THE EFFECT OF A MEDIA SHUTDOWN ON PHILIPPINE NEWS MEDIA CONTENT: A COMBINED SENTIMENT ANALYSIS AND TOPIC MODELLING APPROACH

Matthew R. Maulion[1*], Adrian M. Perez[2]

[1]University of Science and Technology of Southern Philippines, Lapasan, Cagayan de Oro City 9000, Philippines

[2]De La Salle University, 2401 Taft Avenue, Manila, Philippines

*For Correspondence; Tel. +63 9155112804, Email: _matthewr.maulion@ustp.edu.ph_

ABSTRACT – _The ABS-CBN shutdown last May 5, 2020, took a toll in the Philippine media industry. Given this incident, the study intends to determine if significant changes can be detected in the manner that articles were written before and after the ABS-CBN shutdown. Its significance is to determine using natural language processing (NLP) methods if censorship or a sudden event (i.e. a media giant shutdown) can influence the way news is written and published. Articles before and after the shutdown from two primary news sources were chosen, ranging from three months before and after the actual shutdown date. Preprocessing and cleaning were carried out in order to perform sentiment analysis and topic modeling. Results were compared across the two timeframes and across the two sources. Notable shifts in the dominant topics discussed by each source were detected for both timeframes, with topics disappearing and emerging after the shutdown. Furthermore, shifts in sentiment scores were also detected across both sources with evident changes in polarity for some uncovered topics. It is recommended for future work to make the analysis more granular by incorporating page views and conducting an in-depth content assessment through paragraphs and sentences._

Keywords: Sentiment Analysis, Topic Modelling, Natural Language Processing, News Media

1 INTRODUCTION

News media is part of the encompassing coverage of mass media. Its goal is to provide news to a target audience or the general public. This explains the numerous news networks or stations we see on our televisions and the web day by day. News networks serve as an intermediary between the government and the people, allowing them to engage and be proactively involved in society and politics. These news networks must follow and abide by the government's regulations to operate. Broadcast entities in the Philippines must first register with the Securities and Exchange Commission (SEC) in order to incorporate their existence and conduct business. Then, a primary franchise must be obtained from Congress through a passage of a bill [1]. ABS-CBN, a media giant in the Philippines, was not granted renewal of its franchise to continue operations after the expiry of its license last 2020. Through the unfolding of this event, the researchers examined whether the shutdown had an impact on the news articles we consume. ABS-CBN ceased operations last May 5, 2020. This reference point is the groundwork of the various Natural Language Processing (NLP) methods employed in this study. The basis for the selection of GMA as the network of comparison towards ABS-CBN is from a Philippine Media and Public Trust Survey conducted by the Vera Files for 2020 [2]. Through NLP methods, we examined if significant changes can be detected in the manner articles are being written by other mainstream media companies, before and after the ABS-CBN shutdown.

2 METHODOLOGY

2.1 Data Collection – The study used news articles from the ABS-CBN News Channel and GMA Network News and Entertainment. A combined total of 17,460 news articles were collected considering a 3-month date range before and after the shutdown date.

2.2 Data Cleaning and Preprocessing – The cleaning methodology relies mainly on the nature of the dataset. The researchers performed text-cleaning methods that include punctuation/numeric token/stop word removal, part-of-speech tagging, lemmatization, and tokenization. After null removals and the implementation of cleaning and preprocessing methods, a final distribution of the dataset is presented in Table 1.

Table 1. News article distribution after preprocessing

Network	Before shutdown	After Shutdown	Total
ABS-CBN	3,395	3,021	6,416
GMA	3,568	3,654	7,222
Total	6,936	6,675	13,638

2.3 Topic Modeling – This is an unsupervised machine learning technique capable of scanning a set of documents, detecting word and phrase patterns within them, and automatically clustering word groups and similar expressions that best characterize a set of documents. The study used the Latent Dirichlet Allocation (LDA) topic model as it is the most used, state-of-the-art, and simplest method across the computing field [3].

Topic modeling is conducted for both news networks and for both timeframes. However, there is a need to identify the optimal number of topics (k) drawn from the corpus at hand. This is determined through the calculation of the coherence score. It is stated that a set of statements is said to be coherent if they support each other. Thus, a coherent statement set can be interpreted in a context that covers all or most of the statements. Formally, topic coherence refers to a single topic score by measuring the degree of semantic similarity between high-scoring words in the topic. These measurements help distinguish between semantically interpretable topics and topics that are artifacts of statistical inference [4].

2.4 Sentiment analysis – This is a text analytics method that identifies polarity (i.e., a positive or negative opinion) within a specified text. This method measures the attitude, and sentiments, of a speaker/writer based on the computational treatment of subjectivity in a text [5]. This paper utilized the VADER sentiment analysis model which returns a dictionary of scores with each key corresponding to a sentiment value.

The implementation of sentiment analysis is focused on three angles: (1) sentiment analysis with stop words on all articles, (2) sentiment analysis without stop words on all articles, and (3) sentiment analysis without stop words on a per-topic basis. This is implemented for both ABS-CBN and GMA news articles. Mean compound scores for both angles, for both timeframes (i.e., before and after shutdown), and for every news network are computed and compared accordingly.

3 RESULTS AND DISCUSSION
3.1 Determining the Optimal Number of Topics
Upon implementing topic modeling, results show that the optimal number of topics for ABS-CBN should be 9 for both before and after the shutdown, while GMA coherence results dictate their topic number should contain 8 topics before their competitor's shutdown and 10 after the said event (see Fig. 1).
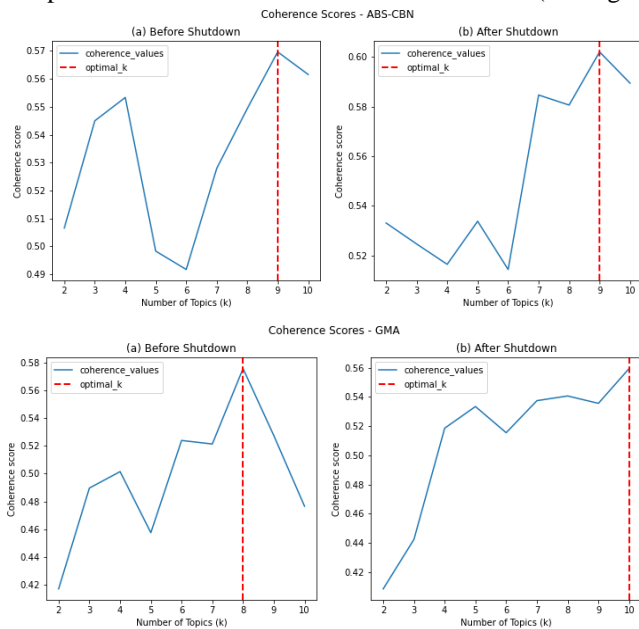

Figure 1. ABS-CBN and GMA LDA Coherence Scores

3.2 Uncovered Topics from both Media Networks
The generated topics were assigned a name by looking at the most salient terms using pyLDAviz. A topic name is composed of one to two words that describe the context or the central theme of the topic. Comparing the derived topic names before and after the shutdown period highlights the topics or news themes that are still widely reported by both networks even after the media giant's shutdown and those that have been less articulated, resulting in the non-appearance of the topic name.

Table 2. ABS-CBN LDA topics

#	Before Shutdown	#	After Shutdown
1	Healthcare	5	Healthcare
2	Quarantine	6	Quarantine
3	Education	8	Education
4	Travel	3	Travel
5	Funds & Aids		
6	President/ Government		
7	Senate/ Government		

8	ABS-CBCN Franchise	9	ABS-CBCN Franchise
9	COVID-19	4	COVID-19
		1	Terrorism
		2	Police/ Military
		7	Foreign Affairs

Table 3. GMA LDA topics

#	Before Shutdown	#	After Shutdown
1	COVID-19	4	COVID-19
2	Travel		
3	Government	10	Government
4	Healthcare	3	Healthcare
5	ABS-CBN Franchise	9	ABS-CBN Franchise
6	Police/ Military	8	Police/ Military
7	Quarantine	6	Quarantine
8	Funds & Aids	5	Funds & Aids
		1	Education
		2	Terrorism
		7	Foreign Affairs

For ABS-CBN, it can be seen in Table 2 that after the network's shutdown, news about Funds & Aid, and the President and Senate (i.e., the government) were less articulated. Also, new topics have emerged, namely, Terrorism, Police/Military, and Foreign Affairs. On the other hand, GMA only has one topic that did not appear after ABS-CBN's shutdown – the Travel topic, as shown in Table 3. Similar to ABS-CBN, new articles were written that contain themes of Education, Terrorism, and Foreign Affairs.

3.3 Uncovered Topics from both Media Networks and their Sentiment Compound Scores
As mentioned in Section 2.4, sentiment analysis will be the first run on all news articles with stop words, and another one for news articles without. It can be inferred based on the result that news articles with stop words tend to polarize to the negative sentiment, and there is no significant difference in the before and after shutdown sentiments for both news networks, as shown in Table 4.

Table 4. VADER compound scores – with stop words

Network	Before Shutdown	After Shutdown
ABS-CBN	-0.115802	-0.105781
GMA	-0.081631	-0.145091

Table 5. VADER compound scores – with stop words

Network	Before Shutdown	After Shutdown
ABS-CBN	-0.004634	0.029926
GMA	0.046802	0.047510

In contrast, sentiment analysis on news articles without stop words yields different results, especially for ABS-CBN. Table 5 shows that the overall sentiment of ABS-CBN articles

shifted from negative to more positive sentiment after the shutdown. On the other hand, GMA news has almost the same average compound score before and after the announcement, but this time on a more positive tone compared to the result with stop words included.

Sentiment analysis was also run on a per-topic basis. Tables 6 and 7 show the per-topic average compound sentiment score for both networks. The compound score ranges from -1 (most extreme negative) and +1 (most extreme positive). Cells are colored green if their score increases and red otherwise. Cells that are left blank mean that the topic names aligned with it do not appear on the other timeframe. Results show that all negatively polarized topics, except COVID-19, shifted to positive in the after-shutdown timeframe for both news networks. Another similarity is that the almost equal sentiment score of the Terrorism topic emerged after the shutdown timeframe – ~-0.5, which is highly negative. Overall, ABS-CBN has a more negative sentiment on this topic, while GMA is leaning towards the positive side.

Table 6. ABS-CBN per topic sentiment

Topic	Before	After
Healthcare	0.2778	0.3345
Quarantine	-0.195	0.0581
Education	0.1709	0.1888
Travel	-0.202	0.1476
ABS-CBN Franchise	0.3712	0.2780
COVID-19	-0.132	-0.267
Funds & Aid	0.0546	
President/ Government	-0.204	
Senate/ Government	0.3329	
Terrorism		-0.505
Police/Military		-0.235
Foreign Affairs		0.5090

Table 7. GMA per topic sentiment

Topic	Before	After
COVID-19	-0.075	-0.167
Government	-0.090	0.3053
Healthcare	0.0140	0.0849
ABS-CBN Franchise	0.5222	0.4528
Police/ Military	-0.2414	0.0791
Quarantine	0.0177	0.1071
Funds & Aid	0.2360	0.1952
Travel	0.0859	
Education		0.3918
Terrorism		-0.534
Foreign Affairs		0.1341

Having speculated that the shutting down of the ABS-CBN network is highly political in nature, it is noticeable that while ABS-CBN's topics about government (i.e., Government, Senate) did not appear as major themes of the after-shutdown news, while GMA's Government topic unfolded to have polarized to the positive sentiment.

4 CONCLUSION AND RECOMMENDATIONS

In this study, topic modeling presented interesting results across news networks. For ABS-CBN, it was observed that after the network's shutdown, articles that focus on funding, financial aid, and government-related affairs were less articulated in the news. Further, the event caused the emergence of new topics, namely, Terrorism, Police/Military, and Foreign Affairs. Interestingly for GMA, only one topic did not appear after ABS-CBN's shutdown (i.e., Travel). Similar to ABS-CBN, GMA news articles contain themes of Terrorism and Foreign Affairs. Contrastingly though, GMA topics still include the topics that were less prevalent in ABS-CBN news articles after the shutdown. Looking at the overall sentiment score for ABS-CBN, it is observed that there is a shift from negative to positive, in contrast with GMA where the news sentiments remain positive even after their competitor's shutdown. Digging deeper into the per-topic sentiments, GMA's government and police/military-related articles have shifted to a more positive sentiment after the shutdown.

In summary, it is apparent that there are changes in the way news is written and reported for both networks. However, it is out of the scope of this study to determine if the change is due to the ABS-CBN shutdown or a result of a natural course of changes in topics.

Though this study primarily sought the changes that can be detected by NLP methods, a more comprehensive study can be made by perhaps integrating the number of page views of each article (within a fixed window of time of its posting) before and after the shutdown to see if even the numbers were affected by the events that unfolded, or whether they were affected by or related to the shifts in topics and sentiments. It is also recommended that the granularity of the articles' time was made more specific to see whether the changes in topics and sentiment were gradual or abrupt.

5 REFERENCES

[1] Tomacruz S. How media outlets get their license to operate. Rappler. Retrieved from https://www.rappler .com/newsbreak/iq/media-outlets-license-operate-press-freedom-sec. (2018)

[2] Tordesillas E. Media and public trust. Vera Files. Retrieved from https://verafiles.org/articles/media-and-public-trust (2020)

[3] Asmussen, C.B., Møller, C. Smart literature review: a practical topic modelling approach to exploratory literature review. J Big Data 6, 93. https://doi.org/10.1186/s40537-019-0255-7 (pp. 4-5) (2019)

[4] Rosner, Frank & Hinneburg, Alexander & Röder, Michael & Nettling, Martin & Both, Andreas. Evaluating topic coherence measures. *Neural Information Processing Systems Foundation (NIPS).* arXiv:1403.6397v1 [cs.LG] (pp. 1-2) (2013)

[5] Taboada, Maite.. Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics.* 2. 10.1146/annurev-linguistics-011415-040518. (pp. 2-4) (2016)