Developing a Commenter Behavior-based Framework for Characterizing YouTube Channels

Shadi Shajari[1] and Nitin Agarwal[1,2]

[1] COSMOS Research Center, University of Arkansas, Little Rock, USA
[2] International Computer Science Institute, University of California, Berkeley, USA
sshajari@ualr.edu, nxagarwal@ualr.edu

Abstract. As a major platform for global content dissemination, YouTube hosts thousands of channels covering diverse topics and attracting varied user engagement. Understanding the behavioral characteristics of these channels is essential for comparative analysis and broader platform insights. While channel-level activity can be studied through multiple signals, the comment section offers a particularly rich source of user interaction data. Patterns in commenter behavior provide an effective basis for characterizing channels and identifying similarities across them. This study introduces a framework for characterizing YouTube channels based on commenter activity, using a dataset of 70 channels that includes 711,301 videos, 13,221,243 commenters, and 129,653,609 comments and covers topics such as U.S. military affairs, geopolitical developments, news, and randomly selected content. The methodology combines co-commenter network construction, clique extraction, feature aggregation, pairwise feature analysis, clustering methods, majority voting, and Euclidean distance measurement to uncover patterns in channel behavior. We define five distinct characterizations based on key commenter activity features such as posting patterns, content diversity, and engagement structure. By relying on unsupervised methods and a combination of structural and behavioral features, this approach enables large-scale, label-free comparison of channels. It provides a flexible framework for identifying groups of channels with similar audience dynamics, offering insights that can support better channel characterization and future studies in social media behavior.

Keywords: Social Network Analysis, Commenter Behavior, Majority Voting, Characterization, Euclidean distance, YouTube

1 Introduction

YouTube is one of the largest video-sharing platforms globally, hosting a vast array of channels that cover a wide range of topics and attract diverse audiences. The platform's comment sections serve as a key space for user interaction, making it a valuable resource for analyzing engagement patterns across different channels. Identifying similarities in commenter behavior provides critical insights into how channels foster audience interaction, shape community dynamics, and exhibit unique communication styles. Despite the richness of this data, systematic methods for comparing channels based on the behavior of commenters have been limited. Existing research has largely concentrated on detecting specific types of anomalous commenter behavior, such as spam, hate speech, or coordinated activity, which contributes significantly to understanding problematic user behavior [8, 9]. However, these studies often fail to address the broader patterns of commenter engagement across channels, leaving a gap in understanding the overall landscape of audience dynamics on YouTube.

In this paper, we propose a comprehensive framework for characterizing YouTube channels based on commenter behavior, focusing on a dataset comprising 70 channels across various topics. This framework examines not only the frequency and timing of interactions but also the underlying diversity in the nature and structure of the comments. The key innovation of our approach lies in the identification of five distinct channel characterizations, each reflecting a distinct pattern of commenter activity. To uncover these patterns, we first perform pairwise feature analysis, followed by the application of unsupervised clustering techniques, including K-Means, Agglomerative Hierarchical Clustering, Mean Shift, Spectral Clustering, Gaussian Mixture Models (GMM), Affinity Propagation, and Fuzzy C-Means, and finally apply majority voting to consolidate the clustering results. These methods allow for large-scale, label-free comparisons of channels, offering fresh insights into channel dynamics. The clustering algorithms group channels based on a variety of behavioral dimensions, including interaction consistency, content variability, and engagement patterns. To enhance the robustness of the clustering results, a majority voting system is used to determine the final cluster for each channel, ensuring that the most consistent patterns across different algorithms are captured. Additionally, Euclidean distance is employed to measure the behavioral proximity between channels, helping to identify which channels exhibit similar engagement profiles. This approach provides a more comprehensive understanding of how different channels engage with their audiences, offering actionable insights for content analysis and platform moderation.

The strength of this work lies in its ability to systematically categorize and compare YouTube channels based on their overall behavioral characteristics rather than focusing on individual anomalies. By leveraging multiple unsupervised clustering methods, the proposed framework ensures robustness and reliability in identifying meaningful channel groupings. Moreover, the use of majority voting across clustering algorithms and the calculation of Euclidean distance between channel behaviors provides a refined, nuanced approach to understanding engagement patterns. This framework not only advances the analysis of commenter behavior but also holds practical implications for channel characterization, audience targeting, and future studies in social media engagement dynamics.

2 Related Work

YouTube's comment sections have been the focus of numerous studies examining challenges that affect the quality of discourse. Common issues identified include spam, bot-generated content, hate speech, and misinformation. Researchers have approached these problems through a variety of methods, including content analysis, sentiment analysis, machine learning, and network-based techniques. Different forms of anomalous behavior pose unique challenges for accurate detection and response, leading to the development of increasingly sophisticated analytical approaches in recent studies. The study by [8] investigated bilingual hate speech detection in YouTube comments. They found that incorporating cross-platform data from Twitter and Gab improved model performance. Their best model achieved an F1-score of 0.74 for English and 0.68 for German content.

Research by [7] investigated "social scam bots" on YouTube, which are automated accounts that mimic real users by posting legitimate-looking comments and accumulating likes to gain prominence. This behavior enables them to evade surface-level detection filters, posing challenges for traditional moderation systems. Hate speech in Portuguese YouTube comments was studied by [3] using a 20k-comment corpus. They focused on both overt and covert hate targeting Afro-descendant, Roma, and LGBTQ+ groups. Covert hate often used irony, fallacies, and sarcasm, making it harder to detect. Building on these efforts, [13] introduced a scoring framework that combines engagement metrics with co-commenter network analysis to detect manipulation across videos. Their approach helps reveal clusters of commenters engaging in coordinated behavior. In a follow-up study, [14] applied Kernel Density Estimation and Gaussian Mixture Models on commenter networks to assign anomaly scores to channels, highlighting the effectiveness of blending textual and network features for identifying subtle coordinated actions. Authors in [10] explored engagement patterns on channels covering political and military content. They found that such channels often display structural anomalies in their commenter networks, such as unusual clustering coefficients and modularity values.

Another contribution by [12] used Graph2vec and UMAP to extract structural embeddings of commenter networks. They identified distinct "commenter mob" patterns, which could be clustered using k-means and hierarchical techniques. A large annotated dataset of YouTube comments countering hate against Jews and Blacks was built by [6]. They identified six types of counterspeech and showed that such comments tend to receive more likes. Their classifiers achieved an F1-score of 0.71 for detecting counterspeech. In [1], rolling window correlation and anomaly detection were applied to capture sudden bursts of suspicious activity, further enhancing the ability to detect inorganic engagement. The study in [11] employed PCA and clustering methods to isolate coordinated behavior at the channel level, offering a high-level view of inauthentic engagement. Lastly, the work in [4] analyzed blackmarket-driven collusion on YouTube. By creating datasets of videos and channels submitted for artificial boosting, they developed models to detect collusive likes, comments, and subscriptions. Their proposed system, CollATe, incorporated temporal commenting patterns and semantic similarity, achieving over 90% true positive rates in detecting collusive behaviors that evade traditional moderation systems.

Together, these studies present a comprehensive view of comment activity on YouTube. By combining machine learning, natural language processing, and network analysis, researchers are uncovering complex behaviors that evade traditional detection systems and pose challenges to platform integrity [2]. However, despite these advancements, there remains a gap in methods for systematically characterizing YouTube channels based on commenter behavior using comprehensive, feature-driven analysis. Most existing research focuses on detecting specific types of anomalies or relies on limited aspects of the data. Our work addresses this gap by taking a broader perspective—analyzing multiple comment-based features collectively to compare channels based on their overall behavioral characteristics. This enables the detection of channels with similar traits that might be overlooked using conventional, single-dimensional approaches.

3 Data Collection

This study utilized the methodology proposed by [5] to collect data through the YouTube Data API. Data was collected on a daily basis and included video postings, comment content, commenter and video identifiers, timestamps, and comment counts. All data collection adhered to the guidelines outlined in the YouTube API documentation and Terms of Service [16]. The combined dataset included 70 YouTube channels covering a range of topics including U.S. military affairs, geopolitical developments, regional news, and randomly selected content. It consisted of 711,301 videos, 13,221,243 commenters, 129,653,609 comments.

4 Methodology

This section outlines the step-by-step process used to analyze commenter behavior across YouTube channels, from constructing the co-commenter network to identifying channels with similar commenter characterizations.

4.1 Co-Commenter Network and Clique Extraction

The co-commenter network was constructed by identifying relationships between commenters based on their activity across multiple videos. Each commenter was represented as a node, and an edge was formed between two commenters if they had both posted comments on at least three common videos. The weight of the edge corresponded to the number of shared videos, ensuring that only meaningful interactions were captured. This approach filtered out organic interactions and could highlight strong connections between commenters who frequently posted comments together, potentially indicating signs of organized interaction. The resulting network provided a structured representation of user interactions, which has been previously utilized in similar analyses [12, 11]. These prior studies suggest that this network structure could be effective in uncovering behavioral patterns and identifying key interaction trends within the commenting community.

To gain deeper insights into user behavior, we extracted cliques from the co-commenter network of each channel. A clique is defined as a fully connected group, meaning that every member of the clique has direct interactions with every other member. We focused on cliques with a minimum of five members, as smaller groups may not provide significant behavioral patterns. These cliques served as the primary units of analysis, enabling a more detailed examination of commenting behavior. Through these group-level interactions, we aimed to uncover and characterize the underlying behavioral patterns of different YouTube channels.

4.2 Feature Calculation for Cliques

We calculated a set of features for each clique based on all the comments posted by its members across all videos. These features were designed to summarize both the content and structure of user interactions. The goal was to create a behavioral commenter profile for each channel that allowed for meaningful comparisons, pattern discovery, and group-level characterization. The features are summarized below:

1. Sentiment Score (S): Measures the emotional tone of comments.
2. Sentiment Variance (SV): Indicate variation in sentiment within a clique.
3. Duplicate Comments (DC): Measures the ratio of duplicate comments in a clique.
4. Vocabulary Uniqueness (VU): Reflects the diversity of words used in comments.
5. Edit Distance (ED): Represents how similar or different the comments are.
6. Length Variability (LV): Shows how much comment lengths differ within the clique.
7. Spam Score (SS): Proportion of comments containing spam or promotional content.
8. Time Gap (TG): Measures the average time between consecutive comments.
9. Toxicity Score (TS): Indicates offensive or harmful language in comments.
10. Published Date (PD): Indicates how early or late comments were posted.

All scores were normalized between 0 and 1, except for toxicity and sentiment scores, which ranged from -1 to 1 to reflect their full polarity. Linguistic and behavioral features such as sentiment, toxicity, spam score, and vocabulary uniqueness were computed using standard NLP tools and rule-based heuristics on comment text. These features capture emotional tone, offensive language, promotional content, and lexical diversity at the clique level. To ensure feature reliability, we required a minimum number of comments per clique. Given that each edge in the co-commenter network links users who commented on at least three shared videos, and that a clique has at least five members, we computed a baseline threshold based on the clique structure. Although cliques could be larger, we used $K = 5$ as the minimum size for analysis.

$$N_{\min} = \binom{K}{2} \times M \times 2 \tag{1}$$

$$\Rightarrow N_{\min} = \binom{5}{2} \times 3 \times 2 = 10 \times 3 \times 2 = 60$$

According to Equation 1, a clique with 5 members must have at least 60 individual comments to meet the minimum threshold for analysis. This ensures that feature extraction is based on sufficient interactions, improving the reliability of behavioral characterization. For a given clique C with N comments ($N_{\min} \leq N$), each behavioral feature is computed independently as follows:

$$F_{cl}^{(j)} = \frac{1}{N} \sum_{i=N_{min}}^{N} F_i^{(j)} \tag{2}$$

Here, $F_{cl}^{(j)}$ denotes the aggregated value of the jth feature (e.g., sentiment, toxicity, edit distance, etc.) for the clique, and $F_i^{(j)}$ is the corresponding feature value for the ith comment. This computation was repeated for all $j \in \{1, 2, \ldots, 10\}$ features, as they are summarized above, resulting in a separate score for each feature for every clique.

4.3 Aggregating Clique Features to Channel-Level

After calculating individual feature scores for each clique within a channel, we further aggregated these values to obtain a unique score per feature for the entire channel. This step captured the channel's overall behavior by summarizing all clique-level features into one representative value per feature. For each feature j, the channel-level score is computed as:

$$F_{ch}^{(j)} = \frac{1}{C} \sum_{k=1}^{C} F_{cl_k}^{(j)} \tag{3}$$

Here, $F_{ch}^{(j)}$ represents the aggregated value of the jth feature for the channel, C is the total number of cliques in the channel, and $F_{cl_k}^{(j)}$ is the jth feature score of the kth clique. This results in a single score for each feature at the channel level, enabling consistent comparison across channels based on shared behavioral dimensions.

4.4 Pairwise Feature Analysis

To better understand channel-level commenter behavior, we aimed to extract some meaningful relationships between selected pairs of behavioral features.

These pairwise comparisons were intended to reveal how different aspects of commenting behavior relate to one another, offering deeper insight into the interaction patterns across channels. By analyzing these combinations, we could uncover behavioral signals that contribute to a more detailed characterization of how users engage with content. The feature combinations include:

1. Toxicity Score (TS) & Spam Score (SS)
2. Spam Score (SS) & Vocabulary Uniqueness (VU)
3. Edit Distance (ED) & Toxicity Score (TS)
4. Published Date (PD) & Spam Score (SS)
5. Duplicate Comment (DC) & Time Gap (TG)
6. Length Variability (LV) & Sentiment Variance (SV)
7. Sentiment Score (S) & Toxicity Score (TS)
8. Length Variability (LV) & Spam Score (SS)
9. Time Gap (TG) & Edit Distance (ED)
10. Published Date (PD) & Duplicate Comment (DC)

These pairwise relationships were treated as meaningful behavioral signatures to capture interactions between features. For example, combining sentiment and toxicity or time gap and duplicate comments can reveal joint behavioral patterns that may not be apparent when features are analyzed individually. While some clusters may be influenced by a dominant feature, the pairwise approach supports more nuanced comparison. The results are visualized using scatter plots, discussed in Section 5.

4.5 Multi-Method Clustering with Majority Voting

To detect patterns across channels and explore possible groupings based on behavioral features, we applied multiple clustering algorithms to each pairwise feature combination, as described in Section 4.4. Specifically, we used seven clustering methods: compact groups (K-Means), hierarchical relationships (Agglomerative Hierarchical), dense regions (Mean Shift), graph-based structures (Spectral), probabilistic overlaps (GMM), exemplar-based assignments (Affinity Propagation), and flexible membership modeling (Fuzzy C-Means). Each algorithm was applied independently to all pairwise combinations, allowing us to examine clustering stability and consistency from different perspectives. This multi-method approach increased robustness, as each technique contributed unique strengths depending on data shape and scale. The resulting cluster outputs helped identify channels with similar behavioral signatures, which were further analyzed. The final cluster assignment for each channel was determined via majority voting [15], selecting the most frequent label across algorithms for each feature pair. Visualization and interpretation of the clusters are presented in Section 5.

Let $\mathcal{X}^{(m)}$ denote the dataset of channels under the m-th pairwise feature combination, where $m = 1, 2, \ldots, M$ and $M = 10$ is the total number of feature pairs considered. For each feature pair $\mathcal{X}^{(m)}$, we applied $K = 7$ different clustering algorithms. Each algorithm, indexed by $k \in 1, 2, \ldots, K$, was configured to produce two clusters, allowing for consistent comparison across methods and feature combinations Let $C_i^{(m,k)}$ represent the cluster label assigned to channel i by the k-th algorithm on the m-th feature combination:

$$C_i^{(m,k)} = \text{Cluster}_k(\mathcal{X}^{(m)}) \tag{4}$$

Since each algorithm may label a channel differently, we determined the final cluster assignment for channel i on feature pair m using majority voting across the 7 algorithms:

$$C_i^{(m)} = \text{Mode}\left(\{C_i^{(m,1)}, C_i^{(m,2)}, \ldots, C_i^{(m,7)}\}\right) \tag{5}$$

Here, Mode{ ... } selects the cluster label (either 0 or 1) that appears most frequently among the seven results. For example, if five algorithms assign channel i to cluster 1 and two assign it to cluster 0, the final label will be cluster 1. This ensures consistency and robustness across diverse clustering methods. Repeating this process for all $M = 10$ feature combinations results in a set of final cluster labels for each channel:

$$\mathcal{C}_i = \{C_i^{(1)}, C_i^{(2)}, \ldots, C_i^{(10)}\} \tag{6}$$

This framework ensured that clustering decisions were stable and reflected consensus across multiple algorithms, providing a more reliable representation of each channel's behavior in pairwise feature spaces.

4.6 Channel Similarity Detection and Characterization

To assess similarities in channel behavior, we computed Euclidean distances between channels in each pairwise feature space. A feature-specific threshold was applied using the 10th percentile of all pairwise distances to identify the closest behavioral relationships. This fixed quantile was chosen to isolate strong similarity signals while reducing noise. Although heuristic, this choice provided consistent filtering across diverse feature distributions; future work can explore adaptive or data-driven alternatives. Channels that consistently fell below this threshold in multiple feature spaces were flagged for further analysis, allowing us to explore their shared behavioral characteristics and distinguish meaningful engagement patterns. Let there be M pairwise feature combinations, and let $\mathbf{x}_i^{(m)}$ represent the position (e.g., 2D coordinates) of channel i in the scatter plot for the m-th feature combination. To measure how similar two channels were in a given feature space, we computed the Euclidean distance between them. For each feature combination m, the Euclidean distance between channels i and j is calculated as:

$$d_{ij}^{(m)} = \|\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)}\|_2 \tag{7}$$

To determine whether a pair of channels are considered similar, we computed a threshold $\theta^{(m)}$ for each feature combination using the 10th percentile of all pairwise distances:

$$\theta^{(m)} = \text{Quantile}_{0.10}\left(\{d_{ij}^{(m)}\}_{i<j}\right) \tag{8}$$

Using this threshold, we defined a binary similarity indicator function $S_{ij}^{(m)}$, which assigns a value of 1 if the distance between channels i and j falls within the most similar 10% of pairs, and 0 otherwise:

$$S_{ij}^{(m)} = \begin{cases} 1, & \text{if } d_{ij}^{(m)} \leq \theta^{(m)} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

To assess overall similarity across all feature combinations, we computed an aggregated similarity score that counted how many times channels i and j are marked as similar:

$$S_{ij}^{\text{total}} = \sum_{m=1}^{M} S_{ij}^{(m)} \tag{10}$$

A higher value of S_{ij}^{total} indicates that two channels tended to remain close across multiple feature combinations, reflecting a more consistent pattern of behavioral similarity. This measure captured how frequently a given pair of channels is considered similar across different feature spaces. While this score is calculated for individual channel pairs, the same principle can be extended to identify larger groups of channels in which all members remain close to one another across several feature combinations.

5 Results

This section presents the outcomes of the proposed methodology, including clustering results across pairwise feature combinations, the identification of similar channel groups, and their behavioral characterizations.

5.1 Clustering Analysis Across Pairwise Feature Combinations

This section presents the results of applying majority voting across 7 clustering methods to the 10 pairwise feature combinations of behavioral features, as described in Section 4.5. Given the large number of plots (10 per method, totaling 70), we present only the majority voting results. This approach consolidates the outcomes from all 7 clustering methods, offering a more concise and clearer visualization of the most consistent groupings. For each pairwise feature combination, clustering was performed using 7 different algorithms, and each channel was assigned a final cluster label based on the most frequent label it received.

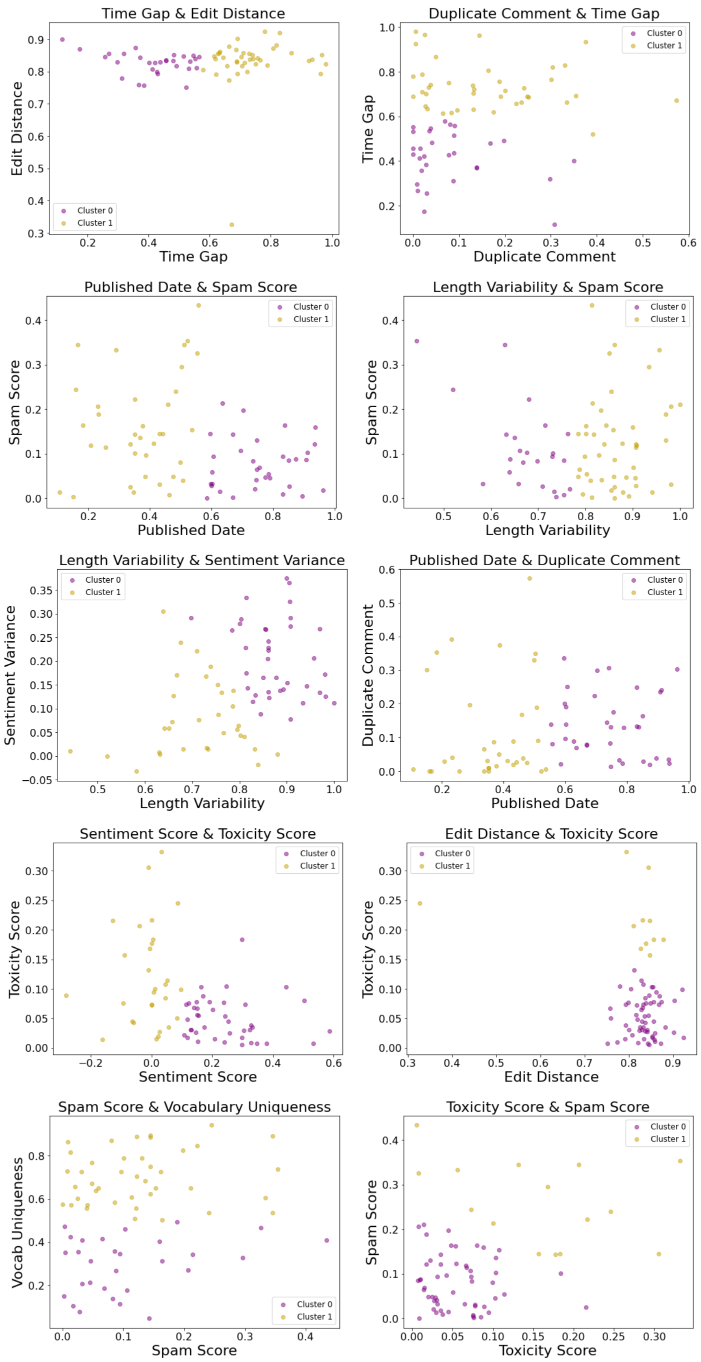

Fig. 1. Final Cluster Assignments for Pairwise Feature Combinations using Majority Voting.

The optimal number of clusters for each clustering method was determined using the Silhouette score. In most cases, 2 clusters were consistently identified as optimal, though some combinations yielded 3. To ensure consistency and enable majority voting, behaviorally similar clusters were merged as needed, resulting in 2 clusters per combination. This setup simplified the majority voting process and enabled consistent assignment of consensus-based cluster labels. Fig. 1 displays the final clustering outcomes for each pairwise feature space. In each plot, channels are positioned according to their normalized feature values and colored by their final cluster label, highlighting how channels group together or separate based on different combinations of comment-based behavioral features.

5.2 Behavioral Characterization of Similar Channel Groups

Building on the majority voting results across pairwise feature combinations, we identified broader patterns in channel behavior by defining generalized behavioral characterizations. Instead of directly clustering channels into final groups, we derived these characterizations from feature patterns that consistently appeared across multiple similarity groupings. Similarity was assessed using Euclidean distances in the 2D scatter plots (as defined in Equation 10), and channels that frequently appeared close together were considered behaviorally aligned. This provided a structured view of how channels engage across different dimensions. From a dataset of 70 YouTube channels, we identified five distinct behavioral characterizations, each reflecting unique structural and linguistic dynamics, as shown in Table 1.

Characterization 1 groups channels that exhibit consistent posting behavior and minimal repetition in comments. This is indicated by high values in Published Date, Time Gap, and Edit Distance, along with low values in Duplicate Comments and Spam Score. These channels demonstrate regular activity and low redundancy in their comment patterns, though they may not necessarily show a wide diversity in linguistic expression.

Characterization 2 includes channels where comment sections reflect a healthier and more constructive interaction environment. This pattern is indicated by a high Edit Distance, suggesting varied wording, along with low Toxicity and Spam Scores that reflect positive and non-disruptive communication.

Characterization 3 identifies channels where the pattern reflects frequent engagement combined with consistent posting behavior. This consistency is indicated by high levels in Published Date, Time Gap, and Edit Distance features. However, these channels also exhibit low levels of Duplicate Comments, Spam Score, and Vocabulary Uniqueness, suggesting limited linguistic diversity despite the regularity of their activity.

Characterization 4 captures channels where structured yet moderately diverse discussions are evident, as suggested by high Edit Distance and significant variability in comment length. These channels are further characterized by low levels of Toxicity and Spam Score, along with moderately high Vocabulary Uniqueness. However, sentiment variability remains limited.

Characterization 5 consists of channels where a balanced environment is reflected, with both structured posting and content variety, showing regular posting and some variety in comments with a flexible and organized style. These channels are characterized by low Toxicity and Spam Scores, along with high values in Published Date and Length Variability. Sentiment variability is relatively low.

Table 1. Summary of Channel Behavior Characterizations and Associated Groups.

#	Characterization	#Channels	Pairwise Features	Features
1	Regular Engagement with Moderate Content Diversity	13	PD & DC PD & SS DC & TG TG & ED	PD = 0.761 DC = 0.110 SS = 0.070 TG = 0.698 ED = 0.839
2	High-Quality and Constructive Interaction	10	TS & SS ED & TS	TS = 0.036 SS = 0.040 ED = 0.822
3	Frequent and Consistent Posting	5	SS & VU PD & DC PD & SS DC & TG TG & ED	SS = 0.051 VU = 0.286 PD = 0.823 DC = 0.052 TG = 0.720 ED = 0.847
4	Structured Interaction with Linguistic Diversity	4	TS & SS SS & VU ED & TS LV & SV LV & SS	TS = 0.021 SS = 0.032 VU = 0.592 ED = 0.838 LV = 0.822 SV= 0.099
5	Balanced Engagement with Content Variation	3	TS & SS PD & DC LV & SV PD & SS DC & TG LV & SS	TS = 0.025 SS = 0.066 PD= 0.775 LV = 0.806 SV = 0.110

These characterizations reveal distinct behavioral profiles shaped by a combination of structural and linguistic commenter features. By examining factors such as timing regularity, content repetition, vocabulary diversity, and sentiment variation, this classification provides a clear view of how YouTube channels foster different patterns of user engagement. Each group highlights a unique mix of engagement styles, offering valuable insights for better channel analysis and future research on online behavioral dynamics.

Fig. 2 presents an example from Characterization 2, illustrating a group of 10 channels that appear close to each other in two pairwise feature combinations: Toxicity Score (TS) & Spam Score (SS) and Edit Distance (ED) & Toxicity Score (TS). In this example, all highlighted channels, shown in red, belong to the same cluster in both plots, indicating consistent grouping across different feature spaces.

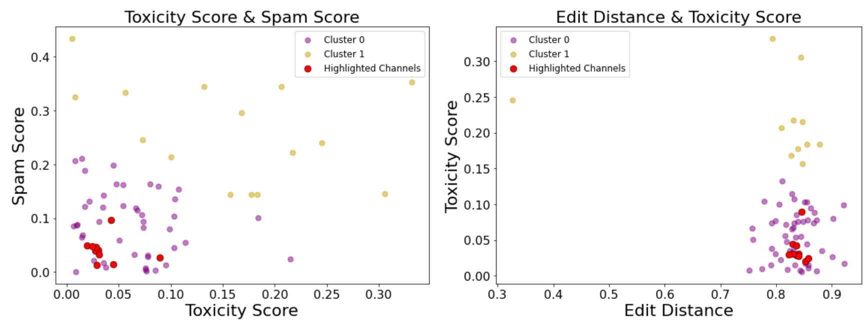

Fig. 2. Example of Highlighted Similar Channels in Characterization 2.

In addition to the quantitative analysis, we conducted a qualitative examination of the channels associated with each characterization. For channels grouped under Characterization 1, we observed that most were related to U.S. political news and other mainstream news outlets. Characterization 2 primarily consisted of channels from the Indo-Pacific region, many of which focused on religious content. Channels under Characterizations 3 and 4 often shared thematic similarities with those in Characterization 1; however, a notable distinction is that these channels frequently used thumbnails featuring prominent world leaders with overlaid text, and their content tended to be highly repetitive across videos. Finally, channels in Characterization 5 were mostly affiliated with news coverage specific to the Indo-Pacific region. These qualitative findings help contextualize the behavioral patterns observed in the clustering results, linking content themes and presentation styles to distinct audience interaction dynamics.

6 Conclusion and Future Work

This study introduces a commenter behavior-driven framework for characterizing YouTube channels, combining pairwise feature analysis, unsupervised clustering, majority voting, and Euclidean distance computation. By measuring distances between channels in various behavioral feature spaces, we quantified their similarity in terms of user interaction. Applying this approach to a dataset of 70 channels that include 711,301 videos, 13,221,243 commenters, and

129,653,609 comments, we identified consistent groupings that revealed shared patterns in comment-based engagement. The resulting five characterizations offer interpretable, high-level behavioral profiles that capture diverse patterns in structural and linguistic engagement. These range from regular and diverse interaction to high-quality, constructive discourse and linguistically rich but structured communication. Rather than focusing on predefined labels or content themes, this framework provides a scalable, unsupervised approach for understanding audience engagement patterns across channels. It offers practical utility for user analysis, engagement monitoring, and strategic communication assessment by identifying which groups of channels attract similar user behaviors.

Future work may incorporate additional contextual signals such as video metadata, channel category, or posting frequency to enrich the framework. While the current characterizations are derived from consistent unsupervised patterns, external validation such as sampling channels from each group for human review of spam, toxicity, and repetition, or testing whether group labels predict moderation outcomes, can help interpret the clusters in real-world terms and increase trust in their practical utility. Temporal analysis using rolling monthly windows can reveal how a channel shifts between behavioral groups (for example, from polite to toxic) during events like elections or viral trends, offering a richer view of audience dynamics. Adding channels from popular categories like music, gaming, or entertainment can test whether the same behavior types hold or if new patterns emerge, strengthening claims of generalizability. Extending the framework to other platforms could also offer cross-platform insights into audience engagement.

Acknowledgments

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078, W911NF-25-1-0147), U.S. Office of Naval Research (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, the Australian Department of Defense Strategic Policy Grants Program, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Donaghey Foundation 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. 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: CySoc. vol. 2024, p. 5th (2024)

2. Bhuiyan, M.M.I., Shajari, S., Agarwal, N.: Resilience and node impact assessment in youtube commenter networks leveraging focal structure analysis. In: The Eleventh International Conference on Human and Social Analytics (HUSO 2025) (2025)
3. Carvalho, P., Caled, D., Silva, C., Batista, F., Ribeiro, R.: The expression of hate speech against afro-descendant, roma, and lgbtq+ communities in youtube comments. Journal of Language Aggression and Conflict **12**(2), 171–206 (2024)
4. Dutta, H.S., Jobanputra, M., Negi, H., Chakraborty, T.: Detecting and analyzing collusive entities on youtube. ACM Transactions on Intelligent Systems and Technology (TIST) **12**(5), 1–28 (2021)
5. 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)
6. Mathew, B., Saha, P., Tharad, H., Rajgaria, S., Singhania, P., Maity, S.K., Goyal, P., Mukherjee, A.: Thou shalt not hate: Countering online hate speech. In: Proceedings of the international AAAI conference on web and social media. vol. 13, pp. 369–380 (2019)
7. Na, S.H., Cho, S., Shin, S.: Evolving bots: The new generation of comment bots and their underlying scam campaigns in youtube. In: Proceedings of the 2023 ACM on Internet Measurement Conference. pp. 297–312 (2023)
8. Shahi, G., Majchrzak, T.: Hate speech detection using cross-platform social media data in english and german language. In: Proceedings of the 20th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST. pp. 131–140 (2024)
9. Shajari, S., Agarwal, N.: Developing a network-centric approach for anomalous behavior detection on youtube. Social Network Analysis and Mining **15**(1), 3 (2025)
10. Shajari, S., Agarwal, N.: Safeguarding youtube discussions: a framework for detecting anomalous commenter and engagement behaviors. Social Network Analysis and Mining **15**(1), 1–24 (2025)
11. 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)
12. 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
13. Shajari, S., Amure, R., Agarwal, N.: Analyzing anomalous engagement and commenter behavior on youtube. AMCIS 2024 Proceedings. 6, https://aisel.aisnet.org/amcis2024/social_comp/social_comput/6 (2024), https://aisel.aisnet.org/amcis2024/social_comp/social_comput/6
14. Shajari, S., Amure, R., Agarwal, N.: Detecting and measuring anomalous behaviors on youtube. In: International Conference on Advances in Social Networks Analysis and Mining. pp. 3–16. Springer (2025)
15. Tuysuzoglu, G., Birant, D., Pala, A.: Majority voting based multi-task clustering of air quality monitoring network in turkey. Applied Sciences **9**(8), 1610 (2019)
16. YouTube API Services Terms of Service : Google developers. https://developers.google.com/youtube/terms/api-services-terms-of-service (nd), accessed: 2023-01-06