheet or BI Tool. The JSON format can also be imported into a BI Tool which can process JSON data. Note that the CSV format includes less fields than the JSON format as it isn't possible to express all the available data in CSV format. Take a look at the Exports Formats & Fields page for full details. We downloaded the CSV format to work on our project.

Fields	Description
Influencer	Name or Social handle of the Author
Source	Social network or News Publication
Country	Country of the author
Reach	Reach is defined as the number of the unique visitors of each source based on monthly activity
AVE	Ave Value assigned to the Source. Advertising Value Equivalency is used to estimate the amount of revenue attributed to an article.
Sentiment	positive, neutral, negative, unknown.

Table 2: Fields used in the Data Analysis are shown below

[Advertising Value Equivalency](#): The logic used while calculating an online advertising value equivalency.

The logic we use when calculating an online advertising value equivalency is : $X * 0.025 * 0.37$

- X (the reach/unique visitor figure)
- * .025 (standard error, assuming that 2.5% of any given audience will view a particular article on average)
- * .37 (37 cents is the dollar value for each visitor). This figure can be adjusted if you want to place more or less value on each viewer.

Data Cleaning:

In our data cleaning process we imported the data that we downloaded from Meltwater Export Console into the RStudio IDE.

We made the below mentioned following operations on the data

- Filtered the rows that have the Source as "facebook", "twitter", "instagram", "youtube", "reddit" and "Tiktok". We didn't consider the other social comments and posts of News articles as the source authentication for these articles seemed challenging with the resources that we have.
- We then converted the Date column from character type to Date Type using the Lubridate library in R
- The rows for the Source: Twitter has Influencer as "NA" and has Tweet Id and Twitter ID. Other Source Types have Influencer Name in the Influencer Column and have no related ID's. So The influencer Name for twitter source was replaced with Twitter ID of the Influencer.
- Later the data was stored as a csv file and used in Google data sStudio for further analysis using Visualization

Analysis:

Our approach was to identify how an influencer is influential based on the below parameters:

- Country wise Influencer wise Reach
- Country wise Influencer wise AVE
- Country wise Influencer-Sentiment wise Reach
- Country wise Influencer-Sentiment wise AVE

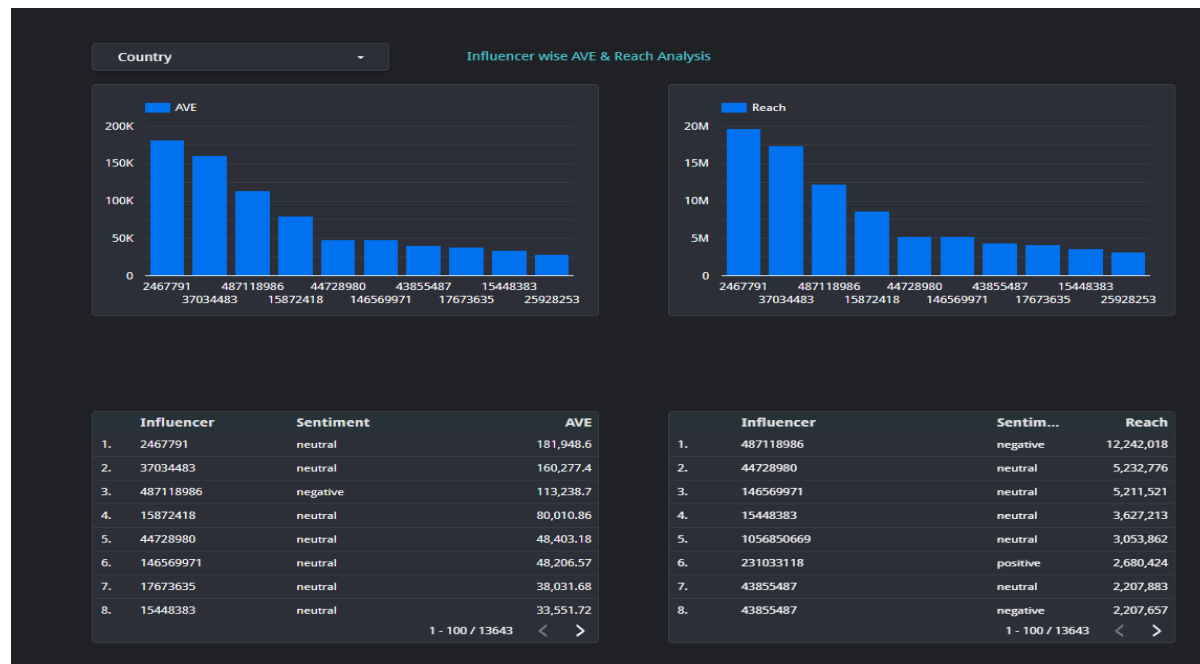


Figure: Influencer wise AVE & Reach Analysis along with Sentiment Scores

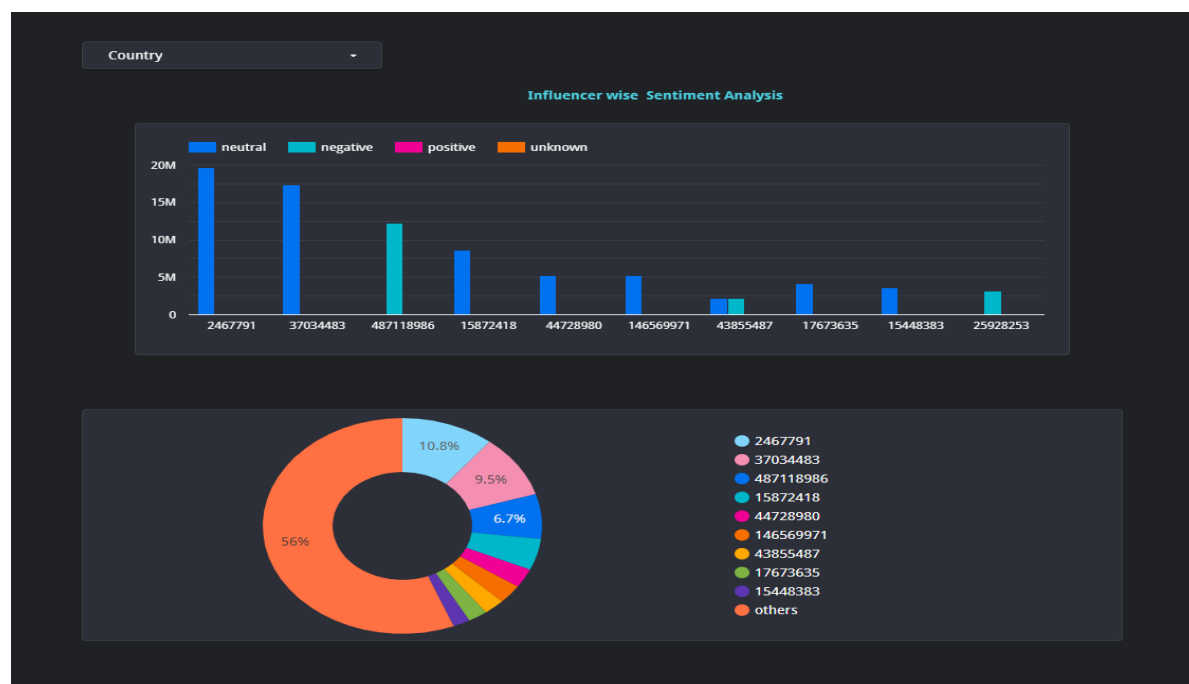


Figure: Influencer wise Sentiment Analysis

Risks:

- Influencer's reputation and Employer's reputation become synonymous.
- Communication and Digital Media Campaign Content should be clearly communicated to have control over the campaign, else it may create conflict.
- Any contract breach, any post misrepresenting the facts about the campaign may pose a serious risk to the employer.
- Litigation on Employers because of Influencers posts on any similar product/service or any postsInsurance Claims Contract Breach

Limitations/Challenges:

- Data is limited
- Authenticating the information
- Accessing the Meltwater Application
- Limited learning Resources
- Getting timely data
- Data Cleaning
- Identifying the potential Sources

Reproducibility:

- Documentation of the approaches (Risks, Challenges, Roadblocks)
- Defined Data Master Catalog for ease of access
- Data Analysis codes - R (Github)
- Defined Product Requirements Document (PRD) for scaling
- Setting up Agile/Scrum based process flow
- Figma - wireframing

Go-to-Strategy:

- We can target negative influencer, preparing a strategy to reduce AVE value
- And use/hire positive influencer to increase the reach, because people are listening to them even without having authentic source
- Add authentic source and factual information in the posts of positive influencers to cleanse out dis/mis-information
- Language, region/location specific outreach planning can be done for social media campaigns, as the parameters are available.
- Monitoring and Evaluation of such campaigns through online activities of the targeted Influencers. Also monitoring of hashtags on social media during the campaign to monitor the impact of the influencer as well as the campaign

Conclusions & Future Recommendations:

Overall, this study introduces methods to identify good influencers based on the key parameters like sentiment score, Advertising Value Equivalency (AVE), Reach. We can make a strategy by monitoring negative influencers and developing a plan to lower AVE value are options. Additionally, employ positive influencers to broaden the audience because people still pay attention to them even when they don't come from a reliable source. Going forward, we will focus on developing the additional features on the dashboard at scale, including more comprehensive positive influencer tweet collection and catalog discovery and collection. We will be interested to see a comparative study to identify the good influencer vs bad influencer based on the reach, sentiment and engagement on the meta platforms vs non meta platforms. Also, we would like to address the following listed below:

- 1) In which platform do more bad actors exist based on the data analysis in the specific time frame?
- 2) What are the most frequent words used by the good influencer in their hashtag/tweet - this could be a word cloud based on the “key phrases”
- 3) How does the positive and negative sentiment trend looks like over the timeframe (maybe we could break this into more specific based on the hit ratio, platform, source)
 - a) Unique users vs AVE vs Sentiment
- 4) Are there any interesting trends where a bad influencer turned to a good influencer
- 5) Location: In which region or country where the most good vs bad influencers are based on the sentiment, reach, AVE and engagement over the timeframe.

Deliverables

Deliverables			
Sr#	Type	Platform	URL
1	Dashboard	Google Data Studio	https://datastudio.google.com/u/0/reporting/a0065597-e807-4113-95c8-1df3858f6155/page/LIH0C
2	Source Code	Github	https://github.com/ThogaruSathvik/Dacss_604_project

Team Structure

Detail Team Structure				
Sr#	Name	Designation	Company	Email Id
1	Levi Braslow	Global Head of Impact, Measurement & Learning,	Purpose	levi.braslow@purpose.com

2	Sarah Barr	Manager - Impact, Measurement, Learning	Purpose	sarah.barr@purpose.com
4	Nathaniel Olin	Associate Director on the Impact, Measurement & Learning team	Purpose	nathaniel.olin@gmail.com
5	Asch Harwood	DACSS 604 Instructor	UMass -DACSS	aharwood@umass.edu
7	Krishna Chaitanya Rao Kathala	DACSS- Graduate Student	UMass -DACSS	kkathala@umass.edu
8	Sathvik Thogaru	DACSS- Graduate Student	UMass- DACSS	sthogaru@umass.edu

Mode of Communication:

Communication Channel			
Sr#	Channel	Handle	Purpose
1	Email	Official email id	Project communication with single or multiple recipients with or without attachments
2	Zoom	Meetings	Recurring Meetings (https://umass-amherst.zoom.us/j/7887511418)

REFERENCES

- [1] Fighting Coronavirus Misinformation and Disinformation, Preventive Product Recommendations for Social Media Platforms
- [2] World Health Organization, “Munich Security Conference Remarks: Transcript of Dr. Tedros Adhanom Ghebreyesus’ speech at the World Health Organization, Munich Security Conference,” February 15, 2020, available at [munich-security-conference](https://www.who.int/news-room/feature-stories/munich-security-conference-remarks).
- [3] Brandy Zadrozny, “These disinformation researchers saw the coronavirus ‘infodemic’ coming,” NBC, News, May 14, 2020, available at [infodemic coming](https://www.nbcnews.com/health/infodemic-disinformation-researchers-saw-coronavirus-coming-ncna102886)

- [4] Hubert Au, Philip N. Howard, and Jonathan Bright, “Coronavirus Misinformation: Weekly Briefings,” Oxford Internet Institute, April 2020, available at [coronavirus-weekly-briefings](#)
- [5] Brandi Collins-Dexter, “Canaries in the Coalmine: COVID-19 Misinformation and Black Communities” (Cambridge, MA: Shorenstein Center on Media, Politics and Public Policy at Harvard Kennedy School, 2020), available at [Covid-19 Misinformation and Black Communities](#)
- [6] Institute for Strategic Dialogue Digital Research Unit, “Covid-19 Disinformation Briefing No.1” (London: 2020), available at [DDR Unit](#)
- [7] Julie Ricard and Juliano Medeiros, “Using Misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil,” Harvard Kennedy School Misinformation Review 1 (2) (2020): 1–6, available at [misinfo review hks harvard](#)
- [8] Alice Marwick and Rebecca Lewis, “Media Manipulation and Disinformation Online” (New York: Data & Society, 2015), available at [datasociety](#)
- [9] Roger McNamee, “Social Media Platforms Claim Moderation Will Reduce Harassment, Disinformation and Conspiracies. It Won’t,” Time, June 24, 2020, available at [social media platforms claim moderation will reduce harassment disinformation and conspiracies it wont](#)
- [10] Ryan, “Analysis: Against Excessive Quarantine Protests”; Carl Miller, “Far-right spreads Covid-19 ‘infodemic’ on Facebook,” BBC News, May 4, 2020, available at [BBC technology](#)