2023-03-17

ANALYZING ONLINE MEDIA PLATFORMS FOR HACKTIVIST GROUP ORGANIZATION AND PROLIFERATION

Quincy Taylor
BYU

Honor Thesis

ANALYZING ONLINE MEDIA PLATFORMS FOR HACKTIVIST GROUP ORGANIZATION AND PROLIFERATION

by

Quincy C. Taylor

Submitted to Brigham Young University in partial fulfillment of graduation requirements for University Honors

Electrical and Computer Engineering Department

Brigham Young University

April 2023

Advisor & Honors Coordinator: Derek Hansen

ABSTRACT

ANALYZING ONLINE MEDIA PLATFORMS FOR HACKTIVIST GROUP ORGANIZATION AND PROLIFERATION

Quincy C. Taylor

Electrical and Computer Engineering

Bachelor of Science

The anonymity and lack of censorship online provides the perfect environment for hacker activists to pursue social change. The expansive reach and democratic access to social media has empowered groups to organize and develop messaging to specifically fit an online audience. As social media become ubiquitous, the reputation and use of messaging application have become mainstream. Due to a self-professed focus on privacy, platforms like Telegram have become the norm for hosting the hacktivist communities. The purpose of this research was to understand the features in Telegram messages that correlate with the most engagement from their audience. As expected, the number of users is the best predictor of number of message views. Other significant predictors of more reactions and/or forwards included certain topics (nationalism, military, and cyber) identified by topic modeling, as well as the inclusion of more text, links, and documents. Interestingly, have a photograph made it more likely a message would be forwarded, but less likely it would be reacted to.

ACKNOWLEDGMENTS

This study was a labor of learning. I am grateful to Dr. Derek Hansen and Dr. Justin Giboney for shepherding me through this research process. Their insights and expertise fueled this work and my learning experience. Beyond this project, they have been mentors throughout my time at Brigham Young University. Dr. Hridoy Sankar Dutta ignited my interest in this topic, while at Cambridge University for the Honors Direct Enrollment Program, for which I am indebted. I would also like to thank Dr. Dennis Eggett for his efforts in the statistical analysis.

Finally, the support and vision provided by my family were instrumental to the completion of this thesis and my university experience. I am forever grateful to them, because they give me eyes to see.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

1. INTRODUCTION

On 5 July 1993, *The New Yorker* published a cartoon with the caption "On the Internet, nobody knows you're a dog." The phrase humorously observed the anonymity of the internet. The anonymity and lack of censorship online provides the perfect environment for hacker activists to pursue social change. The origin of the term "hacktivist" is contested, but recently it has become more common in discussions of geo-political movements and the cyber threat landscape. Hacktivists are individuals who use cyber-attacks to progress a political agenda or ideology (Fowler, 2016). The expansive reach and democratic access to social media has empowered groups to organize and develop messaging to specifically fit an online audience. As social media become ubiquitous, the reputation and use of messaging application have become mainstream. Due to a self-professed focus on privacy, platforms like Telegram have become the norm for hosting the hacktivist communities. For hacktivists, the line between privacy and publicity is fine. The platforms utilized should emphasize privacy for the users and contributors, while guaranteeing easy access to the messaging that is published. The goal of hacktivist groups is to propagate the ideology and gain a following to maximize impact.

Figure 1 Cartoon from The New Yorker Published 5 July 1993

This study will analyze data collected from a channel from the online messaging platform, Telegram, utilized by members of a hacktivist group to understand techniques used to propagate messages within the community. In order to maintain privacy and minimize retaliation from the researched group, the specific identity of the hacktivist group and related identifiers have been anonymized in this study.

1.1 Research Problem

As platforms published predictions for the 2023 trends in cybersecurity, hacktivism emerged as an area of dynamic growth. The technology news outlet, Wired, noted a rise in hacktivism in 2022 (Burgess, 2022). This rise was attributed to the geo-political shifts that were fueled by disinformation campaigns and online movements. This research seeks to contribute a small part to this broader trend in cybersecurity.

This research will assume that the greater engagement through views, forwards, and reactions received by hacktivists groups online indicates success in propelling the claims and messaging of the group. This analysis is important because the questions can lead to a better understanding of the messages and identifying important features that are correlated with increased engagement.

Research Question: What message factors lead to increased engagement in hacktivist Telegram channels?

This question can be broken down to two sub-questions:

1) How does the inclusion of attached data effect the engagement of the message?

   This question is designed to determine how messages with attached content such as images, documents/videos, and website links effect the engagement level after publication. Engagement through forwards, views, and reactions can detect if the message is viral or resonating with the audience. In order to address this question, messages are classified based on the raw message input based on the presence of the external attachments. A Boolean indicator was used to indicate the presence of each of these categories including: link presence, document attachment, and embedded webpage. Messages include different categories of attachment in Telegram.

2) How does the textual content of a Telegram message impact the engagement of the audience?

   Messages with different topics can be identified using a Boolean indicator. This research determined if the textual content of the message and the topics demonstrated have an impact of the engagement it receives from the overall audience. Likewise, the length of a message may impact engagement with it.

Naturally, the textual content is expected to be key to categorizing the messages and detect ideological values of the group. In order to experiment on this idea, the textual content extracted from the channel is used to determine key topics and then the resulting words are treated as indicators of the topic in each message. The classification of the messages was then analyzed to see if the topic would be a predictor of the engagement received by the message. The effect of textual content analysis on the engagement of the message is illustrated in Section 5. This study uses statistical analysis on messages from a sample Telegram hacktivist channel to find out any potential relationship between topic and audience engagement.

1.2 Contributions

Data-driven cybersecurity leverages the expansive amount of publicly available data, or big data, to make security decisions and develop meaningful intelligence about cyber actors and their targets. Previous work has been done to apply techniques of data-driven cybersecurity to the world of hacktivist cybercrime. This study will expand on previous work by utilizing data from Telegram to understand the factors that lead to propagation of hacktivist ideologies.

Previous research tracking the public messages of hacktivist groups have narrowly analyzed the messages on technical forums or other social media platforms. Similarly, this research will center around Telegram and specific channels associated with a prolific hacktivist group. This research will analyze the content of messages and the metadata associated with the messages posted in public Telegram channels. The research will provide an understanding of the characteristics of messages and how they impact readership, forwards, and reactions among members of the hacktivist group.

This research is important because the messaging from hacktivist groups provides us direct insight into the contact hacktivist groups have with their audience. As observed by the popular xkcd comic, hacktivists are highly impacted by the messaging that surrounds them. By understanding how they reach their audience, analysts and commentators are empowered to scale their expectations associated with the activity.

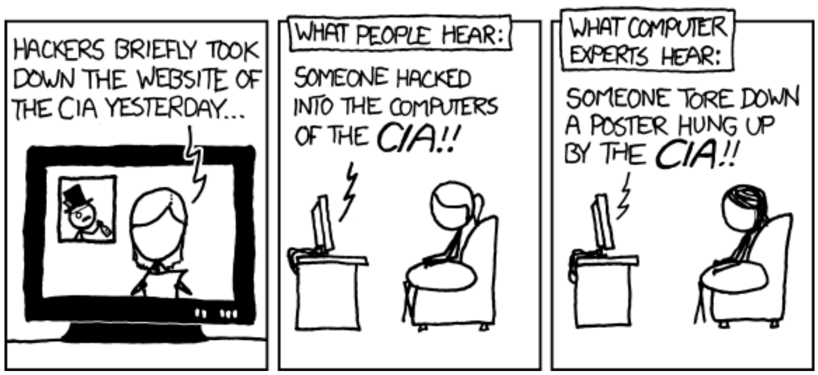

Figure 2 xkcd Comic "CIA"

1.3 Definitions

- Engagement: the audience interaction with a message captured and characterized in this study as views, reactions, and forwards.
- Hacktivist: activists that utilize online disruption and hacking to propagate their ideologies.
- Message factors: functionalities provided through Telegram to build a message.

2. BACKGROUND AND RELATED WORK

A variety of studies exists on maximizing social media engagement, the purpose of Telegram, and the broader use of online platforms by hacktivists. This section will analyze recent research coalescing in this study.

2.1 Why Hacktivism?

Hacktivists are ideologically fueled hacking groups with the motivation and goal of activism often through disruption (Fowler, 2016). Despite stereotypically being associated with rudimentary tactics, techniques, and procedures (TTPs) as seen in Figure 2, hacktivists are nevertheless actors that influence the cyber threat horizon. The claims and efforts made by hacktivist can often lead to loss of availability or reputational damage for their targets. Hacktivist groups have been overlooked as serious threat actors, because they fail to mature and reach the necessary technical capacity of advanced persistent threats (APT) or cybercriminal groups (Djavaherian, 2022). The misalignment between the declared objectives of hacktivists and the measured impact of their actions causes these groups to be dependent on social influence for ideology propagation.

As noted by Loh, hacktivist's value the anonymity of the individual (2022). The article specifically analyzes the effects of their anonymity on the legal status of hacktivist protests. The methods of protests selected in the work meet the criteria of digitality, publicity, nonviolence, and illegality. The protests specifically addressed include DDoS actions, website defacement, leaking, and copyright infringement (Loh, 2022). An important element of the article is the argument that a hacktivist's emphasis on anonymity suggests their unwillingness to accept the legal consequences of the action, therefore disqualifying them from a characterization as social activists. The reality of anonymity amongst hacktivists requires researchers to find alternative means to understand the activity and motivations of these groups. This study aims to contribute to this broad and expanding conversation as hacktivism grows in popularity and potential impact.

The influence of social media has expanded, which allows greater access to data sources by which to study hacktivist groups. Public posts provide a wealth of information directly from the individuals researched. For example, research addresses the identification of key cybercriminals through their participation in Internet Relay Chat (IRC) communities (Benjamin et al., 2016). This data-driven approach to assessing the activity of hacktivist is intended to better understand these anonymous and often elusive communities. Similar to previous work, this study will utilize public messages from self-identified hacktivists to understand better how to achieve increased engagement from their specific audience.

2.2 Why Telegram?

Telegram was created in 2013 by Pavel and Nikoli Durov after leaving the social media platform VKontakte when it was acquired by the company Mail.ru Group in 2014 (Kiselyova, 2014). Telegram has repeatedly emphasized the security of the platform, though recently there have been reports that it may not be as secure as once thought (Loucaides, 2023). Telegram is increasingly popular worldwide. Statista reported that in November 2022, Telegram had over 700 million active users globally (Telegram Messenger, 2022). As shown in Figure3, this number is distributed around the world with the largest populations of Telegram use by geography respectively being India, Russia, the United States, Indonesia, and Brazil (AppMagic, 2022). In many of these states, Telegram has become a full-fledged media platform rather than a simple messaging application. Past research has pointed to Telegram being a source of news and information in Russia and Belarus and also “a means of communication and self-organization in a political crisis” (Bykov, 2021). As the population of use and criticality of information increases, it is clear that Telegram plays a role in the flow of information around the world and the modern digital environment.

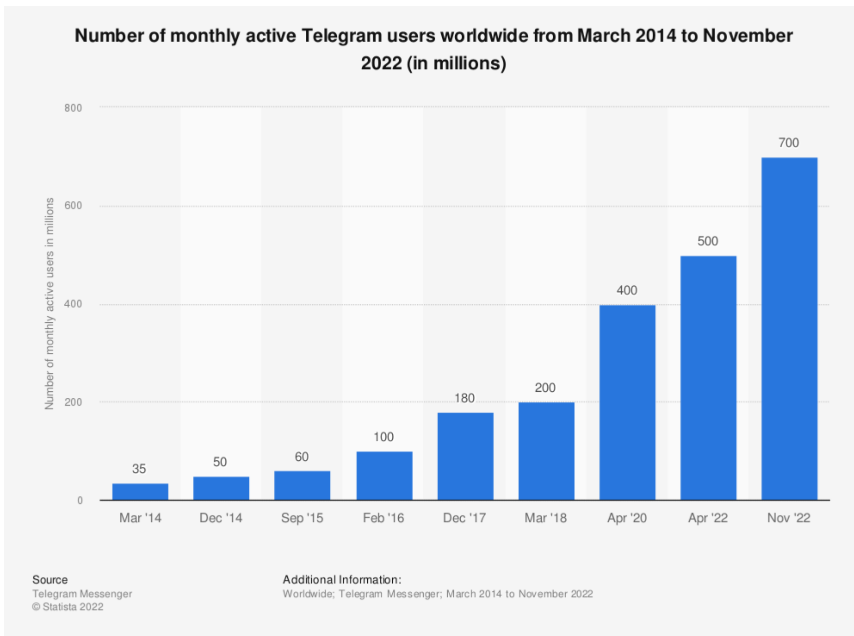

Figure 3 Overall Growth of Telegram from Launch to November 2022 (Statista)

Telegram has marketed itself as emphasizing privacy, though this has been contested because end-to-end encryption is not automatically enabled in the application. Nevertheless, the privacy claim may have led to the vast popularity of the platform. In 2021, *The Economist* magazine noted a sharp increase in membership for Telegram and Signal attributed user concerns for a new privacy policy by the Facebook owned company, WhatsApp (What are Signal and Telegram?, 2021). This announcement likely triggered users to look for alternatives thus creating a sharp increase in installations of the application as illustrated in Figure 4 (Sensor Tower, 2021). This event and its resultant effects demonstrate that the popularity of platforms like Telegram may reflect user interests in anonymity and privacy. The hacktivist value on anonymity is fulfilled through messaging platforms like Telegram.

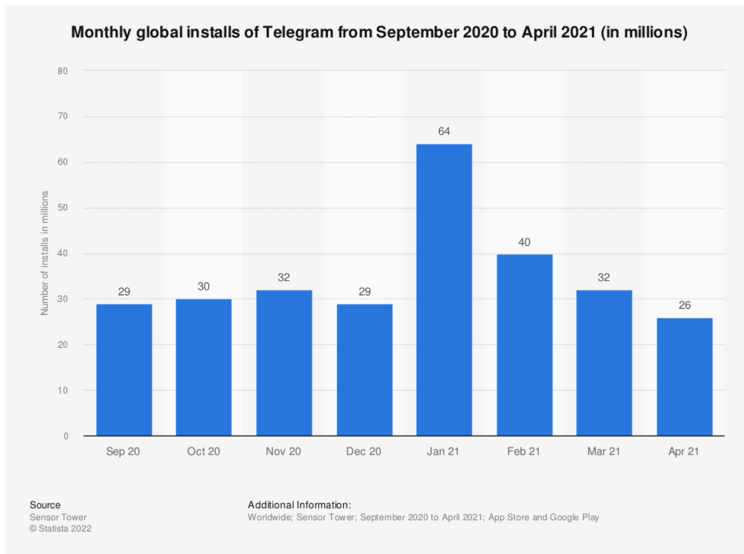

Figure 4 Global Installations of Telegram from September 2020 to April 2021

The functionalities of Telegram include live text, file sharing (video, image, etc.), and channels. Channels facilitate communication to a large audience by providing the ability to broadcast public messages, which aligns it with the communication flow prescribed by hacktivist groups. The administration of a channel is restricted to a limited group which includes the ability to post messages. There is no limitation to the number of subscribers that can follow a channel. Messages can be forwarded to a different user or channel. The functionality of forwarding is similar to retweeting through Twitter. Nevertheless, Telegram's features stand distinct from other micro-blogging platforms, like Twitter, due to the Telegram's ability to forward to a specific group/channel (Dargahi Nobari, 2021).

2.3 How to Increase Engagement?

Telegram differs from other messaging platforms because users are not limited to two-way communication (Dargahi Nobari et al., 2017). Two-way communication is best illustrated through direct or group chats that provide the user the ability to contribute, but the Telegram channel information flow limits contributions to a small group of administrators for broadcast messaging. Compared to groups, channels provide a level of anonymity to the user by limiting communication to broadcast messages by an administrator. The individual anonymity and one-way communication empowers individuals to congregate online, while maintaining a level of privacy. Based on this structure, Nobari has defined a viral message as one which is distributed by the audience of a channel to other users or different groups. Expanding on this definition of a viral message, this study measures three types of engagement with a Telegram message: views (number of people who viewed the message), reactions (number of reactions, such as thumbs up emoji, posted in response to the original message), and forwards (number of times a message was forwarded to another user of group). Instead of deriving virality solely from views as seen in Nobari (2021), this study considers the more active types of engagement including forwarding and reactions facilitated by the Telegram application.

The study of identifying key hackers was addressed in previous work which used a regression model to develop a metric of reputation in hacker forums (Benjamin & Chen, 2012). Many of the same metrics including message length and number of attachments are used in this research. Additionally, the use of a regression model to predict important factors influencing overall engagement was inspired by Chen (2012). In Nobari's 2017 article features of Telegram messages were used to detect advertisement messages and spam. The study concluded that topical linking was key to finding new members and PageRank did not detect the popularity of channels in Telegram. Some of the features identified by Nobari et al. study were used as features in this research including message length, message forwards, and the inclusion of links (2017). Therefore, the features or dependent variables used in the regression model for engagement amongst the hacktivist channel have a basis in work related to message virality and advertising as well as key hacker identification.

3. METHODOLOGY

Scrapers are open-source tools that retrieve and store the data and metadata associated with a site. A python script using the Telethon Library was used to scrape messages from a hacktivist Telegram channel. The message content and metadata, including date, reactions, forwards, media, views, and replies were collected and stored in a CSV for analysis (APPENDIX A). Because the data is available publicly and open to access by anyone, the use of a scraper does not indicate an ethical concern, so and IRB exemption was received.

The dataset is a collection of public messages from a channel self-identified with a hacktivist group with over an estimated 100,000 subscribers. To protect the privacy of the group and also protect the researchers from retaliation by the group, we have anonymized the group and avoid using exact numbers. Data was collected in early 2023. The dataset includes messages from less than a year of activity in the channel with an estimated 3,000 public messages. This period was selected because it represented an overall plateau of channel growth after extreme shifts. The channel has public content that can be accessed by all subscribers without restriction. Throughout the dataset, there are examples of message with a variety of combinations of textual content, photos, media, and links to external sites. For example, there are instances where an image is posted without textual content associated with it or that is just an image, but it only has reactions and views no content. Instances in which posts containing null values due to channel actions, such as pinning posts, rather than message posts were excluded.

Naturally, views consistently outnumber reactions and forwards on a message. This pattern is intuitive because readers can browse the channel increasing the view count without exercising further engagement through reactions and forwarding. Reactions demonstrate a viewer's individual opinion or acknowledgement of the message while maintaining a sense of anonymity in the broader channel. Lastly, forwards involve a reader's propagation of the message to another user or channel. The use of the forward functionality indicates the highest level of engagement, because it can be used to propagate the message beyond the channel membership. The act of forwarding a message demonstrates the viewer level of involvement with the content because they are sharing it with others.

3.1 Message Metadata and Topic Identification

The data sourced from Telegram was stored in a CSV file for convenient organization and access. In order to gain a better understanding of the content of the messages, the content was uniformly translated to English using the Google Translate function built into Google Sheets. The messages were all in another language with few phrases within messages in English previous to the translation. The character length of the content was extracted and the value stored. The links included in the content of the message were extracted. Topic modeling was performed using Latent Dirichlet Allocation (LDA) to identify themes within the messages (Kulshrestha, 2019). The model implemented through Python applied LDA to the collected messages and split them into topics. To facilitate clean analysis, pre-processing the data was necessary.[1] Removing duplicates, URLs, and extraneous punctuation was the first step. To further break down the content of the messages, tokenization allowed for the removal of words 3 characters or less and commonly used words in English called stopwords. The results were then lemmatized and stemmed, which is the process by which words are standardized by tense and person and then reduced to their root form. The genism[2] and nltk[3] libraries facilitated the cleaning of the message text associated with the Hacktivist group. The resulting text was then used to create a dictionary which connected words with the count of their frequency in the set. That dictionary was then used to filter out extremes. Models were then generated to gain a better understanding of what topics are important in reference to the hacktivist group. Words identified by the models were used to identify messages for categorization into three topics. The topics are not mutually exclusive, so some messages may be categorized in more than one topic. Based on the results, three topics were identified as independent indicators. The terms identified with the various topics were used as features of the message in the regression model. Although the titles we use for the

[1] Many methodologies can be used to process the data. The particular methodology employed in this study can be found at https://www.kaggle.com/code/sreejiths0/efficient-tweet-preprocessing/notebook and https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24
[2] Gensim is a Python library used for topic modelling. Additional information on this group can be found at https://radimrehurek.com/gensim/
[3] NLTK is a Python program that utilizes language data to perform natural language processing. Additional information on this group can be found at https://www.nltk.org/

groups are not a perfect representation of all terms, they capture the core element of each topic. They include: nationalism, military, and cybersecurity.

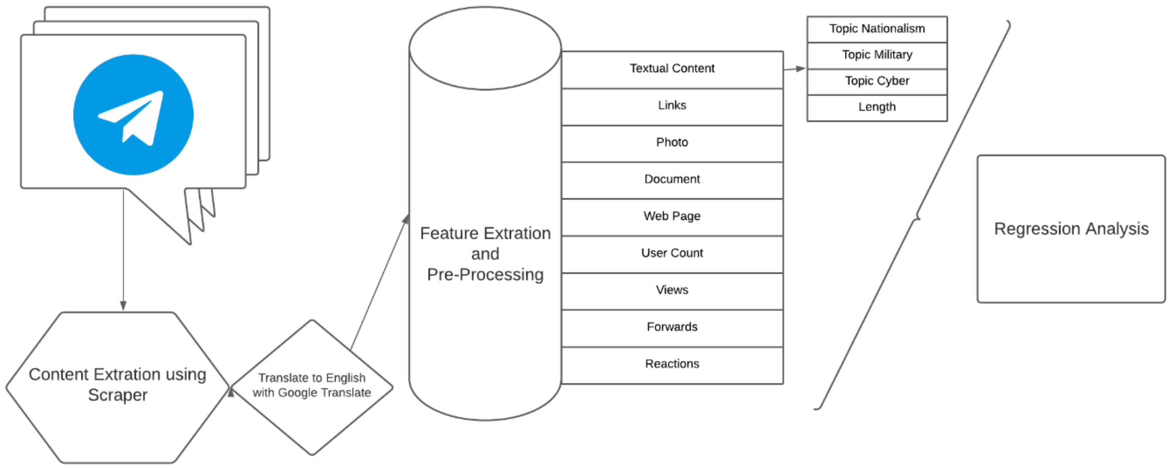

Figure 5 Project Design Diagram

3.2 Data Analysis

To measure the effects of the independent variables on the expected metrics of engagement (dependent variables), we performed a mixed linear regression using R. It was necessary to accommodate for the power log distribution by adding 1 and taking the natural log of the values of forwards and reactions. The views appeared to be normally distributed and was not natural logged. The independent dichotomous variables included: the three topics identified by the topic modeling exercise (nationalism, military, and cybersecurity), presence of links, webpage, photo, and document present. The continuous independent variables included length, user count, and views (for the forward and reaction regression models). We ran a regression model for each independent variable (views, forwards, and reactions). Values with a p-value of less than 0.001 were highlighted in each model as having statistical significance. The adjusted r-squared value was also calculated to determine the variance in the dependent variables that would be predicted by the independent variables.

4. RESULTS

4.1 Overview

The scraper used to retrieve data from the messages on Telegram (Appendix A) allows the researcher to specify the data extracted into a CSV file. The resulting data was then processed to obtain dichotomous variables as previously outlined in Section 5. Table 1 shows the resulting variables after processing and the percentage of the total message count. The variables were used in the regression analyses with the exception of the full Textual Content.

Value	Description	Type
Textual Content	The full text of the Telegram messages scraped including internal links and emojis. The messages were then translated into English to allow for universal analysis.	Text
Topic 1	Messages containing words identified using basic topic modelling indicated as a dichotomous value. The group of words has been labelled Nationalism due to their focus on nation, state, and governance.	Boolean
Topic 2	Messages containing words identified using basic topic modelling indicated as a dichotomous value. The group of words has been labelled Military.	Boolean
Topic 3	Messages containing words identified using basic topic modelling indicated as a dichotomous value. The group of words has been labelled Cyber due to their focus on technical terms and other terms associated with known hacktivist groups.	Boolean
Link Present	This is a dichotomous value indicating if there were links (external or other Telegram channels) included in the messages.	Boolean
Photo	The dichotomous value indicating if a photo was embedded in the message.	Boolean

Document	The dichotomous value indicating the presence of a document in the message.	Boolean
Web Page	The dichotomous value indicating the presence of an embedded web page in the message.	Boolean
User Count	Number of channel subscribers/users at the time of the message post. This value will help account for the overall popularity of the channel as indicated by subscribers.	Numeric
Length	Character length of the textual content in the post.	Numeric
Views	This value serves as both an independent and dependent variable. Views are tracked by Telegram by incrementing the count when a device visits the message daily. The counter is not incremented if visited more than once a day, but is incremented if visited another day.	Numeric
Forwards (Dependent)	Forward value is a dependent variable. Forwards represent the relaying of the message to another Telegram user. The value is incremented when a user forwards the message.	Numeric
Reactions (Dependent)	Reaction value is a dependent variable. Reactions represent the engagement of the user manifested through the appearance of an emoji. The value is incremented when a user reacts to the message. A user can respond with multiple reactions; therefore, a user can be represented multiple times in a reaction count.	Numeric

Table 1 Message Features, Associated Descriptions and Value Types

Figures 7-9 illustrate the distribution of the views, reactions, and forwards. Like most online content, the distribution of the reactions and forwards followed a power law distribution. This logarithmic distribution required correction for the regression analysis. Note that an outlier was excluded from Figure 9 (Distribution of Reactions). The outlier was a value over 10x larger than the next highest value (qualitative analysis in the Discussion). This outlier was excluded to facilitate a more precise analysis.

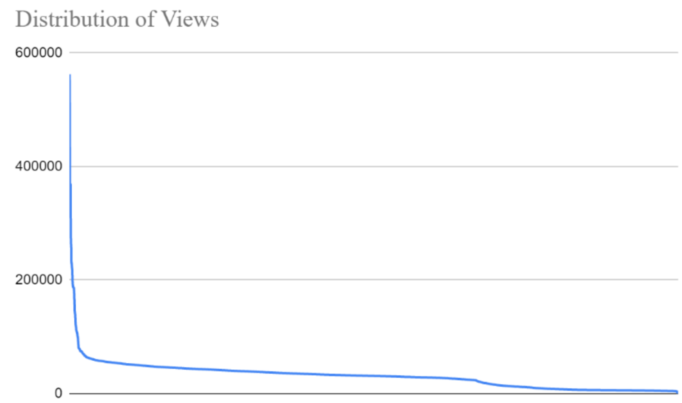

Figure 6 Views Distribution

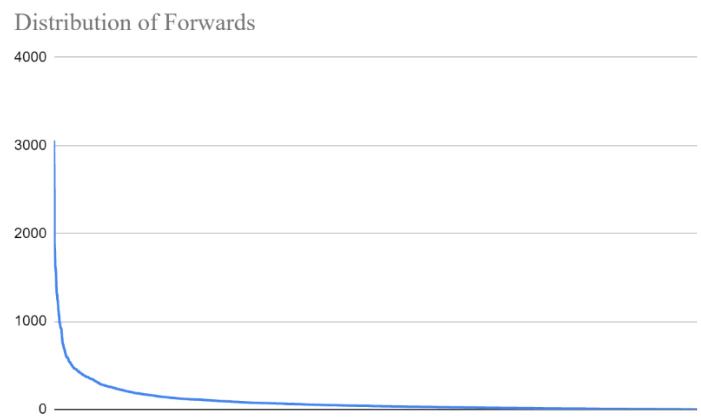

Figure 7 Forwards Distribution

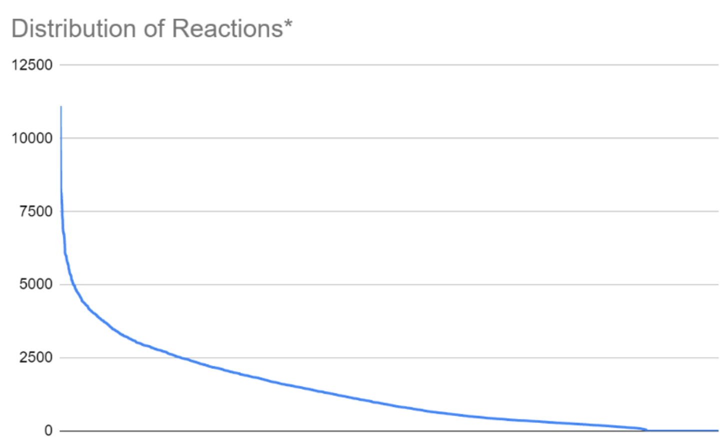

Figure 8 Reactions Distribution

**Note that an outlier was excluded from the distribution*

4.2 Regression Analysis

A regression analysis for each of the three dependent variables (forwards, reactions, and views) determined which independent features had a significant statistical impact. The values for reactions and forwards were logged due to the power law distribution common with online contribution data. The regression model below was used for the three dependent variables. Naturally, the regression model for views excluded the views metric as a feature. The intercepts (one of each of the three models) are shown in Table 2.

$$\begin{aligned} Dependent = \beta_0 + \ \beta_1 Nationalism + \beta_2 Military + \beta_3 Cyber + \beta_4 Link + \beta_5 Photo \\ + \beta_6 Webpage + \beta_7 Document + \beta_8 UserCount + \beta_9 Length \\ + \beta_9 Views + \varepsilon \end{aligned}$$

Dependent Regression Analysis	Intersection Variable
Forwards Regression Model (logged)	1.031
Reactions Regression Model (logged)	4.39
Views Regression Model	-1.003

Table 2 Intersection Variables for Each Regression Analysis

The results of the regression analyses can be found in APPENDIX B. The forwards regression model determined that views, user count, nationalism topic, military topic, length, photo, webpage, and document were all significant features with a p value less than 0.001. Link was also significant, though the p-value was higher with less than 0.01. The topic cyber was the only feature not found to be significantly associated with forwards. The reaction regression model required all features significant with the exception of length with a p value of 0.1. Note that the cyber topic was only significant to a p-value of 0.01. The views regression analysis resulted in fewer significant features. The only feature that was significant with a p-value less than 0.001 was the user count.

The topic cyber also appeared as a potentially significant factor, but much lower with a p-value less than 0.01. The table includes the coefficient estimates for each engagement metric given individual independent factors. The percent difference excluding the log (for reactions and forwards) is indicated in parentheses.

	Views	Reactions - logged	Forwards - logged
Topic Nationalism	-0.045 (-4.5%)	0.151 (16.3%) ***	0.084 (8.74%) ***
Topic Military	-0.165 (-16.5%)	0.069 (7.24%) ***	0.069 (7.2%) ***
Topic Cyber	0.057 (5.7%) *	0.152 (16.4%) ***	0.0732 (7.6%)
Link Present	-0.034 (-3.4%)	0.146 (15.8%) ***	0.039 (4.00%) **
Photo	-0.066 (-6.6%)	-0.290 (-25.14%) ***	0.106 (11.2%) ***
Webpage	0.015 (1.5%)	0.044 (4.54%) ***	-0.048 (-4.64%) ***
Document	-0.056 (-5.6%)	0.044 (4.5%) ***	0.188 (20.7%) ***
User Count	0.775 (77.5%) ***	0.318 (37.4%) ***	0.370 (44.8%) ***
Length ^	-0.045 (-4.5%)	0.148 (16.0%)	0.122 (13.0%) ***
Views	-	0.373 (45.2%) ***	0.467 (59.5%) ***
Adjusted R-Squared	0.2467	0.2491	0.3551

Table 3 Result of the Regression Analysis with Significance Indicators

^ Value increase with the addition of 100 Characters

The coefficient estimate for length is based on the increases for one character. The percentage calculation was determined by scaling it for each 100 characters. In these regression models, the adjusted R-Squared metric indicates the proportion of variance in forwards, views, and reactions that can be explained by the message features. The values produced by the models were all between 0.24 and 0.36, indicating that about a quarter to a third of the variance can be explained by the features in these models.

5. DISCUSSION

5.1 Regression Overview

Views require the least amount of user engagement, because this value is automatically incremented when individuals browse the message. As expected, the primary significant feature that predicted views was user count according to the regression model. Intuitively, the number of users in the channel will influence those actively viewing the posts, likely because of the way in which people use Telegram. For example, notification show new messages and many Telegram users likely read or quickly scan all messages in a channel. This is supported by the high positive relationship between views and users where an increase of one user is associated with a 77.5% increase in views. The views model had the lowest r-squared value which demonstrates the limits of the model to predict views based on past patterns, though approximately a quarter of the variance is explained by the model. As the channel grows and more users frequent the channel, there should be an increased number of views.

The reactions regression analysis produced nine statistically significant features. Notably messages with a topic of nationalism, military, and cyber are 16.3%, 7.2%, and 16.4% more likely to receive a reaction respectively than messages that do not display those themes. These relatively high percentages suggest there is a practically significant increase as well as a statistically significant increase. For a hacktivist channel, these themes are naturally the heart of their ideological messaging. Telegram channels allow people to congregate online around a subject or ideology that elicits emotional or intellectual responses. Reactions provide the opportunity for individuals to display their opinion or feelings surrounding a comment while maintaining a discrete level of anonymity. These topics which could be considered controversial or polarizing provide the substance necessary to provoke a reaction that may be positive or negative. Reactions are also impacted by attachments to the message such as embedded web page, document, or photo. The results of the analysis show that the inclusion of a photo has a -25.14%

impact, which indicates that photos tend to solicit fewer reactions than those without photos. The messages with photos are forwarded more, but have fewer reactions. This could be related to the types of photos that are being shared on the channel. Perhaps these photos are more humorous or can be qualified as memes, at least to those who forward them, but they don't seem to solicit a strong response from the viewers, perhaps because they are off-topic or too fringe. Future work would need to examine the exact reasons.

Reactions require an active response from a viewer. A message was identified with a reaction count more than 10x the value of the next highest value. This outlier represented a message that encouraged users to use reactions to identify their opinion in previously determined categories identified in the message content. This call for engagement from the members of the channel led to a high reaction to view percentage of 57% which is higher than any other message. The nature of the message did not specify the selection of a single reaction; therefore, it is likely that some of those who reacted to it utilized more than one reaction inflating the reaction count to the views.

The forwards model had the highest adjusted R-squared value with 0.3351. Though this value indicates only a modest level of model fitness overall, it is notably the highest out of the three dependent variables. Both forwards and reactions are more likely as the length of the message content increases. Perhaps as the message increases in length, the messages are perceived to be more substantive, leading to more individual action on the part of the viewer. This result contradicts the results of Benjamin and Chen's work, which found that average post length did not appear to be significant contributor to reputation, but rather found that content quality and novelty was more relevant (2012). The length is a continuous variable. An increase in the length of the message by 100 characters is related to an increase of 13% in the forwards. 100 characters represents a significant increase in length. The increase might be related to the perceived merit for a longer message. These messages might be assumed to be more technical or comprehensive therefore instigating engagement from the users. A major feature for increasing the number of forwards was the inclusion of a document. Members are 21% more likely to share through the means of forwarding when an external document or video is included in the message.

Though the R-squared values indicate only a quarter to a third of the variation can be attributed to the features in the regression models, several statistically significant features were identified that incite greater (or less) engagement from subscribers to the hacktivist channel. The topic modeling identified some of the best predictors, indicating the use of content analysis as a viable way to identify important features that predict engagement. In future research, it would benefit the analysis to provide a control channel to recognize if there are differences in activity because of the hacktivist nature of the channel itself. Nevertheless, the features identified provide insights into the components of the messages that increase audience engagement. Future work can examine additional features and see if these preliminary findings hold up with more complete models.

5.2 Telegram Analysis

Based on the nature of Telegram channels, it can be postulated that the views, reactions, and forwards are related. Views are consistently incremented based on users browsing which naturally is correlated with the number of users subscribed to the channel. Views are also a prerequisite for reactions and forwards, which are more active forms of engagement. Forwards have the potential to increase views as a greater number of people gain access to the message than just members of the channel. Reactions like views will be influenced by the number of those with access to the message though the act of reacting to the message will only be performed by a smaller group of people than an automated incrementation of the view count.

The foundation of the methodology used in this study could be expanded to other social network and messaging platforms. Each platform would likely have different features that are inherent in the application to act as independent and dependent features for message analysis.

6. CONCLUSION AND FUTURE WORK

The purpose of this research was to understand the features in Telegram messages that correlate with the most engagement from their audience. The work found that the number of users is by far the best predictor of number of views. Several other factors became critical in predicting more active forms of engagement including reactions and forwards. Interestingly, the inclusion of a photograph made it more likely a message would be forwarded, but less likely it would be reacted to. The existence of all other content types and topic areas were positively associated with reactions and forwards.

6.1 Project Contributions

Understanding the factors that influence engagement in social media channels affects individuals looking to expand their power online and/or connect with their constituents and supporters. For hacktivists, their perceived influence online is a major component of their overall power and success. The contribution of this research is to identify the factors that in fact increase engagement.

This project has identified the following factors as key influences for the increased engagement for a message:

- The views of a message are driven nearly entirely by UserCount, which is the number of channel subscribers at the time of message posting. This result is potentially facilitated by the automated incrementing of the views counter when users browse the content and Telegram push notifications. Some subscribers to the channel likely browse every message in the channel, though how closely they read the messages is not possible to determine.
- The topic of the message matters. Intuitively, messages that heralded an explicit topic, specifically those identified via topic modeling, led to greater engagement compared to messages that did not demonstrate a clear message or did not include textual content. Based on this methodology, the themes are specific to each group, which indicates that they may not be transferable to other groups with significant results. Nevertheless, the isolation of topics and the use of topics as indicators demonstrate a key explanatory power for the administrators of the hacktivist channel. Future research should continue to apply topic modeling to identify channel-specific topics that are related to higher engagement.
- Length leads to more active engagement from the audience through forwards and reactions. In addition to character count, more content generally (links, attachments, etc.) increased the active engagement rate through reactions and forwards.

- The messages with photos are forwarded more often, but have fewer reactions which demonstrates a difference in the active engagement of each of these features.

The results of this study could be used in an explicit attempt to increases engagement on Telegram channels with a similar audience and goal as hacktivist channels.

6.2 Future Work

Further research could consider the growth of the channel over time. A potential research question could be how does the content of the messages affect the growth of the channel categorized by the increase in subscribers? This analysis would take into consideration the relationship between the channel's overall popularity as indicated by the number of its subscribers, and activity, the average number of messages published daily. By using the message dates and channel growth rate as tracked by a platform such as TGStat.[4]

This study addressed the use of Telegram by the hacktivist group, but this approach could be expanded to other social networking or messaging platforms. Telegram presented a unique messaging method through the use of channels, but alternatives could be compared and analyzed in future research. The platforms that are utilized by hacktivists should be prioritized in the study as some platforms include features that are more conducive to hacktivist priorities.

A major limitation to this study is the restricted sample to one Telegram channel associated with a single hacktivist group. In the future, the sample could be expanded to

[4] TGStat is a service that allows companies and researchers understand Telegram channels and groups by tracking posts. https://tgstat.com/

include more than one hacktivist group. A comparative study would allow for continuities and differences to be identified between the results, to determine if there are different techniques for different hacktivist groups. Complementary to increasing the number of hacktivist groups analyzed, a comparison with other Telegram communities could also provide a base line of activity to determine what trends are unique to hacktivists. An expansion of this approach could compare hacktivists with other crime groups that communicate with Telegram channels, but do not qualify as hacktivists. This comparison between activity in the physical and digital world could highlight differences in the communities' priorities and techniques.

The messages used in this study were translated to English using Google Translate to facilitate LDA topic modeling. The translation also allowed for the research team to read the content of the messages. Naturally, translations of message content may not be a perfect parallel due to the nuances of language and limitations of machine translation. Future studies should address this limitation through procedures that maintain the original language and context of the messages when performing topic modeling.

Additionally, a key limitation of this study is the failure to study the sentiment of the messages. The sentiment of a message is an important element in detecting the intention of the message. Consequently, an analysis of the sentiment of each message is a crucial consideration for further study and may increase the overall predictive value of the regression model.

WORK CITED

AppMagic. (September 30, 2022). Number of Telegram app downloads worldwide from January to August 2022, by country (in millions) [Graph]. In Statista. Retrieved March 04, 2023, from https://www-statista-com.byu.idm.oclc.org/statistics/1336855/telegram-downloads-by-country/?locale=en

Benjamin, V., & Chen, H. (2012). Securing cyberspace: Identifying key actors in hacker communities. *2012 IEEE International Conference on Intelligence and Security Informatics*. https://doi.org/10.1109/isi.2012.6283296

Benjamin, V., Zhang, B., Nunamaker, J. F., & Chen, H. (2016). Examining hacker participation length in cybercriminal internet-relay-chat communities. *Journal of Management Information Systems*, *33*(2), 482–510. https://doi.org/10.1080/07421222.2016.1205918

Burgess, M. (2022, December 27). *Hacktivism is back and messier than ever*. Wired. Retrieved March 3, 2023, from https://www.wired.com/story/hacktivism-russia-ukraine-ddos/

Bykov, I. A., Medvedeva, M. V., & Hradziushka, A. A. (2021). Anonymous communication strategy in telegram: Toward comparative analysis of Russia and Belarus. *2021 Communication Strategies in Digital Society Seminar (ComSDS)*. https://doi.org/10.1109/comsds52473.2021.9422858

Djavaherian, B. (2022, September 28-30). The Threat Is Stronger Than the Execution: Realities of Hacktivism in the 2020s. Mandiant. www.virusbulletin.com

Fowler, Kevvie. (2016). Chapter 1 - An Overview of Data Breaches, Editor(s): Kevvie Fowler, Data Breach Preparation and Response, Syngress, Pages 1-26, ISBN 9780128034514, https://doi.org/10.1016/B978-0-12-803451-4.00001-0. (https://www.sciencedirect.com/science/article/pii/B9780128034514000010)

Kiselyova, M. (2014, September 16). *Russia's mail.ru ends vkontakte dispute with $1.5 billion deal*. Reuters. Retrieved March 2, 2023, from https://www.reuters.com/article/us-russia-mail-ru-group-vkontakte/russias-mail-ru-ends-vkontakte-dispute-with-1-5-billion-deal-idINKBN0HB19720140916

Kulshrestha, R. (2019). *A beginner's guide to Latent Dirichlet allocation(lda)*. Medium. Retrieved March 13, 2023, from https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2

Loh, W. (2022). Anonymity, fidelity to law, and digital Civil disobedience. Philosophy & Social Criticism, 0(0). https://doi.org/10.1177/01914537211072886

Loucaides, D. (2023, February 2). *The Kremlin has entered your telegram chat*. Wired. Retrieved March 12, 2023, from https://www.wired.com/story/the-kremlin-has-entered-the-chat/

Sensor Tower. (May 13, 2021). Monthly global installs of Telegram from September 2020 to April 2021 (in millions) [Graph]. In Statista. Retrieved March 03, 2023, from https://www-statista-com.byu.idm.oclc.org/statistics/1239528/monthly-global-installs-of-telegram/

Sutikno, T.; Handayani, L.; Stiawan, D.; Riyadi, M. A.; and Subroto, I. M. I. 2016. Whatsapp, viber and telegram: Which is the best for instant messaging? International Journal ofElectrical & Computer Engineering (2088-8708)6(3).

Telegram Messenger. (November 7, 2022). Number of monthly active Telegram users worldwide from March 2014 to November 2022 (in millions) [Graph]. In Statista. Retrieved March 03, 2023, from https://www-statista-com.byu.idm.oclc.org/statistics/234038/telegram-messenger-mau-users/

What are Signal and Telegram? (2021, January 27). The Economist. Retrieved 2023.

APPENDIX A

The script included below was used to scrape data from Telegram messages.

```
pip install Telethon
pip install pandas
```

Scraper Script

```
from telethon.sync import TelegramClient
import datetime
import pandas as pd

#Excludes the warnings associated with recent function updates.
import warnings
warnings.filterwarnings("ignore")

api_id = '<api_id>'
api_hash = '<api_hash>'

chats = ['<Chat Name>']

client = TelegramClient(None, api_id, api_hash)

df = pd.DataFrame()

for chat in chats:
    with TelegramClient(None, api_id, api_hash) as client:
        for message in client.iter_messages(chat, offset_date=datetime.date.today() ,
reverse=False):
            print(message)
            data = {"Group": chat, "Sender": message.sender_id, "Text": message.text, "Date":
message.date, "Photo": message.photo, "Media" : message.media, "Forwards":
message.forwards, "Views": message.views, "Reactions": message.reactions}

            temp_df = pd.DataFrame(data, index=[1])
            df = df.append(temp_df)

df.to_csv("~/Desktop/<filename>.csv", index=False)
```

APPENDIX B

Thanks to Dr. Justin Giboney and Dr. Derek Hansen for their work analyzing the data collected in this study.

```
Because WebPage and WebPhoto were very highly correlated, we
removed WebPhoto.
We also included Y_Views and UserCount as variables.
This is for data from 4/17/22 forward as we have user count
numbers from that range.

==== Results of Forwards ====

Call:
lm(formula = Y_Forwards.Log ~ Y_Views + UserCount + Link_Present +
    Topic_Nationalism + Topic_Military + Topic_Cyber + Length +
    Photo + WebPage + Document, data =
all_data_w_lt_12k_reactions)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0483 -0.7463  0.0324  0.7909  3.5468

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.031e+00  7.388e-02  13.956  < 2e-16 ***
Y_Views            1.656e-05  7.725e-07  21.433  < 2e-16 ***
UserCount          1.576e-05  8.989e-07  17.537  < 2e-16 ***
Link_Present      -1.235e-01  4.733e-02  -2.610 0.009085 **
Topic_Nationalism  1.986e-01  5.214e-02   3.809 0.000142 ***
Topic_Military     3.558e-01  5.824e-02   6.109 1.10e-09 ***
Topic_Cyber        5.040e-02  5.496e-02   0.917 0.359142
Length             4.824e-04  5.623e-05   8.578  < 2e-16 ***
Photo              9.616e-01  4.529e-02  21.231  < 2e-16 ***
WebPage            4.231e-01  8.198e-02   5.160 2.59e-07 ***
Document           1.295e+00  5.206e-02  24.878  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.137 on 3891 degrees of freedom
Multiple R-squared:  0.3568, Adjusted R-squared:  0.3551
F-statistic: 215.8 on 10 and 3891 DF,  p-value: < 2.2e-16


==== Results of Reactions ====

Call:
lm(formula = Y_Reactions_Count.0sAdded.Log ~ Y_Views + UserCount +
    Link_Present + Topic_Nationalism + Topic_Military +
Topic_Cyber +
    Length + Photo + WebPage + Document, data =
all_data_w_lt_12k_reactions)

Residuals:
```

```
    Min      1Q  Median      3Q     Max 
-8.3385 -0.4668  0.4099  1.1816  3.6415 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)        4.390e+00  1.334e-01  32.914  < 2e-16 ***
Y_Views            1.088e-05  1.395e-06   7.801 7.84e-15 ***
UserCount          2.451e-05  1.623e-06  15.099  < 2e-16 ***
Link_Present       7.688e-01  8.545e-02   8.997  < 2e-16 ***
Topic_Nationalism  5.916e-01  9.413e-02   6.285 3.65e-10 ***
Topic_Military     4.364e-01  1.051e-01   4.151 3.38e-05 ***
Topic_Cyber        3.005e-01  9.923e-02   3.028  0.00248 ** 
Length             1.971e-04  1.015e-04   1.941  0.05227 .  
Photo             -1.652e+00  8.177e-02 -20.201  < 2e-16 ***
WebPage           -1.202e+00  1.480e-01  -8.121 6.13e-16 ***
Document          -5.576e-01  9.399e-02  -5.933 3.24e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.052 on 3891 degrees of freedom
Multiple R-squared:  0.251,  Adjusted R-squared:  0.2491 
F-statistic: 130.4 on 10 and 3891 DF,  p-value: < 2.2e-16



==== Results of Views ====

Call:
lm(formula = Y_Views ~ UserCount + Link_Present + 
Topic_Nationalism + 
    Topic_Military + Topic_Cyber + Length + Photo + WebPage + 
    WebPhoto + Document, data = all_data_w_lt_12k_reactions)

Residuals:
   Min     1Q Median     3Q    Max 
-36772  -9189  -4115   3623 522955 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -1.003e+04  1.525e+03  -6.579 5.38e-11 ***
UserCount          5.618e-01  1.634e-02  34.393  < 2e-16 ***
Link_Present       9.484e+01  9.821e+02   0.097   0.9231    
Topic_Nationalism  7.138e+02  1.082e+03   0.659   0.5096    
Topic_Military    -1.829e+03  1.209e+03  -1.513   0.1303    
Topic_Cyber        2.532e+03  1.140e+03   2.221   0.0264 *  
Length            -3.285e-01  1.167e+00  -0.282   0.7783    
Photo             -1.199e+02  9.399e+02  -0.128   0.8985    
WebPage           -2.105e+03  3.037e+03  -0.693   0.4883    
WebPhoto           1.469e+03  3.257e+03   0.451   0.6519    
Document          -1.440e+03  1.080e+03  -1.333   0.1825    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23590 on 3891 degrees of freedom
Multiple R-squared:  0.2486, Adjusted R-squared:  0.2467 
```

```
F-statistic: 128.7 on 10 and 3891 DF,  p-value: < 2.2e-16
```