Detecting and Measuring Anomalous Behaviors on YouTube

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. In an era where online content can sway public opinion and shape narratives, the influence of YouTube comment sections cannot be underestimated. Anomalous commenter behavior, which often includes organized groups of commenters who collaborate to manipulate public opinion, can disseminate false information and influence discussions, particularly on topics related to sensitive geopolitical and potential economic growth regions such as the Indo-Pacific. Thus, the need for detection mechanisms to assess and quantify anomalous commenter behavior on YouTube channels becomes critical, not only for safeguarding the platform but also for ensuring the quality of information and discussions in this significant area. This study centers on YouTube channels related to the Indo-Pacific region and introduces unsupervised learning methods for detecting anomalous commenter behavior on these channels. Our dataset includes 35 YouTube channels, 308,890 videos, 726,078 commenters, and 1,536,284 comments. The research objective is to assign a normalized score, ranging from 0 to 1, that shows the level of anomalous commenter behavior on each channel. This is accomplished by employing a combination of methods such as Kernel Density Estimation (KDE), Gaussian Mixture Model (GMM), comment activity analysis, and qualitative analysis.

Keywords: Social Network Analysis · YouTube · Anomalous Behavior · Commenter Behavior · Gaussian Mixture Model · Kernel Density Estimation

1 Introduction

YouTube, the largest video-sharing platform globally, has revolutionized the way we access information and entertainment, boasting billions of users and experiencing an incessant influx of content every minute. As a digital giant, YouTube influences narratives, instigates discussions, and provides a global platform for individual voices. While the comment sections on YouTube videos serve as a vital space for interacting with viewers, sharing insights, and expressing opinions, they also raise significant concerns. This open forum is crucial for fostering engagement and building communities that can propagate misinformation, posing potential harm or influencing other communities in conflict areas, such as the

Indo-Pacific region. This issue extends beyond being a mere annoyance; anomalous commenter behavior carries real-world consequences, making it imperative to address this concern not only for YouTube but also for any platform for those who value meaningful and constructive dialogue.

To address the issue of detecting such problematic behavior, this paper focuses on 35 YouTube channels related to the Indo-Pacific region, driven by the geopolitical significance of the area. The term "Indo-Pacific" refers to a vast region encompassing the Indian Ocean and the western and central Pacific Ocean, including the South China Sea. This region holds strategic importance due to its role in global trade, security dynamics, and geopolitical considerations. The selected 35 YouTube channels represent a spectrum of countries and regions within the Indo-Pacific, offering diverse perspectives on politics, religion, and conflict. These channels provide valuable insights, especially within the Indo-Pacific region, where geopolitical complexities and cultural diversities play a pivotal role in influencing global affairs.

In a broader context, "anomalous" for the purpose of this study encompasses a state of doubt or uncertainty regarding the motives, actions, or intentions of individuals, groups, or institutions within a social framework. It conveys a sense or perception that something or someone may lack complete trustworthiness, honesty, or genuineness. This skepticism can be triggered by various factors, including unusual behavior, conflicting information, or a lack of transparency. Specifically, within the scope of this research, we refine our definition of "anomalous" to encompass groups of commenters collaborating to boost engagement on YouTube channels. This behavior is deemed unusual and stands out as it deviates from typical interaction patterns, contributing to our understanding of anomalous activities.

Our unique contribution is an innovative method that assigns an anomalous score ranging from 0 to 1 to each channel, reflecting its level of anomalous commenter activity. A score close to 0 indicates less anomalous behavior, while a score close to 1 suggests more anomalous behavior in the comment section. This score is a numerical measure assigned to channels, indicating the likelihood or degree of anomaly associated with them based on their commenter behavior. It serves as a composite metric derived from various features of the co-commenter network. This is achieved by employing methods such as Kernel Density Estimation (KDE) and Gaussian Mixture Model (GMM) on the 20 network features extracted from the channel's co-commenter network. To evaluate the effectiveness of the proposed methodology, channels with high anomalous scores are further examined by applying comment activity analysis to identify anomalous activities among commenters within a clique. Through qualitative analysis, the video content is examined with a specific focus on the correlation with anomalous commenter behavior.

The subsequent sections of the paper are organized as follows: Section II provides an overview of anomalous activities observed on YouTube, along with the existing knowledge in this domain. Section III delineates the techniques employed for data collection. Section IV describes the research methodology utilized. Section V presents the study's results, encompassing a comprehensive examination of the data and behaviors exhibited by commenters and channels. Lastly, section VI summarizes the research with directions for future research.

2 Related Works

Several approaches have been proposed to examine anomalous behaviors on YouTube. Authors in [16] employed principal component analysis (PCA) alongside clustering techniques to detect suspicious commenter activities on YouTube channels. Similarly, researchers in [8] introduced a method leveraging engagement scores to identify inorganic activities within YouTube networks. Additionally, the study by [9] presented an unsupervised approach integrating multiple layers of analysis to probe coordinated inauthentic behaviors on YouTube.

In [17], researchers explored diverse methods like Graph2vec and UMAP to classify channels based on suspicious commenter behavior and their similarities. Meanwhile, [13] scrutinized YouTube comments on videos combating the stigma surrounding mental illness. In another study, [14] delved into the escalating toxicity within YouTube channels, noting heightened toxic behavior among anti-NATO commenters and stressing the imperative of addressing and understanding online toxicity. Similarly, [1] noted that while social media platforms revolutionize communication, they also spread misinformation. This study analyzed 3,542 YouTube channels, identifying automated engagement boosts and misinformation, leading to some channel suspensions. Additionally, [18] proposed a combined system to assess anomalous behavior on YouTube, revealing widespread anomalies and focusing on enhancing information quality in sensitive geopolitical contexts. Furthermore, [23] explored how multimedia elements in YouTube videos influence user emotions, offering insights for content creators and marketers to enhance engagement.

Authors in [7] employed inverse reinforcement learning and behavioral economics to model and predict YouTube commenting behavior, often adhering to Bayesian utility optimization with rational inattention constraints. Meanwhile, [21] analyzed online sentiment towards TED talks on YouTube, revealing the prevalence of neutral responses, gender-based variations in reactions, and the potential impact of sentiment-based content moderation. The study by [4] explored mechanisms of imitation, intergroup interaction, and communities of co-commenters, focusing on users regularly interacting by commenting on the same posts. Moreover, [20] analyzed the spread of information on the Uyghur issue through blogs and social media, revealing key topics and influential blogs raising awareness about their oppression. Furthermore, [11] found that combining emotion analysis and morality assessment techniques provides a more comprehensive understanding of text data, capturing the complexities of human communication and behavior.

Authors in [24] used a multi-method framework to analyze audio, video, and text modalities in YouTube videos, revealing patterns and themes in South China Sea content. The study also identified potential content amplification techniques.

In addition, [2] examined how colors in YouTube videos impact viewer emotions, revealing significant correlations and providing insights for digital marketing.

While these studies present various methods for analyzing anomalous behaviors and user interactions on YouTube, they each have their strengths and limitations. Techniques like PCA and clustering provide powerful tools for pattern recognition; however, they may lack context-specific accuracy. Methods leveraging engagement metrics offer straightforward indicators, yet they can be susceptible to manipulation. Unsupervised approaches and multi-layer analyses offer comprehensive insights, though they require substantial computational resources and can be complex to implement. Sentiment analysis and studies focusing on specific social issues provide valuable insights; nonetheless, they often do not address the detection of anomalous behavior directly.

The research field has a gap in the exploration of analyzing anomalous commenter behaviors on YouTube, which have the potential to either boost engagement or create artificially viral content. Hence, our emphasis in this paper is on addressing this knowledge gap at the channel level. This study takes a distinctive approach, setting itself apart by categorizing YouTube channels based on their anomalous commenter scores. The novelty of our work lies in the unique categorization of YouTube channels by anomalous commenter scores, providing a fresh perspective on detecting and understanding engagement anomalies at the channel level. This approach not only addresses the existing gap in the research but also offers a practical framework for identifying and mitigating inauthentic activities on YouTube.

3 Data Collection

The data analyzed in this study was obtained using a specialized tool for collecting data from YouTube [10], which utilizes YouTube's API [6]. Videos, comments, and channel data were collected from YouTube using keywords. The relevant keywords were identified by studying coverage in the Indo-Pacific region with further reviews to improve the inclusiveness of the keywords. "Komunis Cina | China pengaruh Indonesia", "Muhammadiyah Cina | China | Tiongkok | Tionghoa", "Kejam Uighur | Uyghur", "Muslim Brother | Indonesia Uighur | Uyghur", etc., are examples of keywords used in the data collection. The dataset included 35 YouTube channels, 308,890 videos, 726,078 commenters, and 1,536,284 comments.

4 Methodology

An overview of our model is provided in this section. Section A explains the creation of a co-commenter network. Subsequently, section B discusses the extraction of co-commenter network features. The process of calculating anomalous scores for channels is outlined in section C.

4.1 Creating Co-commenter Network

The co-commenter network consists of connections between commenters who posted comments on the same video. The weight of the connections between these commenters is determined by the number of shared videos on which both commenters have commented. This weight is considered the threshold for establishing a connection between two commenters in the network and is set at 5 in this research [16]. This implies that the network is formed by linking commenters who have jointly commented on a minimum of 5 videos. Essentially, this criterion helps establish stronger connections in the network, focusing on commenters who have interacted more frequently across a significant number of shared videos.

4.2 Capturing Co-commenter Network Characteristics

In this section, we study the properties of the co-commenter network, exploring features that can provide insights into the network structures, relationships, patterns, and dynamics and any other interconnection behaviors. For this purpose, we extracted a set of 20 network structural features to investigate co-commenter networks, taking inspiration from the work by [9]. These 20 features include parameters such as the number of nodes, edges, unique commenters, and comments and the ratios and averages such as the normalized ratio of co-commenters, average degree, density, and clustering coefficients.

Also, other characteristics related to network modularity and clique structures are considered, such as the number of maximal cliques with at least 5 members, unique commenters in cliques, and various ratios involving commenters in cliques. In addition, other important features that delved into the clique network's structure and connectivity were explored, including the average degree of cliques and clustering coefficients within cliques, etc. Incorporating these 20 network structural features should help gain insights into the roles and influence of individual nodes, the overall connectivity of the network, and potential vulnerabilities or patterns within the network structure.

4.3 Calculating Anomalous Score

This section describes the methods used for calculating the anomalous scores of channels using Kernel Density Estimation (KDE) and Gaussian Mixture Model (GMM) methods. Our initial step was to investigate the distribution of each of the 20 features in depth. We gained insights into the intrinsic characteristics of the co-commenter networks under examination by visualizing the distribution of these features. This preliminary investigation sought to identify any observable patterns, anomalies, or deviations that would be indicative of anomalous behavior. We observed that the size distribution of the cliques in each feature tends towards a Poisson Distribution [19] as shown in Figure 1.

This step extracts the value that minimizes the difference between the observed clique size distribution and the expected Poisson Distribution. This extracted value is then used to replace the number of cliques for each channel,

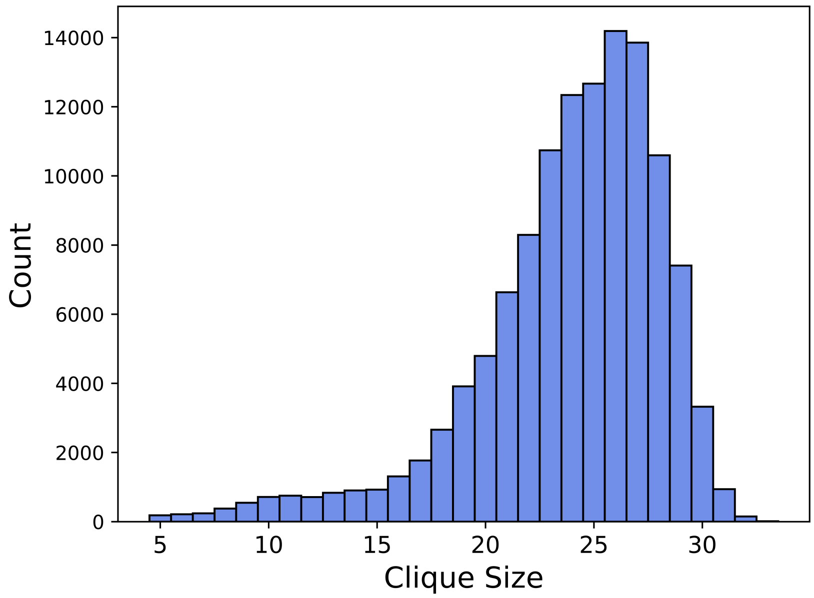

Fig. 1. Size distribution for the number of cliques across some of the channels.

providing a standardized reference for understanding the distribution of cliques across the examined co-commenter networks. This process helps establish a baseline and identify deviations that may indicate unusual or anomalous behavior within the co-commenter network. In the next step, we studied the distribution of the other features using a Kernel Density Estimate (KDE) plot [15]. A KDE plot utilizes a Gaussian kernel to smooth the observations, resulting in a continuous density estimation [22]. The resulting KDE plots reveal that the data can be separated into two distinct distributions as shown in Figure 2. The plot shows the distribution of channels across some of the 20 features. The analysis of the KDE plots for key features, including the "average clustering coefficient", "average cluster coefficient of the cliques", "total comments", and "number of comments by clique/number of nodes", reveals a discernible categorization of channels within the data. The peaks in the distribution of each feature differ but show a similar "bi" distribution. This offers a valuable perspective for deciphering the fundamental details of the co-commenter network. This property can also be observed in the remaining features.

Following that, GMM was used to separate the data in each feature into their respective distributions. The GMM is a probabilistic model in which every data point is produced by combining a finite number of Gaussian distributions whose parameters are unknown [12]. Figure 3 shows the result of applying the model on the average clustering coefficient as an example. A similar analysis was carried out on all other 20 features.

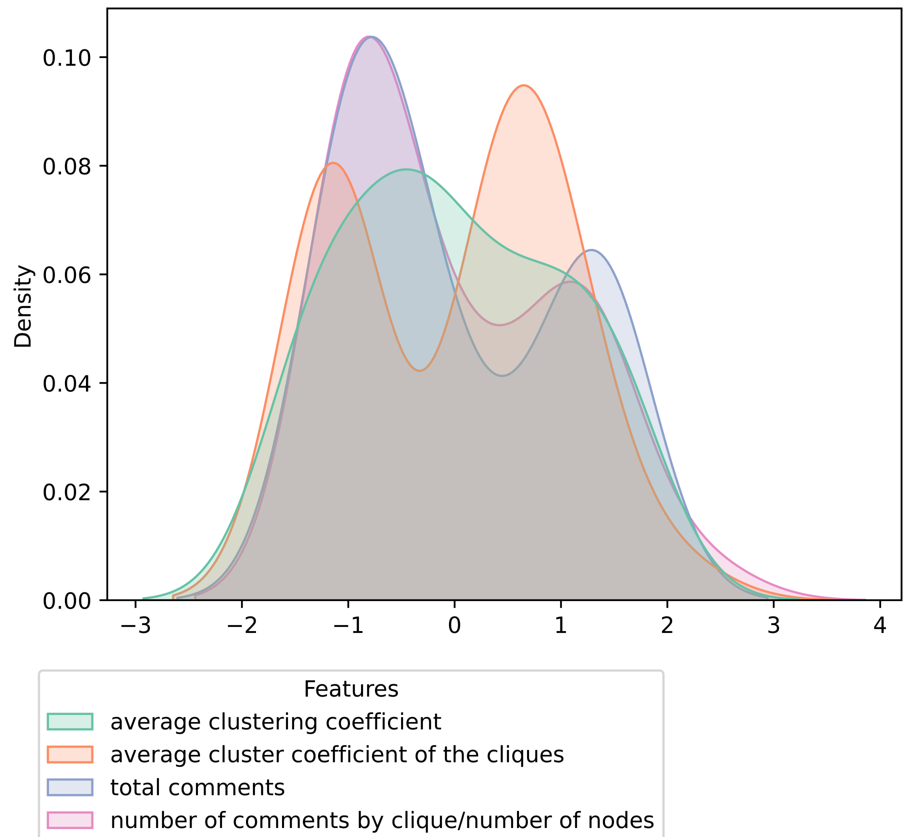

Fig. 2. KDE plots for the features.

The data in each feature was well separated using the Gaussian Mixture model as shown in Figure 3. The result confirms our initial observation in Figure 2, thereby strengthening the idea that channels can be classified according to the collected features. The distance between the two distributions in each feature is a measure of the potential of each feature to separate the channels into distinct categories. Each feature's separability degree is measured using Cohen's effect size (d) [3] given by (1). d measures the effect size that quantifies the difference between the means of two distributions while accounting for their standard deviations.

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{{\delta_1}^2 + {\delta_2}^2}{2}}} \tag{1}$$

Where μ_1 and μ_2 are the mean and δ_1 and δ_2 are the standard deviation of the data in the two distributions. This measure bears a resemblance to a Z score [5]. The d value can be interpreted intuitively as the larger values indicate

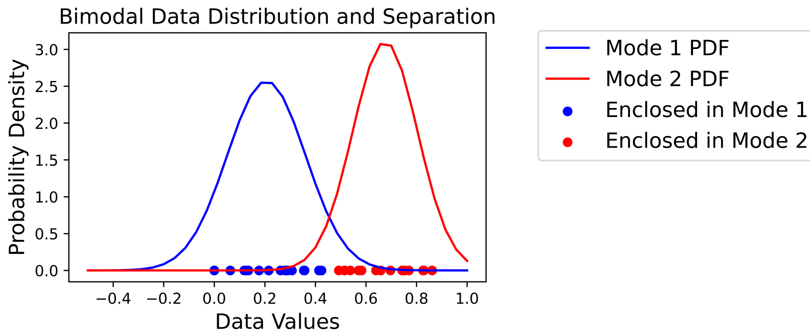

Fig. 3. Separation of data using the Gaussian Mixture Model fitted on average clustering coefficient values.

a larger separation between the means of the two distribution. This particular property of the d value is suitable for our use case. The d value is calculated for each feature, and it is used as the weight of each feature in calculating the anomalous scores, as presented in (2).

$$\mathcal{S} = \sum_{i=0}^{n} d_i \cdot x_i \tag{2}$$

Where $\mathcal{S}$ is the anomalous score, d_i is the Cohen value (weight) for the feature i, and x_i is the value of the feature i for the channel.

5 Result

This section presents the results of assigning anomalous scores to each channel, as detailed in section A. To evaluate the effectiveness of our findings, we performed comment activity analysis on the largest clique found in each of the top three most anomalous channels, as discussed in section B. Furthermore, in section C, we conducted an in-depth examination of the video contents in the three most anomalous channels to provide a comprehensive qualitative assessment.

5.1 Assigning Anomalous Scores to Channels

This section presents the results of the data analysis conducted on both features and the calculation of the anomalous score. Figure 4 shows channel IDs on the x-axis and the anomalous scores on the y-axis in descending order. Notably, a distinction exists between channels with relatively higher and lower anomalous scores. For example, the channel "Towards Eternity" ("UCPubBVD-Czu7IWWnitlkEsNw") holds the highest anomalous score of 0.71, marking it as the most anomalous channel in this dataset. The second highest is the "Invoice Indonesia" ("UCDg1bjGQOONSRmRSOyTYMdA") channel with a anomalous score of 0.70, followed by the "Draft Popular" ("UCQW2tTNJ40V2T6MqV1eMZ0w") channel with a anomalous score of 0.69.

These findings suggest that certain network properties, such as the average clustering coefficient of the cliques, significantly influence the calculation of anomalous scores. Additionally, the total comments and the total number of commenters in the network emerge as crucial indicators of each channel's possible anomalous nature. The presented results not only highlight the channels with the highest anomalous scores but also emphasize the importance of specific network properties in determining the overall level of anomalous behaviour associated with each channel.

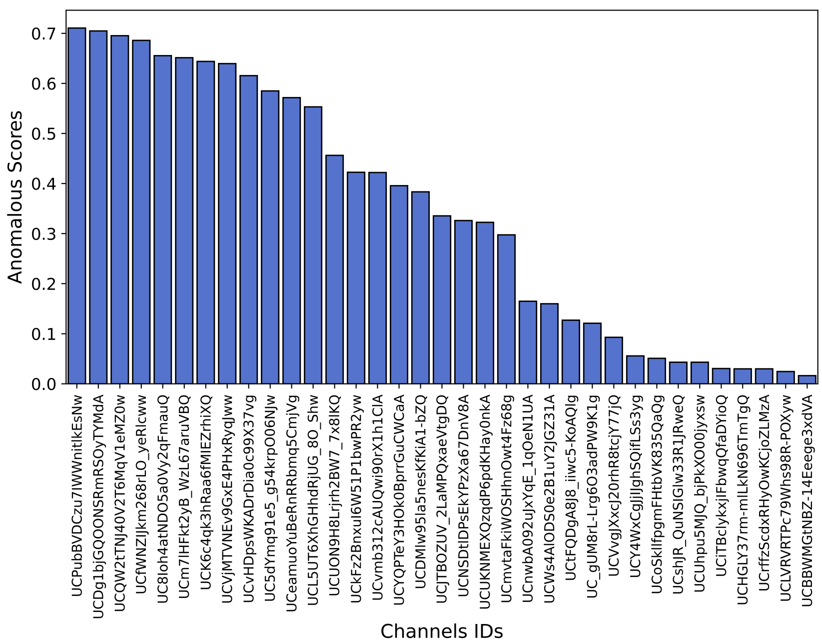

Fig. 4. Anomalous scores for 35 channels.

Table 1. Top Three Anomalous Channels Statistics.

Channel	Subscriber	# Cliques	Anomalous score
Towards Eternity	1.8M	368	0.71
Invoice Indonesia	1.45M	22934	0.70
Daftar Populer	15.4M	62072	0.69

Figure 5 presents the weight (coefficient) assigned to each feature obtained through (1). The coefficients assigned to each feature represent the weights associated with each feature in the calculation of anomalous scores, as outlined in (2). Each feature is assigned a specific weight, indicating its significance in categorizing the channels based on anomalous levels. For example, it is obvious that "total comments", "average cluster coefficient of cliques", and "total commenters" stand out as the most important features used in calculating the anomalous score. Furthermore, Figure 5 reveals the distribution of weights (coefficients) assigned to different network features in determining anomalous scores. This variation in weights highlights the interplay between features, where some contribute more significantly to the overall score than others. Also, these features collectively contribute to assessing anomalous levels for each channel and aid in understanding the relative influence and interplay of individual features. Table 1 presents the three most anomalous channels, indicating the number of cliques (each with a minimum of 5 members), subscribers, and anomalous scores for each.

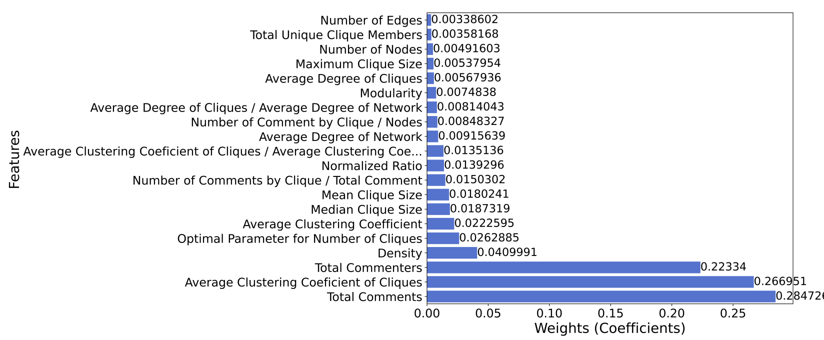

Fig. 5. The 20 network structure features and their corresponding weights.

5.2 Comment Activity Analysis

In this section, we extensively analyze the comment activity within a specific clique with the three most anomalous channels, identified based on their anomalous scores. Due to the large number of cliques for each channel, we analyze the behavior of clique commenters in the largest clique.

The temporal dynamics of the clique's engagement are visualized through a time-series plot, depicting the number of comments over the timeline. Figures 6, 7, and 8 sequentially present the commenter activity for the channels "Towards Etenity", "Invoice Indonesia", and "Daftar Populer". The biggest clique sizes for these channels are 8, 10, and 5, respectively. These figures provide a thorough understanding of the interaction pattern of the commenters and the comment dynamics within the identified YouTube channel clique.

The associated plots distinctly illustrate a consistent upward trajectory in these activities over time, suggesting a noticeable trend in anomalous behavior observed within each channel. Additionally, with the channel "Towards Eternity" as a most anomalous channel, mob behavior is evident, with a group of commenters increasingly posting comments over a specific time period and then disappearing. This proves the existence of mobs within the channel.

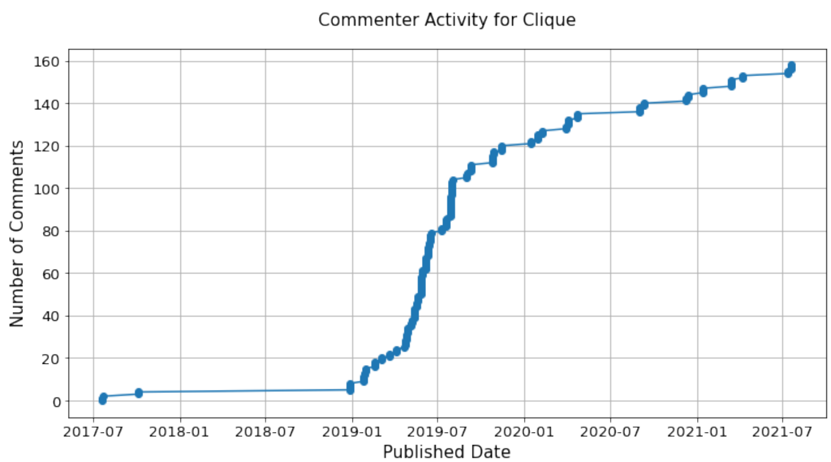

Fig. 6. Clique commenter activity for the channel "Towards Eternity".

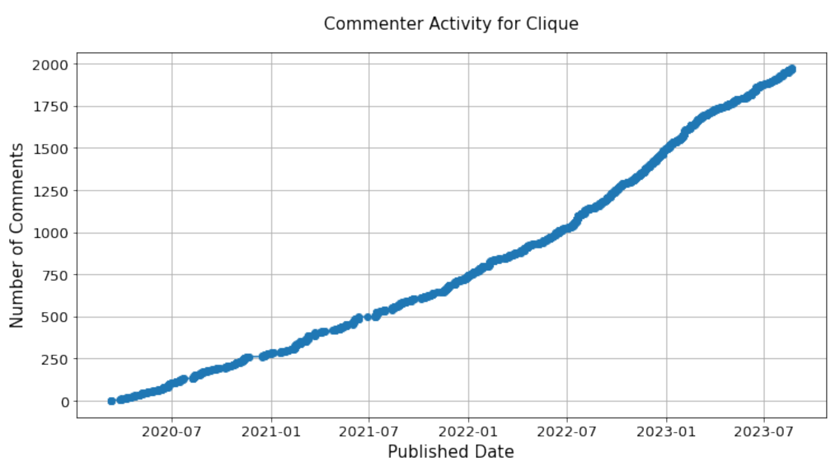

Fig. 7. Clique commenter activity for the channel "Invoice Indonesia".

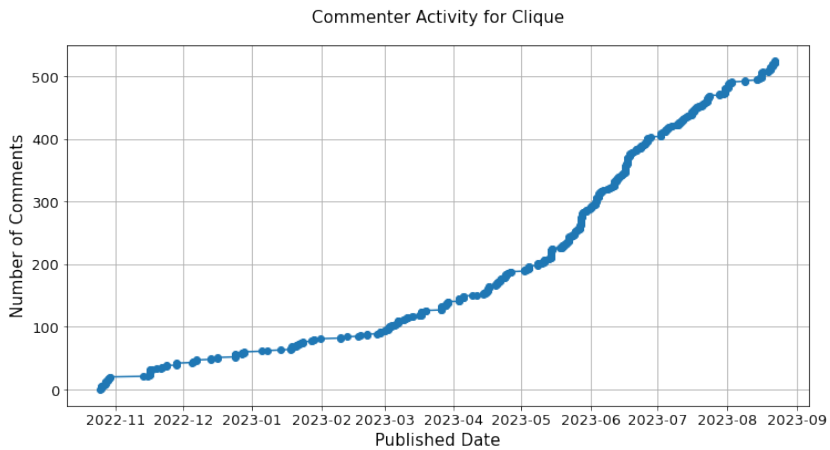

Fig. 8. Clique commenter activity for the channel “Daftar Populer”.

5.3 Qualitative Analysis

Upon careful examination of the three anomalous channels, a distinct pattern emerged. Channels predominantly featuring content related to religion, war, and political themes were associated with higher scores, indicating a heightened level of anomalous behavior among commenters. This correlation underscores the significance of thematic context in assessing and interpreting anomalous activities within the analyzed channels. In our in-depth exploration of clique commenter activity within section B, focusing on the top three anomalous channels, we carefully selected two videos as prime examples for illustration.

Figure 9 presents two video examples where a clique of 8 commenters actively and synchronously engaged in posting comments on videos associated with the “Towards Eternity" channel. This pattern persisted as we examined Figure 10, which depicts a similar scenario unfolding in the realm of “Invoice Indonesia" where a clique of 10 commenters collaboratively participated in the commentary. Figure 11 extends this narrative, capturing a parallel occurrence within the channel “Daftar Populer" where a group of 5 commenters demonstrated synchronized engagement. These types of behaviors raise questions about the motives behind such collective comment activities and prompt further investigation into the dynamics of these online interactions within the identified channels.

6 Conclusion and Future Works

In this paper, we present an effective methodology for identifying anomalous commenter behavior within 35 YouTube channels associated with the Indo-Pacific region using the structural properties of co-commenter networks. Our approach employs a network analysis-based method, utilizing 20 network structural features to generate an anomalous score for each channel. Delving further

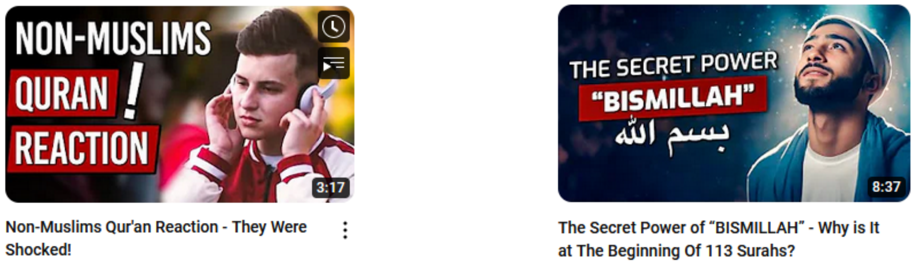

Fig. 9. Sample videos of channel "Towards Eternity".

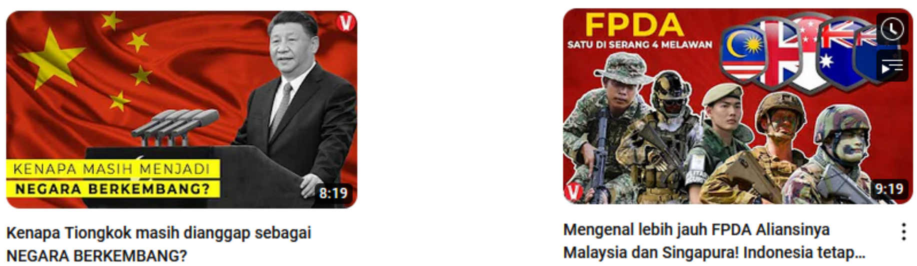

Fig. 10. Sample videos of channel "Invoice Indonesia".

into channels with elevated anomalous scores, we applied comment activity analysis to uncover instances of anomalous behavior among commenters, emphasizing a nuanced understanding of user engagement dynamics. Our qualitative analysis findings indicate that channels with higher scores tend to feature videos predominantly centered around themes of religion, war, or political news. This insight underscores the relevance of thematic content in understanding and addressing anomalous behaviors within the context of YouTube channels focused on the Indo-Pacific region.

In future studies, we aim to enhance the predictive capabilities of our model for larger datasets and explore the relationship between channel characteristics and anomalous activities. This includes examining how different features of a channel, such as the frequency of posts, the types of content shared, and the engagement levels of followers, correlate with anomalous behavior. By understanding these relationships, we hope to fine-tune our model to better detect and predict such activities. Furthermore, we plan to conduct a comprehensive examination of content creators operating within anomalous channels. This involves analyzing their posting patterns, content themes, and engagement dynamics. By gaining deeper insights into the behavior and strategies of these creators, we can develop more targeted interventions and strategies to mitigate their impact.

In addition to these goals, a comparative study with existing work will be undertaken. We will benchmark our model against state-of-the-art approaches in anomalous activity detection on social media platforms. By comparing the

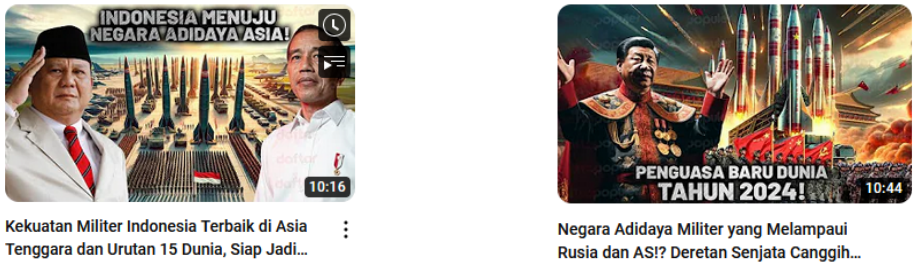

Fig. 11. Sample videos of channel "Daftar Populer".

performance, scalability, and accuracy of our model with established methodologies, we aim to identify strengths and areas for improvement. This comparative analysis will not only validate the effectiveness of our approach but also provide insights into the advancements made in this field. Such a study is essential for situating our work within the broader research landscape and ensuring that our contributions are both innovative and impactful.

Through these future endeavors, we anticipate strengthening the overall efficacy of our approach and contributing valuable knowledge to the ongoing efforts to maintain a secure and trustworthy online environment. By continuously improving our model and learning from both our own research and existing studies, we aim to develop a robust solution for detecting and addressing anomalous activities on YouTube and other social media platforms.

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. 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
2. Cakmak, M.C., Shaik, M., Agarwal, N.: Emotion assessment of youtube videos using color theory. In: 9th International Conference on Multimedia and Image Processing. (2024). https://doi.org/10.1145/3665026.3665028
3. Cohen, J.: Statistical power analysis for the behavioral sciences. Routledge (2013)
4. Ferreira, C.H., Murai, F., Silva, A.P., Almeida, J.M., Trevisan, M., Vassio, L., Mellia, M., Drago, I.: On the dynamics of political discussions on instagram: A network perspective. Online Social Networks and Media **25**, 100155 (2021)
5. Glass, G.V., McGaw, B., Smith, M.L.: Meta-analysis in social research. (No Title) (1981)
6. Google Developers: Api reference | youtube data api. URL: https://developers.google.com/youtube/v3/docs (2021), visited on January 21, 2022
7. Hoiles, W., Krishnamurthy, V., Pattanayak, K.: Rationally inattentive inverse reinforcement learning explains youtube commenting behavior. Journal of Machine Learning Research **21**(170), 1–39 (2020)
8. Hussain, M.N., Tokdemir, S., Agarwal, N., Al-Khateeb, S.: Analyzing disinformation and crowd manipulation tactics on youtube. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 1092–1095. IEEE (2018)
9. Kirdemir, B., Adeliyi, O., Agarwal, N.: Towards characterizing coordinated inauthentic behaviors on youtube. In: ROMCIR@ ECIR. pp. 100–116 (2022)
10. 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)
11. Mbila-Uma, S., Umoga, I., Alassad, M., Agarwal, N.: Conducting morality and emotion analysis on blog discourse. In: International Conference on Collaboration Technologies and Social Computing. pp. 185–192. Springer (2023)
12. McLachlan, G.J., Rathnayake, S.: On the number of components in a gaussian mixture model. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **4**(5), 341–355 (2014)
13. McLellan, A., Schmidt-Waselenchuk, K., Duerksen, K., Woodin, E.: Talking back to mental health stigma: An exploration of youtube comments on anti-stigma videos. Computers in Human Behavior **131**, 107214 (2022)
14. Obadimu, A., Mead, E., Hussain, M.N., Agarwal, N.: Identifying toxicity within youtube video comment. In: Social, Cultural, and Behavioral Modeling: 12th International Conference, SBP-BRiMS 2019, Washington, DC, USA, July 9–12, 2019, Proceedings 12. pp. 214–223. Springer (2019)
15. Scott, D.W.: Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons (2015)
16. 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)

17. 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
18. 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
19. Sharma, D., Singh, B., Raja, M., Regin, R., Rajest, S.S.: An efficient python approach for simulation of poisson distribution. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). vol. 1, pp. 2011–2014. IEEE (2021)
20. Umoga, I., Mbila-Uma, S., Alassad, M., Agarwal, N.: Analyzing blogs about uyghur discourse using topic induced hyperlink network. In: International Conference on Complex Networks and Their Applications. pp. 412–423. Springer (2023)
21. Veletsianos, G., Kimmons, R., Larsen, R., Dousay, T.A., Lowenthal, P.R.: Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on youtube ted talk comments. PloS one **13**(6), e0197331 (2018)
22. Waskom, M.L.: Seaborn: statistical data visualization. Journal of Open Source Software **6**(60), 3021 (2021)
23. 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)
24. Yousefi, N., Shaik, M., Agarwal, N.: Characterizing multimedia information environment through multi-modal clustering of youtube videos. In: In 4th International Conference on SMART MULTIMEDIA, Hilton Redondo Beach, Los Angeles, USA (2024)