Stimulating sustainable behaviour via social media: using text analytics to identify engagement enhancing tweet characteristics for environmental NPOs

MASTER MARKETING

Author: Larissa Elena van Boxem

Student number: 477257

Supervisor: dr. D.S. Kolesnyk

Co-reader: Prof. dr. M.G. de Jong

Abstract

This study uses Twitter data from ten well-known environmental non-profit organizations (NPOs) to reveal what topics and tweet characteristics enhance online engagement. Several tweet characteristics such as the use of hashtags, media, user mentions, sentiment and tweet novelty are derived from the data, and topics are obtained using Latent Dirichlet Allocation (LDA). Ordinary Least Squares (OLS) are used to identify topics and tweet characteristics that enhance online engagement, where likes, retweets, and comments are used as proxies for online engagement. Finally, Markov Chains are used to develop a text generator for highly engaging tweets. The results show that hashtag use and tweet novelty negatively affect engagement while the use of images enhances engagement levels. Engagement enhancing topics are found to be, among others, the fossil fuel industry and the war in Ukraine. The study reveals that Markov Chains can provide a sound basis for generating tweet content ideas but human supervision remains imperative. Furthermore, the results suggest that engagement enhancing tweet characteristics may differ for environmental NPOs given their sensitive nature.

Contents

1 Introduction

The Earth's climate has changed drastically over the past few years, with carbon dioxide levels reaching unprecedented heights (NASA, 2022). This has resulted in rapid changes in global temperatures that have caused events such as warming oceans, shrinking ice sheets, rising sea levels and ocean acidification. These changes and the increasing evidence for climate change have given rise to many environmental non-profit organizations (NPOs) that aim to protect the environment against human forces.

The importance and influence of these environmental NPOs have been emphasized in many studies. One way in which NPOs exert influence is by participating in global politics, where they sometimes play an advisory role or engage in important conversations (Betsill & Corell, 2001; Becker, 2016). However, research shows that the overall trust in the government and these environmental organizations is relatively low (Brewer & Ley, 2013). Bryce (2007) argues that the public's trust in NPOs depends on their (positive) experiences with the organization. Environmental organizations therefore use social media as a means to gain trust and followers, promote sustainable behaviour, spread their messages, and communicate with benefactors. Greenberg & MacAulay (2009) studied the presence and communication of environmental NPOs on social media and found that most organizations are not leveraging the full potential of social media for constituting engagement. It is therefore of great importance to understand how environmental organizations can effectively promote their messages through social media in order to constitute positive experiences and gain trust with the public.

To optimize effective communication, this study aims to identify topics, message structures and message characteristics that have a positive effect on social media engagement. It does so by scraping Twitter data from ten popular, well-established environmental NPOs and applying text analytics to analyze the effect of tweet characteristics and tweet topics on social media engagement. The data also allows us to develop an algorithm for generating engaging tweets that can serve as a basis for new tweet content. To summarize the aim of the study, the following research question is formulated:

How can environmental NPOs enhance online engagement on Twitter based on tweet content and other tweet characteristics?

In the next Section, the existing literature is reviewed to define NPOs, their purposes and operations. Furthermore, research on social media usage, definitions of social media engagement and drivers of social media engagement of environmental NPOs is discussed. In addition, previous research on tweet characteristics that enhance online engagement are highlighted. In Section 3, the data are introduced and some descriptive statistics are presented. This section also zooms in on words and topics in the tweet data by performing word frequency analyses and by identifying neighbouring words. In Section 4, the methods are discussed, with the emphasis being on topic extraction with Latent Dirichlet Allocation (LDA), and how this method can be applied in regression models. This Section also provides some details on how tweet novelty is determined. Furthermore, the text generation algorithm based on maximum likelihood estimation with Markov Chains is explained. The results of the models and algorithms proposed in Section 4 are summarized and discussed in Section 5. Besides the results of the regression models, this section also presents some examples of how Markov Chains can be used as a basis for creating engaging tweet content. Finally, the main conclusions and limitations of the study are outlined in Section 6.

The main purpose of this study is to gain valuable insights into how environmental NPOs can effectively communicate their sustainable messages in order to enhance online engagement and reach a wider audience. This is of great importance given the increasing influence NPOs are given in the global political sphere, and the fact that environmental issues are becoming more and more severe. It is therefore now more important than ever for people to become aware of the possible consequences of their actions and how this can affect pressing issues likes global warming.

This study also carries scientific relevance as no previous study was found utilizing LDA for tweet data in the context of environmental NPOs in order to identify topics that constitute higher engagement. Furthermore, previous studies show contradicting findings and only few concrete and actionable insights are provided, especially in relation to the use of specific words and topics.

2 Theoretical framework

2.1 Defining non-profit organizations

According to Dolnicar & Lazarevski (2009), the non-profit sector lacks a clear definition. Gonzalez et al. (2002) define NPOs as "any organisation without a financial objective, under private control, which aims to generate a social benefit for a specific sector of society". Salamon & Anheier (1992) developed a general definition of the non-profit sector for comparative research. They argue that the most useful definition emphasizes the basic structure and operation of the NPO. In specific, the sector is defined as a collection of organizations that are formal, private, non-profit distributing, self-governing, and voluntary. Formal relates to the organization being institutionalized and showing some degree of institutional reality such as regular meetings and officers. Private means that the organization is institutionally separate from the government, and non-profit distributing means that no profits, not even partially, are assigned to owners or directors. Self-governing relates to the organization being equipped to control their own activities; they are entirely independent and not controlled by any outside entity. The final characteristic of an NPO is that the organization involves some degree of voluntary participation. Salamon & Anheier (1992) emphasize that not all activities or income must stem from voluntary actions; some voluntary input suffices to categorize an organization as voluntary.

Jegers (2008) criticizes in his book "Managerial Economics of Non-Profit Organizations", the definition from Salamon & Anheier (1992). He argues that using the term "organization" already implies its formal, private, and self-governing characteristics. He therefore suggests that only focusing on the non-profit distributing aspect of NPOs is sufficient to capture the essence of these organizations and distinguish them from other types of organizations in the economic landscape.

Jegers (2008) defines environmental NPOs as one of the seven main groups in the classification of non-profits. These NPOs can again be divided into two subgroups: (1) NPOs focusing on the environment and (2) NPOs focusing on animal welfare. This study includes both subgroups of NPOs.

2.2 Social media usage of environmental NPOs

The popularity of social media has given rise to valuable opportunities for non-profits over the past decade. Social media allows NPOs to easily communicate with stakeholders, followers and the media (Lovejoy & Saxton, 2012). Furthermore, social media channels help NPOs to spread their messages to advocate sustainable behaviour. Previous studies, however, show that NPOs are not using social media platforms to their full potential. Tölkes (2018) emphasizes the lack of understanding on how to design effective sustainability messages. Font & Villarino (2015) point out the lack of persuasiveness in sustainability communication. They state that companies do not understand the potential benefits of constructing messages in such a way that it positively affects sustainable behaviour.

As previously mentioned, this study focuses on Twitter data to study how environmental NPOs can effectively communicate and spread their sustainable messages. Twitter is a micro-blogging platform that allows its users to share real-time messages of up to 280 characters. Over the years, the platform has proven its potential for propagating news (Castillo et al., 2013). Twitter is used more frequently for sharing information, and while all social media platforms aim to foster social interactions, Twitter is more effective in doing so due to its simple user interface and its conversational nature (del Mar Galvez-Rodriguez et al., 2016). Guo & Saxton (2014) also emphasize the usefulness of Twitter in the NPO sector, and research on the American Red Cross' use of Twitter and Facebook by Briones et al. (2011) showed that these platforms are very effective for constituting online engagement.

Guo & Saxton (2014) study how NPOs are using social media to engage in advocacy work. They use Twitter data from 188,501 organizations and use content analysis to examine the prevalence of different advocacy tactics. They found that around 50% of all Tweets did not involve any advocacy tactic, indicating that the majority of NPOs are not utilizing social media to its full potential. The results of this research are in line with a study from Lovejoy & Saxton (2012), who also found, using Twitter data, that the largest NPOs are not using Twitter to maximize stakeholder involvement. They instead continue to use Twitter as a one-way communication channel.

2.3 Social media engagement

2.3.1 Defining consumer engagement in the online sphere

Patterson et al. (2006) construct a multi-level definition of consumer engagement based on different aspects of the relationship between a customer and the organization. They define it as "the level of a customer's physical, cognitive and emotional presence in their relationship with an organization". Hollebeek (2011) defines brand engagement as "the level of a customer's motivational, brand-related and context-dependent state of mind characterized by specific levels of cognitive, emotional and behavioral activity in brand interactions". Brodie et al. (2013) define a working definition of consumer engagement in the online brand community. Their definition is based on a virtual brand community with interactive experiences between the organization and members of the brand community. They also argue that consumer engagement is context-dependent, and the intensity of engagement fluctuates. In line with Patterson et al. (2006), they also define consumer engagement as being a multidimensional concept with not only cognitive and emotional, but also behavioural dimensions. While all three definitions slightly differ, they all point out that consumer engagement is built on two aspects: a behavioural and psychological aspect, where the latter fosters the former.

Dolan et al. (2016) point out that social media is one of the most important channels for brands to communicate and engage with their customers and followers. However, the customer experience with the brand through social media channels is simply one touch-point of the overall customer experience. Dolan et al. (2016) adopt the definition from Van Doorn et al. (2010) but with a slight adaptation to conceptualize social media engagement: "Social media engagement behaviours go beyond transactions, and may be specifically defined as a customer's behavioural manifestations that have a *social media* focus [adapted], beyond purchase, resulting from motivational drivers."

In the case of non-profit organizations, consumer engagement is not about behaviours beyond purchase, but rather about behaviours that result solely from motivational drivers because followers relate to the values and practices of the organization. It is not about relationship building that translates itself into increased sales, but rather into increased support and donations. Hou & Lampe (2015) point out that because social media effectiveness cannot be measured through metrices such as return on investment (ROI), it is hard to measure or quantify the effectiveness of social media spending for non-profit organizations.

In this study, the extent to which a tweet generates online interaction in the form of the number of comments, likes, and retweets is used as a proxy for social media engagement.

2.3.2 Drivers and implications of social media engagement

Dessart (2017) aims to understand the drivers of social media engagement and does so by approaching social media engagement from a three-dimensional construct of affective, cognitive and behavioural dimensions. Cognitive engagement refers to the overall mental activity relating to the brand, and affective relates to feelings of joy, enthusiasm and excitement when engaging with the brand. The behavioural dimensions relate to customers sharing, learning, and endorsing brand-related items and messages. Dessart (2017) focuses its research on online brand communities on social media as a form of engagement. The study finds that online interaction propensity, meaning the willingness to communicate with others, attitudes towards community participation and product involvement, are positively related to social media engagement. Of all three predictors, community engagement constitutes the highest brand engagement. This is in line with research by Algesheimer et al. (2005) who emphasize that a brand community fosters stronger brand attitudes.

Dessart (2017) also finds that high social media engagement positively affects brand relationships by stimulating brand loyalty and brand trust. This is especially relevant in the case of non-profit organizations, as brand trust is vital to the success of the organization and the extent to which the organization will reach its goal. Roggeveen & Grewal (2016) point out that online firm activities do not only help to keep the brand top-of-mind, but also have the ability to turn the online audience into active advocates.

Roggeveen & Grewal (2016) identify multiple drivers of social media usage and engagement. The first is the need of people to connect with each other, also called the connected effect. The study argues that "humans always need to be connected to other people, and social media provides them with a new, easy and engaging way to do so". A second driver is convenience, meaning that it is now easier than ever to contact others and seek information. The latter immediately relates to the third driver of social media engagement, namely information seeking. Social media has the ability to provide (relevant) information and participate in conversations. By participating in conversations or posting something that invokes reactions from others, people obtain the feeling they can exert influence.

Roggeveen & Grewal (2016) argue that Twitter allows users to actively participate in highly dynamic conversations with back-and-forth interactions. The study draws a comparison to the tendency that when shoppers spend more time in a store, they buy more. From a marketing-related perspective: when users spend more time on social media interacting with a brand, they buy it more. However, the study also recognizes that still too little organizations adopt this style of communication. Instead, many firms limit themselves to a one-way communication style.

Because of the many benefits of social media engagement, it is of great importance for organizations to understand what drives it. The truth is, many organizations already have a lot of engagement data, but they forget to apply it in order to obtain valuable insights. This is unfortunate, as social media posts contain rich information that allows a company to understand its customers better (Roggeveen & Grewal, 2016).

2.3.3 Tweet characteristics affecting online engagement

This Section zooms in on specific tweet characteristics that in previous studies were found to enhance engagement. Besides general tweet characteristics such as the use of hashtags and media, the focus also lies on textual aspects such as the use of specific topics, words and text sentiment.

Bhattacharya et al. (2014) studied the effect of multiple tweet characteristics on engagement of US Federal Health Agencies, where engagement was measured as the number of retweets. They collected 164,104 tweets from 25 Federal Health Agencies. The findings show that hashtags, URLs and user-mentions are positively related to the number of retweets. No effect was found of tweet sentiment on the number of retweets. Han et al. (2019) also found no effect of tweet sentiment on user engagement. Brady et al. (2017) found that the inclusion of sentiment in tweets, especially negative sentiment, seem to increase in popularity based on the number of retweets. The study found that for polarizing issues such as climate change, the presence of emotional language fosters tweet diffusion. In specific, they found that every additional moral-emotional word enhances tweet distribution by approximately 20%.

Ibrahim et al. (2017) found, using Twitter data from major brands and their customer service accounts, that the presence of a link in tweets improves customer sentiment, mainly because it provides the customer with helpful and relevant information. Bhattacharya et al. (2014) also found a positive effect of link usage on online engagement. Malhotra et al. (2012) found, however, that embedding a link in your tweet does not result in more retweets. Lee (2015) supports this finding, whereas Enge (2014) found a small enhancing effect of link inclusion in tweets. Z. Liu et al. (2012) also found empirical evidence for link inclusion being positively related to online engagement on Twitter. Knowing the effects of link inclusion on tweet engagement is of particular interest since previous research show that the majority of tweets includes a link (Boyd et al., 2010).

Wadhwa et al. (2017) analyzed tweets from the American Journal of Neuroradiology to determine what tweet characteristics are associated with higher engagement rates. They define engagement by the total number of times a user interacted with a tweet. This could be, among others, in the form of link clicks, likes, retweets, follows etc. The study found that the presence of images and hashtags positively affects engagement. Furthermore, the study finds that posting in the morning constitutes higher user engagement than in the afternoon or evening. A study by Li & Xie (2020) also found that including images in social media content has a significant positive effect on engagement. Zor et al. (2021) studied how online engagement on Twitter changes throughout the day for both entertainment (vice content) and scientific news (virtue content). The study finds that in the morning and afternoon, engagement is higher for virtue content, while in the evening the engagement shifts more towards vice content. Given these results, a higher online engagement of tweets posted by environmental NPOs in the morning and afternoon is expected.

DeMasi et al. (2016) focused specifically on the effect of hashtag use on Twitter and how this can drive online engagement. They distinguish between different types of hashtags and find that hashtags that are more relatable to the real world are more engaging. Lahuerta-Otero & Cordero-Gutiérrez (2016) find that influencers on Twitter use more hashtags than non-influencers. According to Enge (2014), hashtags enhance engagement because they organize topics with tags that apply at a global level.

Another factor to take into account is the length of tweets. Enge (2014) found that longer tweets have a more positive effect on online engagement (measured by the number of retweets) than shorter tweets. Lee (2015) also found that longer tweets constitute higher engagement, with the ideal tweet being between 120 to 140 characters (the maximum number of characters in 2015). In 2017, Twitter increased the maximum number of characters for a tweet from 140 to 280. In contrast to Lee (2015), Lahuerta-Otero & Cordero-Gutiérrez (2016) found that the word count of influencers is fewer than those with a smaller social media presence. The study also finds that these influencers are more likely to strongly express their (positive or negative) feelings on the platform. Lahuerta-Otero & Cordero-Gutiérrez (2016) also emphasized user mentions as another possible factor that can constitute engagement on Twitter. This is because when a user is mentioned, he/she receives a notification, can join the conversation, or like and/or retweet the tweet.

Another strategy to foster engagement is by directly asking questions to the public. According to dialogic theory for communication in public relations by Kent & Taylor (2002), one of the principles of creating open dialogues is to ask questions as well as answer questions. Asking questions opens the potential for engagement and brand loyalty in the long run.

Vosoughi et al. (2018) emphasize that tweet novelty can affect the number of people a tweet reaches. They argue that novelty attracts human attention and also encourages information sharing. Novel information is also regarded as more valuable and surprising and therefore constitutes higher engagement. The study finds that online engagement in the form of retweets is higher for tweets that are more novel. The study utilizes three different measures of information distance and averages the result to determine the extent to which a tweet can be regarded as novel. This study will adopt a similar, but slightly simplified version of this method to derived tweet novelty. The methods behind the derivation of tweet novelty are introduced and explained in Section 4.

Fownes et al. (2018) review what is currently known of the discussion on climate change on Twitter. They touch upon a few topics that are frequently discussed in the context of climate change. They state that popular topics include the causes of climate change, for example the fossil fuel or the cattle industry, and the solutions for climate change, for example through politics and global laws. Another popular topic touches upon the effects of climate change, such as global warming, floodings and deforestation. These findings provide important indications of what topics could possibly constitute higher engagement.

Based on the above-mentioned findings of the effects of different tweet characteristics on online engagement, the following effects are expected:

Table 1: Tweet characteristics affecting engagement and expected effects

Variable	Hashtag	Link	Question	Time of day	Characters	Sentiment	Media use	Mentions	Novelty
Expected effect	+	?	+	$\sim$	+	?	+	+	+

With regards to the *Time of the day* variable, a positive effect on engagement is expected for tweets posted in the morning and afternoon, and a negative effect for tweets posted in the evening or night. Media use relates to a tweet that includes either a video or an image. Because of the contradicting findings on the effects of link inclusion in tweets and the fact that many previous studies did not find an effect of tweet sentiment on online engagement, the expected effects of these variables are unknown. For most factors in the Table above, a positive effect on online engagement is expected.

3 Data

This study uses tweet data from the Twitter accounts of ten well-known environmental non-profit organizations. The ten non-profits and corresponding follower counts as of 14 April 2022 can be found in Table 2. Tweet data are scraped with Octoparse, and the data are processed and analyzed in R (version 4.0.3). Since this research aims to study the effectiveness of the tweets generated and posted by the non-profits themselves, retweets are removed from the data, resulting in a total of 1,574 tweets. Because retweets are removed from the data and certain accounts retweet more frequently than others, the number of tweets scraped per non-profit fluctuates quite a bit (see Table 2). The Table shows that the two largest accounts are the World Wide Fund for Nature (WWF), with almost 4 million followers, and Greenpeace with almost 2 million followers. To limit the effects of follower growth over time in the analysis, all tweets are scraped to a maximum of 1 year to the date of scraping.

Table 2: Environmental non-profits, number of scraped tweets, and follower count on 14 April 2022.

Non-profit name	**Number of followers**	**Number of tweets scraped**
WWF (World Wide Fund for Nature)	3.9M	139
Greenpeace	1.9M	93
The Nature Conservancy	1M	179
National Wildlife Federation	602.7K	163
Climate Reality	605.1K	224
Sierra Club	385.6K	118
350 dot org	393K	65
Climate Power	240.1K	170
Friends of the Earth (Action)	228.1K	221
Earthjustice	231.1K	206

The average number of likes, comments and retweets for the ten organizations can be found in Table 3. The Table shows that more followers does not necessarily constitute higher engagement levels. While the two biggest accounts (WWF and Greenpeace) have significantly higher engagement numbers than all other organizations, the smaller accounts such as Earthjustice and Friends of the Earth (Action) seem to have much higher engagement levels than the accounts from Sierra Club and 350 dot org, which have over 100 thousand followers more. This indicates that there could be some engagement enhancing effects at the organizational level that distinguish organizations from each other. For example, smaller organizations could have a stronger and closer brand community than others. Another possibility is that certain organizations tweet, in general, about more popular topics than others. It is therefore important to take these organizational effects into account when predicting online engagement.

Table 3: Environmental NPOs and average number of tweet likes, comments, and retweets

Non-profit name	**Mean likes**	**Mean retweets**	**Mean comments**
WWF (World Wide Fund for Nature)	163.40	77.03	6.19
Greenpeace	156.4	84.13	10.38
The Nature Conservancy	41.38	14.03	0.77
National Wildlife Federation	28.84	9.73	0.66
Climate Reality	79.80	36.55	3.74
Sierra Club	14.04	6.58	1.35
350 dot org	21.74	10.85	0.53
Climate Power	18.34	8.70	0.82
Friends of the Earth (Action)	38.85	22.94	1.79
Earthjustice	79.31	33.15	2.02

Since the Table shows that the biggest accounts have by far the highest engagement levels, controlling for the effects of follower count remains important. This is done by dividing the number of likes/retweets/comments for a specific tweet by the number of followers of the corresponding account, and multiplying the resulting number by 100,000. The last step is done to avoid very small engagement numbers. The results are called *normalized* engagement levels. Based on the distribution of these normalized likes, retweets, and comments in the data, tweets are classified into high, medium and low engagement classes. This allows for further comparison as to what tweet characteristics are frequently associated with high engagement levels. For the classification, the first quartile (the value under which 25% of the data are found) of the distribution is used as the approximate boundary for low engagement tweets, and the third quartile (the value under which 75% of the data are found) as the boundary for high engagement levels. All values between the first and third quartile of the distribution are classified as medium. Since the first quartile for the normalized number of comments was 0, all tweets with a value equal to 0 are classified as low. This variable also contains relatively more low-engagement observations than the other two engagement proxies.

Table 4 shows the number of high, medium, and low engagement tweets for all three engagement proxies. The tweet classification allows us to create a subset of engaging tweets that will serve as the input of a tweet generation algorithm that can write engaging tweets based on one or two words as input. This is done with the help of Markov Chains, which is explained in detail in Section 4. For this subset, a tweet is included if at least two of the three engagement proxy's are classified as high, resulting in 399 tweets being used as input data for the tweet generation model.

Table 4: Number of high, medium and low engagement tweets for all three engagement proxies

Engagement proxy	**High**	**Medium**	**Low**
Likes	393 (25%)	740 (47%)	441 (28%)
Retweets	395 (25%)	770 (49%)	409 (26%)
Comments	477 (30%)	403 (25%)	698 (44%)

From the Twitter data and based on previous research discussed in Section 2.3.3, several characteristics are derived which are used as predictors of engagement. These are: the time of posting (either night, morning, afternoon or evening), the sentiment of the tweet (either positive, negative or neutral), whether the tweet contains a link, an image or a video, whether the tweets contains hashtags, whether the tweet contains a question to the public or a user mention, the character count of the tweet, and tweet novelty. Furthermore, using natural language processing and Latent Dirichlet Allocation (LDA), tweet topics are derived. This method and how tweet novelty is measured, is discussed in detail in Section 4.

The descriptive statistics of all categorical predictor variables can be found in Table 5. The *Time of day* variable is categorized as follows: tweets posted between 00:00 and 06:00 are categorized as *night*, from 06:00 till 12:00 tweets are categorized as *morning*, and from 12:00 till 18:00 as *afternoon*. Finally, all tweets posted between 18:00 and 00:00 are categorized as *evening*.

Sentiment scores are derived using the Bing sentiment lexicon from B. Liu et al. (2005). With this dictionary, words are assigned a score between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment. To categorize tweets as either negative, positive or neutral, the categorization from Ibrahim et al. (2017) is applied. Here, a tweet is categorized as neutral if the sentiment score falls between -1 and 1, and as positive when the score is higher than 1. Consequently, a tweet is categorized as negative if the sentiment score is lower than 1.

Table 5 shows that the great majority of all tweets includes a link (78.6%). Only around 20% of all tweets includes an image, and only 11% a video. The majority of the tweets do not involve the use of hashtags (56.8%) and only 10% of the tweets include a question to the public. Around 30% of the tweets mention another twitter user. Most tweets are posted in the evening (37.9%) and in the afternoon (36.1%). Furthermore, the great majority of the tweets is categorized as neutral (74.5%), and from the non-neutral tweets, the majority is positive (14.9%). The average number of characters in the data is 190.74 with a standard deviation of 63.38. The shortest tweet in the data is only 11 characters whereas the longest tweet is 280 (the maximum number of characters).

Table 5: Descriptive statistics predictor variables

Variable	**N**	**Share (in %)**
Link		
Yes	1237	78.6%
No	337	21.4%
Image		
Yes	308	19.6%
No	1266	80.4%
Video		
Yes	174	11.1%
No	1400	88.9%
Hashtag		
Yes	680	43.2%
No	894	56.8%
Question		
Yes	156	9.9%
No	1418	90.1%
User mention		
Yes	476	30.2%
No	1098	69.8%
Time of day		
Night	246	15.6%
Morning	164	10.4%
Afternoon	567	36.1%
Evening	597	37.9%
Sentiment		
Positive	235	14.9%
Negative	167	10.6%
Neutral	1172	74.5%

Tweet novelty is derived from relative differences between a tweet and the tweet posted before that by a single account. To quantify tweet novelty, the average of the cosine-similarity measure and Kullback-Leibler (KL) divergence is used. How these two measures differ from each other and how they are calculated is explained in Section 4. Table 6 presents the average tweet novelty scores for all ten NPOs. The higher the score, the more novel a tweet can be regarded. This therefore allows us to identify which organizations tweet about more diverse topics.

Table 6 shows that the account with one of the highest average engagement levels, WWF, actually has the lowest relative novelty score. Greenpeace, another organization that scored very high on engagement levels, is ranked fourth when comparing novelty scores. One of the smaller accounts that stood out on engagement scores, Climate Reality, is also low on the list, indicating that tweet novelty may not have such as strong effect on engagement levels as previous research suggests.

Table 6: Mean tweet novelty per organization

Organization	Mean novelty
Friends of the Earth (Action)	3.32
Earthjustice	3.13
Sierra Club	3.00
Greenpeace	2.92
National Wildlife Federation	2.91
350 dot org	2.83
Climate Reality	2.78
Climate Power	2.69
The Nature Conservancy	2.48
WWF	2.46

As mentioned before, three variables are used as a proxies of social media engagement: the number of retweets, the number of comments, and the number of likes a tweet has generated. The descriptive statistics corresponding to these variables can be found in Table 7. The Table shows that in general, tweets are liked more often (mean 69.72), than commented on (mean 2.59) or retweeted (mean 29.09). The maximum number of likes in the data is also significantly higher than the maximum number of retweets and comments.

Table 7: Descriptive statistics engagement variables

Variable	N	Mean	St. Dev	Min	Max
Retweets	1574	29.09	57.84	0	872
Comments	1574	2.59	9.52	0	308
Likes	1574	69.72	195.38	0	5141

3.1 Word frequency analysis

To obtain an idea of the most frequently occurring topics in the tweet data, the words with the highest frequencies are visualized in the word cloud in Figure 1. Here, words are increasing in size with frequency.

The figure shows that, naturally, many tweets involve environmental topics, with words such as "wildlife", "nature", "planet" and "pollution". The Figure also shows that environmental organizations tweet about political topics, which can be derived from words such as "congress", and "Biden". Socio-economic topics occur frequently as well, with words such as "community", "economy", and "people".

As mentioned before, tweets are categorized as either positive, negative, or neutral based on the sentiment score chart from Ibrahim et al. (2017). With this categorization, word frequencies for positive and negative tweets can be visualized with comparison clouds. Comparison clouds visualize words that are most discriminating for a certain category. They are based on relative frequencies and words with the highest difference in relative frequency across categories are visualized larger than words with smaller differences in relative frequency. The comparison cloud can be found in Figure 2. Here, green words correspond to tweets with positive sentiment, and red words to tweets with negative sentiment.

Figure 1: Most frequent words in tweet data.

Figure 2 shows that positive tweets are more about the protection of the environment, the conservation of wildlife, and terms such as supporting and helping, while negative tweets are often about crises, threats and pollution. While they may touch upon the same topics, it is evident that the same concerns are formulated in completely different ways. In the next section, a simple exploratory analysis is performed to get an idea of how, among other things, tweet sentiment is related to the number of likes, retweets and comments.

Figure 2: Most discriminating words for positive and negative tweets.

Similar Figures can be made for tweets that are classified in the high and low engagement category. For this Figure, tweets are categorized as high (low) if at least two of the three engagement measures is classified as high (low). In Figure 3 the most discriminating words for high (in green) and low (in red) engagement tweets can be seen. The Figure shows that some topics associated with high engagement levels are about politics (with words such as "administration", "biden", and "federal"), endangered species ("wolves", "extinction", "amazon") and the fossil fuel industry ("fuels", "fossil", "emissions" etc.). In the low engagement category, the topics are less evident and more general words such as "nature", "world", "join" and "learn" can be seen. For example, where the high engagement tweets are about extinction in specific, the low engagement tweets are simply about wildlife and nature.

Figure 3: Most discriminating words for high and low engagement tweets.

3.2 Exploratory analysis

In order to obtain an idea of the expected effect of tweet characteristics on online engagement, the average number of likes, retweets and comments for different levels of the predictor variables are analyzed. As previously mentioned, given that the number of likes, retweets and comments are partly influenced by follower counts, the three columns to the right show the results when taking the normalized engagement levels. The results are summarized in Table 8. The Table shows that the average engagement numbers change significantly when controlling for the number of followers. The three columns on the left suggest that tweets without a link correspond to higher engagement levels than tweets including a link. When controlling for the number of followers, the results show that the difference between tweets with and without a link becomes much smaller. Very small differences in average engagement numbers are also found for the variable *Image*, when controlling for the number of followers. Based on the three right columns, the results suggest that including a video in a tweet, using hashtags, and asking questions negatively affects engagement. The same result is found for the variable *User mention*. The results also suggest that, when controlling for follower count, tweets posted during the night correspond to the highest engagement numbers. Finally, the results also suggest that tweets with negative sentiment constitute higher engagement.

Table 8: Average number of likes, retweets and comments for different predictor variables

Variable	Likes	Retweets	Comments	Likes norm.	Retweets norm.	Comments norm.
Link						
Yes	60.76	26.89	2.52	12.18	5.68	0.45
No	102.60	37.18	2.86	13.66	5.78	0.48
Image						
Yes	94.77	34.31	2.71	13.32	5.65	0.45
No	63.63	27.83	2.56	12.30	5.71	0.46
Video						
Yes	119.56	48.00	5.55	9.79	4.26	0.49
No	63.53	26.74	2.22	12.83	5.88	0.45
Hashtag						
Yes	73.73	29.73	2.69	6.91	3.01	0.24
No	66.67	28.61	2.51	16.74	7.74	0.62
Question						
Yes	62.61	27.06	2.43	8.83	4.11	0.33
No	70.50	29.32	2.61	12.90	5.87	0.47
User mention						
Yes	39.08	17.19	1.91	9.27	4.45	0.36
No	83.00	34.25	2.89	13.90	6.24	0.49
Time of day						
Night	73.13	32.22	2.89	17.39	8.16	0.60
Morning	120.35	52.74	4.08	7.26	3.48	0.23
Afternoon	76.44	28.43	2.79	12.01	5.14	0.43
Evening	48.03	21.94	1.87	12.38	5.82	0.49
Sentiment						
Positive	80.00	25.73	2.58	10.35	3.79	0.40
Negative	75.10	37.47	3.08	18.30	9.75	0.74
Neutral	66.90	28.57	2.52	12.10	5.51	0.43

To identify tweet topics, words are visualized with multidimensional scaling (MDS) in Figure 4. In the MDS map in the Figure, words that are in close proximity of each other, appear together more often in the tweet data. This allows us to identify topics that NPOs tweet about. Neighbouring words that possibly present certain topics are grouped together in the circles in the Figure. The Figure shows 7 different groups of words that could possibly indicate a tweet topic. On the bottom left the words "youth", "movement", and "earthday" indicate that environmental organizations tweet about Earth Day and the events surrounding this day, especially for young people who join the movement against climate change. Going upwards, a topic related to politics in relation to pollution can be found. Right from this topic a more general topic about protecting the environment is found, with words such as "action", "planet", and "protecting". In the middle of the Figure, two smaller topics are identified: the left one is about communities and possibly about their access to clean water, and the right one is about the conservation of wildlife. On the far right, a topic about endangered species is identified, with many animal names appearing in this topic. Finally, on the far left, a big topic about the fossil fuel industry and the war in Ukraine is identified.

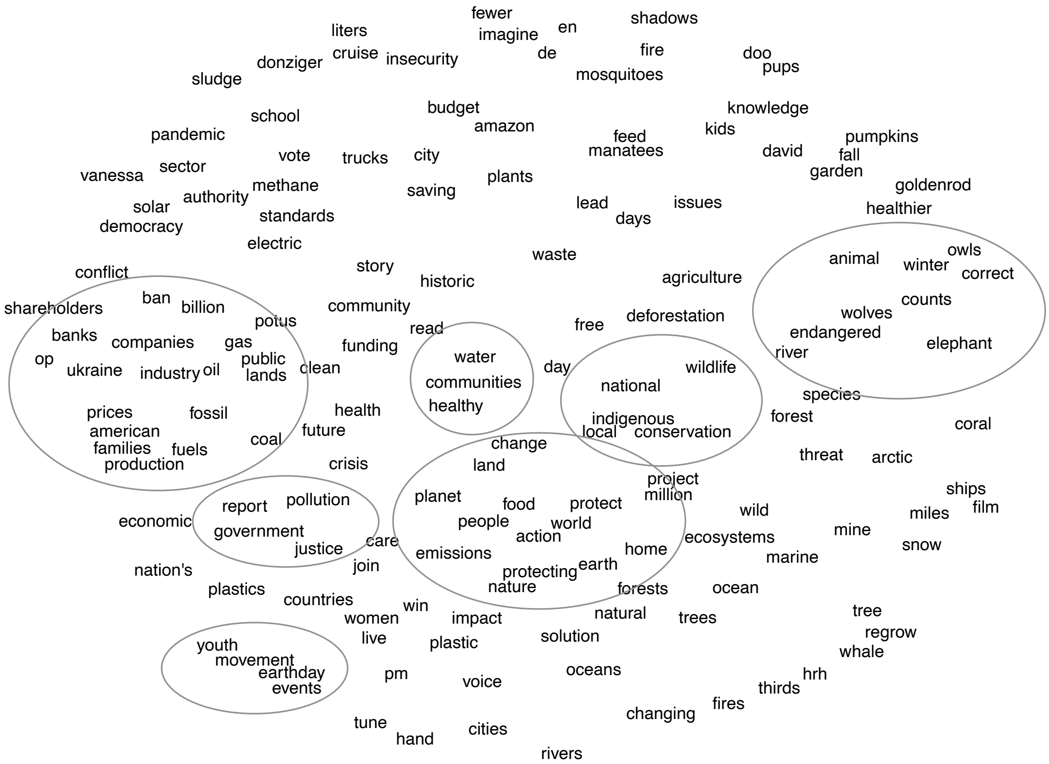

Figure 4: Neighbouring words in tweet data (after text cleaning).

4 Methodology

4.1 Latent Dirichlet Allocation (LDA)

LDA is a popular method for topic extraction. With LDA, a single text or document corresponds to multiple topics based on topic probabilities. All words belong to all topics, but the words are given different probabilities or weights for different topics. Topics are therefore characterized by a certain distribution of words (Blei et al., 2003). For example, the word "wildlife" may be given a weight of 40% in topic 3 and only a weight of 1% in topic 1. Because all topics contain all words it is important to only consider the words with the heighest weights for each topic (Kumar et al., 2014).

LDA assumes the following process for a single document:

1. Decide on the number of words
2. Obtain topic probabilities
3. For each word in the document:
   (a) Choose a specific topic
   (b) Choose a term from the topic
4. Collect words to obtain full document

It is important to distinguish between terms and words. A term can correspond to multiple topics whereas a single word in a sentence cannot. The final output of the model is a dirichlet distribution that generates a vector of numbers that are all between 0 and 1, and add up to 1. For example, assume that K denotes the number of topics and β_k the term distribution for topic k. θ_n presents the multinomial topic distribution for document n. Now assume that $K = 2$, where topic 1 is about university life and topic 2 about spare time, and we have the terms *lecture*, *school*, *party*, and *friends*. If $\beta_1 = (0.4, 0.4, 0.01, 0.19)$, $\beta_2 = (0.01, 0.01, 0.59, 0.39)$, and $\theta_n = (0.9, 0.1)$, we can conclude that this specific document is for 90% about university life and for 10% about spare time. We can also see that topic 1 (about university life) has higher term probabilities for the terms *lecture* and *school*, and topic 2 (about spare time) weighs highest on the terms *party* and *friends*.

A single document can thus cover many topics. The number of topics a single document corresponds to is determined by α, the prior distribution of θ_n where $\theta_n \sim Dirichlet(\alpha, ..., \alpha_n)$. α covers what you expect of the distribution of θ_n. There are as many α values as elements in the vector. α also introduces a variance because the assignment of topics will not always be the same. If we denote the expected probability of a certain topic k as:

$$E[p_k] = \frac{\alpha_k}{\sum_j \alpha_j},$$

then the variance of p_k can be denoted as:

$$Var[p_k] = \frac{E[p_k](1 - E[p_k])}{1 + \sum_j \alpha_j}.$$

The formula above shows that when the sum of alpha's is large, the variance decreases. This determines how many topics correspond to a certain document. In practice, a smaller α_k is preferred, since this decreases the variance and documents are only about few topics.

For estimating LDA, the number of topics K needs to be fixed. A common method to estimate the optimal number of topics is to use the perplexity measure. The goal here is density estimation where a high likelihood on a held-out, validation set is desired. The perplexity measure is decreasing in the likelihood of the validation set, therefore the lowest perplexity measure indicates the optimal number of topics (Blei et al., 2003). The perplexity measure can also be used to determine the optimal value of alpha. The optimal value of alpha is the value that minimizes perplexity on the validation set. The model that results from the optimal alpha and number of topics allows for evaluating the most prominent topic per document and the topic probability corresponding to that document. The latter is important because the highest topic probability can vary significantly across documents. These topics and corresponding probabilities can be used in further analyses to evaluate their influence on online engagement. How this will be implemented is specified in the next subsection.

4.2 Model specification

4.2.1 Ordinary Least Squares

This study focuses on three proxies for engagement: likes, retweets, and comments. Therefore, the model proposed below will be applied three times, resulting in three different regression models. To control for the effect that certain Twitter accounts have more followers than others, the measurements of engagement are divided by the number of followers as specified in Section 3. Ordinary least squares (OLS) will be used to predict engagement and identify the main predictors of engagement.

The following methodology follows from Weisberg (2005): A simple OLS model consists of a mean function,

$$Y = \beta_0 + \beta_1 x,$$

and the variance function,

$$Var(Y) = \sigma^2.$$

Here, β_0 reflects the value Y when x is equal to zero and β_1 indicates the rate of change in Y when x changes by one unit. Because the variance of Y is $\sigma^2 > 0$, the observed value of Y will typically not be equal to the expected value of Y. To account for this difference, an error term e is added to the equation,

$$Y = \beta_0 + \beta_1 x + e_i.$$

This error term reflects unknown, and thus unobserved, parameters that affect Y but are independent of x. The error term differs between individuals/case and we assume that the error terms between individuals are independent as well. Furthermore, the errors are assumed to be normally distributed.

The OLS estimators β_0 and β_1 are the values for which the residual sum of squares (RSS) are minimized:

$$RSS(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2.$$

The variance σ^2 is essentially the average of the squared size of the error term. The estimator of σ^2 can therefore be obtained by averaging the squared residuals, thus dividing RSS $= \sum \hat{e}_i^2$ by its degrees of freedom. The degrees of freedom of the error term are equal to number of observations minus the number of parameters in the mean function Weisberg (2005). Therefore, if the parameters in the mean function are equal to 2, the estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{RSS}{n-2}.$$

This function is also called the residual mean square. When taking the square root of σ^2, we obtain the standard error of the regression. This standard error is in the same unit as the response variable.

In the multiple linear regression model, the mean function depends on more than one variable. Important is that the variable that is added to the equation explains a part of Y that is not already explained by the other variable(s). The general multiple linear regression model with response variable Y and predictor variables $X_1, ..., X_z$ takes on the following form:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_z X_z.$$

4.2.2 Deriving tweet novelty

As mentioned in Section 2, Vosoughi et al. (2018) found that tweet novelty constitutes higher engagement since new information invokes feelings of surprise and is regarded as more valuable. Tweet novelty is derived from the topic distributions and is calculated as the information distance between a certain tweet and the tweet posted before that by the same account. Vosoughi et al. (2018) average three measures of information distance to calculate tweet novelty. This study utilizes two of these to derive tweet novelty: Kullback-Leibler (KL) divergence and cosine similarity. In line with Vosoughi et al. (2018), the formula that is used to measure information uniqueness (IU) with cosine similarity is

$$IU(z_{pt}, z_{y(t-1)}) = 1 - \cos(z_{pt}, z_{y(t-1)}),$$

where z_{pt} is the topic distribution of tweet p in time t and $z_{y(t-1)}$ is the topic distribution of tweet y in time $t-1$. A higher value of IU corresponds to the content of tweet p being more unique compared to the content of tweet y.

The second measure of tweet novelty is KL divergence, which measures how one probability distribution diverges from a second probability distribution (Vosoughi et al., 2018). The KL divergence is also known as the relative entropy between two probability density functions (Hershey & Olsen, 2007). The KL divergence is calculated as follows:

$$D(f||g) = \int f(x) log \frac{f(xt)}{g(x(t-1))} dx.$$

Here, $f(x)$ is the probability density function in time t and $g(x)$ the probability density function in time $t-1$. The KL divergence satisfies self-similarity, meaning that $D(f||f) = 0$, and $D(f||g) = 0$ only if $f = g$. The measure also satisfies positivity, meaning that $D(f||g) \leq 0$ (Hershey & Olsen, 2007). A lower KL divergence indicates smaller differences between the documents, meaning that a higher KL divergence indicates that the tweet is more novel compared to the previous one.

As mentioned before, to obtain the final measure of novelty, the two distance measures introduced above are averaged and the result is used as a predictor of online engagement.

4.2.3 Model implementation

This study utilizes the multiple linear regression model with several predictors of engagement, among which all of the tweet characteristics mentioned in Section 3. Organizational fixed effects are included to control for organizational effects that stay constant over time, such as certain environmental organizations having stronger brand communities. In doing so, fixed effects control for any observed and unobserved differences across organizations. The study also controls for the effect of the number of likes, comments and retweets of the previous tweet by including three lag variables. By including the main topic and corresponding topic probability, we can evaluate what topics constitute higher engagement, and whether this engagement differs between higher/lower topic probabilities. To do so, an interaction effect between the main topic and the corresponding topic probability will be added in a second model. The first model therefore looks as follows:

$$\begin{aligned}
&Engagement_{it} = \beta_0 + \beta_1 * Organization_i + \beta_2 * Hashtag_{it} + \beta_3 * Question_{it} + \beta_4 * Video_{it} + \beta_5 * Image_{it} \\
&+ \beta_6 * Link_{it} + \beta_7 * Mention_{it} + \beta_8 * Sentiment_{it} + \beta_9 * \textit{Time of day}_{it} + \beta_{10} * \textit{Character count}_{it} \\
&+ \beta_{11} * \textit{Tweet novelty}_{it} + \beta_{12} * Likes_{j,t-1} + \beta_{13} * Comments_{j,t-1} + \beta_{14} * Retweets_{j,t-1} \\
&+ \beta_{15} * \textit{Main topic}_{it} + \beta_{16} * \textit{Topic probability}_{it},
\end{aligned}$$

where *Engagement* is the number of normalized likes, retweets, or comments of tweet i in time t. The three lag variables relate to the normalized number of likes, comments and retweets of tweet j in time *t-1*. The second model including the interaction effect between the main topic and the corresponding topic probability looks as follows:

$$\begin{aligned}
&Engagement_{it} = \beta_0 + \beta_1 * Organization_i + \beta_2 * Hashtag_{it} + \beta_3 * Question_{it} + \beta_4 * Video_{it} + \beta_5 * Image_{it} \\
&+ \beta_6 * Link_{it} + \beta_7 * Mention_{it} + \beta_8 * Sentiment_{it} + \beta_9 * \textit{Time of day}_{it} + \beta_{10} * \textit{Character count}_{it} \\
&+ \beta_{11} * \textit{Tweet novelty}_{it} + \beta_{12} * Likes_{j,t-1} + \beta_{13} * Comments_{j,t-1} + \beta_{14} * Retweets_{j,t-1} \\
&+ \beta_{15} * \textit{Main topic}_{it} + \beta_{16} * \textit{Topic probability}_{it} + \beta_{17} * \textit{Main topic}_{it} * \textit{Topic probability}_{it}.
\end{aligned}$$

In the models above, the main topic is an unordered factor variable, which is internally coded as a dummy variable by R. This means that when the number of topics generated by LDA is equal to 10, the model includes 10 - 1 beta's. Furthermore, the variables *Hashthag*, *Question*, *Video*, *Image*, *Link* and *Mention* are dummy variables, taking on the value 1 if the corresponding characteristic is present, and 0 otherwise. As previously specified in the data section, *Sentiment* is a variable that can take on three levels; positive, negative, and neutral. *Time of day* can take on four levels; morning, afternoon, evening and night.

Given that three different proxies for engagement are used (likes, retweets, and comments), a total of six models are created. To evaluate the models, the data is divided into a train and test set of respectively 70% and 30% of the data and the performance on the out-of-sample data (test set) will be evaluated.

4.2.4 Text generation with Discrete Time Markov Chains

The following methodology follows from Spedicato et al. (2016): A discrete time Markov Chain (DTMC) is a sequence of random variables characterized by the Markov property, which states that the probability of moving to a future state X_{t+1} is only dependent on the current state X_t. The state space X_n exists of a set of possible states $S = s_1, s_2, s_n$ that can be finite. The state space can be anything: weather conditions, stock performances, and in this study, letters or words (Maltby et al., 2022). The chain moves between states and the probability of moving from state s_z to state s_i is denoted as p_{zi}. A DTMC is time-homogeneous when the transition probability does not change over time, meaning that the probability of moving from state s_z to s_i always remains p_{zi}. The distribution of the transition probabilities from one state to another can be presented into a transition probability matrix P.

To illustrate, imagine a situation in which there are two states: s_1 and s_2. The probability of moving from state s_1 to state s_2 is $\alpha = 0.25$ and the probability of moving from s_2 to s_1 is $\beta = 0.95$. This information can be summarised in a transition diagram:

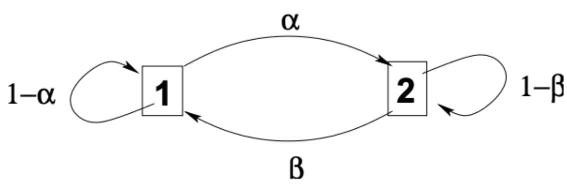

Figure 5: Transition diagram (Konstantopoulos, 2009).

The transition probability matrix P will then look as follows:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} 0.75 & 0.25 \\ 0.95 & 0.05 \end{pmatrix}$$

It can be seen that the rows in the transition matrix sum to one. Furthermore, all entries are always non-negative. Figure 5 is also called a random walk, and it is how every Markov Chain can be illustrated. The arrows show in which directions the "walker" can move.

When it comes to using Markov Chains for text generation, the Markov Chain is build as such that every word in the corpus is connected to every other word in the corpus with corresponding transition probabilities Pernicano (2021). For example, if the initial word is *the*, the Markov Chain assigns a probability to every other word in the corpus that shows how likely it is to follow the initial word. The quality of the Markov Chain is therefore very dependent on its input data. For example, when taking a famous novel as input data to randomly generate a new story, the sentences are likely to make much more sense than when using Whatsapp text data as input. For a novel, the probability of the word *the* being followed by itself or another word that makes little sense is very small and likely close to 0%, while for Whatsapp data, this may be much higher given the frequent occurrence of spelling and formulation errors. That is one of the main limitations of using Markov Chains in the context of text generation.

As mentioned in Section 3, tweets are classified into low, medium and high engagement categories based on the distribution of the normalized number of likes, retweets and comments. For the Markov Chain input data, at least two of the three engagement proxies must be classified as high. Two Markov Chain text generators are then developed with one having unigrams (a single word) as input, and the other having bigrams (a pair of words) as input. The output of the text generators can provide a basis for successful tweet content and also allow us to gain more insights into what words and topics are associated with high engagement levels.

5 Results

5.1 Exploratory analysis

To evaluate the effects of the previously specified tweet characteristics on online engagement, a set of t-tests and ANOVA (Analysis of Variance) tests are used to determine whether differences in engagement levels for different types of tweets are statistically significant. The t-test is used to compare the sample means of two groups (for example, tweets with and without a link), while ANOVA is used to compare the means of three or more groups (for example, tweets with neutral, positive, and negative sentiment). The t-test is therefore used for the variables *Hashtag*, *Question*, *Video*, *Image*, *Link* and *Mention*, while ANOVA is used for the variables *Sentiment* and *Time of day*. The normalized engagement proxies are used to identify significant differences. The results can be found in Table 9.

Table 9: T-test and ANOVA results

Variable	Hashtag	Question	Link	Video	Image	Mention	Sentiment	Time of day
Likes	0.000***	0.019**	0.373	0.039**	0.541	0.000***	0.005***	0.001***
Retweets	0.000***	0.049**	0.052	0.006***	0.936	0.004***	0.000***	0.001***
Comments	0.000***	0.145	0.467	0.779	0.911	0.051*	0.021**	0.065*

Note: *p<0.1; **p<0.05; ***p<0.01.

The results show that including one or more hashtags in a tweet significantly affects the mean of all three engagement measures, compared to tweets without hashtags. The same result is found for the variable *Mention*, meaning that including user mentions is found to affect the average number of likes, retweets, and comments. The results also shows that the average engagement levels differ for tweets with different sentiment scores (either positive, negative or neutral). The final variable that is found to significantly affect all three engagement proxies is *Time of day*, indicating that the timing of your tweet can either positively or negatively influence your engagement numbers. Table 9 shows that including a video only significantly affects the average number of likes, and retweets, but not the number of comments. Including a link or an image is not found to have a significant effect on engagement for all three engagement measures.

5.2 LDA

In this Section, the LDA results are presented and discussed. The topics resulting from this analysis will be, as previously mentioned, used as predictor variables for an OLS model of which the results are presented in Section 5.3. As outlined in Section 4, the first step of LDA is to determine the number of topics K. This is done with the perplexity measure that was introduced in Section 4.1. The optimal number of topics is the number at which the perplexity measure is minimized. This number is found to be 22 for the data. Second, the optimal value of alpha (α) needs to be determined. Given $K = 22$, the optimal value of α is found to be 0.1. Running the model with the optimal values K and α generates a topic distribution per term as well as a topic distribution per tweet. The ten words with the highest weights for all 22 topics are presented in Table 10. All topics are given a name based on these ten words.

Table 10: Topics and corresponding top ten words with highest topic probabilities

(a)

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**	**Topic 11**
join	potus	deforestation	people	pollution	coal	justice	healthier	oil	nature	ocean
earth	act	power	world	communities	read	women	local	fossil	action	plastic
day	congress	free	future	health	project	history	amazing	gas	cop	pollution
hear	build	action	planet	water	learn	environmental	chance	fuel	world	people
love	bold	plastics	nature	protect	plants	month	week	industry	leaders	stop
solutions	pass	dollarvalue	earth	trucks	monarch	jackson	birds	dollarvalue	climateaction	oceans
video	families	crisis	protect	reduce	power	court	authority	ceos	naturenow	marine
calling	action	petition	leaders	air	bees	judge	activists	companies	global	global
we're	crisis	tackle	join	plant	air	whm	totalenergies	prices	water	million
week	clean	helped	healthy	electric	fired	nwf	stopeacop	crisis	naturepositive	save
Earth day	Politics	?	Earth's future	Communities	?	WHM	Crude oil	Fossil fuels	Nature	Ocean pollution

(b)

Topic 12	**Topic 13**	**Topic 14**	**Topic 15**	**Topic 16**	**Topic 17**	**Topic 18**	**Topic 19**	**Topic 20**	**Topic 21**	**Topic 22**
emissions	food	people	clean	change	public	indigenous	communities	report	wildlife	species
time	recent	live	fossil	impact	protect	community	pipeline	food	national	wildlife
gas	read	greenhour	fuels	happening	communities	environmental	continue	ipcc	answer	protect
methane	news	wildlife	future	fires	forests	communities	stop	global	largest	wolves
greenhouse	students	kids	can't	life	administration	black	valley	solutions	correct	protection
planet	major	landscape	economy	story	lands	people	mosquitoes	sustainable	winter	endangered
action	ban	trees	time	warming	biden	conservation	local	crisis	world	conservation
agriculture	oil	understand	independence	drought	waters	production	gmo	impacts	message	habitat
warming	conflict	benefits	real	floods	ecosystem	national	environmental	planet	park	salmon
bold	land	powered	transition	storms	carbon	blackhistoymonth	cycle	growpositive	garden	river
Global warming	?	?	Energy transition	Climate change	Politics	Black history month	Pipeline	?	Wildlife park	Endangered species

Table 10 shows that while some topics have a very clear theme, others contain many different words that do not seem to have a clear relation to each other. These topics with many different words are therefore not given a global topic. It can also be seen that certain topics have many overlapping themes. For example, topics 2 and 17 both relate to politics and political practices in relation to the environment. Topics 8, 9, 12 and 15 are all about fossil fuels, global warming and the energy transition. There are also many general topics that contain words such as "action", "change", "impact", and "protect". These are for example topics 3, 4, 10, and 16. The final frequently occurring topic is the topic about wildlife and endangered species, which are topics 21 and 22.

Given that the aim of the study is to provide actionable insights on how to stimulate online engagement, the number of topics are reduced to 10. The optimal number of topics according to the perplexity measure is thus ignored to be able to provide results that are easier to interpret and yield more valuable insights. The ten topics and their global themes can be found in Table 11.

Table 11: Final topics and global themes

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
justice	wildlife	clean	food	nature	oil	species	fossil	oil	emissions
environmental	protect	act	people	world	project	wildlife	fuel	gas	water
women	environmental	future	plastic	action	drilling	learn	oil	billion	air
read	wolves	congress	ocean	cop	fuel	national	gas	war	communities
black	conservation	build	global	people	salmon	protect	dollar	prices	pollution
community	forests	action	life	join	pipeline	extinction	industry	dollarvalue	health
month	protection	economy	planet	future	crisis	change	stop	ukraine	clean
history	trees	senate	pollution	planet	activists	day	power	pump	coal
jackson	restore	change	waste	leaders	stop	birds	crisis	companies	trucks
court	species	biden	sealife	critical	threat	sea	ipcc	profits	methane
Social change	Protect forests	Politics	Ocean pollution	Induce action	Oil drilling	Protect wildlife	Fossil fuel industry	War in Ukraine	Pollution

Table 11 shows that compared to the previous scenario with $K = 22$, the topics are now much more distinct. While there still remains some overlap in words (for example in topics 2 and 7), the main theme is clearly different. In the next Section, these topics are used to analyze how they affect engagement levels.

To be able to accurately capture the effect of a certain topic on online engagement, a balanced distribution of topics is needed. This distribution is visualized in Figure 6. The Figure shows that topics 3 and 5 are the most frequently occurring topics in the data set. The least occurring topic is topic 9, with only 100 tweets having this as their main topic. This figure also indicates that with the 22 topics the perplexity measure suggested, the number of tweets corresponding to a specific topic would be very low, which would have limited the power of the analysis significantly.

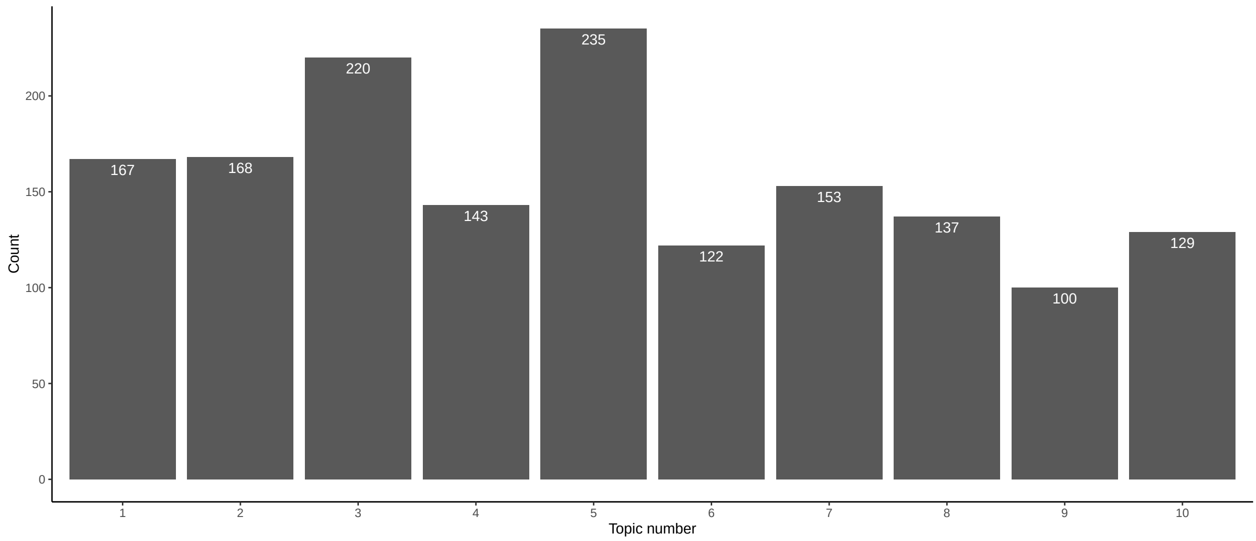

Figure 6: Topic distribution.

For every tweet, the main topic and the corresponding topic probability is derived. The distribution of these topic probabilities are visualized in Figure 7. The Figure shows that most tweets have high maximum topic probabilities (0.9 or higher), while very few tweets have low topic probabilities (0.4 or lower). Some clear peaks are visible around the topic probabilities of 0.45, 0.85 and 0.95. The latter corresponds to approximately 90 tweets having a very high maximum topic probability. The relatively small number of tweets with low maximum topic probabilities indicates that only focusing on the main topic of the tweet may already provide valuable insights into what topics stimulate online engagement, since the majority of the tweets relate mainly to one topic.

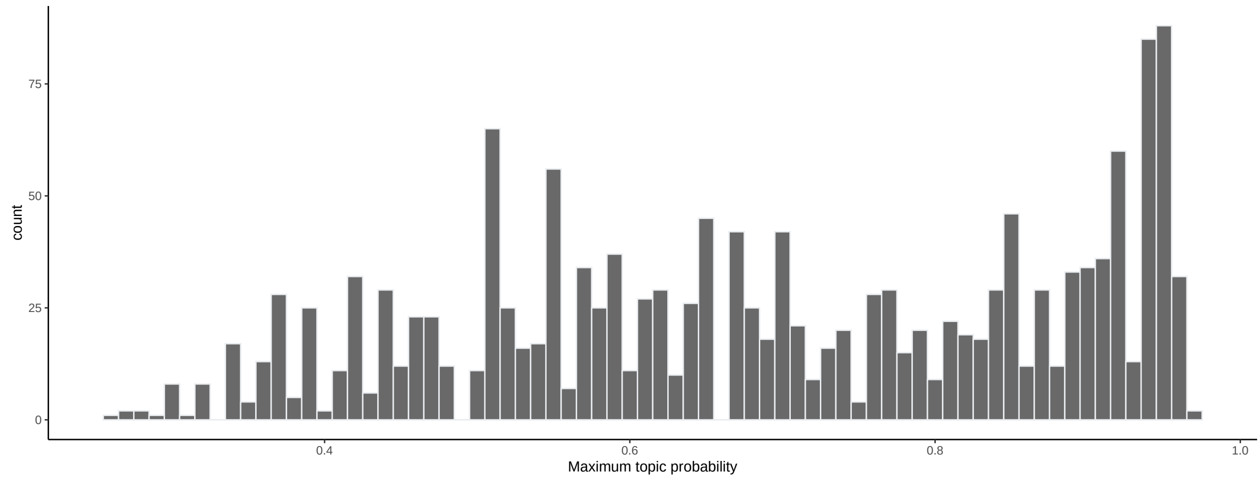

Figure 7: Distribution of maximum topic probabilities.

5.3 OLS

This section presents the results of the OLS analysis to predict online engagement and quantify the effects of different tweet characteristics. A total of six models are tested and evaluated. Three models for every measure of engagement, and three additional models including an interaction effect between the main topic and the corresponding maximum topic probability. To account for the effect that some accounts have more followers than others, the engagement proxies are divided by the number of followers in the same way as described in Section 3. In addition, all models include organizational fixed effects to control for factors that differ between NPOs and affect online engagement levels, such as strong brand communities or brand advertisements.

Table 12 shows that many variables are not found to have a statistically significant effect on the number of likes, retweets and comments. In contrast to previous findings, including a link, video or asking a question in the tweet is not found to enhance online engagement. The time of posting and character count is also not found to significantly affect engagement. The results do show that including an image positively affects the number of likes of a tweet. In contrast to what previous studies found, including a hashtag is found to negatively affect the number of likes, retweets and comments. When focusing on tweet sentiment, the results show that a tweet with neutral or positive sentiment decreases engagement compared to a tweet with negative segment, which is in line with findings from Brady et al. (2017).

The previous number of likes and retweets is found to significantly affect the number of likes, but not the number of retweets and comments. In contrast to previous studies, user mentions are not found to positively affect online engagement: a statistically significant negative effect on two of the three engagement proxies is found.

Table 12: OLS results

	Dependent variable:		
	Likes	Retweets	Comments
	(1)	(2)	(3)
Link	-2.279	-0.660	-0.001
	(2.186)	(1.067)	(0.131)
Video	3.221	1.187	0.081
	(2.823)	(1.377)	(0.170)
Image	3.980^{*}	1.599	0.159
	(2.121)	(1.035)	(0.127)
Hashtag	-5.921^{***}	-2.643^{***}	-0.291^{***}
	(1.761)	(0.859)	(0.106)
Question	0.564	0.395	0.075
	(2.476)	(1.208)	(0.149)
Time of day: morning	-1.479	-0.047	0.021
	(3.216)	(1.569)	(0.193)
Time of day: afternoon	-1.539	-0.392	0.058
	(2.420)	(1.181)	(0.145)
Time of day: Evening	-2.000	-0.802	0.046
	(2.336)	(1.140)	(0.140)
Likes_{t-1}	0.202^{**}	0.064	0.0001
	(0.098)	(0.048)	(0.006)
Comments_{t-1}	0.508	0.502	0.061
	(0.822)	(0.401)	(0.049)
Retweets_{t-1}	-0.355^{*}	-0.127	-0.003
	(0.202)	(0.098)	(0.012)
Mention	-3.241^{*}	-1.426^{*}	-0.174
	(1.755)	(0.856)	(0.105)
Sentiment: neutral	-2.646	-2.234^{*}	-0.186
	(2.451)	(1.196)	(0.147)
Sentiment: positive	-5.406^{*}	-4.099^{***}	-0.141
	(3.048)	(1.487)	(0.183)
Topic 2	5.855^{*}	2.009	0.093
	(3.323)	(1.621)	(0.200)
Topic 3	2.853	1.806	0.313
	(3.193)	(1.558)	(0.192)
Topic 4	6.290^{*}	2.991^{*}	0.230
	(3.346)	(1.632)	(0.201)
Topic 5	3.871	1.912	0.149
	(3.276)	(1.598)	(0.197)
Topic 6	2.873	0.811	0.007
	(3.579)	(1.746)	(0.215)
Topic 7	4.000	2.271	0.035
	(3.588)	(1.750)	(0.216)
Topic 8	10.351^{***}	5.124^{***}	0.684^{***}
	(3.494)	(1.705)	(0.210)
Topic 9	7.474^{**}	5.828^{***}	0.546^{**}
	(3.803)	(1.855)	(0.229)
Topic 10	1.739	0.703	0.066
	(3.535)	(1.724)	(0.212)
Topic probability	5.119	2.012	0.094
	(4.104)	(2.002)	(0.247)
Character	0.008	0.005	0.00003
	(0.014)	(0.007)	(0.001)
Novelty	-1.014	-0.494	-0.070^{*}
	(0.663)	(0.324)	(0.040)
Constant	7.602	4.136	0.241
	(6.752)	(3.294)	(0.406)
Fixed effects	Yes	Yes	Yes
Observations	1,092	1,092	1,092
R^2	0.208	0.188	0.084
Adjusted R^2	0.181	0.161	0.054
Residual Std. Error (df = 1056)	23.835	11.628	1.432
F Statistic (df = 35; 1056)	7.904^{***}	6.980^{***}	2.779^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01		

When looking at the topics that were derived from the data, the results show that topics 8 and 9 positively affect online engagement, compared to the topic 1. Table 11 showed that topic 8 is about the negative effects of the fossil fuel industry, and topic 9 is about the war between Ukraine and Russia and how this affects the economy. The topic on marine pollution (topic 4) is also found to positively affect the number of likes and retweets. The topic on the protection and conservation of the forests and its species (topic 2), is only found to positively affect the number likes. No significant effect of the maximum topic probability and character count on online engagement is found.

Tweet novelty was introduced as a predictor variable since previous research found that tweet novelty can enhance engagement. The study argued that novel tweets are often regarded as more valuable and therefore encourage information sharing (Vosoughi et al., 2018). A positive effect of tweet novelty on engagement was therefore expected. Table 12, however, shows a negative effect of tweet novelty on online engagement. While this effect is only statistically significant for the number of comments, it suggests that tweet novelty may either not play such an important role in enhancing engagement for environmental NPOs, or tweet novelty is less appreciated by the public for these specific organizations.

To summarize, where previous studies found a significant effect of link inclusion, video content, the inclusion of questions, character count, and the time of posting, this study does not. For the latter, it should be noted however, that time differences were not taken into account and it could very well be that the demographics of the majority of the followers of the accounts correspond to different time zones. This would make it difficult to accurately capture the effect of the time of posting. This study also finds negative effects of the use of hashtags, user mentions, and tweet novelty on online engagement, whereas previous studies found these factors to positively influence engagement.

Table 13 shows the results of the second model that includes the interaction effect between the different topics and the maximum topic probability. The results show that when including these interaction effects, the main effects of the topics that were previously found to be statistically significant, become insignificant. The model shows, however, that the interaction effects for topics 9 and 10 are found to be statistically significant when the outcome variable is the number of retweets. This means that the effect of these topics differ for different topic probabilities. In other words, how topics 9 and 10 affect engagement is conditional on the height of their maximum topic probability.

It should be noted that the models with the same dependent variables have the exact same R^2 score. This means that adding the interaction terms does not allow us to explain more of the variation in the dependent variable. The two models with the outcome variable *Likes* have the highest R^2 of 0.208. The models with the number of retweets and comments as output variable have an R^2 of respectively 0.188 and 0.084, meaning that the variables in the model can only explain 18.8% and 8.4% of the variation in the dependent variable.

Table 13: OLS results with interaction effects*

	Dependent variable:		
	Likes	Retweets	Comments
	(1)	(2)	(3)
Link	-2.195	-0.567	0.016
	(2.208)	(1.091)	(0.130)
Video	3.521	0.082	0.001
	(2.836)	(1.404)	(0.168)
Image	3.941*	0.940	0.112
	(2.134)	(1.067)	(0.127)
Hashtag	-5.888***	-3.816***	-0.341***
	(1.774)	(0.849)	(0.101)
Question	0.550	-0.426	0.029
	(2.491)	(1.252)	(0.150)
Time of day: morning	-1.745	-1.818	-0.117
	(3.232)	(1.547)	(0.185)
Time of day: afternoon	-1.656	-1.548	-0.062
	(2.437)	(1.182)	(0.141)
Time of day: evening	-2.040	-1.614	-0.012
	(2.354)	(1.172)	(0.140)
Likes_{t-1}	0.204**	0.145***	0.004
	(0.099)	(0.049)	(0.006)
Comments_{t-1}	0.529	-0.071	0.028
	(0.832)	(0.417)	(0.050)
Retweets_{t-1}	-0.369*	-0.147	-0.002
	(0.203)	(0.101)	(0.012)
Mention	-3.422*	-2.207**	-0.231**
	(1.769)	(0.875)	(0.105)
Sentiment: neutral	-2.866	-2.705**	-0.227
	(2.466)	(1.247)	(0.149)
Sentiment: positive	-5.311*	-4.294***	-0.157
	(3.067)	(1.547)	(0.185)
Topic 2	-5.970	-3.876	-0.239
	(11.414)	(5.768)	(0.689)
Topic 3	-5.457	-0.028	0.167
	(10.529)	(5.310)	(0.634)
Topic 4	-11.313	-4.841	-0.328
	(11.418)	(5.784)	(0.691)
Topic 5	-5.847	-1.727	-0.332
	(10.958)	(5.532)	(0.661)
Topic 6	1.470	0.524	-0.413
	(12.530)	(6.335)	(0.757)
Topic 7	-4.336	-1.565	-0.209
	(12.684)	(6.424)	(0.768)
Topic 8	6.682	2.769	0.502
	(12.644)	(6.334)	(0.757)
Topic 9	-10.478	-8.625	-0.915
	(16.165)	(8.190)	(0.978)
Topic 10	-18.668	-9.560	-0.590
	(13.048)	(6.604)	(0.789)
Topic probability	-8.897	-3.012	-0.344
	(11.312)	(5.701)	(0.681)
Character	0.011	0.009	0.0005
	(0.014)	(0.007)	(0.001)
Novelty	-1.083	-0.296	-0.050
	(0.671)	(0.339)	(0.040)

* *Note:* *Table continued on the next page.*

	Dependent variable:		
	Likes	Retweets	Comments
	(1)	(2)	(3)
Topic 2 * topic probability	18.047	10.464	0.556
	(16.402)	(8.280)	(0.989)
Topic 3 * topic probability	12.868	0.923	0.025
	(15.307)	(7.664)	(0.916)
Topic 4 * topic probability	26.684	11.925	0.834
	(16.434)	(8.310)	(0.993)
Topic 5 * topic probability	14.886	3.556	0.523
	(15.779)	(7.885)	(0.942)
Topic 6 * topic probability	1.918	0.633	0.640
	(18.831)	(9.531)	(1.139)
Topic 7 * topic probability	12.868	4.850	0.303
	(17.504)	(8.843)	(1.056)
Topic 8 * topic probability	6.537	5.420	0.312
	(17.615)	(8.884)	(1.061)
Topic 9 * topic probability	25.647	20.190^{*}	1.939
	(21.162)	(10.724)	(1.281)
Topic 10 * topic probability	30.585	17.704^{*}	1.122
	(18.719)	(9.465)	(1.131)
Constant	16.519^{*}	10.190^{**}	0.918^{*}
	(9.631)	(4.420)	(0.528)
Fixed effects	Yes	Yes	Yes
Observations	1,092	1,092	1,092
R^2	0.208	0.188	0.084
Adjusted R^2	0.181	0.161	0.054
Residual Std. Error (df = 1056)	23.835	11.628	1.432
F Statistic (df = 35; 1056)	7.904^{***}	6.980^{***}	2.779^{***}
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

It should be noted that the results in Table 12 and 13 differ quite from the results presented in Section 5.1, where many variables, such as *Video* and *Question*, did show a significant effect on (at least one of) the outcome variables. A possible explanation for this is multicollinearity among predictor variables. This assumption is tested with the Variance Inflation Factor (VIF) and correlations. All VIF levels are below 4 and only a medium-high correlation of approximately -0.35 between the variables *Link* and *Video* was found. All other correlations were below 0.3. This indicates that there could be some multicollinearity issues in the models but the effect should be relatively limited.

The results presented in Tables 12 and 13 are based on the training data (70%). All models are applied to the test data to evaluate the prediction power of the models. This is done with the Root Mean Squared Error (RMSE). The larger the RMSE, the poorer the fit and predictive power of the model. The RMSE values for all models on the test set can be found in Table 14.

Table 14 shows that the models with the number of comments as outcome variable correspond to the lowest RMSE. This indicates that these models have the highest predictive power on out-of-sample data. However, it should be noted that the average number of comments is also significantly lower than the number of likes and retweets in the data. A lower RMSE is therefore expected. For all models, the RMSE is relatively high, indicating low predictive power for the out-of-sample data.

Table 14: RMSE values for all OLS models

Model	Outcome variable	RMSE
Model 1	Likes, no interaction	22.67
Model 2	Retweets, no interaction	11.02
Model 3	Comments, no interaction	1.13
Model 4	Likes, interaction	22.74
Model 5	Retweets, interaction	11.49
Model 6	Comments, interaction	1.15

5.4 Tweet generation with Markov Chains

For the text generators, a Markov Chain is fitted using maximum likelihood estimation (MLE). The text generator requires several inputs: the number of sentences to generate, the number of words the generated sentences will comprise of, and the word(s) the sentences will begin with. As mentioned in Section 4, two text generators are developed: one which requires a single word (unigram) as input, and one which requires two words (bigram) as input. For the Markov Chain to be able to fit the probability distribution, it is important that the input word that is specified also exists in the data. Therefore, the data set used for the Markov Chain is slightly different from the data set used for the LDA analysis. The Markov Chain input data does not omit stop words such as ”the” and ”we”, as these words are critical for sentence generation.

As stated before, tweets are classified into low, high and medium engagement classes based on the distribution of the normalized number of likes, retweets, and comments. Tweets that are part of the input data for the text generators are tweets of which at least two of the three engagement proxies are classified as high. Figure 8 shows the output that follows from the text generator developed based on unigrams. The input word that is specified is “The”, the number of sentences to generate was set at five and the number of words for each sentence was set at fifteen (resulting in sixteen words in total).

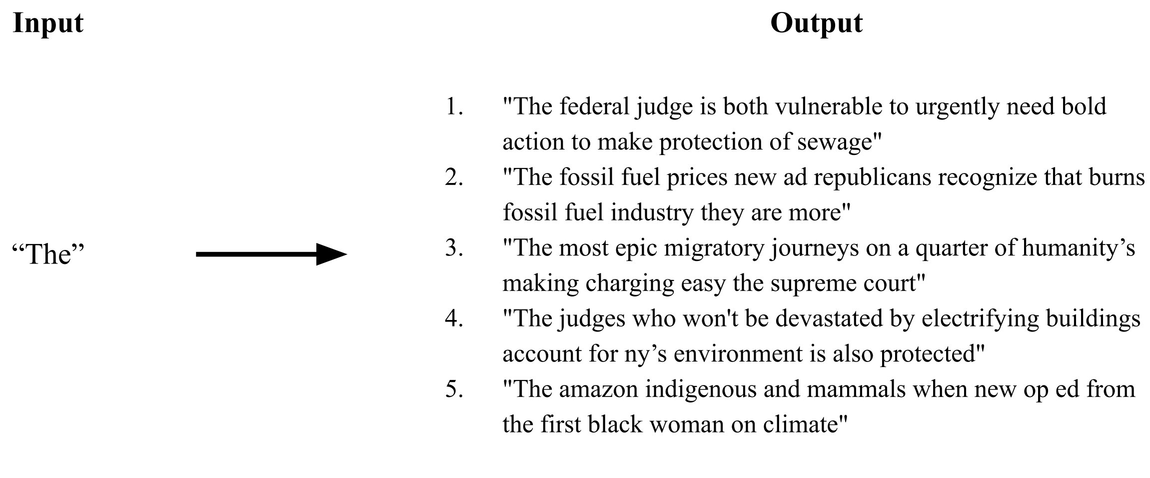

Figure 8: Example of generated text with a single input word.

Figure 8 shows that the text generator has trouble generating well-formulated sentences. It should therefore not be used as an underlying model to automatically generate tweets, but more as a source of inspiration for identifying new tweet topics and words that can enhance engagement. When it comes to adjectives for example, the text generator clearly shows that engaging tweets use strong adjectives such as “urgently”, “epic”, and “bold”. Some topics that come across are fossil fuel prices, politics and cultural topics (“the first black woman” for example).

Figure 9 shows an example of the output of the text generator for bigrams, where the input words are “we are”. It can be seen that four of the five sentences start with “We are calling”, indicating that the bigram “we are”, is very frequently followed by either the bigram “calling on” or “calling for”. The Figure also shows that sentences 1 and 4 are relatively accurate in the beginning but then become less accurate at the end. It can also be seen that sentences are never really completely finished. The algorithm is unable to detect how a sentence is finished given that it is solely based on the likelihood of one word following another, and not on the sentence as a whole. Therefore, as mentioned before, this algorithm cannot be used to automate the tweet content generation process without human supervision.

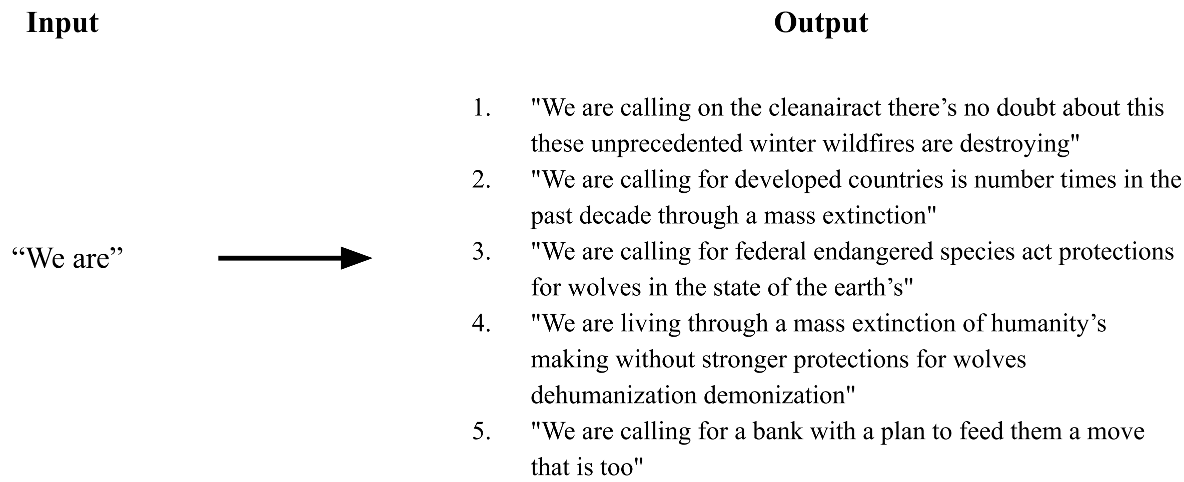

Figure 9: Example of generated text with two input words.

While text generation with Markov Chains can provide a sound basis for identifying engaging tweets and generating fresh content ideas, it is evident that the algorithm in itself is flawed and incomplete. Weak input generates weak output and when the input data is limited, many similar sentences will be generated. However, when the input data comprises of many tweets with high engagement levels that also touch upon the right topics, this algorithm could be a very valuable source for content ideas for environmental NPOs.

6 Conclusions and discussion

6.1 Main findings and implications

This study aims to identify tweet characteristics and topics that enhance online engagement for environmental NPOs on Twitter. It does so by focusing on the following research question: *How can environmental NPOs enhance online engagement on Twitter based on tweet content and other tweet characteristics?*

In order to answer the research question, LDA is used to identify and extract topics, and OLS are used to measure the effect of these topics and other tweet characteristics on online engagement. In addition, a text generator based on Markov Chains is developed to generate tweets and identify specific words and sentence structures that are associated with highly engaging tweets. Online engagement is measured with the use of three proxies: the number of retweets, likes and comments. Tweets are classified into high, low and medium engagement classes based on the distribution of the normalized engagement proxies. A tweet is used as input data for the text generator if at least two of the three engagement proxies is classified as high.

To obtain more insights into the expected effect of different tweet characteristics, t-tests and ANOVA tests were performed, allowing us to see whether the number of retweets, likes and comments differ between tweets with and without these characteristics. The results showed that the inclusion of hashtags, questions, videos and user mentions significantly affects at least one of the engagement proxies. Furthermore, tweet sentiment and the time of the day at which the tweet was posted was also found to significantly affect online engagement.

In the second part of the analysis, LDA was used to extract topics from the tweet data and identify the main topic and the corresponding maximum topic probabilities for every tweet. While the optimal number of topics was initially found to be 22, the number of topics was reduced to 10 to allow for better interpretable results and more valuable insights. Including these main topics and their maximum topic probabilities as predictor variables in the OLS models allowed us to see what topics correspond to enhanced online engagement. Two types of models were tested for every predictor of online engagement: one with an interaction effect between the main topic and the corresponding maximum topic probability, and one without an interaction effect between these two. The OLS models without interaction effects showed a positive and significant effect of four different topics on most of the engagement proxies. These four topics are about (1), the fossil fuel industry, (2) the war in Ukraine, (3) the protection of forests and (4) the pollution of the ocean.

Contrary to findings of Bhattacharya et al. (2014), a negative effect of hashtag use and user mentions was found. This indicates that environmental NPOs should refrain from using hashtags and tagging other people in their tweets. A possible explanation for this is that tweets with many hashtags could be regarded as 'too much' or overwhelming. Since the hashtag dummy only indicates whether or not at least one hashtag is present, this assumption cannot be tested but should be tested in future research. Furthermore, a tweet of which the content is weak, remains weak even when hashtags are being used. A third possible explanation is that hashtags could enhance engagement, but only if the hashtags are relevant for the tweet content. Controlling for the relevancy of hashtags could perhaps have shown a different effect than what the results show now.

The results also showed a negative effect of tweet novelty on online engagement, contradicting the study of Vosoughi et al. (2018). A possible explanation for this is that environmental NPOs are niche organizations and even between environmental NPOs, there are fundamental differences. For example, WWF is an environmental NPO that works in the field of wilderness preservation, and Climate Reality is an organization that focuses mainly on educating people on climate change and advocating solutions against the climate crisis. It could therefore be that novel tweets, in the sense that these tweets fall outside their niche, can be regarded as unusual and not fitting with the brand. Based on this study, we would therefore advise organizations to refrain from tweeting novel content that falls out their usual scope of topics.

Unlike studies from Ibrahim et al. (2017), Wadhwa et al. (2017), Zor et al. (2021) and Kent & Taylor (2002), no significant effect of link inclusion, video content, the inclusion of questions and the time of posting was found. In line with Brady et al. (2017), the results showed that tweets with negative sentiment are associated with higher engagement levels than tweets with neutral and positive sentiment. Brady et al. (2017) emphasized that especially for polarizing issues such as climate change, the presence of emotional language fosters tweet sharing. This emotional language for the topic of climate change was mainly anger, the study finds. Based on the results of this study, environmental organizations would therefore benefit from using strong, negative, or angry statements in their tweets to enhance online engagement.

In the final part of the results, a text generator based on Markov Chains was developed. The results showed that although the output was imperfect, the generator could still serve as a sound basis to identify topics, words, and sentence structures that are associated with highly engaging tweets. The output of the generator, for example, showed that strong adjectives such as "epic" and "urgently" are associated with higher engagement levels and that certain sentence structures such as "we are calling for", could provide a strong basis for new tweet content.

The study also showed that many previous findings on engagement enhancing tweet characteristics hold, but only before controlling for the number of followers. Section 3 showed that the average number of likes, retweets, and comments for tweets with different tweet characteristics changes significantly when dividing by the number of followers, and that many major differences are either canceled out or switched around. For example, the average number of likes for tweets including a hashtag prior to controlling for follower count was much higher than of those without a hashtag. When using the normalized likes for comparison, the results turn around: tweets without hashtags seem to have much higher engagement numbers. This shows that follower count is a very important determinant of online engagement.

It should be noted that the findings of this study cannot be generalized to other types of organizations but mainly provide implications for environmental non-profit organizations. It could very well be that the use of hashtags and video content can boost engagement for commercial organizations who are simply looking to promote their products. Because environmental organizations operate in a socially and politically sensitive environment, certain marketing techniques may not be as efficient as for commercial organizations. This may also be the case for commercial companies selling eco-products, as they need to convey a completely different message to the public than environmental NPOs. Where environmental organizations need to make the public aware of the pressing issues surrounding climate change and what we need to do to stop it from becoming worse, companies selling eco-products need to make the public aware of how their product can benefit the environment as well as the customer. Where the former is set in a more negative context, the latter is set in a positive context. Eco-product companies should convey positive messages to induce purchase behaviour whereas environmental organizations focus on conveying negative messages to make people change their behaviour.

The study thus reveals that tweet characteristics and content matter, but more research is needed given the contradicting findings on this topic, and the fact that there is still very little research on what basic tweet characteristics enhance engagement for environmental organizations. This study does, however, show that certain topics result in higher engagement levels than others, and that "trending" topics such as the current events between Ukraine and Russia and its impact on our economy induce tweet interaction. The study also implies that topics related to endangered species and animals constitute higher engagement, possibly because these topics invoke more direct feelings of anger and sadness, compared to hearing about more general events, such as rising sea levels and higher temperatures. For sustainability marketers, it could therefore be a good strategy to focus their tweet content more on emotional topics and topics surrounding popular events that have important implications for the environment.

6.2 Limitations

This study also has some important limitations that should be taken into account when interpreting the results. First of all, it should be noted that the model is incomplete and can only explain little variation in the dependent variables. While a low R^2 is problematic for prediction purposes, it makes sense in the context of social media. The success of a social media post involves much more than just text, hashtags and whether a photo was included or not. It is about the quality, colours, and context of all those factors, but also momentum. For example, if Greenpeace would introduce another major protest action, resulting in big news headlines, this is very likely to drive their online engagement.

Another limitation relates to the data. As discussed before, no effect of the time of tweeting on engagement was found. This could be the result of time differences that could not be taken into account because the demographics of the followers of the accounts are unknown. Another important limitation is omitted variable bias. There are several factors that could affect online engagement which the model does not control for, such as post frequency. Due to software limitations and retweets being removed from the data, controlling for post frequency was not possible. Another important limitation is that geographical factors were not taken into account. Fownes et al. (2018) found that tweets on climate change are much more popular in some countries than others. Furthermore, the content of those tweets varies between locations as well. Even within countries, significant differences can be found.

Another limitation that results from software limitations is the fact that not all posts from all ten accounts were scraped from the exact same time frame. While the differences in time frames are relatively small (no differences larger than 1 year), this could still have affected the results. However, it can be assumed that major events related to climate change enhanced engagement across all environmental organizations during that period of time, therefore, differences in tweet characteristics would still be able to explain differences in engagement levels.

Another limitation is the fact that even within the study, contradicting results were found. While t-tests and ANOVA tests showed that many variables did have a significant (or insignificant) effect on the number of likes, retweets, and comments, the OLS yielded different results. A possible explanation for this is that when control variables are added to the regression, other variables lose power. On top of that, there could also be multicollinearity issues as mentioned previously. However, this problem is assumed to be limited as multiple measures showed that the correlations between variables were relatively low.

As mentioned before, the results are also contradicting to previous studies. A possible explanation is that engagement enhancing tweets characteristics are different for environmental non-profit organizations, individuals or other, commercially oriented, organizations. Environmental organizations could be more prone to criticism since they are becoming more and more institutionalized (Berny & Rootes, 2018). Furthermore, given that environmental organizations are often in the middle of political, sensitive debates, their strategies to obtain followers and gain the public's trust through social media may be completely different.

For future research, it would therefore be interesting to zoom in more on the text part of the tweets to see what specific words or n-grams can drive engagement, as this may be a much more important driver of engagement for environmental NPOs than general tweet aspects such as hashtags and user mentions. The latter should, of course, still be taken into account as these are important controls that can possibly explain part of the variance in engagement measures, but since the topics that environmental NPOs tweet about are often of a more sensitive nature, choosing the right words is of great importance.

Another recommendation for future research is to take into account image and video quality, such as the topic of the media item and its colours. Previous research has shown that colours play a critical role in influencing emotions (Yu & Egger, 2021). Li & Xie (2020) found that high quality images and images that are professionally shot consistently lead to higher online engagement. Therefore, to get a more complete model of factors influencing online engagement, these are definitely factors that should be included in future research. Finally, as mentioned before, future research should take into account post frequency, geographical differences and time effects when data is scraped over a long period of time.

References

Algesheimer, R., Dholakia, U. M., & Herrmann, A. (2005). The social influence of brand community: Evidence from european car clubs. *Journal of marketing*, *69*(3), 19–34.

Becker, E. (2016). The influence of environmental ngos in the global society. *Butler Journal of Undergraduate Research*, *2*(1), 17.

Berny, N., & Rootes, C. (2018). *Environmental ngos at a crossroads?* (Vol. 27) (No. 6). Taylor & Francis.

Betsill, M. M., & Corell, E. (2001). Ngo influence in international environmental negotiations: a framework for analysis. *Global environmental politics*, *1*(4), 65–85.

Bhattacharya, S., Srinivasan, P., & Polgreen, P. (2014). Engagement with health agencies on twitter. *PLoS One*, *9*(11), e112235.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *2010 43rd hawaii international conference on system sciences* (pp. 1–10).

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313–7318.

Brewer, P. R., & Ley, B. L. (2013). Whose science do you believe? explaining trust in sources of scientific information about the environment. *Science Communication*, *35*(1), 115–137.

Briones, R. L., Kuch, B., Liu, B. F., & Jin, Y. (2011). Keeping up with the digital age: How the american red cross uses social media to build relationships. *Public relations review*, *37*(1), 37–43.

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of business research*, *66*(1), 105–114.

Bryce, H. J. (2007). The public's trust in nonprofit organizations: The role of relationship marketing and management. *California Management Review*, *49*(4), 112–131.

Castillo, C., Mendoza, M., & Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research*.

del Mar Galvez-Rodriguez, M., Caba-Pérez, C., & López-Godoy, M. (2016). Drivers of twitter as a strategic communication tool for non-profit organizations. *Internet Research*.

DeMasi, O., Mason, D., & Ma, J. (2016). Understanding communities via hashtag engagement: A clustering based approach. In *Tenth international aaai conference on web and social media*.

Dessart, L. (2017). Social media engagement: a model of antecedents and relational outcomes. *Journal of Marketing Management*, *33*(5-6), 375–399.

Dolan, R., Conduit, J., Fahy, J., & Goodman, S. (2016). Social media engagement behaviour: a uses and gratifications perspective. *Journal of strategic marketing*, *24*(3-4), 261–277.

Dolnicar, S., & Lazarevski, K. (2009). Marketing in non-profit organizations: an international perspective. *International marketing review*.

Enge, E. (2014, Dec). *Twitter engagement unmasked: A study of more than 4m tweets.* Perficient Blogs. Retrieved from `https://blogs.perficient.com/2014/12/11/twitter-engagement-umasked/`

Font, X., & Villarino, J. (2015). Sustainability marketing myopia: The lack of sustainability communication persuasiveness. *Journal of Vacation Marketing*.

Fownes, J. R., Yu, C., & Margolin, D. B. (2018). Twitter and climate change. *Sociology Compass*, *12*(6), e12587.

Gonzalez, L. I. A., Vijande, M. L. S., & Casielles, R. V. (2002). The market orientation concept in the private nonprofit organisation domain. *International Journal of Nonprofit and Voluntary Sector Marketing*, *7*(1), 55–67.

Greenberg, J., & MacAulay, M. (2009). Npo 2.0? exploring the web presence of environmental nonprofit organizations in canada. *Global Media Journal: Canadian Edition*, *2*(1).

Guo, C., & Saxton, G. D. (2014). Tweeting social change: How social media are changing nonprofit advocacy. *Nonprofit and voluntary sector quarterly*, *43*(1), 57–79.

Han, X., Gu, X., & Peng, S. (2019). Analysis of tweet form's effect on users' engagement on twitter. *Cogent Business & Management*.

Hershey, J. R., & Olsen, P. A. (2007). Approximating the kullback leibler divergence between gaussian mixture models. In *2007 ieee international conference on acoustics, speech and signal processing-icassp'07* (Vol. 4, pp. IV–317).

Hollebeek, L. D. (2011). Demystifying customer brand engagement: Exploring the loyalty nexus. *Journal of marketing management*, *27*(7-8), 785–807.

Hou, Y., & Lampe, C. (2015). Social media effectiveness for public engagement: Example of small nonprofits. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3107–3116).

Ibrahim, N. F., Wang, X., & Bourne, H. (2017). Exploring the effect of user engagement in online brand communities: Evidence from twitter. *Computers in Human Behavior*, *72*, 321–338.

Jegers, M. (2008). *Managerial economics of non-profit organizations.* Routledge.

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public relations review*, *28*(1), 21–37.

Konstantopoulos, T. (2009). Markov chains and random walks. *Lecture notes*.

Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics.* Springer.

Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. the use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, *64*, 575–583.

Lee, K. (2015, Nov). *What analyzing 1 million tweets taught us.* Retrieved from `https://thenextweb.com/news/what-analyzing-1-million-tweets-taught-us`

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? an empirical study of image content and social media engagement. *Journal of Marketing Research*, *57*(1), 1–19.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on world wide web* (pp. 342–351).

Liu, Z., Liu, L., & Li, H. (2012). Determinants of information retweeting in microblogging. *Internet Research*.

Lovejoy, K., & Saxton, G. D. (2012). Information, community, and action: How nonprofit organizations use social media. *Journal of computer-mediated communication*, *17*(3), 337–353.

Malhotra, A., Malhotra, C. K., & See, A. (2012). How to get your messages retweeted. *MIT Sloan Management Review*, *53*(2), 61–66.

Maltby, H., Pakornrat, W., & Jackson, J. (2022). *Markov chains.* Brilliant.org. Retrieved from `https://brilliant.org/wiki/markov-chains/`

NASA. (2022, Jan). *Climate change evidence: How do we know?* NASA. Retrieved from `https://climate.nasa.gov/evidence/`

Patterson, P., Yu, T., & De Ruyter, K. (2006). Understanding customer engagement in services. In *Advancing theory, maintaining relevance, proceedings of anzmac 2006 conference, brisbane* (pp. 4–6).

Pernicano, G. (2021). *Text generation with markov chains: An introduction to using markovify.* Towards Data Science. Retrieved from `https://towardsdatascience.com/text-generation-with-markov-chains-an-introduction-to-using-markovify-742e6680dc33`

Roggeveen, A. L., & Grewal, D. (2016). *Engaging customers: The wheel of social media engagement* (Vol. 33) (No. 2). Emeral Group Publishing.

Salamon, L. M., & Anheier, H. K. (1992). In search of the non-profit sector. i: The question of definitions. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *3*(2), 125–151.

Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., Yadav, D., & Cordón, I. (2016). The markovchain package: a package for easily handling discrete markov chains in r. *Accessed Dec*.

Tölkes, C. (2018). Sustainability communication in tourism–a literature review. *Tourism Management Perspectives*, *27*, 10–21.

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of service research*, *13*(3), 253–266.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, *359*(6380), 1146–1151.

Wadhwa, V., Latimer, E., Chatterjee, K., McCarty, J., & Fitzgerald, R. (2017). Maximizing the tweet engagement rate in academia: analysis of the ajnr twitter feed. *American Journal of neuroradiology*, *38*(10), 1866–1868.

Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons.

Yu, J., & Egger, R. (2021). Color and engagement in touristic instagram pictures: a machine learning approach. *Annals of Tourism Research*, *89*, 103204.

Zor, O., Kim, K. H., & Monga, A. (2021). Tweets we like aren't alike: Time of day affects engagement with vice and virtue tweets. *Journal of Consumer Research*.