

Understanding Information Operations using YouTubeTracker

Thomas Marcoux and Nitin Agarwal

*College of Engineering and Information Technology
University of Arkansas at Little Rock
Little Rock, Arkansas 72204, USA
{txmarcoux & nxagarwal}@ualr.edu*

Adewale Obadimu

*College of Engineering and Information Technology
University of Arkansas at Little Rock
Little Rock, Arkansas 72204, USA
{amobadimu}@ualr.edu*

Abstract - YouTube is the second most popular website in the world. Over 300 hours worth of videos are uploaded every single minute and 5 billion videos are watched every day - almost one video per person worldwide. Because videos can deliver a complex message in a way that captures the audience's attention more effectively than text-based platforms, it has become one of the most relevant platform in the age of digital mass communication. This makes the analysis of YouTube content and user behavior invaluable to information scientists but also communication researchers, journalists, sociologists, and many more. There exists a number of YouTube analysis tools but none of them provide in-depth qualitative and quantitative insights into user behavior or networks. To that direction, we introduce YouTubeTracker - a tool designed to gather YouTube data and gain insights on content and users. It can help identify leading actors, networks and spheres of influence, emerging popular trends, as well as user opinion. The analysis of user engagement and social networks even reveals suspicious and inorganic behaviors (e.g., trolling, botting) causing algorithmic manipulations.

Index Terms - Youtube, YouTubeTracker, Social media, information operations, bots, disinformation, misinformation.

I. INTRODUCTION

YouTube has offered a platform for professionals and amateurs to provide their content to a previously unattainable audience. That new audience was offered the possibility to interact with the content they resonate with through commenting, but also rating and sharing through social medias. This turned passive consumers into active actors of the content they enjoy - a major shift in online behavior. One would think such a phenomenon would elicit much curiosity from the data science community, but YouTube has failed to garner as much scrutiny as other social media giants such as Facebook and Twitter - largely due to being a video-based platform. But there is much to be learned from YouTube data. Besides analysis of movements such as traffic, likes, dislikes, etc. over time, comments are a deep source of insight on user behavior. Comments' data can be mined to shed light on user interests, networks and overlaps between communities, or content consumption behaviors.

Because this is a challenging process, there is a lack of systematic research that would help us analyze content engagement, user behavior, and more. In an attempt to provide analysts the tools they need to perform various research

(behavioral, political analysis, sociology,...), we present YouTubeTracker [15]. Next, we discuss the state of the art in analyzing YouTube data.

II. STATE OF THE ART IN YOUTUBE ANALYSIS

YouTube is a video sharing platform that provides an unparalleled ability for hosting and sharing video content [1]. YouTube also provides a great deal of customization, and opportunities to solidify user's branding and content engagement across various platforms [1]. According to Alexa [2], an Internet traffic monitoring service, YouTube is the second most popular website and accounts for 20% of web traffic. Research by [3-4] suggests that around 300 hours of videos are uploaded every minute and 1 billion hours of videos are watched each day. Another study by Cha et. al. [5] found that 60% of YouTube videos are watched at least 10 times on the day they are posted. The authors in [5] highlight that if a video does not attract viewership in the first few days after upload, it is unlikely to attract viewership later on. Although YouTube provides a means for users to track their content engagement, there is a lack of systematic research on YouTube due to the dearth of analytical tools that can analyze YouTube data. Some of the noteworthy analytical tools for YouTube include: channelmeter [6], vidooly [7], socialreport [8], quintly [9], ranktrackr [10], socialbakers [11], rivaliq [12], cyfe [13], and dasheroo [15]. However, none of these tools provide in-depth qualitative and quantitative insights into various behavioral patterns on YouTube. Recognizing the need for creating useful tools for extracting actionable knowledge from YouTube, we developed YouTubeTracker [15]. In the subsequent section, we discuss the capabilities of this tool.

III. YOUTUBETRACKER

Described below are some of the features and analytical capabilities of YouTubeTracker.

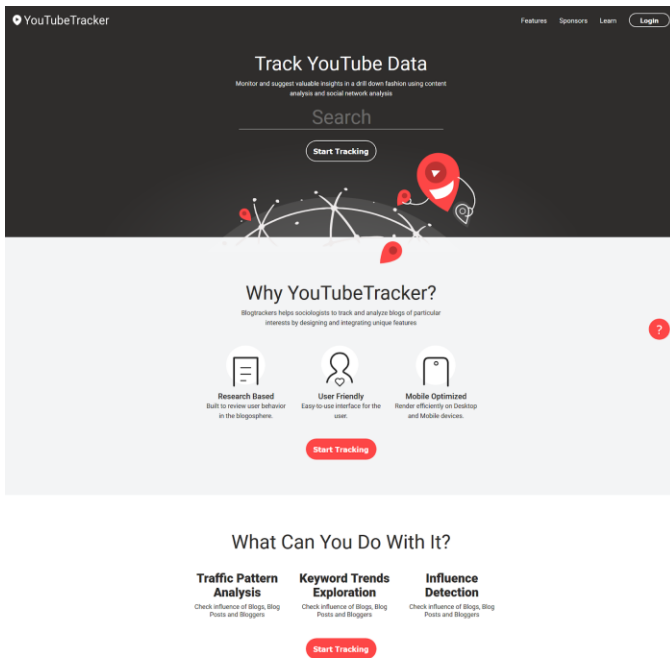


Fig1. YouTubeTracker Home

comments and subscribers over time. This can track viewer's interest in a specific topic over time in an easily understandable manner, which lets analysts measure interest in current events.



Fig2. YouTubeTracker Search Page

1. Tracker Feature Tracker is a concept that helps users curate a collection of content for a topic of interest¹. A tracker could comprise of a set of channels or videos grouped under one topic or theme chosen by the user. Users can feed content of interest to their tracker or dynamically add content they discover while browsing. This allows the user to analyze an ensemble of items and discover overarching patterns. On the tracker dashboard, the user is presented with a bird's eye view of their selected content. The total number of videos and channels are displayed, along with the sum of all likes, dislikes, views, subscribers and comments - along with a timeframe of the channels' activity. The social media footprint study informs the user on the network of their trackers. The distribution of different social medias sites used across the videos and channels of the tracker are displayed in a bar chart.

2. Posting Frequency reveals posting activity in all or some channels, as well as the top contributors and their location. This can reveal unusual trends, such as large user engagement for small/young channels - which can be an indication of botting

3. Content Analysis provides advanced features such as language and category distribution - as well as prominent comment analysis. In this screen, users can see what commenters are the most active actors of their community, what their impact is, or what topics the community discusses.

4. Content Engagement: shows an overview of the type of interaction users have with the selected channels and videos. Line charts display the number of views, likes, dislikes,

IV. RESULTS

In one initial report using this technology, we analyzed content relevant to NATO's 2018 Trident Juncture exercise that was held October 2017 to November 2018. The tracker consisted of official NATO channels and anti-NATO videos published on YouTube during that period. A total of 1324 videos were analyzed among which 96 videos were categorized as NATO-owned, 390 videos were hostile, and 838 videos were earned - that is, supportive of NATO but not NATO-owned. Below are the exact numbers we found.

videos	1324
views	7947124
likes	169988
dislikes	10624
comments	28127
commenters	15491
likes on comments	77324

¹ This is a common concept across our family of tools such as Blogtrackers (<http://blogtrackers.host.ualr.edu>).

replies to comments	22014
---------------------	-------

Table 1. YouTubeTracker Findings

We looked for signs of algorithmic manipulation - usually dense sets of activity like commenting and liking - seeking to promote specific pieces of content to reach organic audiences.

We found that hostile videos received higher user engagement (views, comments, etc.) on average than NATO owned or earned videos. While NATO owned and earned videos had entirely organic engagement, hostile videos exhibited strong indications of robotic activities. For instance, not only were most liked and replied to comments posted on hostile videos, but translating high-engagement comments revealed robotic speech patterns. Google translate was used for Russian, French, and German comments. Resulting in odd sentences that a human translator confirmed to be unusually worded. This could be a case of computer generated comments.

NATO owned videos had mostly positive comments, whereas, comments on hostile videos had exceptionally high negative sentiment towards the exercise, NATO, and the US. Discussions in Russian language tend to be most liked and most replied. Several of these discussions exhibited strong signs of inorganic or robotic activity – a tactic typically used to drive the content up into YouTube’s recommendation algorithms. While most videos were posted on channels located in the United States. Videos posted on channels from Russia were largely hostile.

V. CONCLUSION AND FUTURE WORKS

In this paper, we went through the lack of, and need for in-depth, quantitative behavior analysis tools for YouTube. To answer that demand, we provide YouTubeTracker, which is an evolving tool. For the sake of consistency we used mockup wireframe designs but this tool is available now [15] and being rapidly developed and improved - including its design and capabilities. We recognize the importance of user-friendly design as our goal is to provide a valuable tool not only for technically-adept users but also for business owners or influencers that would benefit from getting a deeper understanding of their audience.

ACKNOWLEDGMENT

This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, and the Jerry L. Maulden/Entergy Endowment at the University of

Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

REFERENCES

- [1] YouTube, L. L. C. "YouTube." Retrieved 27 (2011): 2011.
- [2] "Youtube.com Traffic, Demographics and Competitors - Alexa." [Online]. Available: <https://www.alexa.com/siteinfo/youtube.com>. [Accessed: 03-Jun-2018]
- [3] "Youtube Statistics - 2018," MerchDope, 26-Apr-2018.
- [4] Y. Press, "YouTube for Press," YouTube. [Online]. Available: <https://www.youtube.com/intl/en-GB/yt/about/press/>. [Accessed: 13 - Dec-2017]
- [5] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems.," IEEEACM Trans. On, vol. 17, no. 5, pp. 1357–1370, 2009.
- [6] <https://channelmeter.com/>
- [7] <https://vidooly.com/>
- [8] <https://www.socialreport.com/>
- [9] <https://www.quintly.com/>
- [10] <http://ranktrackr.com/>
- [11] <https://www.socialbakers.com/>
- [12] <https://www.rivaliq.com/>
- [13] <https://www.cyfe.com/>
- [14] <https://www.dasheroo.com/>
- [15] Hussain, Muhammad Nihal, et al. "Analyzing Disinformation and Crowd Manipulation Tactics on YouTube." 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2018.
- [16] <http://youtubetracker.host.ualr.edu/>