UMassAmherst

Data Analytics and Computational Social Science Program

PURPOSE

Executive Summary 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

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

Introduction:

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.

Client Description:

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.

Website: www.purpose.com

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 "messengers" 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

Resources Used:

Miro	Brainstorming, Ideation
Meltwater API	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
Google Data Studio	Building dashboards and visualization
GitHub	Code hosting platform for version control and collaboration. https://github.com/ThogaruSathvik/Dacss_604_project
Platforms explored	PowerBI, Tableau.

Timeline: Jun 11, 2022 - Aug 16, 2022

Team: Krishna Chaitanya Rao Kathala Sathvik Thogaru

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.

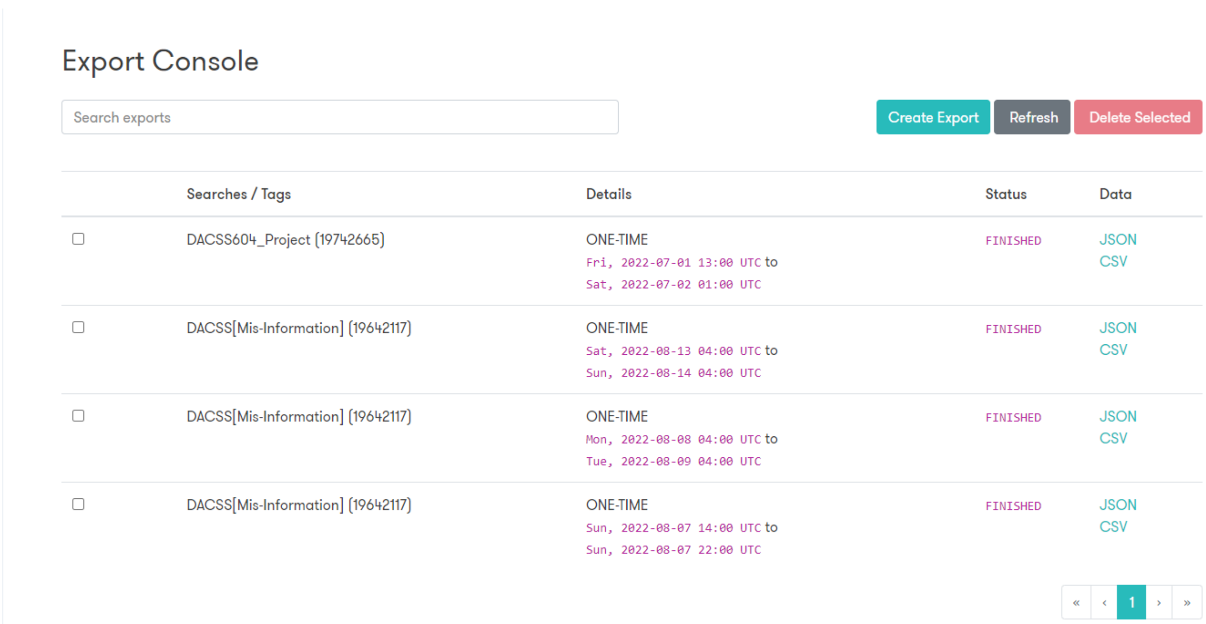

Figure: Snapshot of the Meltwater Export console and the results

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

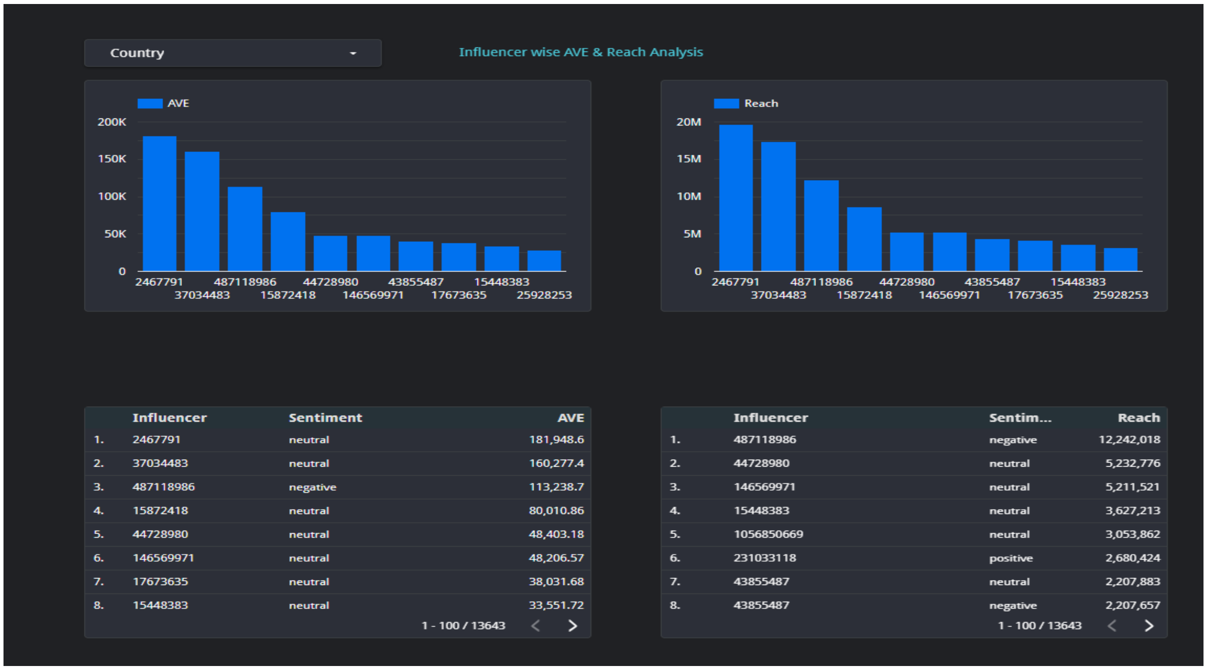

Figure: Influencer wise AVE & Reach Analysis along with Sentiment Scores

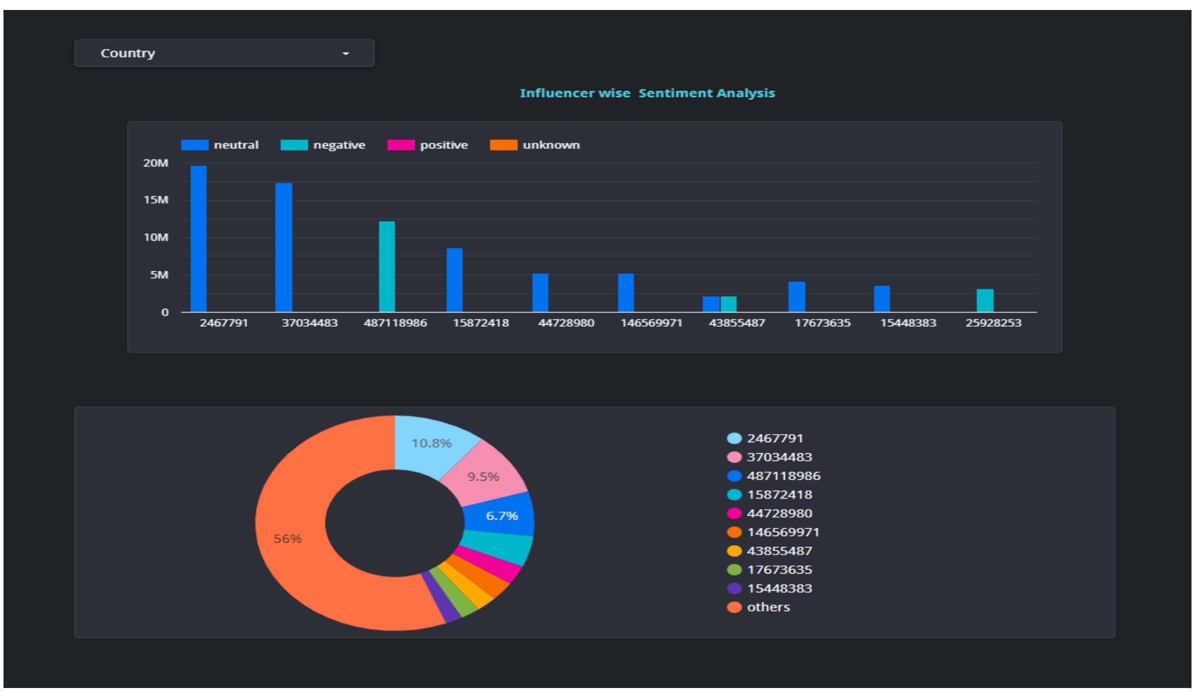

Figure: Influencer wise Sentiment Analysis

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.