Goal - Social Media Analytics Dashboard

Description: Build a Dashboard to Present Your Data Analysis After Collecting Social Media Data from a Single or Multiple Social Platforms

Hello Folks,

We are glad to have you apply to join us here to embark on a journey of collectively learning and conducting research as an independent group. This document lays out a hackathon-esque article search dashboard we want you to design in order to identify and visualize networks of accounts that promote a certain link, hashtag, keyword, or topic on a social media platform as an exercise for understanding your ability to collect data and build neat visualizations of social media datasets. We have framed this assignment as a mini-research task involving you searching for unreliable media links shared on a social platform of your choice. The plots that you incorporate and the technology you use will be an important learning experience for you and in all likelihood will be central to the work you do at SimPPL.

Let's dive into the details of your first research question: "How does news sharing from unreliable media providers vary across social platforms?"

Task Objectives

1. Examine, visualize, and identify patterns in media reliability scores from this list of unreliable news providers https://iffy.news/disinfo-dashboard

2. Collect platform-specific engagement metrics including likes, shares, and the number of replies to each of the posts containing an article from these providers to understand user interaction with these posts.

3. Learn to tell a story with a graph, building intuitive and engaging data visualizations.

Submission Instructions

Share a github link of your repo with us at simppl.collabs@gmail.com with the title "Applicant for SimPPL".

Bonus: If you host your jupyter notebook or JS dashboard, we consider it a significant improvement over applicants that haven't hosted their solution.

Please ensure you include:
1. A link to a video recording of your dashboard hosted on YouTube or Google Drive.
2. A detailed README file or a wiki (with screenshots)

Both of these make it easier for us to run your code and evaluate the assignment.

Suggested Action Items

You will need to build a basic data collection platform to collect data from some fringe social networks within platform terms of service, in order to answer the question. Think of this as recreating a crowdtangle-like dashboard (shown in Figure 1) for curating the news feed from the different social media platforms that are considered fringe platforms.

CrowdTangle is a useful external tool originally built to analyze posts on Meta mainly for marketing purposes. It's shutting down soon, but it's a nice example for you to use as a reference. As a hackathon project, we want you to build a similar tool for more general-purpose use-cases as part of our research exploration in the interest of creating replicable tools for trust and safety research. This assignment is designed to provide you with a neat project you can showcase regardless of your selection into SimPPL.

Some examples that may guide you in learning more about smaller platforms including some so-called fringe platforms (since there are existing data collection tools for these platforms already), in order of priority for the kinds of platforms we'd like to see you analyze.

Note: Pick one or two platforms at max.

Platform Name	**How Many Applicants Have Attempted This (so you can try something new :))**
Gab	0
ShareChat	0
Koo	0
4chan	0
Telegram	1
Bitchute	1
Rumble	0
If you can collect data from this platform we will be 98% likely to hire you: Moj	0
Threads	0
Twitter / X	1
Mastodon	0
VK	0
Instagram	0

Examples of how the CrowdTangle Dashboard looks:

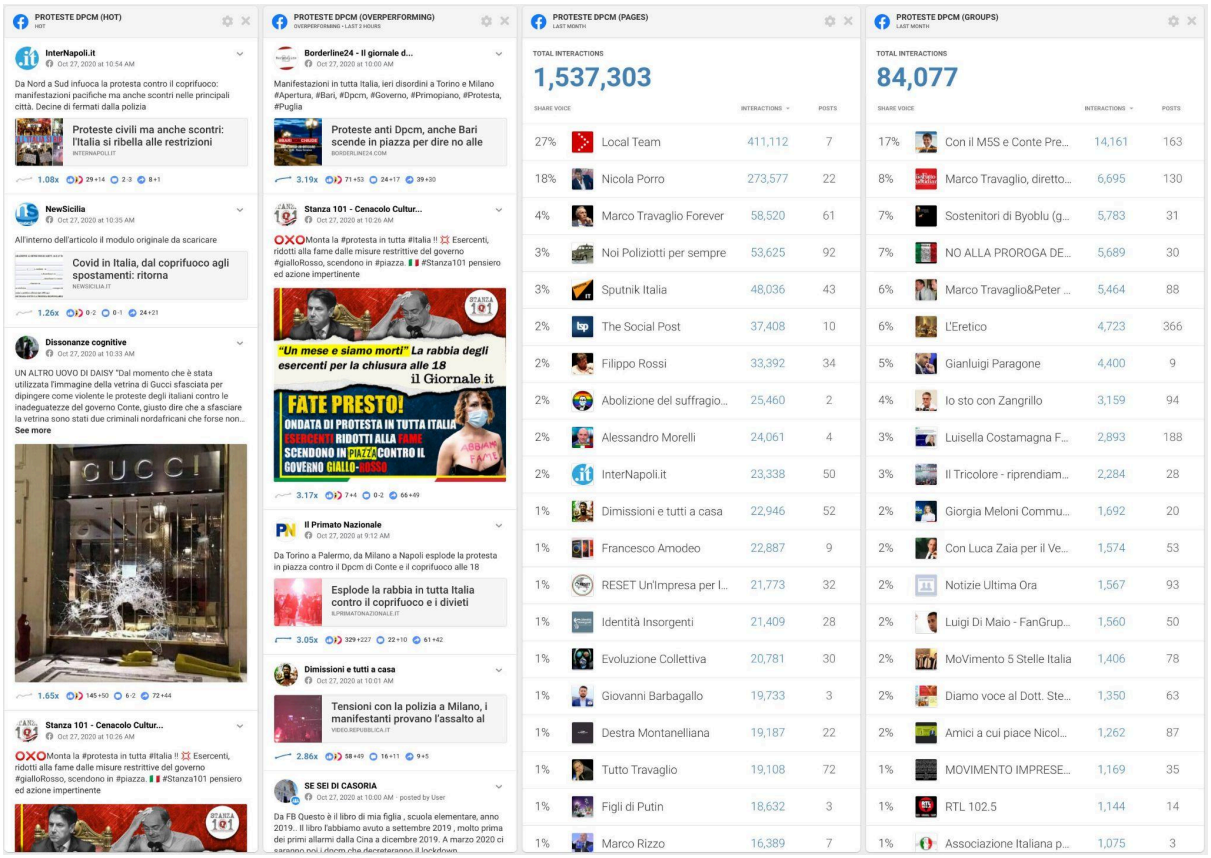

Fig 1. The CrowdTangle Dashboard now slated to be shutdown by Meta (formerly Facebook)

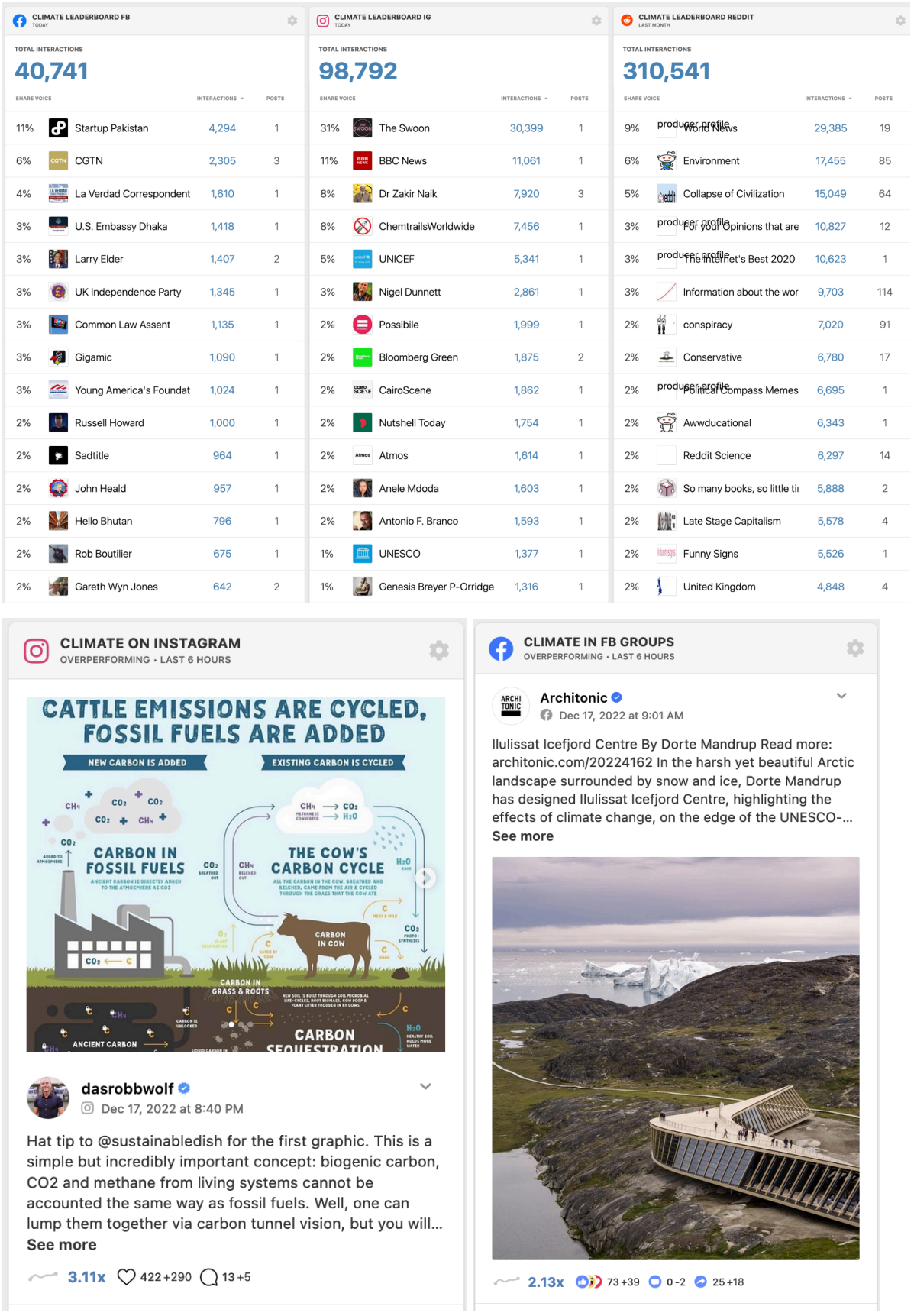

Fig. 2: CrowdTangle had a lot of interesting engagement analytics available on the dashboard.

Why do we care about this?

We have tools for curating and analyzing data from Reddit and Twitter and we built https://parrot.report (see this video in case the site is under maintenance—which it is right now) to study the sharing of news from certain unreliable Russian media providers. To ramp you up towards understanding how to go about this, and to expand your understanding of the broader social media ecosystem, we would like to extend our analysis to other publicly available platforms listed above to get a broader range of viewpoints from different (apolitical/politically biased) groups. In the long run, this research intends to accomplish the following objectives:

- Track different popular trends to understand how public content spreads on different social media platforms.
- Identify posts containing misinformation with the use of fact-checking mechanisms.
- Analyze the trends across lots of accounts over time and report on how information spreads.

Suggested Instructions

Step 1: Popular / Trending Message Tiles

In this task, we would like you to gather and display popular or trending posts from different social media platforms that are relevant to a specific query for the last 7 days from the current date.
We suggest the following details to be displayed for each post (as shown in Figure 2):

- Account posting that information
- Post content
- Interactions for that post (no. of likes, comments, shares, etc)

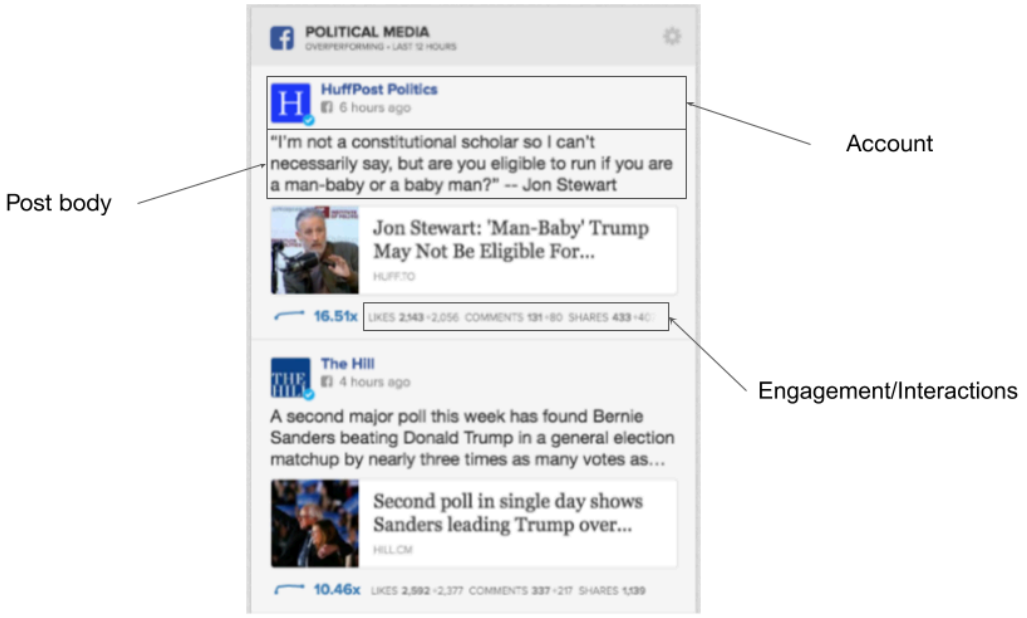

Figure 2: Example post from Facebook tile of Crowdtangle dashboard

Step 2: Influence boards

In this task, we would like to aggregate the collected data from step 1 to display the cumulative trends and stats of engagements across different social media platforms. We specifically care about the following details listed below (refer figure 3 for an example):

- Most interactive public accounts of that forum/group (relevant to the query) (in descending order of interactions).
- Interactions: No. of interactions (e.g. likes, reactions, comments, shares) or views it received.
- Share voice: Percentage share of no. of interactions in that forum/group.
- Posts: No. of posts by accounts in that forum for the past 7 days.
- Rate: Percentage share of posts by accounts in that forum for the past 7 days.

Note: Graph shown in Figure 3 below is not mandatory and interested folks can try it after the above requirements are finished.

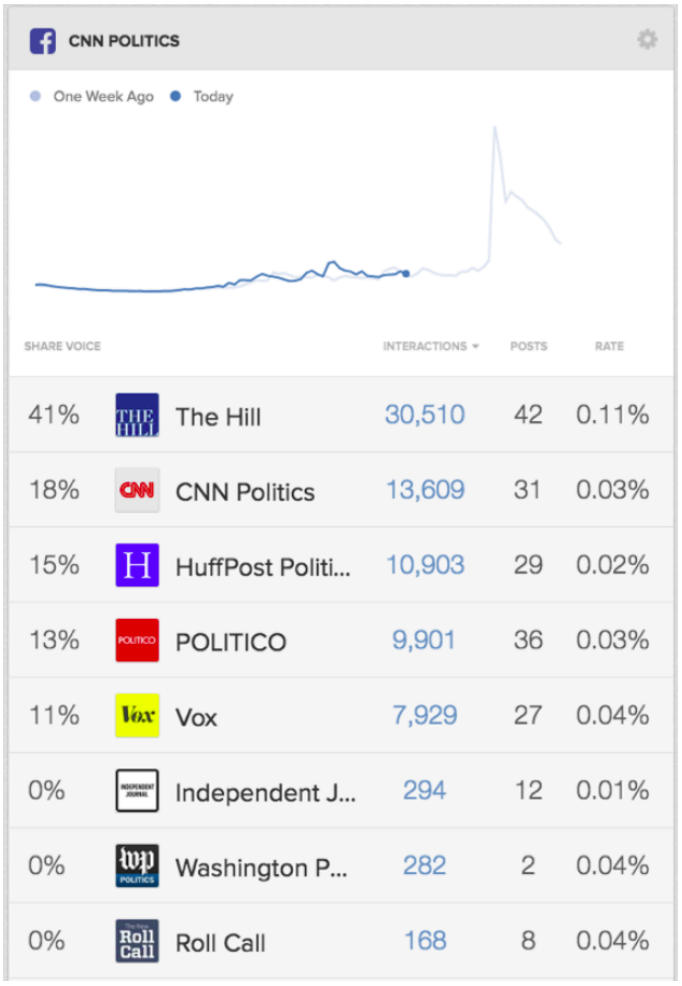

Figure 3: Example influence board for a query from Crowdtangle dashboard

Questions

To be clear the Step 1 and 2 above are just examples. You can implement those but you could also think of going beyond those or going down a different route for your assignment. If you have any doubts with regards to the above task, you can reach out to Swapneel anytime on LinkedIn or via email (simppl.collabs@gmail.com). We hope you have a good time doing this research and development.

Completing your Assignment

1. Send us an email at simppl.collabs@gmail.com letting us know which platform you're attempting to build this search-and-visualize dashboard for.
2. Submit a completed assignment (code, video recording, and link to dashboard) to us also via email with the title "SimPPL Applicant Assignment Completed". Please use the same email ID you applied to us with.
3. Share a link to a github repository (your choice, public or private–but if private, share it with Github ID: SwapneelM so he can review it).

Note

Remember, you don't have to build the perfect system. It can be a simple, elegant, and effective querying platform. All we want to see is the ability to search for a keyword, hashtag, link, or piece of input in any format you prefer, and the ability to identify posts on a platform that match the input either immediately or asynchronously once data is available. We will be testing the functionality of your system and the intuitiveness of the plots you have developed. Telling a single thoughtful story is much better than telling multiple broken ones–go for quality, not quantity :)

4. Presentation matters, please ensure your assignment is possible to easily understand for someone who may be a non-expert on social media and non-technical at running complex queries.
5. Design and UX matters. Thinking through how a user may utilise your system is more important than adding yet another feature to make it more technically complex.

References

- https://github.com/ChrisStevens/garc
- https://github.com/whymath/sharechat-scraper
- https://tgstat.ru/en
- https://start.me/p/0Pqbdg/osint-500-tools?locale=el
- https://github.com/ScriptSmith/instamancer (broken now, but still useful to refer to)