Uncovering Structural Consistency in 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. YouTube channels express their identity and engage audiences through titles, descriptions, transcripts, and category tags. While past research has focused on user engagement, the internal consistency of content across a channel's videos remains underexplored. This study presents a framework for characterizing channels based on semantic similarity among core content features. Using a dataset of 120 channels and 131,729 videos, pairwise similarity scores were computed across six feature combinations. Each pair was clustered using three unsupervised algorithms, with results merged through majority voting to form stable channel groupings. The analysis revealed recurring alignment patterns, resulting in three behavioral characterizations that reflect diverse strategies in metadata coherence, narrative flow, and thematic structure. The method is scalable, language-agnostic, and requires no labeled data. It provides a structured approach to understanding how content formatting shapes a channel's presentation, offering a foundation for future work on content strategy, visibility, and transparency in online video platforms.

Keywords: Social Network Analysis, YouTube, Large Language Models, Unsupervised clustering, Majority Voting

1 Introduction

YouTube has become a central platform for digital content creation, attracting contributors from education, news, politics, and entertainment. While extensive research has examined user engagement and recommendation algorithms, less is known about how content is internally structured within individual channels. Each video typically includes multiple textual elements such as the title, description, transcript, and category, which shape how viewers interpret, access, and trust the content. These components influence both engagement and discoverability. When metadata diverges from actual content, it can distort user understanding, especially in sensitive or underrepresented domains, leading to narrative misrepresentation or algorithmic bias. Although prior work has analyzed user behavior and anomalous commenting patterns, relatively few studies have addressed how consistently a channel's metadata aligns with its video content [12]. Such inconsistencies can alter audience expectations, affect perceived credibility, and influence how algorithms rank or recommend videos. Given the growing scale of user-generated content, there is a need for automated, annotation-free methods to detect organizational patterns in video archives. To fill this gap, we introduce a content-centric framework that analyzes YouTube channels based on semantic similarity among their key video components. Using a dataset of 120 channels and 131,729 videos, we computed semantic similarity scores using a language model for all six pairwise combinations among the four content fields. We generated similarity matrices and applied three unsupervised clustering algorithms. A majority voting scheme was used to derive stable channel clusters. Channels that repeatedly clustered together across pairwise feature combinations were then grouped into broader content behavior profiles, yielding three distinct channel characterizations based on how metadata aligns with actual video material.

This framework offers a novel perspective on content organization in YouTube channels. It supports scalable analysis without requiring labeled data or internal platform signals. The method is applicable across content domains and enables comparative studies on metadata design, editorial strategies, and channel positioning. By focusing on internal structure rather than external interaction, our approach complements research on user behavior and engagement, contributing a new dimension to platform based media studies [11].

2 Related Work

As YouTube solidifies its role as a leading platform for video sharing and community building, research has expanded to examine diverse content attributes, including textual, visual, behavioral, and network-based elements. Recent studies have enhanced our understanding of channel dynamics by analyzing both video content and user interactions, particularly through comment and transcript analysis. For instance, [6] and [3] used large language models to extract transcript features like emotional tone and narrative depth, linking them to engagement and psychological profiling. Complementary work has focused on detecting behavioral anomalies in commenting activity. Studies by [14, 13] used Graph2Vec, dimensionality reduction, and clustering to uncover coordinated commenting, while later work introduced composite anomaly scores combining behavioral and engagement signals, particularly in politicized contexts [15, 16]. These approaches also examined how formatting and thematic coherence influence channel identity. Other research explores how content and production traits affect user engagement. For example, [7] found that video length, subtitles, and topical focus shape interaction on athlete-run channels. The study by [5] exposed reliability gaps in medical and educational videos using DISCERN and GQS criteria. Network-based studies have further illuminated platform-wide patterns. For instance, [9] used time-sliced clustering to reveal shifts in content genres over time, while [10] highlighted the rise of scripted, commercialized content for children between 2016 and 2018 driven by recommendation algorithms. Similarly, [2] highlighted algorithmic amplification of emotionally charged or polarized content in political domains. Together, these findings offer rich insight into YouTube's content ecosystem. However, most prior work focuses on isolated signals, such as behavior, engagement, or quality, without assessing the internal consistency across a channel's own content components. To address this gap, we propose a semantic similarity based framework that characterizes channels through the alignment of titles, descriptions, transcripts, and categories.

3 Data

To collect content features, we followed the approach in [4], using the YouTube Data API to retrieve video-level metadata daily. This included video IDs, titles, descriptions, and category tags, in accordance with YouTube's Terms of Service [17]. We also used the Apify YouTube Transcripts Scraper API [1] to collect publicly available transcripts, capturing both metadata and textual content. Our initial crawl collected over 131,729 videos from 120 carefully selected YouTube channels spanning topics such as U.S. military activity, international politics, regional news, and public-interest programming. This variety allowed us to capture differences in language use, topic complexity, and audience focus, from specialized discussions to general-interest content. Due to processing constraints, we applied stratified sampling to preserve the original popularity distribution across videos while reducing data volume.

4 Methodology

This section presents a step-by-step framework for examining how YouTube channels organize their content, starting with measuring similarity between different video features and ending with finding groups of channels that share similar content patterns.

4.1 Content Feature Processing

This section describes the methodology used to measure semantic alignment across a video's key textual elements. We focus on four fundamental components: title, description, transcript, and assigned category. These components differ in length, linguistic style, and level of detail, making them useful for evaluating how consistently a channel conveys its thematic intent. All four elements were preprocessed using a standardized pipeline, including lowercasing, punctuation and special character removal, whitespace cleanup, and lemmatization when applicable. To ensure consistency in comparisons, non-English content was translated before analysis. Both translation and similarity scoring were performed using OpenAI's GPT-4o-mini model via API [8]. This model enabled context-aware semantic comparisons between content fields of varying lengths, outperforming traditional embedding or keyword-based methods. It was particularly effective in aligning short metadata (e.g., a title or category) with longer text such as a transcript. For instance, it could assess alignment between a full transcript and a brief label like "News & Politics," a task that often challenges conventional models due to input imbalance or loss of context. As a result of this process, six base pairwise similarity metrics were computed for each video based on the following feature combinations:

- Title–Description
- Title–Transcript
- Title–Category
- Description–Transcript
- Description–Category
- Transcript–Category

Each of these pairwise similarity values was normalized to fall within the range [0,1], where higher scores reflect stronger semantic alignment. These normalized similarity scores were then used to build a content profile for each channel. For every video v in a channel c and feature combination k, we denote:

$$S_{c,v}^{(k)} \in [0,1]$$

To represent the overall behavior of a channel, these values were averaged across all N_c videos in channel c:

$$\bar{S}_c^{(k)} \;=\; \frac{1}{N_c} \sum_{v=1}^{N_c} S_{c,v}^{(k)} \tag{1}$$

This process produces a single averaged score, denoted by $\bar{S}_c^{(k)}$, for each foundational feature k within a given channel c. These aggregated values enable consistent and interpretable comparison across channels. To emphasize the most meaningful differences in content alignment, we narrowed our focus to the most distinctive and analytically valuable pairwise feature comparisons for the subsequent analysis. The selected combinations include:

1. **Title–Category vs. Title–Transcript:** Helps determine whether a title better represents the spoken content or the assigned topic label.
2. **Description–Category vs. Description–Transcript:** Evaluates whether the description provides an accurate summary of the video content or primarily mirrors the assigned category label.
3. **Transcript–Category vs. Title–Description:** Compares how well the transcript aligns with its category label, relative to the cohesion between title and description.
4. **Title–Category vs. Description–Category:** Examines the consistency of different metadata elements with the given category tag.
5. **Title–Transcript vs. Description–Transcript:** Evaluates whether the title and description provide a unified narrative about the video content.
6. **Transcript–Category vs. Description–Transcript:** Measures whether the transcript aligns more closely with its category or with its accompanying description.

These six selected pairwise feature combinations were used in the clustering phase, as they provided the most meaningful contrasts across features that differ in purpose, length, and specificity. Focusing on these combinations enhanced the interpretability of the resulting channel characterizations, avoiding unnecessary complexity while retaining analytical depth.

4.2 Consensus Clustering Using Multiple Unsupervised Models

To uncover structural patterns in channel content, we performed clustering separately on each of the six pairwise feature combinations using K-Means, hierarchical agglomerative clustering, and spectral clustering. Each method was applied independently, producing cluster assignments for each pairwise similarity space. Let $\mathcal{X}^{(m)}$ represent the dataset of channels embedded in the m-th pairwise feature space, where $m \in \{1, 2, \dots, 6\}$. For each algorithm $k \in \{1, 2, 3\}$, a clustering label $C_i^{(m,k)}$ is assigned to channel i. To consolidate the results, we adopt a majority voting strategy across the three clustering outputs to obtain a final label $C_i^{(m)}$ for each channel:

$$C_i^{(m)} = \text{Mode}\left(\{C_i^{(m,1)}, C_i^{(m,2)}, C_i^{(m,3)}\}\right) \tag{2}$$

By applying this procedure across all six pairwise feature combinations, we obtained a comprehensive set of cluster assignments corresponding to each channel:

$$\mathcal{C}_i = \{C_i^{(1)}, C_i^{(2)}, \dots, C_i^{(6)}\} \tag{3}$$

This approach captures a channel's alignment across semantic dimensions, with majority-based consolidation improving clustering stability and reducing algorithm-specific bias.

4.3 Content-Based Channel Characterization

To evaluate alignment within each cluster, channels were mapped to a two-dimensional similarity space for each pairwise feature combination. Their pairwise closeness was then quantified using Euclidean distance:

$$d_{ij}^{(m)} = \|\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)}\|_2 \tag{4}$$

where $\mathbf{x}_i^{(m)}$ and $\mathbf{x}_j^{(m)}$ are the coordinates of channels i and j in feature space m. To retain only the most meaningful similarities, we applied a quantile-based filter. For each feature space m, we determined the 10th percentile distance threshold $\theta^{(m)}$ and labeled channel pairs as similar if $d_{ij}^{(m)} \leq \theta^{(m)}$. This binary similarity was defined as:

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

We then aggregated these binary labels across all M feature combinations to compute a cumulative similarity score:

$$S_{ij}^{\text{total}} = \sum_{m=1}^{M} S_{ij}^{(m)} \tag{5}$$

Higher total scores indicate greater consistency in content alignment across multiple feature spaces. To form groups, we constructed a graph where channels were nodes and edges connected highly similar pairs. Channels linked directly or through intermediate neighbors were merged into connected components, resulting in groups of channels that exhibit consistent content similarity without relying solely on labeled supervision.

5 Results

This section presents the findings derived from the framework, including the clustering patterns observed across different pairwise feature combinations and the identification of channel groups that repeatedly co-occur as nearest neighbors across various pairwise feature combinations.

5.1 Channel Clustering via Pairwise Feature Combinations

This section details how YouTube channels were clustered based on their content similarity scores across multiple feature combinations. Rather than including all 18 plots produced by applying three unsupervised clustering methods to six different pairwise feature combinations, we present only the aggregated results obtained through majority voting. This approach emphasizes the most consistent and interpretable patterns while minimizing repetition. For each pairwise feature combination, clustering was performed separately using three distinct unsupervised algorithms. To assign a final cluster to each channel, we implemented a voting mechanism that selected the most frequently occurring label among the three clustering outcomes.

The optimal number of clusters for each pairwise feature combination was identified using the Silhouette score, which typically indicated two clusters, though some cases initially suggested three. In such cases, clusters displaying similar behavioral traits were combined to enforce consistency. This merging ensured that all feature combinations ultimately produced exactly two clusters, creating a uniform structure that allowed the majority vote to operate on a stable and comparable basis across all pairwise similarity score spaces. Figure 1 presents the final clustering results obtained through the majority voting process. Each plot displays channels mapped onto a two-dimensional space using their normalized content similarity scores, with colors indicating the final cluster assignments. These visual representations reveal distinct groupings of channels that exhibit comparable content similarity patterns, making the structural separation across various pairwise feature combinations visually apparent.

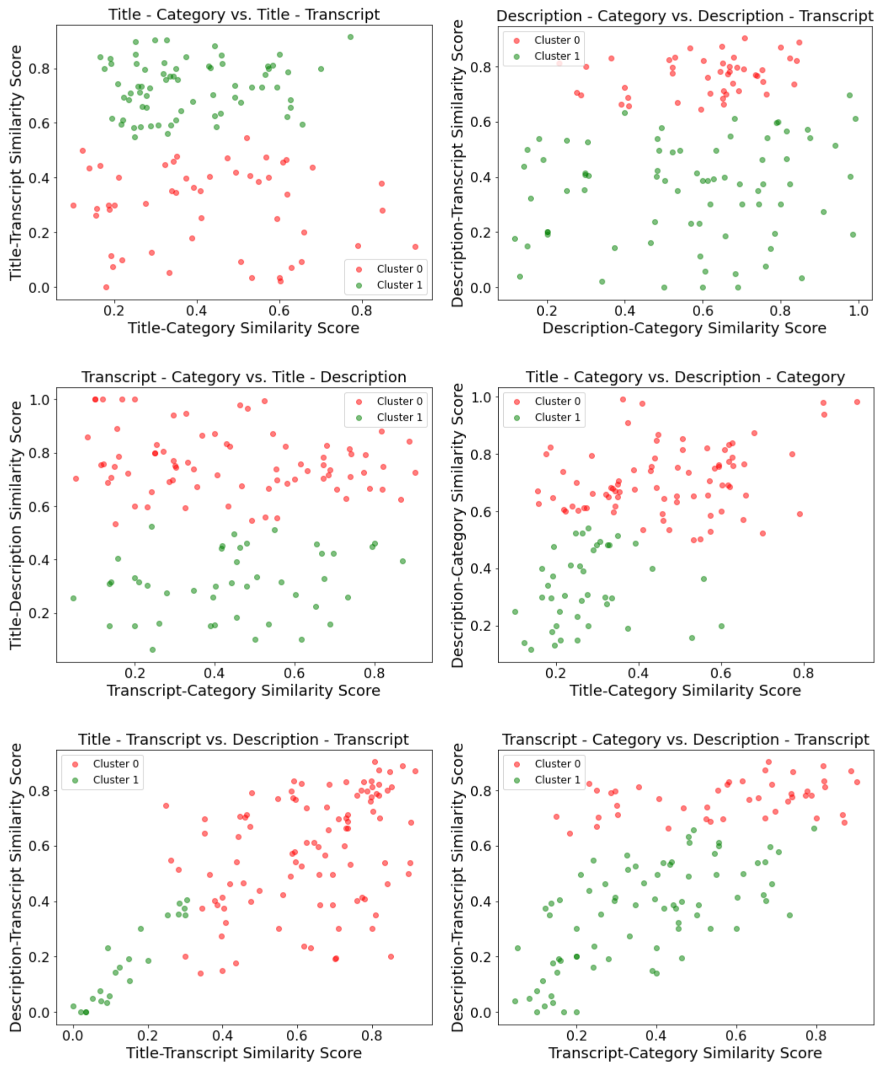

Fig. 1. Channel clusters based on majority voting.

5.2 Characterizing Semantically Aligned Channel Groups

After establishing stable clusters via majority voting, we analyzed recurring proximity patterns across feature combinations to identify broader content behavior profiles. Channels were assigned to these profiles by selecting those consistently within the top 10% of closest Euclidean distances in the two-dimensional similarity plots, as defined in Equation 4. Channels frequently appearing as nearest neighbors across multiple feature spaces were grouped under shared characterizations, reflecting similar content strategies. Applied to Applied to

120 YouTube channels, this approach yielded stable groupings summarized in Table 1, where #PFC indicates the number of pairwise feature combinations in which channels consistently co-occurred. The numbers listed below refer to the assigned identifiers for each pairwise combination, as defined in Section 4.1.

Table 1. Summary of Channel Characterizations Based on Feature Similarity.

#Characterization	# Channels	# PFC	Average Similarity Score
1	22	2, 5	Description–Category = 0.74 Description–Transcript = 0.78 Title–Transcript = 0.83
2	16	2, 3, 5, 6	Description–Category = 0.75 Description–Transcript = 0.80 Transcript–Category = 0.71 Title–Description = 0.73 Title–Transcript = 0.77
3	12	1, 2, 3, 4, 5, 6	Description–Category = 0.55 Description–Transcript = 0.42 Title–Category = 0.32 Title–Transcript = 0.53 Transcript–Category = 0.58 Title–Description = 0.68

Each of the three characterizations captures a distinct alignment pattern based on the most representative feature combinations, as described below:

Characterization 1: This group includes 22 channels showing consistent behavior across two pairwise feature combinations. They display strong alignment between descriptions, transcripts, and categories, with titles also reasonably matching spoken content. This indicates deliberate content structuring. Such patterns are typical of educational or informational channels, particularly those focused on public-interest topics like U.S. military or policy, where metadata accuracy and content clarity support transparency and credibility.

Characterization 2: This group includes 16 channels with consistent alignment across four pairwise feature combinations. They show strong semantic consistency across titles, descriptions, transcripts, and categories. While minor variations exist, the overall coherence reflects a deliberate thematic or brand-driven strategy. Such behavior is common among channels focused on international politics, geopolitical commentary, or advocacy, where message consistency is key to audience trust and narrative control.

Characterization 3: This group consists of 12 channels showing consistent behavior across all six pairwise feature combinations, but with low similarity between key elements such as titles and categories, and descriptions and transcripts. The metadata often misaligns with spoken content, leading to a fragmented semantic structure. This pattern is typical of regional news or multi-topic channels, where content spans diverse subjects and labeling practices are less standardized.

6 Conclusion, Limitations, and Future Work

This paper introduced a new approach for characterizing YouTube channels by analyzing the semantic alignment of core content elements: titles, descriptions, transcripts, and categories. We computed similarity scores across six pairwise feature combinations and applied multiple unsupervised clustering methods. Using majority voting, we identified stable channel clusters with consistent internal content structures. Channels with similar alignment patterns were grouped into three high-level categories, each reflecting a distinct strategy for organizing and presenting content. These categories revealed varying consistency between surface-level metadata and underlying video content, pointing to diverse editorial approaches. Our findings show that content behavior can be meaningfully distinguished through structural alignment alone, without relying on engagement metrics or user interaction data. This contributes a new dimension to YouTube research and complements prior work on commenting behavior and algorithmic exposure [12]. By combining large language models with unsupervised learning, our method enables scalable, language-independent analysis and offers insight into how internal inconsistencies may affect visibility and representation.

The current framework has limitations. It considers only four textual components and excludes visual or contextual elements such as thumbnails or platform-specific dynamics. The pairwise comparison method assumes equal importance across features, which may not apply universally. The dataset's focus on English or translated content may also limit cultural and linguistic coverage.

Future work will expand the dataset to include non-English channels, apply weighted feature importance, and incorporate multimodal signals such as audio and thumbnails. We also aim to integrate user interaction data, including comment analysis, to build a more comprehensive understanding of how content structure influences engagement and audience perception.

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. Apify Technologies: Youtube transcripts scraper (2024), https://apify.com/mtrunkat/youtube-transcripts-scraper, accessed via Apify API on 2024-06-01

2. Cakmak, M.C., Agarwal, N., Oni, R.: The bias beneath: analyzing drift in youtube's algorithmic recommendations. Social Network Analysis and Mining **14**(1), 171 (2024)
3. Choi, G.E., Pyun, M., Yoon, S.H., Kim, Y., Shin, H., Lee, S.Y.: Exploring the relationship between youtube video characteristics and a viewer's mental health traits among young adults. Frontiers in psychiatry **15**, 1364930 (2024)
4. 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)
5. Kuka, G.I., Gursoy, H.: Content quality and reliablity of youtube™ videos as a source of information about good oral hygiene practices in adults. PeerJ **12**, e18183 (2024)
6. Mishra, A.N., Sengupta, P., Biswas, B., Kumar, A., Coussement, K.: Leveraging machine learning and generative ai for content engagement: An exploration of drivers for the success of youtube videos. Journal of Business Research **193**, 115330 (2025)
7. Moon, B., Oh, T.: A study on the characteristics of sports athletes' youtube channels and user reactions. Behavioral Sciences **14**(8), 700 (2024)
8. OpenAI: Gpt-4o model card (2024), https://platform.openai.com/docs/guides/gpt, accessed via API at https://platform.openai.com/docs/guides/gpt
9. Paolillo, J., Ghule, S., Harper, B.: A network view of social media platform history: Social structure, dynamics and content on youtube (2019)
10. Paolillo, J., Harper, B., Boothby, C., Axelrod, D.: Youtube children's videos: development of a genre under algorithm. In: Proceedings of the Annual Hawaii International Conference on System Sciences (2020)
11. Shajari, S., Agarwal, N.: Developing a network-centric approach for anomalous behavior detection on youtube. Social Network Analysis and Mining **15**(1), 3 (2025)
12. 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)
13. 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)
14. 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
15. Shajari, S., Amure, R., Agarwal, N.: Analyzing anomalous engagement and commenter behavior on youtube. AMCIS 2024 Proceedings (2024), available at https://aisel.aisnet.org/amcis2024/social_comp/social_comput/6
16. Shajari, S., Amure, R., Agarwal, N.: Navigating the anomalies: A comprehensive analysis of youtube channel behavior. In Proceedings of the 16th International Conference on Social Networks Analysis and Mining (ASONAM 2024), Calabria, Italy (2024)
17. YouTube API Services Terms of Service : Google developers. https://developers.google.com/youtube/terms/api-services-terms-of-service (nd), accessed: 2023-01-06