Detecting and Characterizing Inorganic User Engagement on YouTube

Oluwaseyi Adeliyi, Ishmam Solaiman, Shadi Shajari, Ugochukwu Onyepunuka, Nitin Agarwal

COSMOS Research Center, University of Arkansas - Little Rock, USA
{odadeliyi,iasolaiman,sshajari,uponyepunuka,nxagarwal}@ualr.edu

Abstract

Social media platforms' growth has been nothing short of astounding, reshaping the way we communicate, share, and connect. The accessibility and user-friendly nature of platforms like YouTube, Facebook, Twitter, and others have enabled millions of individuals to express themselves, engage with content, and establish communities. However, alongside positive aspects, the rise of social media has also brought misinformation, which can spread like wildfire across these platforms. This research focuses on identifying and characterizing YouTube accounts that use automated or semi-automated means to boost their user engagement statistics and to amplify their misinformation. We assess statistics from 3542 channels, combining techniques including rolling window correlation analysis, anomaly detection, rule-based classification, and clustering. We found varying levels of inorganic activity across the channels, with some of those exhibiting high levels being suspended by YouTube. We also discovered that some channels that boost their content inorganically, publish videos of a similar discourse.

Introduction

YouTube is one of the most popular and most visited websites in the world today, making the platform a prime target for misinformation and online coordinated campaigns (Del Vicario et al. 2016; Dutta et al. 2021). As a result, researchers have studied these campaigns on the platform using a variety of computational techniques to identify the tactics and characterize the dynamics of online coordinated campaigns (Carley 2020; Galeano et al. 2020). Previous studies have revealed that these campaigns are driven by YouTube channels that utilize inorganic means to boost their user engagement metrics. We define inorganic as any use of automated or semi-automated methods (Ferrara et al. 2016) such as bots or paid services to boost engagement, rather than organic user-generated appraisals (Giglietto et al. 2020).

Misinformation thrives in the fertile grounds of social media due to a combination of factors, and automated accounts, or bots, play a significant role in this phenomenon. These bots are software programs designed to mimic human behavior, often posting, sharing, and engaging with content at an unprecedented speed and scale. The manipulation of these metrics not only distorts the perceived popularity and credibility of content but also exploits the platform's algorithms, increasing the likelihood of misinformation reaching a wider audience. This artificial inflation creates an illusion of legitimacy and community support, which can significantly impact user perception and the dissemination of information. By conflating inorganic user engagement with misinformation propagation, this research aims to uncover the symbiotic relationship between these elements, demonstrating how automated and semi-automated operations contribute to the spread of misinformation, thereby undermining the integrity of online discourse.

This support could be in the form of comments, views, and subscribers obtained inorganically, and these actions are prohibited on YouTube according to their terms of service (YouTube 2023). While similar research has been conducted on the detection of automated account activity on platforms like Twitter (Al-khateeb, Anderson, and Agarwal 2021; Khaund et al. 2021), it is vital that these behaviors are also detected on YouTube in a timely manner to curb the spread of online misinformation campaigns on the platform. The algorithm-driven nature of YouTube's recommendation system plays a pivotal role in determining which videos are shown to users. When a video has artificially inflated views and comments, it can lead to the algorithm recommending that video to more users on the platform.

This phenomenon can amplify the reach of misinformation and contribute to the creation of echo chambers, where users are exposed to content that reinforces their existing beliefs. This means that even content with dubious credibility can gain undue prominence, potentially swaying public opinion and perpetuating falsehoods. Given the critical role of YouTube as a platform for information dissemination, our research endeavors to go beyond mere analysis and aims to contribute to the development of practical tools for early detection and intervention of inorganic user engagement on the platform. By delving into the time series analysis of user engagement metrics, we seek to uncover patterns indicative of inorganic activity, especially those orchestrated by automated or semi-automated means.

In this study, we detect and characterize inorganic behaviors and patterns within YouTube channels and combine a series of techniques, including the time series analysis

of user engagement statistics and supervised/unsupervised machine learning methods (Kirdemir, Adeliyi, and Agarwal 2022). Using data from YouTube (Google Developers 2021) and Social Blade (Social Blade 2023), We analyze a dataset comprising 3542 Indo-Pacific channels and identify inorganic behaviors exhibited by the channels in the dataset. The reason we are focusing on these channels is that they span across various countries and regions within this area, addressing topics ranging from politics to religion and conflicts. These channels provide invaluable insights and diverse perspectives, particularly regarding the Indo-Pacific, enriching our understanding of religious matters, conflicts, and political dynamics between countries such as Indonesia and China, as well as other countries in this area.

Literature Review

Many research studies have extensively examined the behaviors exhibited on YouTube across different dimensions. Recent research has put forth detection models and introduced CollATe, a neural architecture designed to identify videos that solicit collusive comments (Dutta et al. 2021). Researchers have also conducted a study that pinpointed networks involved in sharing political news, each driven by distinct motivations. Giglietto et al. (2020) underscores the influence of inauthentic entities on the diversity of news sources, thereby illuminating manipulation strategies. Similarly, authors have also conducted an in-depth investigation into the impact of automated accounts on Online Social Networks (OSNs), analyzing the behavioral patterns of such accounts, their coordination strategies, and the associated challenges in detection (Khaund et al. 2021).

These studies rely on network analysis and the interaction between entities such as commenters. Another example is an approach rooted in social network analysis to detect "commenter mobs" on YouTube, where groups of commenters engage in activities aimed at artificially inflating video engagement using methods such as Principal Component Analysis, Graph2vec, and UMAP (Shajari, Agarwal, and Alassad 2023; Shajari, Alassad, and Agarwal 2023a,b). In terms of user engagement metrics, Hussain et al. (2018) suggested a strategy involving the examination of video metadata, comment analysis, and user engagement to identify and counteract malicious activities effectively. In this study, we utilize user engagement metrics to uncover inorganic activity within YouTube channels. Yousefi et al. (Yousefi, Shaik, and Agarwal 2024) presented a model that combines audio, visual, and textual aspects to characterize YouTube videos. Additionally, their earlier work (Yousefi, Cakmak, and Agarwal 2024) delves into emotional aspects of video content, investigating how text, visual colors, and audio features interact to evoke emotions in YouTube users.

A recent study investigated fake engagement services' underground markets and social media management panels, revealing a diverse range of customizable services to boost engagement for content (Nevado-Catalán et al. 2023). Some researchers have also introduced graph-based techniques to detect deceptive practices on YouTube (Beutel et al. 2013). While it might be difficult to track channels that purchase appraisals from black markets, the user engagement statistics could give researchers insights into the growth patterns of channels. For example, some studies have focused on comments data, using hash-based techniques to recognize pairs of near-duplicate comments on YouTube (Kin Wai, Horawalavithana, and Iamnitchi 2021).

In this research, we analyze data spanning the comments, views, subscribers, and videos of channels to detect inorganic activity. We highlight the specific period where the activity is detected, and we also go further to explain the behavior within the channels that exhibit these behaviors, including the nature of content that they publish. This approach is vital for detecting not only channels that have exhibited inorganic activity, but those that are likely to exhibit such behaviors in the future. Our research offers policymakers and platform owners a means to proactively detect and thwart inorganic, coordinated behavior, which is often used to promote misinformation on social media platforms such as YouTube (Kirdemir, Adeliyi, and Agarwal 2022).

Methodology

The data collection process and the methodology utilized to detect inorganic activity within YouTube channels are detailed in this section. Our approach for detecting and characterizing inorganic user engagement is predicated on the understanding that such activities are not merely violations of platform policies but are instrumental in the propagation of misinformation. By artificially enhancing engagement metrics, inorganic activities skew the informational landscape, enabling content with dubious credibility to proliferate. This approach allows us to not only identify channels engaged in these practices but also to examine the content they promote, offering insights into how misinformation leverages inorganic support to gain visibility and influence. Through this lens, our analysis extends beyond the technical detection of inorganic behaviors, probing the interconnectedness of these tactics with the broader dynamics of misinformation spread on YouTube. This methodology entailed conducting a time series analysis of each channel's daily engagement metrics. Figure 1 provides a high-level summary of our method described in detail below.

Data Collection

The data used in this analysis were collected by COSMOS Research Lab with their YouTube data collection tool that utilizes YouTube's API (Google Developers 2021). Videos, comments, and channel data were collected from YouTube using the keywords provided by the subject matter experts from Arizona State University. The relevant keywords were identified by studying coverage in the Indo-Pacific region with further reviews to improve the inclusiveness of the keywords. The narratives have different date ranges for their data. Table 1 below shows a snippet of the keywords used for each narrative.

This resulted in a list of 5000 channel IDs with videos relating to the narratives. From the list of 5000 Indo-Pacific channel IDs collected, we were able to get daily engagement statistics (subscriber counts, number of views, and number

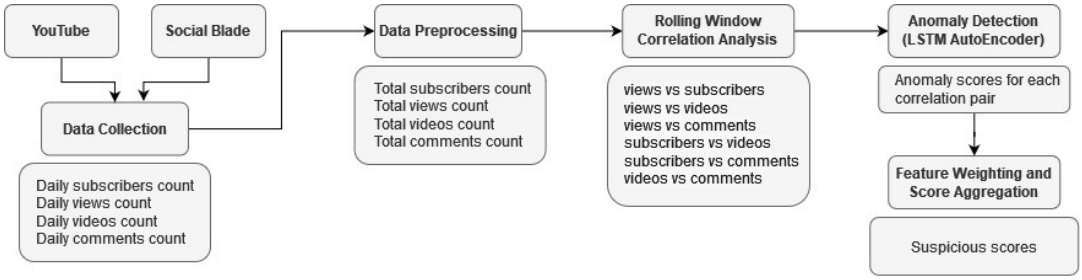

Figure 1: High-level analysis workflow diagram.

S/N	Narrative	Date	Keywords
1	China	2018–2022	'Chinese Communism' | 'China influences Indonesia' | 'Dominating China' | 'China' economy in Indonesia' | 'Muhammadiyah China' | 'China' | 'Tiongkok' | 'Tionghoa', | 'Nahdlatul Ulama China' | 'China' | 'South China sea' | 'Tionghoa PKI China'
2	Uyghur	2018–2022	'Uyghur Oppression' | 'Free Uyghur'| 'Uyghur Cruelty' | 'Muslim brother' | 'Indonesia Uyghur' | 'Uyghur'| 'Muslim brother' | 'Uyghur'

Table 1: Narrative Extraction Sample.

of videos) on 3542 channels using Social Blade's API. For our analysis, we considered only data between September 2018 and February 2022. Finally, we preprocessed the data by eliminating missing values in the engagement metrics and computing the total number of views, subscribers, and comments. The engagement metrics that were utilized in discovering channels that exhibit inorganic activities included the total views, total subscribers, total comments, and total video counts for each channel.

Rolling Window Correlation

In this step, we looked at the correlation between engagement metrics as an indication of inorganic activity, e.g., a channel with decreasing subscribers or videos but increasing views and comments and vice versa. For this reason, we used rolling correlation, which allowed us to calculate and visualize changes in correlation between variables over time. To decide the optimal window for the rolling correlation analysis, we grouped the data into rolling windows of 5 days and 100 days, then computed the pairwise correlation between total videos, total views, total comments, and total subscribers for each window. The results from the rolling window of 100 days were more significant than the 5 day window, so we proceeded to use a rolling window of 100 days in our analysis. The output from this step comprised start and end dates and the values of correlation pairs (i.e., views and videos, views and subscribers, subscribers and videos, views and comments, subscribers and comments, videos and comments) for each window.

Anomaly Detection

In this step, we trained a long short-term memory (LSTM) model on the output of the rolling window correlation analysis. The model is composed of 4 LSTM layers of 128,64,64 and 128 hidden units with a dropout rate of 0.2 and mean squared error (MSE) as the loss function and uses the Adam optimizer. we represented the loss output from the computation as the anomaly confidence score (or anomaly score) to account for deviation between the model's prediction and actual data. We observed a trend of higher anomalous scores being directly correlated to higher levels of activation across the 6 indicators. Furthermore, the average of the anomaly score served as the threshold for capturing anomalous periods in a channel's correlation pair. This yielded a list of anomalous data points for each correlation pair as shown in Table 2 below. An anomalous chart was created for each channel's correlation pair by plotting their anomaly score on a two-dimensional line plot as shown in Figure 2. From the peaks we observe an unusually high amount of engagement across the 6 indicators representing potential periods of inorganic activities. The down trend further supports this as the anomaly score returns to its normal range after the 3 month period.

Rule-based Classification

A rule-based classification algorithm described by Kirdemir et al. (2022) was used to score the anomalous periods based on the results of the anomaly detection on each correlation pair (also referred to as indicators). These indicators are derived from stats such as videos, comments, subscribers and views by taking the ratio of pairs such as (views & subscribers), (views & comments), (views & videos), (videos & comments), (subscribers & comments), & (subscribers & videos). Each indicator's minimum correlation and maximum anomaly scores were used to determine their inorganic scores ranging from 1 (lowest) to 5 (highest).

Each of these indicators is examined separately to evaluate how it differs from anticipated or typical trends. Greater variances from expected patterns might suggest anomalous activity. For example, if there's a notable rise in comments while the viewership of videos drops significantly, it could prompt questions about the consistency of engagement rates.

Start Date	End Date	Duration	Average Views	Average Subscribers	Minimum Correlation	Maximum Anomaly Score
2019-10-23	2020-03-10	139 days	208,501,054	393,125	0.877	0.976
Start Date	End Date	Duration	Average Views	Average Subscribers	Minimum Correlation	Maximum Anomaly Score
2020-08-14	2020-11-29	107 days	274,492,331	6,107	-0.341	1.042

Table 2: Channel statistics showing views versus subscribers and views versus videos anomaly data.

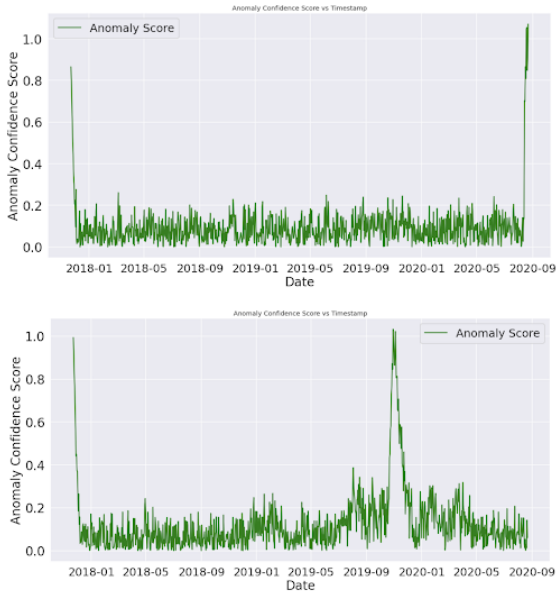

Figure 2: 2-D plots showing a channel's anomaly chart for its views versus subscribers indicator (top) and views versus videos indicator (bottom).

To aggregate the scores from each indicator and arrive at a single inorganic score, the semantics of each indicator were defined and used to assign weights to each indicator, as inspired by (Kirdemir, Adeliyi, and Agarwal 2022). These weighted scores were then aggregated to produce a single inorganic score for each observation with a range of (0,1).

Principal Component Analysis and Clustering

To determine where each data point lay on a 2-D plot, we utilized Principal Components Analysis (PCA). PCA is a dimensionality-reduction technique that minimizes information loss while reducing the number of variables in a data set. The applied PCA reduced the dimensions in the dataset from 16 features to 2 principal components, which we visualized on a scatter plot. To identify channels that exhibited similar engagement behaviors, we utilized density-based spatial clustering of applications with noise (DBSCAN) to group channels into clusters. DBSCAN is an unsupervised clustering method used with large data containing noise or outliers. An epsilon value of 0.1 was used, with 7 chosen as the minimum number of samples.

Suspended Channels

Our research revealed that some YouTube channels that exhibited inorganic activity in our research were later suspended by YouTube, as some of the patterns and behaviors captured violated the platform's terms of service. Using the YouTube API, we collected the active status for all 3542 channels in our dataset to verify which were active channels and which had been suspended. We found 99 channels were suspended having a score between 0.6 to 1 which can serve as a proxy for validation of our model's predictive capability.

Topic Modeling

After clustering our data, the video titles were preprocessed, and an LDA topic model was trained on them, with the optimal number of topics programmatically determined based on the coherence score. The goal of topic modeling in our methodology is to understand the type of information being disseminated by the channels, aiming to identify patterns of inorganic engagement across videos, thereby revealing potential strategies to manipulate discourse and spread misinformation. For example, if certain themes such as political unrest or health misinformation are prevalent among channels with high inorganic scores, it suggests a targeted approach to exploit YouTube's algorithms and user engagement for specific agendas. In our findings, the cluster-wise topic distribution and their respective keywords are discussed.

Analysis and Findings

In Figure 3, the scatter plot shows the spread of data across the first two principal components, with the colors indicating the intensity of the inorganic activity of each data point. The deeper colors show higher inorganic activity, and the lighter colors indicate lower inorganic activity. The scatter plot helps in understanding concentration of inorganic activity in clusters. It is observed that the darker points are in close proximity to each other suggesting focal points of inorganic activity. This plot helps us pick out channels with higher inorganic behavior scores which we can see in Table 3 highlighting channels with the highest inorganic scores in the dataset.

From this dataset of 3542 Indo-Pacific channels, the channel with the highest inorganic score was *Breaking News TV* with a score of 0.72. The anomalous period for this data point occurred between 24 March 2019 and 8 October 2019. The videos published by this channel seem to contain a mixture of clips from TV news channels and military demonstrations with videos focusing mainly on military incidents involving the USA, Russia, China, and North Korea. *Breaking News TV* has now been suspended by YouTube, alongside 98 other channels within our dataset.

We also noticed that the channel had made a resurgence since our first data collection, now identified as *Military*

Channel ID	Channel name	Inorganic score	Period of inorganic activity
UCN_qIhm7BAq9Qxa6j9XMVXA	Breaking News TV	0.72	24 March 2019–8 October 2019
UCrGZO3wJ20CWiy36Fdu6vdw	Badminton Talk	0.68	24 April 2019–8 November 2019
UCkQSMH1vP1KcxSPArUYLLQ	viral_makkodak	0.67	7 February 2020–4 September 2020
UCM2XWLz9zUcYZRJNl2dsjnw	DCTV Heroes	0.66	23 May 2019–18 December 2019
UCmvtaFkiWOSHhnOwt4Fz68g	SAFA News	0.64	26 November 2019 – 22 June 2020
UCBZ05ULEN_e-8keaj_AN4vA	Donate4Free	0.61	22 August 2020–11 March 2021
UCJTBOZUV_2LaMPQxaeVtgDQ	SMART MULTIMEDIA VIDEO	0.61	14 May 2020–12 December 2020

Table 3: Top 7 channels showing inorganic activity.

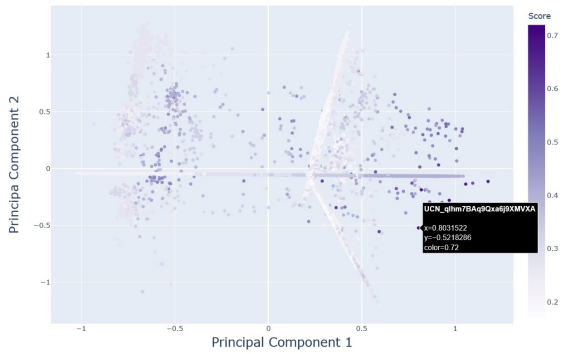

Figure 3: Scatter plot indicating channels and their inorganic scores.

News - BnTV. We identified this by the channel logo, as shown in Figure 4 and Figure 5. Before *Breaking News TV* was suspended, it had a total of 137,000 subscribers and 7,214,981 views as of January 2022, compared to 220 subscribers and 13,630 views for *Military News - BnTV* in August 2023. We also discovered that *Breaking News TV* was created in February 2017 while *Military News - BnTV* was created in October 2017 shortly after *Breaking News TV* was removed from YouTube, this suggests that YouTube had a reason for removing the channel, which can support our methodology. However, the new channel still features videos with similar narratives about the military, politics, and war.

The channel with the second highest inorganic activity in our dataset was *Badminton Talk*. This is a sports channel focused on badminton, specifically in Indonesia. From the channel page, we see videos with hundreds and thousands of views, as shown in Figure 6. However, during the period where inorganic activity was detected for this channel, a video with about a million views was published. On further analysis, we discovered that the video was a promotional video and was the default or featured video on the channel page, as shown in Figure 7. Only two videos in the channel's history have had up to a million views, and the featured video is one of them. The high number of views on the two specific videos compared to the rest of the channels and subscribers of the channels caused our model to pick up this channel's activity as inorganic, causing a high score in the indicators. While we cannot ascertain the reason for the unusually high number of views on that video, it could have been promoted as an advertisement or inorganically.

Cluster	NC	HIS	AIS	NSC
0	3,024	0.66	0.30	98
1	62	0.38	0.29	0
2	360	0.51	0.32	0
3	17	0.54	0.49	0
4	5	0.72	0.54	1

Table 4: DBSCAN cluster data depicting NC (Number of Channels), HIS (Highest Inorganic Score), AIS (Average Inorganic Score), and NSC (Number of Suspended Channels).

To understand the similarities in engagement trends within these channels, we extracted the periods with the maximum inorganic score within each channel. This was done because the initial output from our analysis had over 57,000 observations comprising multiple inorganic periods within each channel. Keeping only the maximum inorganic periods reduced this number to 3,539 observations, after which we used the DBSCAN clustering algorithm to group the data into clusters as shown in Figure 8 and Table 4. The data in the plot above is distributed across both principal components, depending on which indicators each data point shows some anomalies. Data points within each cluster show similar behavior in terms of their indicators where anomalies are detected.

Cluster 0

This cluster comprises the largest number of observations, with many channels that exhibit high inorganic activity. Channels within this cluster show anomalies across all six indicators with consistent inorganic activity within the views and subscribers statistics, indicating possible inorganic activity within the channels' views and subscribers. The total number of suspended channels here is 98, and most of the channels in our dataset appear in this cluster, with 25% of the channels having inorganic scores ranging from 0.4 to 0.66. Some of the channels within this cluster include *SMART MULTIMEDIA VIDEO* and some other military channels such as *Armageddon TV* (also suspended from YouTube), *The Military TV*, *Secret Military News*, and *Fortress Defense*.

Table 5 below shows the top five topics in this cluster, their keywords, and their distribution. The most dominant topic here, with about 60% of the distribution, comprises keywords such as *china*, *russia*, *taiwan*, *south*, *sea*, and *war*, indicating videos with discourse around the South China Sea as well as tensions between China and Taiwan. In prior stud-

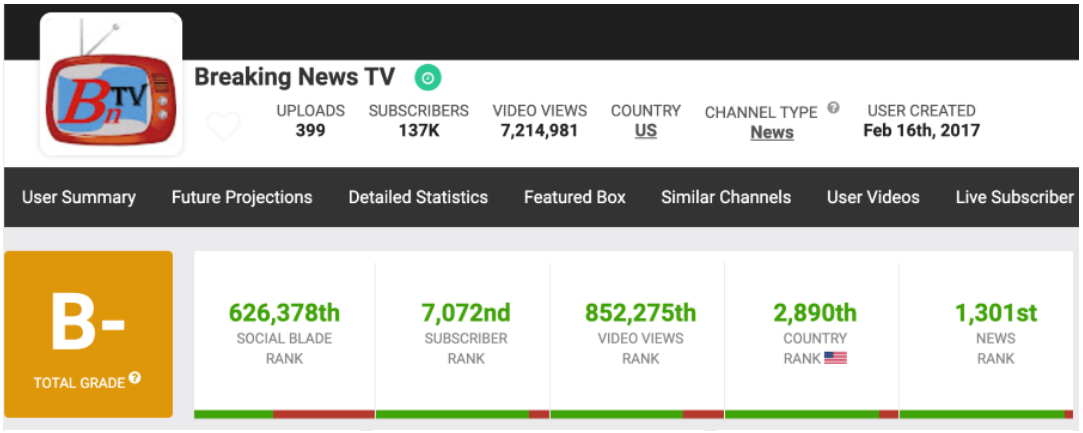

Figure 4: *Breaking News TV* statistics on Social Blade.

Topic ID	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Distribution
11	china	russia	taiwan	war	sea	hingga	south	bantuan	0.5805
17	dan	ini	indonesia	china	berita	yang	untuk	terbaru	0.0851
4	covid	china	vaksin	perang	new	selatan	begini	attack	0.0435
13	china	world	shanghai	missile	city	border	jet	amid	0.0413
3	dengan	china	tak	abdul	ustadz	beijing	somad	bikin	0.0351

Table 5: Cluster 0 topics and their keyword distribution.

Topic ID	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Distribution
0	china	uyghur	berita	uighur	muslim	covid	tiongkok	tahrir	0.6583
1	china	berita	covid	dan	feb	maritime	pagi	pancasila	0.3417

Table 6: Cluster 1 topics and their keyword distribution.

Topic ID	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Distribution
27	polisi	kasus	untuk	news	akan	ibu	waspada	tim	0.1658
29	yang	jadi	indonesia	dalam	dana	team	express	center	0.0804
18	dan	jokowi	hari	terkait	city	menteri	tol	laut	0.0777
32	ini	soal	orang	gelar	papua	aiman	exclusive	perempuan	0.0555
22	live	motor	china	conference	ceremony	rusia	putin	spokesperson	0.0533

Table 7: Cluster 2 topics and their keyword distribution.

Topic ID	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Distribution
5	cara	alessia	video	lyric	best	live	china	sub	0.5256
0	shengchi	beauty	hui	huang	zheng	皮囊之下	ugly	sub	0.1538
3	sub	eng	dream	stand	与君歌	china	cara	alessia	0.1026
1	alessia	cara	live	best	china	zheng	huang	ugly	0.0897
2	china	shengchi	beauty	hui	eng	sub	皮囊之下	ugly	0.0769

Table 8: Cluster 3 topics and their keyword distribution.

ies (Kirdemir, Adeliyi, and Agarwal 2022), channels with news, military, and politics-related content have been known to show more inorganic activity, and we see very similar activity in this cluster.

Cluster 1

This cluster comprises channels that exhibit the least inorganic activity, with an average inorganic score of 0.29 and the highest score being 0.38. Channels within this cluster show anomalies across four indicators; however, the scores indicate relatively lower levels of inorganic activity. These channels include *Indie Media Kepri*, *Spartan Tutorials*, and *Tsaqofah TV*, among others. As shown in Table 6, the only two topics in this cluster show keywords highlighting China, the repression of Uyghurs, and Covid, and there is no discourse around war or military.

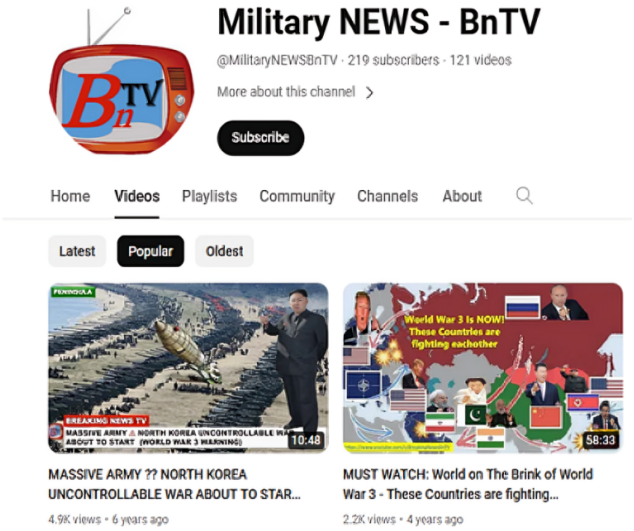

Figure 5: Snapshot of *Military News – BnTV*'s channel page on YouTube.

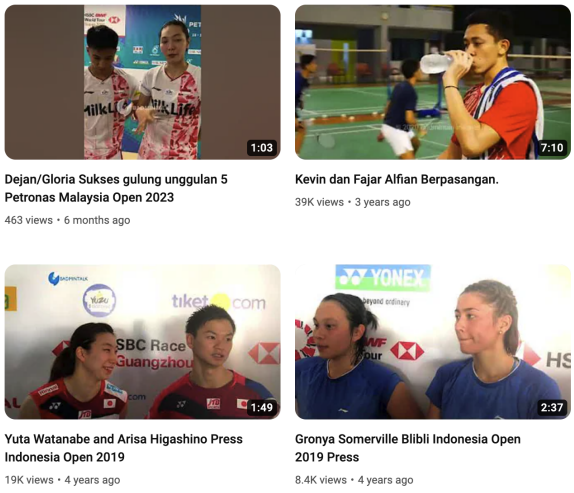

Figure 6: Snapshot of *Badminton Talk*'s channel page on YouTube.

Cluster 2

Channels within this cluster also show anomalies across all six indicators with an average inorganic score of 0.32. Channels within this cluster exhibit inorganic activity within the subscribers, videos, and comment statistics, with the highest score being 0.51. These channels include some news channels such as *NTD Indonesia*, *Andalus*, and *WION*, among others. As shown in Table 7, the top five topics in this cluster have a moderate distribution spread, with keywords including *news* and *polisi*. While there is little presence of military-related content, we do see news as one of the top keywords, indicating the presence of news and media-related chan-

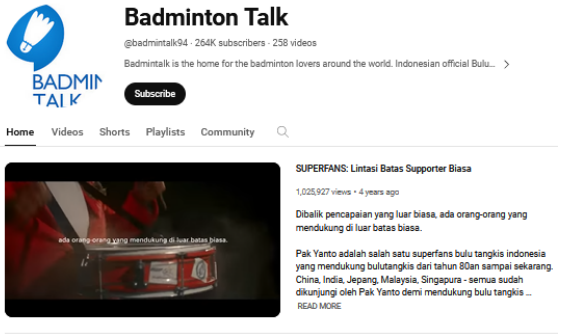

Figure 7: Snapshot of *Badminton Talk*'s featured video.

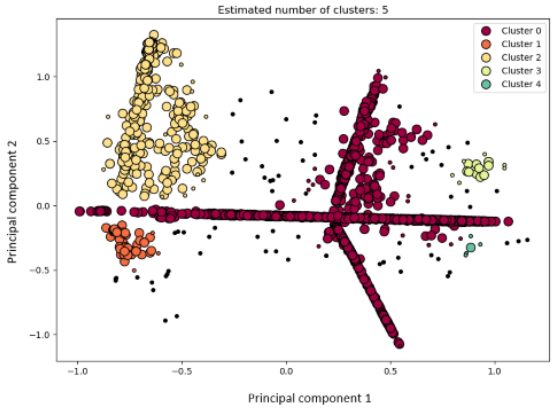

Figure 8: DBSCAN clustering of channels.

nels, which are known to show inorganic activity (Kirdemir, Adeliyi, and Agarwal 2022), possibly because they tend to publish viral content, among other reasons. This could also explain the higher inorganic score in this cluster as compared to cluster 1.

Cluster 3

The channels in this cluster show anomalies across three indicators, with a higher average inorganic score of 0.49. Inorganic activity is detected within the views, subscribers, and comments statistics, with the highest score being 0.54. Some of the channels include *TNS Media*, *AlessiaCaraVEVO*, *FutureVEVO*, *Geeks Spaces*, and more. From Table 8, we see that Topic 5 has the highest distribution with keywords such as *video*, *lyric*, and *live*. While this cluster has a smaller number of channels, the keywords indicate media-related content, which also tend to show higher inorganic activity due to their growth patterns, hence the relatively higher score in this cluster.

Cluster 4

The channels with the highest inorganic activity are present in this cluster, with an average inorganic score of 0.54 and the highest score being 0.72. While the channels in this cluster also show anomalies across all six indicators, similar to cluster 0, strong inorganic activity is detected within

the views, subscribers, and comments statistics. The suspended channel in this cluster is *Breaking News TV*, and other channels in this cluster include *viral_makkodak* and *Badminton Talk*. Only three videos were extracted from this cluster with one of the titles as *"DRONE MILIK CHINA DITEMUKAN DI SULSEL"*, which translates as "*CHINESE DRONE FOUND IN SULSEL*", indicating military-related content. Another video title reads *"Viral!! Aksi Solidaritas Bela Islam Uyghur 27 Desember 2019 Ptk Kalbar"*, which translates as "*Viral!! Solidarity Action to Defend Uyghur Islam 27 December 2019 Ptk Kalbar*".

Conclusion and Future Work

In this study, we aimed to detect and characterize inorganic activity within a dataset comprising 3542 channels. From our analysis, we discovered that some of the channels that exhibit highly inorganic activity publish a similar form of content, and we also discovered that clusters with higher inorganic scores comprise channels that publish more news-, politics-, and military-related content. The inorganic scores generated in this study are based on the user engagement statistics, and we intend to study other dimensions, such as the commenter activity within the channels, to provide a multidimensional approach to scoring inorganic activity.

Furthermore, we verified the channels that have been suspended by YouTube within our dataset. Verifying the account status of these channels was done to obtain some form of ground truth for validating channels that exhibit inorganic activity. While an account could be suspended for several reasons, inorganic activity could also lead to suspension according to YouTube's policies (YouTube 2023). As this study progresses, we will also explore other effective means to validate inorganic activity within channels. The channel with the highest inorganic score within our dataset—*Breaking News TV*—was among 98 others that had been suspended, with an alternate channel already having been created long before the channel was suspended. This could be a common strategy for similar channels that grow their engagement statistics inorganically, providing them with a fallback option when their channels are suspended.

We also discovered that inorganic activity was predominant within the views and subscriber statistics of channels in our dataset, indicating that channels that grow their appraisals inorganically may focus more on their view and subscriber count, as these are key metrics for monetization and recommendation on YouTube. Building on this analysis, we envisage the creation of bots programmed to monitor channels for signs of inorganic activity in real-time. This proactive approach to detection holds the potential to thwart misinformation propagation strategies at their inception, providing invaluable assistance to policymakers and platform moderators in maintaining the integrity of the platform. The behaviors captured by the channels that exhibited inorganic activity can be integrated into data collection software such as YTCDB (Solaiman and Agarwal 2023) to further supplement metrics in aiding inorganic behavior detection and be used to collect larger datasets taking into account engagement and behavior metrics of channels.

Acknowledgements

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

Al-khateeb, S.; Anderson, M.; and Agarwal, N. 2021. Studying the role of social bots during cyber flash mobs. In *Social, Cultural, and Behavioral Modeling: 14th International Conference, SBP-BRiMS 2021, Virtual Event, July 6–9, 2021, Proceedings 14*, 164–173. Springer.

Beutel, A.; Xu, W.; Guruswami, V.; Palow, C.; and Faloutsos, C. 2013. Copycatch: stopping group attacks by spotting lockstep behavior in social networks. In *Proceedings of the 22nd international conference on World Wide Web*, 119–130.

Carley, K. M. 2020. Social cybersecurity: an emerging science. *Computational and mathematical organization theory*, 26(4): 365–381.

Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Stanley, H. E.; and Quattrociocchi, W. 2016. The spreading of misinformation online. *Proceedings of the national academy of Sciences*, 113(3): 554–559.

Dutta, H. S.; Jobanputra, M.; Negi, H.; and Chakraborty, T. 2021. Detecting and analyzing collusive entities on YouTube. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(5): 1–28.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The rise of social bots. *Communications of the ACM*, 59(7): 96–104.

Galeano, K. K.; Galeano, L.; Mead, E.; Spann, B.; Kready, J.; and Agarwal, N. 2020. The role of youtube during the 2019 Canadian federal election: a multi-method analysis of online discourse and information actors. *J. Future Conflict*, 2: 1–22.

Giglietto, F.; Righetti, N.; Rossi, L.; and Marino, G. 2020. It takes a village to manipulate the media: coordinated link

sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society*, 23(6): 867–891.

Google Developers. 2021. API Reference — YouTube Data API. URL: https://developers.google.com/youtube/v3/docs. Visited on January 21, 2022.

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)*, 1092–1095. IEEE.

Khaund, T.; Kirdemir, B.; Agarwal, N.; Liu, H.; and Morstatter, F. 2021. Social bots and their coordination during online campaigns: A survey. *IEEE Transactions on Computational Social Systems*, 9(2): 530–545.

Kin Wai, N.; Horawalavithana, S.; and Iamnitchi, A. 2021. Multi-platform information operations: Twitter, facebook and youtube against the white helmets. In *Proceedings of The Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media (ICWSM)(SocialSens' 21). Atlanta, USA*.

Kirdemir, B.; Adeliyi, O.; and Agarwal, N. 2022. Towards Characterizing Coordinated Inauthentic Behaviors on YouTube. In *ROMCIR@ ECIR*, 100–116.

Nevado-Catalán, D.; Pastrana, S.; Vallina-Rodriguez, N.; and Tapiador, J. 2023. An analysis of fake social media engagement services. *Computers & Security*, 124: 103013.

Shajari, S.; Agarwal, N.; and Alassad, M. 2023. Commenter Behavior Characterization on YouTube Channels. In *Proceedings of the Fifteenth International Conference on Information, Process, and Knowledge Management (eKNOW 2023)*, 59–64.

Shajari, S.; Alassad, M.; and Agarwal, N. 2023a. Characterizing Suspicious Commenter Behaviors. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, 631–635.

Shajari, S.; Alassad, M.; and Agarwal, N. 2023b. Detecting Suspicious Commenter Mob Behaviors on YouTube Using Graph2Vec. *arXiv preprint arXiv:2311.05791*.

Social Blade. 2023. Getting started with the Business API (Matrix) - Socialblade.com. URL: https://socialblade.com/business-api. Visited on August 22, 2023.

Solaiman, I. A.; and Agarwal, N. 2023. Multiagent-based Youtube Content Discovery Bot. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, 450–453.

Yousefi, N.; Cakmak, M. C.; and Agarwal, N. 2024. Examining Multimodal Emotion Assessment and Resonance with Audience on YouTube.

Yousefi, N.; Shaik, M.; and Agarwal, N. 2024. Characterizing Multimedia Information Environment through Multimodal Clustering of YouTube Videos. *Proceedings of the 4th International Conference on SMART MULTIMEDIA*.

YouTube. 2023. Fake engagement policy - YouTube Help. URL: https://support.google.com/youtube/answer/3399767?hl=en&ref_topic=9282365. Visited on November 14, 2023.

Ethical Statement

In conducting our research on the detection and characterization of inorganic activity within YouTube channels, we are steadfast in our commitment to upholding ethical principles. Our aim is to contribute responsibly to the understanding of online misinformation dynamics, disseminating our findings with accuracy and integrity. We recognize the potential impact of our work on the digital ecosystem and pledge to approach dissemination with caution, avoiding sensationalism or misrepresentation. Through ongoing reflection and dialogue, we seek to foster ethical awareness and responsible conduct in the realm of online research and technology development, ultimately striving to promote integrity and transparency in digital platforms and beyond.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? No
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? No
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes
   (b) Have you provided justifications for all theoretical results? Yes
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? No
   (e) Did you address potential biases or limitations in your theoretical framework? Yes
   (f) Have you related your theoretical results to the existing literature in social science? Yes

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes

3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? No
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, it is described in methodology
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA