Understanding Information Operations using YouTubeTracker

Thomas Marcoux
College of Engineering and
Information Technology
Little Rock, USA
txmarcoux@ualr.edu

Nitin Agarwal
College of Engineering and
Information Technology
Little Rock, USA
nxagarwal@ualr.edu

Obadimu Adewale
College of Engineering and
Information Technology
Little Rock, USA
amobadimu@ualr.edu

Muhammad Nihal Hussain
College of Engineering and
Information Technology
Little Rock, USA
mnhussain@ualr.edu

Katrin Kania Galeano
College of Engineering and
Information Technology
Little Rock, USA
kkaniagalea@ualr.edu

Samer Al-Khateeb
Creighton University
Omaha, USA
sxalkhateeb@ualr.edu

ABSTRACT

YouTube is the second most popular website in the world. Over 300 hours worth of videos are uploaded every 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 platforms in the age of digital mass communication. This makes the analysis of YouTube content and user behavior invaluable not only to information scientists but also communication researchers, journalists, sociologists, and many more. There exists a number of YouTube analysis tools but few of them provide an in-depth qualitative and quantitative insights into user behavior or networks from massive aggregated data. Towards that direction, we introduce YouTubeTracker - a tool designed to gather YouTube data and gain insights on content and users. This tool can help identify leading actors, networks and spheres of influence, emerging popular trends, as well as user opinion. This analysis can also be used to understand user engagement and social networks. This can help reveal suspicious and inorganic behaviors (e.g., trolling, botting) causing algorithmic manipulations. Utility of the YouTubeTracker application is demonstrated via a case study on NATO's 2018 Trident Juncture Exercise.

CCS CONCEPTS

• **Human-centered computing** → *Social networking sites.*

KEYWORDS

youtube, youtubetracker, social media, information operations, bots, disinformation, misinformation

ACM Reference Format:

Thomas Marcoux, Nitin Agarwal, Obadimu Adewale, Muhammad Nihal Hussain, Katrin Kania Galeano, and Samer Al-Khateeb. 2019. Understanding Information Operations using YouTubeTracker. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI '19 Companion)*, October 14–17, 2019, Thessaloniki, Greece. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3358695.3360917>

1 INTRODUCTION

YouTube provides a platform for professionals and amateurs to share their content to a previously unattainable audience. This idea allows the consumer of digital contents to interact with the content they resonate with not only through comments, but also rating and sharing through social media. 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. However, there is much to be learned from YouTube data. Besides the longitudinal analysis of content engagement such as traffic patterns, amount of likes and dislikes on a specific video over a period of 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. Due to the humongous amount of streaming data on YouTube, 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, etc.), we present YouTubeTracker. In the subsequent section, we briefly highlight some of the state of the art technologies in YouTube analysis, we then discuss some of the features and capabilities of our YouTubeTracker application.

2 STATE OF THE ART IN YOUTUBE ANALYSIS

YouTube is a video sharing platform that provides an unparalleled ability for hosting and sharing video content. YouTube also provides a great deal of customization, and opportunities to solidify user's branding and content engagement across various platforms. According to Alexa [12], YouTube is the second most popular website and accounts for 20% of web traffic. Research [11, 15] suggests that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WI '19 Companion, October 14–17, 2019, Thessaloniki, Greece

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6988-6/19/10... \$15.00

<https://doi.org/10.1145/3358695.3360917>

around 300 hours of videos are uploaded every minute and 1 billion hours of videos are watched each day. Another study [14] found that 60% of YouTube videos are watched at least 10 times on the day they are posted. The authors in [14] 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 [1], vidooly [2], socialreport [3], quintly [4], ranktrackr [5], socialbakers [6], rivaliq [7], cyfe [8], and dasheroo [9]. 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. Next, we discuss the capabilities of this tool.

3 YOUTUBETRACKER

YouTubeTracker is an application that provides valuable insights in a drilled down version from YouTube data. In this section, we describe some of the features and analytical capabilities of YouTubeTracker. Figure 1 shows the landing page as shown in <http://youtubetracker.host.ualr.edu/> and Figure 2 shows our main search page..

YouTubeTracker is an application that provides valuable insights in a drilled down version from YouTube data. In this section, we describe some of the features and analytical capabilities of YouTubeTracker. Figure 1 shows the landing page as shown in <http://youtubetracker.host.ualr.edu/> and Figure 2 shows our main search page.

Figure 1: YouTubeTracker Home

1. Tracker Feature Tracker is a concept that helps users curate a collection for analysis based on a topic of interest. 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. Figure 3 describes the tracker dashboard - which 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 tracker. 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 media sites used across the videos and channels of the tracker are displayed in a bar chart.

Figure 2: YouTubeTracker Search Page

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 a fairly recent channel - which can be a strong indication of bots generating artificial content engagement.

3. Content Analysis provides advanced features such as language and category distribution - as well as prominent comment analysis. This feature also allows users to 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. The line charts display the number of views, likes, dislikes, 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.

Figure 3: YouTubeTracker Tracker Dashboard Page

3.1 Technical Infrastructure

Our web app is based on the Django web framework - used to query the YouTube API and store results to the database. As described in Figure 4, we use a MySQL server to house the data collected

Figure 4: YouTubeTracker Technical Infrastructure

through the API. Our visual analysis tools use D3.js. The Django framework extracts data from the MySQL database processes it, and visualizes it using D3.js on the Web Application front end.

3.2 Data extraction and Youtube policies

In order to allow our users to run their own analysis on the data they collect through Trackers, we are working on a feature that them to download the data they track in a JSON format as shown in Figure 5. One challenge associated with this is that, due to YouTube's privacy policies, we must keep up with any deletion of videos or channels as any related data is subject to the privacy policy and must be deleted from our database within 30 days.

Figure 5: YouTubeTracker Admin Console - Export Page

4 CASE STUDY: 2018 TRIDENT JUNCTURE EXERCISE

In one of our preliminary analyses that leveraged this tool, we analyzed content relevant to NATO's 2018 Trident Juncture exercise [10, 13] 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. As shown in Table 1, a

Table 1: Data Statistics for NATO's TRIDENT JUNCTURE Exercise (2018) as reported from YouTubeTracker Database.

YouTube Metrics	Count
videos	1,324
views	7,947,124
likes	169,988
dislikes	10,624
comments	28,127
commenters	15,491
likes on comments	77,324
replies to comments	22,014

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.

We observed signs of algorithmic manipulation - usually dense sets of activity like commenting and liking - seeking to promote specific pieces of content to reach organic audiences. One such commenter flash mob is shown in Figure 6.

Figure 6: Commenter flash mobs

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. Some examples of possibly artificial comments are shown in Table 2.

NATO owned videos had mostly positive comments, whereas, comments on hostile videos had exceptionally high negative sentiment towards the exercise, NATO, and the US. Figure 7 shows that pro-NATO videos are mainly talking about positive aspects such as "Alliance", "United", "Appreciation", "Love", "Respect", "Freedom", "Country", and "Military".

Table 2: Some of the most replied-to comments for NATO's TRIDENT JUNCTURE Exercise (2018) as reported from YouTubeTracker Database.

Original comment	Translated comment
А что ВЫ Думаега по поводу этих учений? Поддержите Лайком ролик! Единственным соперником НАТО был СССР. Современная Россия не соперник альянсу. Не та армия, не те люди, не то правительство. Ну и название,прям Рен Тв	What are YOU Thinking about these teachings? Support the movie with Like! The only rival of NATO was the USSR. Modern Russia is not an alliance rival. Not the army, not the people, not the government. Well, the name, straight Ren Tv

Figure 7: Topic modeling of pro-NATO channels

However, anti-NATO videos are talking about negative topics such as "Russia", "Soldier", "Attack", "America", "Profanity", "Fake News", and "Insult" - as shown by Figure 8.

Figure 8: Topic modeling of anti-NATO channels

A deeper analysis of the negative content from anti-NATO videos revealed the following leading themes (Figure 9). These narratives are the result of running topic modeling on the most influential blogs. This gives us further information on the main points of NATO critiques.

Figure 9: Leading anti-NATO narratives

Discussions in Russian language tend to be most liked and most replied to. 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 - see Figure 10.

Figure 10: Location Analysis

5 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. The tool is available now and being rapidly developed and improved - including its design and capabilities. We recognize the importance of a 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. YouTubeTracker's utility is demonstrated via a case study on NATO's 2018 Trident Juncture exercise. Analysis of the YouTube data collected through YouTubeTracker application shows signs of robotic content and algorithmic manipulations by

adversarial actors. Location analysis and target audience analysis deliver key insights for an analyst. Topic modeling and analysis helps reveal stark differences between pro-NATO videos and anti-NATO videos and several leading anti-NATO themes.

ACKNOWLEDGMENTS

This research is funded in part by the U.S. National Science Foundation IIS-1636933, ACI-1429160, 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, N00014-19-1-2336), 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 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] [n. d.]. <https://channelmeter.com/>
- [2] [n. d.]. <https://vidooly.com/>
- [3] [n. d.]. <https://www.socialreport.com/>
- [4] [n. d.]. <https://www.quintly.com/>
- [5] [n. d.]. <http://ranktrackr.com/>
- [6] [n. d.]. <https://www.socialbakers.com/>
- [7] [n. d.]. <https://www.rivaliq.com/>
- [8] [n. d.]. <https://www.cyfe.com/>
- [9] [n. d.]. <https://www.dasheroo.com/>
- [10] [n. d.]. <https://www.nato.int/cps/en/natohq/157833.htm>
- [11] [n. d.]. YouTube for Press. Retrieved 13 - Dec-2017 from <https://www.youtube.com/intl/en-GB/yt/about/press/>
- [12] [n. d.]. Youtube.com Traffic, Demographics and Competitors - Alexa. Retrieved 03-Jun-2018 from <https://www.alexa.com/siteinfo/youtube.com>
- [13] et al. Hussain, Muhammad Nihal. 2018. Analyzing Disinformation and Crowd Manipulation Tactics on YouTube. IEEE/ACM ASONAM.
- [14] P. Rodriguez Y. Ahn M. Cha, H. Kwak and S. Moon. 2009. Analyzing the video popularity characteristics of large-scale user generated content systems, Vol. 17. IEEEACM Trans. On, 1357–1370.
- [15] MerchDope. 2018. Youtube Statistics - 2018.