2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Analyzing Disinformation and Crowd Manipulation Tactics on YouTube

Muhammad Nihal Hussain, Serpil Tokdemir, Nitin Agarwal
Department of Information Science
University of Arkansas, Little Rock, AR, USA
{mnhussain, sxtokdemir, nxagarwal}@ualr.edu

Samer Al-khateeb
Department of Journalism, Media, and Computing
Creighton University
Omaha, Nebraska, USA
Sameral-khateeb1@creighton.edu

***Abstract*—YouTube, since its inception in 2005, has grown to become largest online video sharing website. It's massive user-base uploads videos and generates discussion by commenting on these videos. Lately, YouTube, akin to other social media sites, has become a vehicle for spreading fake news, propaganda, conspiracy theories, and radicalizing content. However, lack ineffective image and video processing techniques has hindered research on YouTube. In this paper, we advocate the use of metadata in identifying such malicious behaviors. Specifically, we analyze metadata of videos (e.g., comments, commenters) to study a channel on YouTube that was pushing content promoting conspiracy theories regarding World War III. Identifying signals that could be used to detect such deviant content (e.g., videos, comments) can help in stemming the spread of disinformation. We collected over 4,145 videos along with 16,493 comments from YouTube. We analyze user engagement to assess the reach of the channel and apply social network analysis techniques to identify inorganic behaviors.**

Keywords— YouTube, disinformation, fake news, inorganic activities, bots, spam, conspiracy theories, deviant, malicious behaviors, social network analysis

I. INTRODUCTION

With the proliferation of smart devices, mobile applications, and social network platforms, the social side effects of these technologies have become more profound, especially in social and political disintegration. Several journalistic and few academic investigations have reported that modern communication platforms such as social media (e.g., Twitter, Facebook, blogs, YouTube, etc. and the "deep web" channels) are strategically used to coordinate cyber propaganda campaigns. YouTube is one of the fastest growing medium for behavioral steering, with various production styles catering to specific demographics (e.g., teens and youth) subjecting them to conspiracy theories, disinformation campaigns, and radicalization strategies. Prolific linking of YouTube videos in tweets, blogs, Telegram posts, etc. has tremendously helped frame the discourse and is considered an extremely successful information operation tactic. Moreover, due to the afforded anonymity and perceived less personal risk of connecting and acting online, adversarial cyber campaigns are becoming increasingly common among socio-technically competent "hacktivist" groups to provoke hysteria, influence mass opinions, stoke civil unrest, effect civil conflict, or even coordinate (cyber)-attacks. Such deviant behaviors are categorized as the new face of transnational crime organizations that could pose significant risks to social, political, and economic stability.

In August 2016, while a vigorous national debate was underway on whether Sweden should enter a military partnership with NATO, officials in Stockholm suddenly encountered an unsettling problem: a flood of distorted and outright false information on social media, confusing public perceptions of the issue [1]. The claims were alarming: If Sweden, a non-NATO member, signed the deal, the alliance would stockpile secret nuclear weapons on Swedish soil; NATO could attack Russia from Sweden without government approval; NATO soldiers, immune from prosecution, could rape Swedish women without fear of criminal charges. They were all false, but the disinformation spilled into the traditional news media. As the Swedish defense minister, Peter Hultqvist, traveled the country to promote the pact in speeches and town hall meetings, he was repeatedly grilled about the bogus stories. A largely unstated message of these propaganda campaigns is that European governments lack the competence to deal with the crises they face, particularly immigration and terrorism, and that their officials are all American puppets. Speaking during the 75th anniversary of the Soviet Information Bureau, Mr. Dmitry Kiselyev said the age of neutral journalism was over. "*If we do propaganda, then you do propaganda, too,*" he said, directing his message to Western journalists, in an interview on the state-run Rossiya 24 network. He continued, "*While the business of "persuasion" is more expensive now, if you can persuade a person, you don't need to kill him.*". Such disinformation dissemination efforts leverage a variety of social media tactics, especially the use of videos in disseminating agitation propaganda.

YouTube being the largest video sharing website, is used to host radicalizing contents by deviant actors. According to Alexa [2], an Internet traffic monitoring service, YouTube is the second most popular website and accounts for 20% of web traffic. Around 300 hours of videos are uploaded every minute and 1 billion hours of videos are watched each day [3-4]. YouTube users can upload new videos or engage with existing videos (view, like, dislike or comment) on videos. In this paper, we analyze data from YouTube to identify disinformation dissemination and crowd manipulation tactics deployed on YouTube. For our analysis, a YouTube channel riddled with

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

conspiracy theory videos was identified. We analyzed video posting behavior and user engagement to detect inorganic and deviant behaviors. More details about our research methodology and analysis are presented in Section 3. Next, we present literature review.

II. LITERATURE REVIEW

Steady rise in YouTube's popularity has attracted a surge of research. Most research on YouTube has been qualitative and give insights into various behavioral patterns. Early studies on YouTube focused on users' preferences [5] and user engagement [6-7]. Cha et. al. [7] found that 60% of videos are watched at least 10 times on day they are posted and if a video does not attract viewers in subsequent days, it is unlikely to attract viewers later on [7]. Another study [8] identified patterns in user types that can be used to predict users' behaviors. Authors in [9] developed a model that leveraged sentiment analysis to identify polarity of each comment and predicted if it would get likes or dislikes. Thelwall et. al. [10] analyzed comments from different types of dance videos using comment term frequency comparison to identify words that belonged to specific dance styles. This study also identified words that are strongly associated with each gender by aggregating comments by gender and analyzing them separately. Finally, the authors built a semantic network of dance styles based on comment similarity and found semantic relations among different dance styles.

The rise in popularity of social media sites, such YouTube, has made them particularly vulnerable for abusive behaviors from bots and troll accounts that post spam comments in large volumes[11]. O'Callaghan et. al. [12] built a co-commenter network using comment similarity and applied network analysis techniques to identify spam bots. Authors also found these comments posted by these spam bots are often associated with orchestrated campaigns that can remain active for long periods of time. However, our work is different from aforementioned works as we analyze user engagement to study disinformation dissemination and crowd manipulation strategies.

III. METHODOLOGY

Here we describe the datasets used in this research and the overall research methodology along with results and findings.

A. Data Collection

We collected data using YouTube Data API [13] and Web Content Extractor[14].

1) YouTube Data API

Using YouTube Data API, we extracted title, published date and description of 4,145 videos from the channel of interest. Due to privacy reasons and YouTube's data usage agreement, we cannot divulge the name of the channel that hosted the videos. We extracted the number of views, likes, dislikes, comments and comment text for all the videos published by the channel. For each comment, we extracted commenter ids, commenters' name, comment text, likes and replies on the comments, and published date. We also extracted video id, title, channel id, published date and description information of all the related videos for every video published by the channel. The data is summarized in Table 1.

2) Web Content Extractor

We used the video id obtained from YouTube Data APIs to construct URL of these videos and trained Web Content Extractor (WCE) to collect the *views, likes* and *dislikes* information for each video. We also used WCE to setup a crawler that extracts number of *subscribers* and *total views* information of the channel on an hourly basis.

TABLE I. YOUTUBE DATA STATISTICS.

Videos	4,145
Channel Subscribers	26,369
Views	8,233,922
Likes	30,153
Dislikes	7,603
Comments	16,493
Commenters	9,661
Likes on comments	20,829
Replies on comments	8,19

B. Analysis

We analyzed the channel's information dissemination and user engagement to study crowd manipulation. Analysis was done in two steps: (1) metadata analysis to study overall information dissemination and user engagement, and (2) commenter-behavior analysis to study commenter or user behavior on the channel.

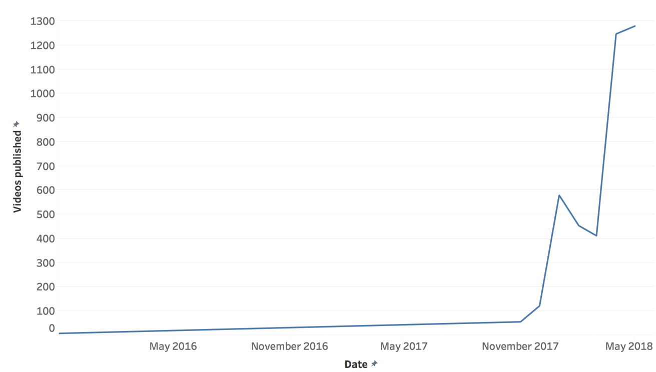

Figure 1. Video Postings Trend.

1) Metadata Analysis

To study the activity and content engagement trends of the channel, we analyzed its video posting behavior since its existence, i.e., from August 1, 2015 to May 31, 2018 (see Figure 1). Although this channel had been on YouTube since August 2015, an unusual spike in video posting was observed around April and May 2018. A highly correlated spike in the number of views was observed for this channel. While one can post as many videos as he/she likes, getting unusually high number of views as soon as a video is posted is rather odd, especially for an unknown YouTuber. This is a strong indicator of a sophisticated information campaign run by this channel.

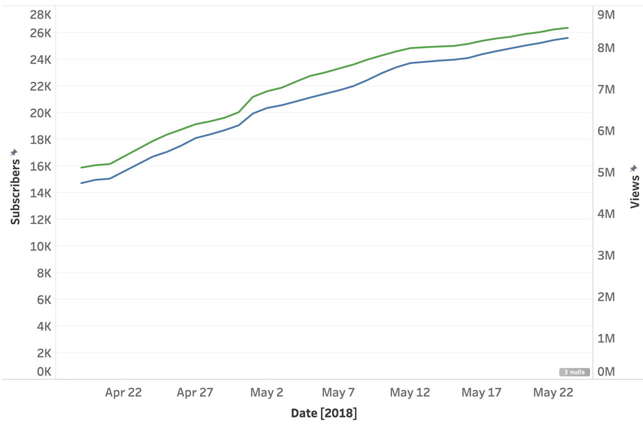

Figure 2. Channel Subscriber and Viewership Trends. Blue trend line represents viewership and green trend line represents subscribers

Next, we zoom into the channel's high period of activity to assess the reach of the channel. In other words, does the influence and popularity of the channel change over time? To answer this question, we study the trend for the number of subscribers and total viewership for all the videos posted on the channel from April 15, 2018 to May 20, 2018 (Figure 2). A consistent increase in popularity of the channel was observed.

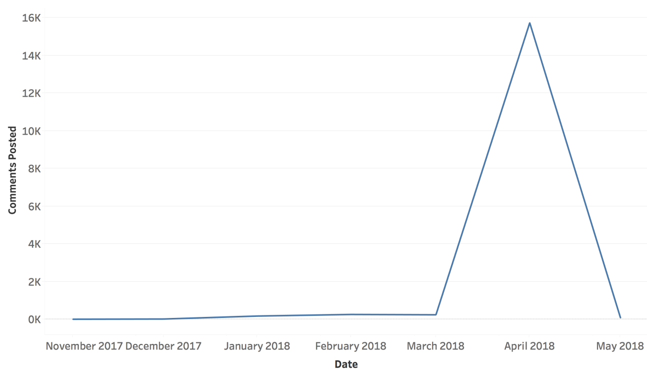

Figure 3. Comments Postings Trend.

To study the user engagement, we analyzed users' response by generating comment posting trend for the channel from November 2017 to May 2018 (see Figure 3). A spike in user activity, similar to video posting (from Figure 1) was observed. Steady increase in popularity (from Figure 2) but sudden increase in comments (from Figure 3) with increase in video posting (from Figure 1) suggests that this channel either had a significant user base which were waiting for videos to be posted or the comments/commenters are fabricated or inorganic. Either of these cases are alarming and provide objective indicators to warrant further investigation into the channel's activities.

2) Commenter Behavior Analysis

To study commenter behavior, we built video-commenter network, shown in Figure 4. In Figure 4, two types of commenting behaviors were observed. Peripheral groups of commenters where many users comment on one video and a core groups of commenters where the users comment on multiple videos numerous times. Chances of identifying unusual or inorganic behaviors are higher in the second type of behavior, i.e., core user segment.

Figure 4. Video-Commenter Network. Red nodes are commenters that commented on videos in green nodes.

To investigate the core-commenters, we generated co-commenter network by connecting two commenters if they commented on the same video (see Figure 5). Co-commenter network revealed unique behaviors where few commenters formed dense communities and commented on the same video while other commenters acted as a broker to bridge connections between these communities. To further investigate this behavior, we analyzed the comments in the biggest cluster (indicated as region 1 in Figure 5) and its surrounding broker nodes (indicated as regions 2, 3, and 4 in Figure 5). Comments posted by the commenters in the big cluster were related to the World War III conspiracy theories and had words like Russia, USA, Israel, Syria and Iran. Whereas, comments from the broker commenters had few words related to World War III conspiracy but also mentioned words related to conspiracies pushed during the 2016 US elections and other popular/global events, thereby muddling the discourse with other conspiracy theories.

To investigate inorganic commenting behavior, we constructed a commenter-comment network. In figure 6 and 7, the blue nodes indicate comments (comment text) posted by commenters in red nodes. Commenter-comment network revealed two types of inorganic behaviors: (1) *bot-like behavior*, where an identical comment is posted by large groups of commenters (see Figure 6) and (2) *spam comments*, where the same comment is posted numerous times by a commenter (see Figure 7). Upon further examination, we found that these comments contained spam URLs often promoting a product, or links to unrelated videos in an attempt to boost viewership, or links to spurious websites, etc.

Figure 5. Co-Commenter Network.

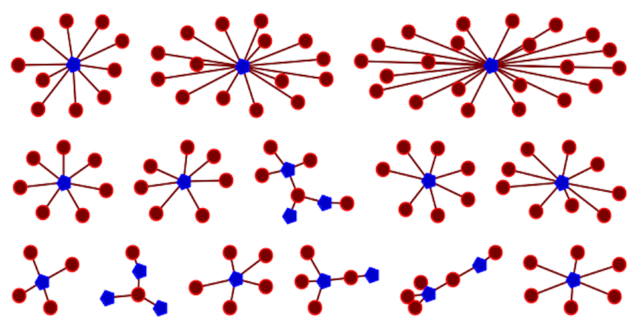

Figure 6. Commenter-Comment Network (bot like behavior). Blue nodes indicate comment text that was posted by the commenters in red nodes.

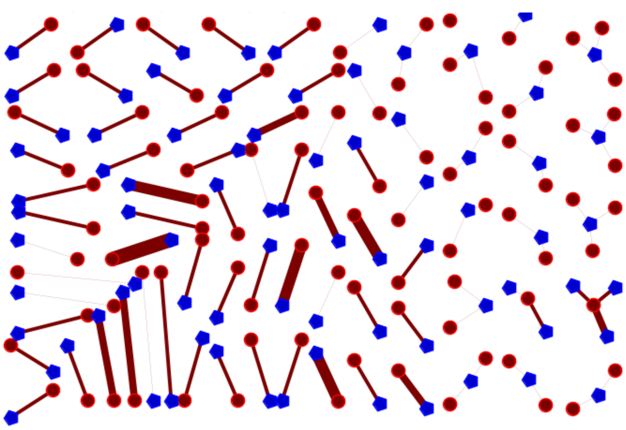

Figure 7. Commenter-Comment Network (spam comments). Blue nodes indicate comment text that was posted by the commenters in red nodes.

IV. Conclusion and Future Work

Social media along with its many benevolent uses provides users freedom to express their opinion without revealing their identity. This feature, if left unchecked, can be used for radicalization, hysteria propagation, crowd manipulation, and propaganda dissemination. It is important to monitor and detect these deviant acts early on to thwart efforts to subvert democracy.

In this paper, we analyzed video posting patterns of a conspiracy theory-riddled YouTube channel. We used social network analysis to study commenters' behaviors and identified commenter communities that exhibited inorganic or bot-like behaviors. In future, we plan to further investigate these communities by analyzing semantics of the comment text along with the identified network structure. Another prospective research direction is to investigate suggested related videos to study algorithmic bias and manipulation techniques.

Acknowledgement

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), and the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock.

References

[1] N. MacFarquhar, "A Powerful Russian Weapon: The Spread of False Stories," *The New York Times*, 20-Jan-2018.

[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] K. Purcell, "The State of Online Video," *Pew Research Center: Internet, Science & Tech*, 03-Jun-2010. .

[6] X. Cheng, J. Liu, and C. Dale, "Understanding the characteristics of internet short video sharing: A YouTube-based measurement study," *IEEE Trans. Multimed.*, 2011.

[7] 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.

[8] M. Maia, J. Almedia, and V. Almedia, "Identifying user behavior in online social networks," in *Proceedings of the 1st Workshop on Social Network Systems*, 2008, pp. 1–6.

[9] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro, "How useful are your comments?: analyzing and predicting youtube comments and comment ratings," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 891–900.

[10] M. Thelwall, "Social media analytics for YouTube comments: potential and limitations," *Int. J. Soc. Res. Methodol.*, vol. 21, no. 3, pp. 303–316, 2018.

[11] A. Sureka, "Mining User Comment Activity for Detecting Forum Spammers in YouTube," 2011.

[12] D. O'Callaghan, M. Harrigan, J. Carthy, and P. Cunningham, "Network Analysis of Recurring YouTube Spam Campaigns.," in *ICWSM*, 2012.

[13] "YouTube Data API," *Google Developers*. [Online]. Available: https://developers.google.com/youtube/v3/. [Accessed: 03-Jun-2018].

[14] Newprosoft, "Web Content Extractor," *Newprosoft Web Data Extraction Software*, 2004. [Online]. Available: http://www.newprosoft.com/web-content-extractor.htm. [Accessed: 11-Dec-2017].