Navigating the Anomalies: A Comprehensive Analysis of YouTube Channel Behavior

Shadi Shajari*, Ridwan Amure*, and Nitin Agarwal

COSMOS Research Center, University of Arkansas at Little Rock, Little Rock, Arkansas, USA
{sshajari,raamure,nxagarwal}@ualr.edu

Abstract. YouTube, a prominent online platform hosting millions of users globally, is experiencing a concerning surge in unusual activities. This paper introduces a scoring method that integrates engagement metrics and commenter behavior features to evaluate anomalies in YouTube channels. By examining engagement indicators such as comments, views, and subscriber numbers and with analyzing frequent commenter groups, the study employs various techniques including Cosine Similarity and Principal Component Analysis, along with average and maximum aggregation methods. Our dataset includes 71 channels, 642,952 videos, 12,425,587 commenters, 123,882,200 comments, 83,396,188,807 views, and 139,985,870 subscribers. Focusing on channels related to the US Air Force, news, and the Indo-Pacific region, the research offers insights into anomalous activities on YouTube. Our objective in this study is to establish a secure online space while maintaining the credibility of YouTube.

Keywords: Social Network Analysis · Anomalous Behavior · Cosine Similarity · Principal Component Analysis

1 Introduction

In the vast world of online content consumption, YouTube stands out as the top choice for billions of users globally. However, within its extensive library of videos and channels, a concerning pattern has emerged: the rise of anomalous behaviors that undermine the trustworthiness of user engagement. Anomalous behavior on YouTube channels encompasses actions that deviate from expected norms, potentially violating YouTube's rules and terms of service or hinting at deceptive practices. A typical scenario occurs when commenters collaborate to artificially boost the number of comments on channels by posting comments together on the videos, thereby faking higher engagement. Moreover, anomalies can manifest in various forms, such as sudden increases or decreases in engagement levels, unusual patterns in viewer interaction, or disparities between the numbers of views, subscribers, and comments.

This paper describes a methodology for examining anomalous engagement and commenter behavior across 71 YouTube channels. It investigates two distinct analytical methods, Cosine Similarity and Principal Component Analysis, in addition to average and maximum aggregation techniques. Our research focuses on channels related to the US Air Force, news, and the Indo-Pacific region, which are often associated with deceptive practices due to their contentious topics and content. Ultimately, our goal is to establish comprehensive scoring methods for detecting and evaluating anomalous behaviors within YouTube's platform. These scores aim to provide a thorough understanding of anomalous behaviors on YouTube channels. The following subsequent sections are organized as follows: The Related Works section provides an overview of anomalous behaviors detected on YouTube, along with existing knowledge in this area. The Data Collection section describes the techniques utilized for collecting data. The Methodology section delves into the research approach utilized. The results section presents the study's findings, offering a comparative examination of the two methods. Lastly, the Conclusion and Future Works section wraps up the research, offering recommendations for future exploration.

2 Related Works

A range of studies have explored anomalous behavior on YouTube. The study by [8] focused on detecting coordinated inauthentic activities, finding that channels with such behavior exhibit fewer but higher magnitude peaks in their anomaly patterns. Researchers in [4] discussed the impact of Content Delivery Network (CDN) server selection on user experience and presented an Anomaly Detection system for detecting unexpected cache-selection changes in CDN traffic. These studies collectively underscore the need for robust detection and mitigation strategies for anomalous behavior on YouTube.

The actions of suspicious commenters have been closely examined, as shown by [13] utilizing graph2vec and UMAP techniques. The study by [12] suggested approaches like Principal Component Analysis (PCA) to identify and understand groups of commenters disseminating inaccurate opinions regarding the US military. Authors in [2] analyzed YouTube channels to identify accounts using automated methods to boost engagement and spread misinformation. The study by [17] examines how text, audio, and colors in YouTube videos affect user emotions, finding text impacts news videos and colors influence movie trailers. Insights help content creators enhance audience engagement. Research on YouTube engagement behavior has identified several key factors. According to [3], more evenly interactive commenters and creators who reply to comments and present reactions can increase their video engagement. Research conducted by [10] proposed a continuum of user engagement, with commenting and conversational behaviors being key components. Analysis by [5] highlighted the role of commenters in boosting channel influence and shaping perceptions, particularly in the context of spreading disinformation. Researchers in [7] identified different motives for user engagement, with social interaction being a strong predictor for commenting.

The literature review highlights diverse studies on online behavior and deceptive practices on platforms like YouTube, noting the challenge of lacking standardized metrics for detecting anomalies. This inconsistency affects platform integrity and user trust, emphasizing the need for uniform detection methodologies to enhance social media surveillance.

3 Methodology

This section describes the methods and datasets used in the study to develop the anomalous scores.

3.1 Data Collection

The data analyzed in this study was collected by a specialized tool for collecting data from YouTube [9], which operates through YouTube's API [6]. This tool collected daily video postings, the number of comments, and comment-specific data (e.g., comment content, commenter ID, video ID, and timestamps). Additionally, to collect daily subscriber counts and the number of daily video views, we utilized the Social Blade API [15]. Data were collected from 71 channels, encompassing a total of 642,952 videos, 12,425,587 commenters, 123,882,200 comments, 83,396,188,807 views, and 139,985,870 subscribers.

3.2 Suspended Channels

We found through our research that certain YouTube channels showing anomalous activity were subsequently suspended by the platform due to their behaviors breaching YouTube's terms of service. Utilizing the YouTube API, we collected data on the active status of all channels in our dataset to distinguish between active channels and those that had been suspended. We found that 7 channels had been suspended out of 71 due to violating YouTube's community guidelines [1].

3.3 Estimating Anomalous Scores

The data collection and feature extraction methodology for this study was inspired by the approaches described in [8] and [14]. In brief, our method involved collecting data from 71 YouTube channels as described above and extracting two sets of features that were descriptive of each channel. These features were expected to be informative for distinguishing anomalous channels. Our feature set, 26 in total, comprised two primary categories: structural features and engagement features.

- Structural Network Features (20): These features were extracted from the co-commenter network comprising connections between commenters who have commented on the same video, where the weight of the connection indicates

how many videos they have commented on together[12], and these features captured the underlying structure of the network, including the number of nodes, number of edges, average clustering coefficient, modularity, etc.
- Engagement Features (6): These features quantified the level of engagement between users and content, and included metrics like views and subscribers alongside videos and comments. Specific engagement features included ratios of numbers of views to subscribers, views to videos, views to comments, subscribers to videos, subscribers to comments, and videos to comments.

Two approaches were used to combine the features extracted from the engagement trend and co-commenter network. Each method is described in the following section.

Anomalous Score Estimation using Cosine Similarity We employed cosine similarity [11] shown in equation 1 to measure the similarity between each active channel and each suspended channel where A and B are the feature vectors. Cosine similarity is a metric that calculates the normalized dot product of two vectors. In this context, the features of the active and suspended channels represent the vectors being compared. A higher cosine similarity score indicates a greater degree of similarity between the channels, suggesting that the active channel exhibits characteristics closer to those of the known anomalous channels.

$$similarity(\mathfrak{S}) = \frac{AB}{||A||\;||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\,\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

Cosine similarity was calculated between an active channel and each suspended channel, resulting in an array of similarity scores. To obtain a single, representative anomalous score for each active channel, we calculated the average and maximum of all the similarity scores in the array. This aggregated value reflects the overall similarity of the active channel to the set of suspended channels.

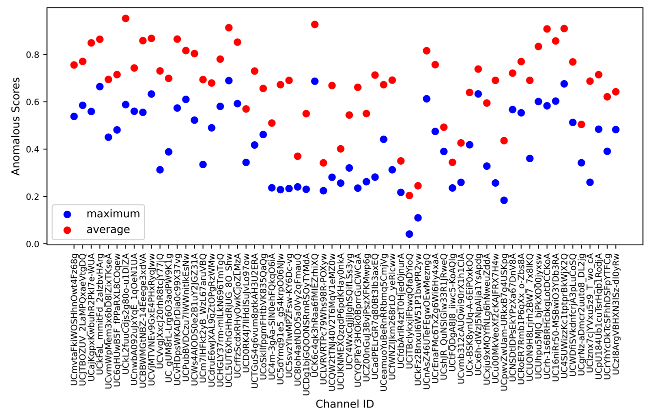
Anomalous Score Estimation Using Principal Component Analysis This subsection details the second method employed for anomalous score estimation. This approach leverages Principal Component Analysis (PCA) [16] to identify underlying factors that contribute to anomalous channel behavior and utilizes these factors to calculate a weighted anomalous score. To achieve this, we extracted coefficients associated with each feature in the principal components. These coefficients represent the relative contribution of each feature to the respective component. By scaling these coefficients to a range of 0 to 1 and applying a weighted average along with the corresponding structural network and engagement feature values for each channel, we calculated an anomalous score for each channel.

4 Result

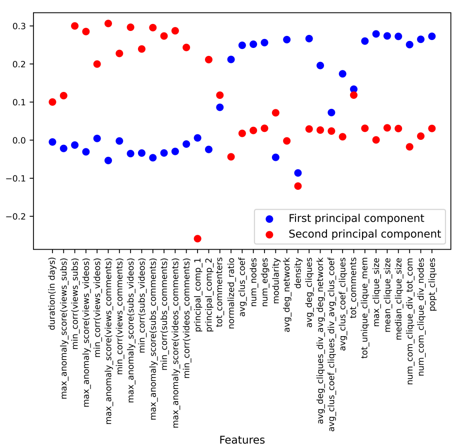
This section presents and discusses the results obtained from applying the two methods for anomalous score estimation on YouTube channels. The analysis aims to evaluate the effectiveness of each method in identifying potentially anomalous channels. Figure 1a illustrates the distribution of anomalous scores calculated using the cosine similarities approach. Notably, the figure compares two approaches for aggregating the individual similarity scores obtained for each active channel: average and maximum. When using the "average" aggregation function, the majority of channels exhibit scores below 0.6. This suggests that, on average, active channels possess only a moderate degree of similarity to the set of suspended channels. Conversely, applying the "maximum" function results in a significant shift, with over 70% of channels scoring higher than 0.6. This indicates that at least one suspended channel shares a high degree of similarity with a large portion of the active channels when considering individual comparisons.

These observations highlight the potential benefits of each aggregation strategy. The "maximum" approach may be better suited for scenarios where identifying even a single close resemblance to a known anomalous channel is crucial, potentially improving the true positive rate (correctly identifying anomalous channels). However, this strategy might also lead to a higher number of false positives (mistakenly identifying legitimate channels as anomalous) due to the possibility of capturing outliers within the individual similarity scores. On the other hand, the "average" approach provides a more balanced distribution, potentially reducing the number of false positives. However, it may be less effective in capturing channels that exhibit strong similarity to just a few suspended channels.

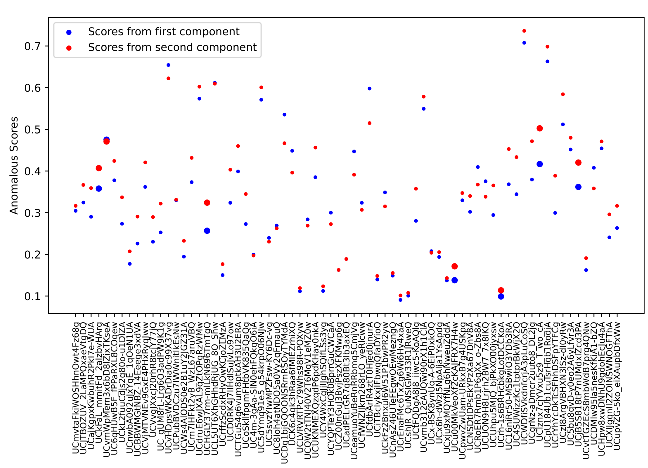
In the PCA analysis, we note that the first principal component was dominated by the features from the co-commenter network while the second component was dominated by the features from the engagement trend analysis as seen in Figure 1b. Despite this opposing behavior, the weighted average calculated using the first and second components produced a similar score distribution as seen in Figure 1c. Finally, to evaluate the effectiveness of the weighted score in identifying anomalous channels, we compare the scores with a pseudo-ground truth. As expected, the results in Figure 1c demonstrate that the suspended channels are assigned higher anomalous scores compared to the active channels. This observation validates the ability of the weighted score to distinguish between legitimate and potentially anomalous channels. Our analysis aimed to identify effective methods for estimating the anomalous behaviour of YouTube channels. Here, we examine a particularly interesting case: Channel UCL5UT6XhGHhdRjUG_8O_Shw (KH Buya Syakur Yasin MA) stands out as a particularly anomalous case. Across all three methods, it received consistently high scores. Notably, the maximum similarity score reached a very high value of 0.912954, indicating a strong resemblance to known anomalous channels. Even the average similarity score remained well above 0.6 (at 0.689398), suggesting a consistent pattern of anomalous characteristics. Scores derived from both



(a) Channels and their anomalous scores using maximum and average similarities to other suspended channels. The blue points represent the scores using the maximum function and the red points represent scores obtained using the average function.



(b) Coefficient of features extracted from the first and second components in the PCA analysis.



(c) Channels and their anomalous scores using features importance from the first and second components. The suspended channels are shown with large dots.

Fig. 1: Analysis of channels and their anomalous scores using different methods

principal components (0.611628 for PC1 and 0.609410 for PC2) further reinforce this anomalous behavior.

5 Conclusion and Future Works

This paper presents a methodology for detecting and evaluating anomalous activities on YouTube channels, focusing on the US Air Force, news, and the Indo-Pacific region channels. By integrating engagement metrics and commenter behavior, we devised scoring methods to quantify the degree of anomalous behavior exhibited by each channel. Our analysis, spanning 71 channels, highlights various forms of anomalies, from commenter behavior to irregularities in engagement levels. This approach provides a robust method for identifying potential anomalous activities, while different aggregation strategies provide nuanced insights into anomaly detection. In future research, we aim to delve deeper into the content analysis of comments and videos to distinguish between "good" and "bad" anomalies. By examining the semantic content and sentiment of comments, along with the themes and context of videos, we intend to develop more nuanced anomaly detection methods. Additionally, we plan to explore discrepancies in engagement levels, examining how fluctuations in likes, shares, and other forms of interaction can serve as indicators of anomalous activity on YouTube channels.

Acknowledgment

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189, W911NF-23-1-0011, W911NF-24-1-0078), 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, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). 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 Community Guidelines & Policies - How YouTube Works — youtube.com. https://www.youtube.com/howyoutubeworks/policies/community-guidelines/ (2024), [Accessed 05-05-2024]

2. Adeliyi, O., Solaiman, I., Shajari, S., Onyepunuka, U., Agarwal, N.: Detecting and characterizing inorganic user engagement on youtube. In: Workshop Proceedings of the 18th International AAAI Conference on Web and Social Media (2024). https://doi.org/10.36190/2024.01
3. Byun, U., Jang, M., Baek, H.: The effect of youtube comment interaction on video engagement: focusing on interactivity centralization and creators' interactivity. Online Information Review **47**(6), 1083–1097 (2023)
4. D'Alconzo, A., Casas, P., Fiadino, P., Bar, A., Finamore, A.: Who to blame when youtube is not working? detecting anomalies in cdn-provisioned services. In: 2014 International Wireless Communications and Mobile Computing Conference (IWCMC). pp. 435–440. IEEE (2014)
5. Galeano, K., Galeano, R., Agarwal, N.: An evolving (dis) information environment–how an engaging audience can spread narratives and shape perception: A trident juncture 2018 case study. Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities pp. 253–265 (2020)
6. Google Developers: Api reference | youtube data api. URL: https://developers.google.com/youtube/v3/docs (2021), visited on January 21, 2022
7. Khan, M.L.: Social media engagement: What motivates user participation and consumption on youtube? Computers in human behavior **66**, 236–247 (2017)
8. Kirdemir, B., Adeliyi, O., Agarwal, N.: Towards Characterizing Coordinated Inauthentic Behaviors on YouTube. In: ROMCIR@ ECIR. pp. 100–116 (2022)
9. Kready, J., Shimray, S.A., Hussain, M.N., Agarwal, N.: Youtube data collection using parallel processing. In: 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). pp. 1119–1122. IEEE (2020)
10. Ksiazek, T.B., Peer, L., Lessard, K.: User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. New media & society **18**(3), 502–520 (2016)
11. Lahitani, A.R., Permanasari, A.E., Setiawan, N.A.: Cosine similarity to determine similarity measure: Study case in online essay assessment. In: 2016 4th International conference on cyber and IT service management. pp. 1–6. IEEE (2016)
12. Shajari, S., Agarwal, N., Alassad, M.: Commenter behavior characterization on youtube channels. In: Proceedings of the eKNOW International Conference on Information, Process, and Knowledge Management, Venice Italy. pp. 59–64 (2023)
13. Shajari, S., Alassad, M., Agarwal, N.: Characterizing suspicious commenter behaviors. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining. pp. 631–635 (2023). https://doi.org/10.1145/3625007.3627309
14. Shajari, S., Amure, R., Agarwal, N.: Analyzing anomalous engagement and commenter behavior on youtube. AMCIS 2024 Proceedings. 6 (2024), https://aisel.aisnet.org/amcis2024/social_comp/social_comput/6
15. Social Blade: Getting started with the business api (matrix) - socialblade.com. URL: https://socialblade.com/business-api (2023), visited on August 22, 2023
16. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology **61**(3), 611–622 (1999), https://academic.oup.com/jrsssb/article-abstract/61/3/611/7083217, publisher: Oxford University Press
17. Yousefi, N., Cakmak, M.C., Agarwal, N.: Examining multimodal emotion assessment and resonance with audience on youtube. In: Proceedings of the 9th International Conference on Multimedia and Image Processing (ICMIP), IEEE, accepted for presentation (2024)