Analysis of Communities formed on YouTube Comment Sections

Thiago Amado Costa[1], **Humberto Torres Marques Neto**[1]

[1]ICEI – Pontifícia Universidade Católica de Minas Gerais (PUC-MG)
Belo Horizonte, Minas Gerais - Brasil

thiago.amado@sga.pucminas.br, humberto@sga.pucminas.br

Abstract. *YouTube is the largest video streaming platform on the web, attracting billions of users daily who watch and engage with content through comments. A significant portion of these users consists of children and teenagers, who frequently interact with one another in the comment sections, forming active communities. However, these communities can also experience negative interactions between users, including instances of online bullying, for example. To explore these communities, the YouTube Data API v3 was used to collect comments from videos produced by creators targeting this demographic. Using these comments, co-commenter networks were constructed. Topic modeling and sentiment analysis were then applied to gain deeper insights into the content and dynamics within these communities.*

1. Introduction

YouTube has grown exponentially since its launch in 2005. With billions of users, it is now the largest video streaming platform on the web, and a major source of online entertainment content. As highlighted by [Alqahtani et al. 2023], many children and teenagers use the platform as an alternative to traditional entertainment sources, such as television, with the majority of their browsing time spent on YouTube.

One of the ways kids and teenagers interact on the platform is through the comment section. This feature allows users to share their thoughts on the video, the content creator, or even other commenters. These interactions can form online communities, and the analysis of these communities can provide valuable insights into the dynamics of user interaction, providing a deeper understanding of the platform's social ecosystem.

The study by [Alqahtani et al. 2023] demonstrates that many of the current developments focus on identifying inappropriate content targeting young children, with advanced machine learning and deep learning techniques being used to improve the detection of these harmful videos, with the goal of enhancing children's safety on the platform. However, there is limited research that focus on the analysis of the social network communities formed on the comment section of YouTube creators who create content aimed at children and teenagers.

This study aims to analyze the communities formed within the comment sections of YouTube creators who create content aimed at children and teenagers, as well as the content of the comments themselves. To achieve this, 10 Brazilian YouTube creators were selected based on their subscriber count and total views, with the communities in the comment sections of each creator's 50 most recent videos analyzed individually.

2. Related Work

A social network community can be defined as a set of members of that specific network that interact with each other, similar to real-world communities but in an online environment. These online communities are shaped based on how users interact on each platform, and the analysis of such communities can uncover hidden relationships between users and how they engage([Nooribakhsh et al. 2024]).

A recent research by [Kirdemir and Adeliyi 2023] uses community analysis to investigate coordinated inauthentic campaigns on YouTube, focusing on characterizing suspicious behaviors through a multi-step analysis of engagement trends and co-commenter networks. These co-commenter networks are built by connecting two commenters if they commented on the same video. By analysing these co-commenter networks and channel engagement trends, they were able to identify patterns indicative of manipulation, such as increasing views paired with decreasing subscribers, and highlight channels exhibiting coordinated commenting behaviors. [Shajari et al. 2023] follows this methodology to explore the problematic behaviors of commenters on YouTube, specifically focusing on ”commenter mobs” that manipulate engagement metrics to distort public perception. By analyzing 20 targeted channels through social network analysis, the research fills gaps in understanding how suspicious commenter activities boost engagement, which has been insufficiently addressed in prior literature. Employing a co-commenter network model and clustering techniques, the study identifies distinct groups with varying levels of suspiciousness and highlights collusion among commenters across channels. Key findings reveal central figures driving discussions and coordinated efforts to amplify specific narratives, suggesting these channels significantly contribute to the spread of misinformation.

With a similar methodology, [Hussain et al. 2018] investigates disinformation tactics used on a specific conspiracy theory channel, distinguishing its approach by focusing on user engagement patterns rather than just spam detection. Data was collected using the YouTube Data API, monitoring metrics like views, likes, dislikes, and comments, revealing patterns indicative of potential manipulation. The analysis identified two commenter groups, peripheral and core, with the second group exhibiting higher instances of inorganic behaviors. A co-commenter network analysis revealed clusters around specific conspiracy topics and identified bot-like and spam actions.

[Shekhar et al. 2021] Investigates the prevalence and nature of abusive comments on YouTube, particularly focusing on the impact of hate speech on users, especially teenagers, by utilizing exploratory data analysis and topic modeling. The study employed Latent Dirichlet Allocation (LDA) for topic modeling and sentiment analysis using the TextBlob library to assess the emotional tone of comments. Findings revealed that certain YouTubers received significantly harsher comments, with varying sentiment levels indicating the need for interventions to protect young creators from severe online harassment. The study concludes by recommending measures such as disabling comments on particularly abused videos and emphasizes the importance of understanding the dynamics of cyberbullying.

In the context of children's content on YouTube ([Alqahtani et al. 2023]), machine learning and deep learning techniques are used to analyze video, audio, and user comments to detect inappropriate material. These methods often rely on natural language processing for text analysis and computer vision for visual content detection. Despite

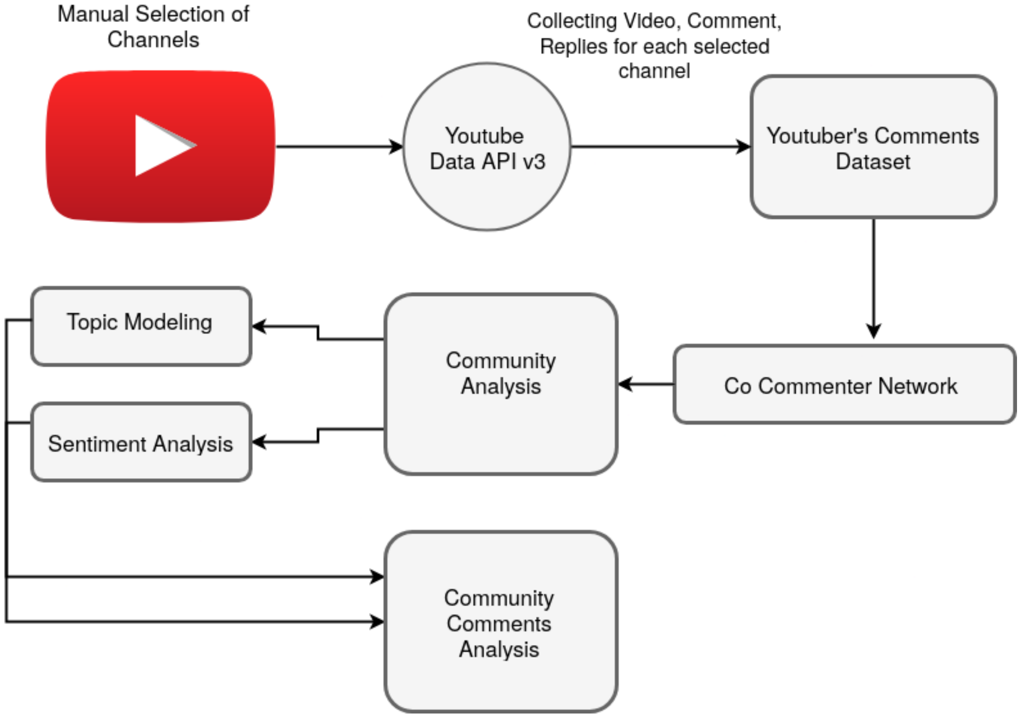

Figure 1. Methodology

recent advancements, the challenge of accurately detecting inappropriate content targeted at this demographic using machine learning and deep learning approaches still persists.

This study differs from the previously mentioned works, by exploring the co-commenter network methodology combined with the Topic Modeling and Sentiment Analysis, specifically within the context of YouTube creators producing content for children and teenagers.

3. Methodology

This study uses a similar methodology as previous works. It consists of a data collection phase, a community analysis phase, and a content analysis phase, as ilustrated in Figure 1.

3.1. Data Collection

Initially, the YouTube Data API v3 was used to gather comprehensive channel metadata for each specific YouTuber. Subsequently, we retrieve the IDs and titles of 50 videos from each channel. Finally, we construct the YouTuber's comments dataset by aggregating all comments and replies associated with each video, as well as the author information and the number of likes of each comment. For this, 10 different popular YouTube channels were chosen, with content varying from gaming and vlogging, with the number of subscribers ranging from 8 million to 46 million. The number of comments collected per channel ranges approximately 11,000 to 43,000.

3.2. Community Analysis

To analyze the communities that emerge within each YouTuber's comment section, the datasets constructed in the previous phase is used to construct the co-commenter networks. Each commenter is connected to another if they commented on the same video, and the weight of the edge is increased if they commented on more than one video together. To maintain only the stronger connections, we filtered out co-commenters who commented on less than 10 videos together, as done by [Shajari et al. 2023] and [Kirdemir and Adeliyi 2023].

To identify the communities of each co-commenter network the Louvain Algorithm was used. This agorithm was chosen because of its high efficiency and performance on real-world networks without ground truth, as shown by [You et al. 2020].

Following the method developed by [Kirdemir and Adeliyi 2023], a set of metrics were computed for each co-commenter network, including average degree, number of nodes and edges, average coefficient, modularity, coverage, the number of communities identified with the Louvain Algorithm. Graph clique metrics were also computed, including the number of maximal cliques that have at least five members, average clique size, median clique size, average degree of clique members, average clustering coefficient of clique members. To discover the similarities between the chosen channels, these metrics were used with KMeans and Hierarchical Clustering, along with Principal Component Analysis to reduce the complexity while maintaining important features.

3.3. Comments Analysis

To further understand user behavior, we implement topic modeling and sentiment analysis. For the topic modeling phase, we employ the BERTopic model ([Grootendorst 2022]), while the sentiment analysis is conducted using the multilingual XLM-roBERTa-base model([Barbieri et al. 2022]). Initially, the comments are processed and cleaned, masking user handles to preserve anonymity, removing emojis, links and stop words Then, topic modeling is utilized to extract themes associated with the communities identified in the previous phase. Finally, the sentiment analysis model evaluates the polarity of the content, offering deeper insights into the emotional tone of the discussions and facilitating a more comprehensive analysis of the communities.

4. Results

In this section, the results are shown and analyzed. Considering that a modularity value above 0.3 is a good indicator of significant community structures in a network [Clauset et al. 2004], The louvain algorithm successfully identified significant community structures in all the YouTube creator's co-commenter networks.

The KMeans and Hierarchical Clustering methods were applied with the community metrics described. Both methods identified 4 distinct clusters, as can be seen in Figure 2.

The most distinguishable channels are on clusters 1 (Rezendeevil) and 3 (Felipe Neto). Both creators had very large co-commenter networks, with Rezendeevil's network having 13.808 nodes and 149.033 edges, and Felipe Neto having 24.987 nodes and 125.165 edges. The creator Rezendeevil, who has the lowest modularity also had the highest average degree of 21.5, number of maximal cliques and average clique size. Similarly, Felipe Neto had a lower modularity score of 0.64, a high number of maximal cliques, and a high average degree of 10. This indicates that both creators have high interconnected and engaged communities. The higher average degree indicates that Rezendeevil's followers are more engaged than Felipe Neto's, while the higher modularity score shows that

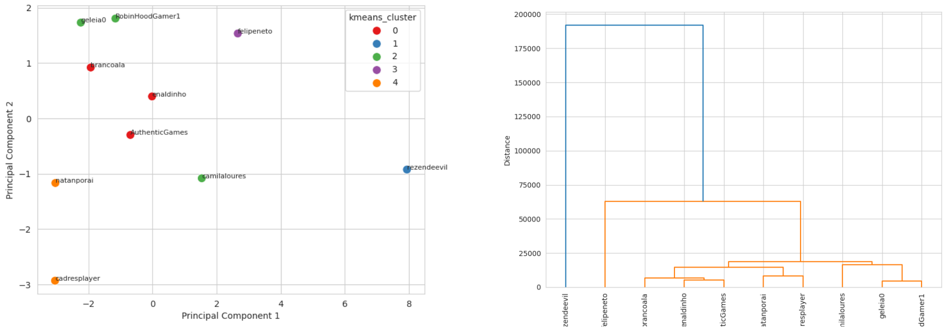

Figure 2. KMeans and Hierarchical Clustering

Felipe Neto's community structure is more distinct. This can be seen through Figure 3 and Figure 4, with each community colored.

A per community sentiment analysis showed that most comments on Rezendeevil's videos were labeled as positive by the XLM-roBERTa-base model, which is consistent with a highly engaged community of fans. Most positive comments focus on the video content or specific individuals featured in the video, along with compliments directed towards the creator. Through topic analysis, a recurring theme labeled "quero" was identified, with comments such as "Eu quero" and "Eu quero muito uma." (translated as "I want" or "I really want one"). These comments predominantly come from members of communities 5 and 6. A more detailed investigation of the videos receiving these comments may reveal potential signs of direct or implicit advertising aimed at children and teenagers.

Unlike the previous creator, Felipe Neto had more negative comments than positive comments. By analyzing the polarity per community, it was possible to identify that community 12, that has the smallest number of positive comments, is also the community with the highest clustering coefficient of approximately 0.23 , lowest conductance of approximately 0.06 and highest density of 0.21. This can be an indicative of a close-knit organized community that frequently comment on the same content spreading negative and hateful comments. Topic Modeling also shows recurring themes labeled "Brasil", "Hipocrisia" ("Hypocrisy"), "Dinheiro, Pobres" ("Money, Poor"), with comments such as "é tanta burrice por metro quadrado é um absurdo como tem gente que acredita nestas hipocrisias" ("it's such stupidity per square meter, it's absurd how many people believe in these hypocrisies."), "Esse Felipe merda é uma das desgracas que so existe aqui no Brasil, ..." ("This Felipe piece of trash is one of the disasters that only exists here in Brazil, ..."). The other communities had lower densities, lower average clustering coefficients and higher values of conductance, with a wider variety of topics and higher number of positive comments.

References

Alqahtani, S. I., Yafooz, W. M. S., Alsaeedi, A., Syed, L., and Alluhaibi, R. (2023). Children's safety on youtube: A systematic review. *Applied Sciences*, 13(6).

Barbieri, F., Espinosa Anke, L., and Camacho-Collados, J. (2022). XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In *Proceedings*

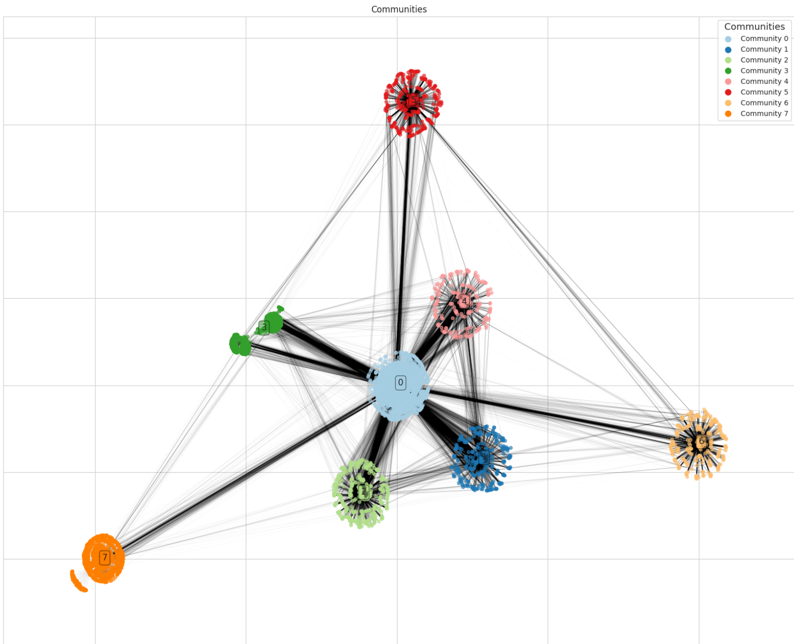

Figure 3. Rezendeevil's Communities

of the Thirteenth Language Resources and Evaluation Conference, pages 258–266, Marseille, France. European Language Resources Association.

Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 70:066111.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Hussain, M. N., Tokdemir, S., Agarwal, N., and Al-Khateeb, S. (2018). Analyzing disinformation and crowd manipulation tactics on youtube. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1092–1095. IEEE.

Kirdemir, B. and Adeliyi, O. (2023). Towards characterizing coordinated inauthentic behaviors on youtube. In *The 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval (ROMCIR 2022) held with the 44th European Conference on Information Retrieval (ECIR 2022)*.

Nooribakhsh, M., Fernández-Diego, M., Gonzá lez Ladrón-De-Guevara, F., and Mollamotalebi, M. (2024). Community detection in social networks using machine learning: a systematic mapping study. *Knowledge and Information Systems*, pages 1–55.

Shajari, S., Agarwal, N., and Alassad, M. (2023). Commenter behavior characterization on youtube channels.

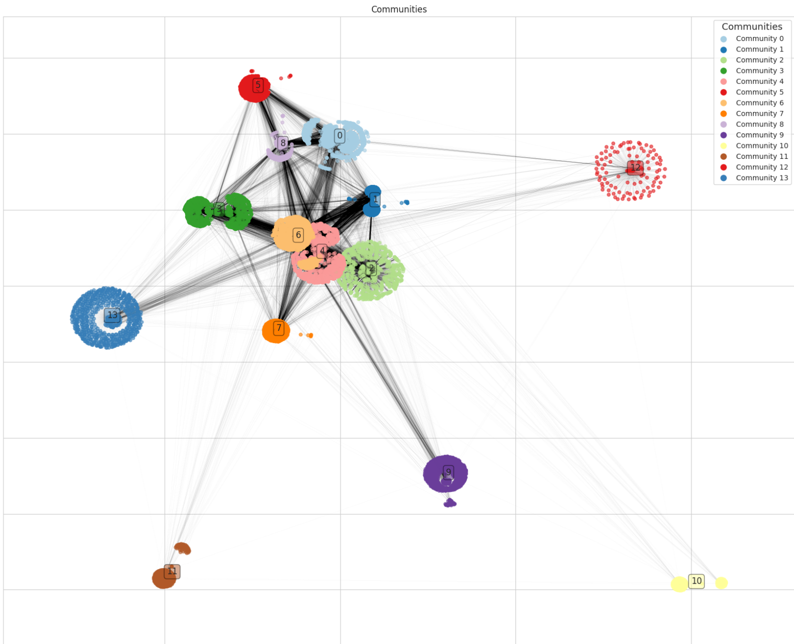

Figure 4. Felipe Neto's Communities

Shekhar, S., Akanksha, and Saini, A. (2021). Utilizing topic modelling to identify abusive comments on youtube. In *2021 International Conference on Intelligent Technologies (CONIT)*, pages 1–5.

You, X., Ma, Y., and Liu, Z. (2020). A three-stage algorithm on community detection in social networks. *Knowledge-Based Systems*, 187:104822.