

Article

Analysis of Popular Social Media Topics Regarding Plastic Pollution

Phoey Lee Teh ^{1,*}, Scott Piao ^{2,†}, Mansour Almansour ^{2,†}, Huey Fang Ong ^{3,†} and Abdul Ahad ^{1,*}¹ School of Engineering and Technology, Sunway University, Bandar Sunway 47500, Malaysia² School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK; s.piao@lancaster.ac.uk (S.P.); m.s.almansour@lancaster.ac.uk (M.A.)³ School of Information Technology, Monash University Malaysia, Bandar Sunway 47500, Malaysia; ong.hueyfang@monash.edu

* Correspondence: phoyleet@sunway.edu.my (P.L.T.); abdul.a22@imail.sunway.edu.my (A.A.)

† These authors contributed equally to this work.

Abstract: Plastic pollution is one of the most significant environmental issues in the world. The rapid increase of the cumulative amount of plastic waste has caused alarm, and the public have called for actions to mitigate its impacts on the environment. Numerous governments and social activists from various non-profit organisations have set up policies and actively promoted awareness and have engaged the public in discussions on this issue. Nevertheless, social responsibility is the key to a sustainable environment, and individuals are accountable for performing their civic duty and commit to behavioural changes that can reduce the use of plastics. This paper explores a set of topic modelling techniques to assist policymakers and environment communities in understanding public opinions about the issues related to plastic pollution by analysing social media data. We report on an experiment in which a total of 274,404 tweets were collected from Twitter that are related to plastic pollution, and five topic modelling techniques, including (a) Latent Dirichlet Allocation (LDA), (b) Hierarchical Dirichlet Process (HDP), (c) Latent Semantic Indexing (LSI), (d) Non-Negative Matrix Factorisation (NMF), and (e) extension of LDA—Structural Topic Model (STM), were applied to the data to identify popular topics of online conversations, considering topic coherence, topic prevalence, and topic correlation. Our experimental results show that some of these topic modelling techniques are effective in detecting and identifying important topics surrounding plastic pollution, and potentially different techniques can be combined to develop an efficient system for mining important environment-related topics from social media data on a large scale.

Keywords: plastic pollution; environment; topic modelling; structural topic model; natural language processing; social media; public opinion



Citation: Teh, P.L.; Piao, S.; Almansour, M.; Ong, H.F.; Ahad, A. Analysis of Popular Social Media Topics Regarding Plastic Pollution. *Sustainability* **2022**, *14*, 1709. <https://doi.org/10.3390/su14031709>

Academic Editor: Ilias Gialampoukidis

Received: 1 December 2021

Accepted: 13 December 2021

Published: 1 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The global production of plastic has increased significantly to more than 380 million tonnes annually (See report at <https://ourworldindata.org/where-does-plastic-accumulate>, accessed on 30 November 2021), which leads to the massive accumulation of plastic waste, like plastic bags, containers, toys, foams, pre-production gears, plastic fragments, and microplastics [1,2]. By 2050, plastic waste is expected to increase to 12 billion tons if countermeasures are not taken [3,4]. Due to littering, the illegal dumping of rubbish, and poorly managed waste dumps, plastic waste could be everywhere in our natural environment and oceans. In addition, plastic waste is often washed away by rain or blown away by the wind, causing widespread environmental pollution [5]. Plastic pollution undoubtedly has detrimental effects, including global warming, water and soil contamination by chemicals released by plastic, affecting human health, killing animals and plants, and threatening all species' survival. Therefore, all stakeholders ought to make concerted efforts and share responsibility in solving the plastic pollution crisis, including the government, industries,

civil society, and consumers. Nevertheless, public behaviour does not demonstrate major changes in the reduction of plastic usage. For instance, in supermarkets, consumers still use plastic bags to pack their vegetables, take out food or drinks, and carry them in plastic containers and utensils. This global problem can only be solved effectively if the public is happy to commit to alternatives. For some members of the public, this may lead to inconvenience, such as travelling with self-bottled shampoo rather than retail shampoo packs. To date, the vast majority of people are not fully aware of the need for civic engagement. Detecting popular trending topics and public sentiments concerning particular issues can reveal the level of civic engagement and reflect the public acceptance of policy changes. However, traditional research methods, such as surveys and interviews, are time-consuming and costly. In recent years, the increase of social media content is becoming an essential source of value creation [6]. Trending topic and sentiment analysis that leverages social media has attracted considerable research attention in many communities. Such studies include the trending topic analysis with topic detection [7], examination of public opinion by observing the voluntary compliance of energy consumption [8], exploitation for sustainability tourism [9], predicting elections [10], and traffic management [11].

As a major social media platform, Twitter allows the public to freely share their contents, and so it is a handy source for analysing public opinion. In this study, we aim to understand public sentiments or opinions towards plastic pollution challenges, and thereby better inform future solutions and policy decision-making. Two research questions were formed:

1. What are popular topics related to plastic pollution discussed in the Twitter-sphere?
2. How are those topics related?

This paper is organised as follows: In Section 2, a brief overview on related work is presented. In Section 3, a discussion on related methodology is presented. In Section 4, data materials and techniques used for text analysis and the topic modelling results extracted from our data and analysis of the results are presented. In Section 5, the implication and application of our work are discussed. Finally, the conclusion is presented in Section 6.

2. Literature Review

Numerous studies have used them as a reference point for sustainable development since their formulation [12–15]. This is because governments want quantitative measurements to analyse their performance in relation to a variety of difficulties, ranging from natural resource management to pollution control.

When researchers seek evidence to aid in government decision-making on a particular subject, one method that researchers use is to examine the opinion of individuals or users on social networks [16–18]. Even though social media platforms such as Twitter and Facebook are growing in popularity and increasing the number of individuals who have a voice on issues such as the environment, these platforms have demonstrated their limitations as a single means of policymaking, because not all individuals are represented by social media [19–21]. Numerous studies, however, have used Twitter and hashtags to ascertain the population's topical feelings about a particular issue associated with environmental or global health [22]; other studies, for example, those examining public perception in social media following a natural disaster, analyzed the key terms retrieved from the remarks [23].

Amongst many other pertinent studies, authors in [23] evaluated how Twitter members interacted after an earthquake or a flood when they were unable to make phone calls. Additionally, in [24], the authors studied the hashtag #characterbuildings, which was used to communicate Twitter users' perspectives on New Zealand's sustainable housing. Similarly, in [25], the authors utilized publicly available tweets to explore political, social, physical, and historical elements that enable hotel facilities to enhance their sustainability decision-making and management.

In [24], the authors established a model of environmental social awareness based on social media with the goal of equipping individuals with the required information, abilities, and attitudes to mitigate unfavorable environmental impacts. Additionally, in [26], the authors utilized

the social platform to assess public attitudes toward environmental health, gather information about community concerns, and interact with residents about environmental dangers.

2.1. Sentiment Analysis towards Social Network

Sentiment analysis is a popular technique for determining the emotional impact of messages released on a particular subject [26]. It is derived from the utilisation of Big Data, and is used to perform research on the impact of events on social media, to evaluate consumer sentiment towards services and products, and to comprehend online communication [27]. Sentiment analysis is essentially the process of acquiring topic-based Twitter remarks.

Sentiment analysis can be used in conjunction with other technologies to elicit the most critical aspects under investigation. For example, Pak and Paroubek [28] created a variety of approaches for Twitter analysis. By and large, the study considers the connotations of commonly occurring words, which convey a range of emotions, and can be classed as good, bad, or neutral. For instance, Palomino et al. [22] employed sentiment analysis to ascertain public opinion and analyse associated speech in order to gain a better understanding of the natural environment's effect on people's mental health and well-being. On the other hand, Ekenge et al. [26] conducted a topic-based sentiment analysis of Twitter remarks, and illustrated that this social networking site is useful for gauging public sentiment with respect to environmental health, obtaining information about the community's most pressing issues, and communicating about environmental risks. In general, past research has revealed that social networking sites are an excellent venue for doing this type of study. Although some of the difficulties associated with sentiment analysis are due to the context in which the content is examined, sarcasm, or irony, this technology may be trained to mitigate these impacts and shortcomings. To be precise, we can predict ironies and sarcasms by utilising machine learning and data mining algorithms. Table 1 shows summary of important research with respect to social network.

Table 1. Summary of important research with respect to social network.

Research	Social Network	Aim	Analysis of Unit
Palomino et al. [22]	Twitter	Identifying public opinion and examining discourse connected to nature-deficit disorder and other nature-health notions in order to better understand the impact of the natural environment on people's well-being and health.	HashTag
Ekenge et al. [26]	Twitter	By monitoring the public response on the Twitter social network to comments about rising water levels, by determining the issues that most concerned the community about environmental risks.	Posts
Saura et al. [25]	Twitter	To determine the good, neutral, and negative elements reported by hotel guests in Spain	Posts
Chisholm and O'Sullivan et al. [29]	Twitter	To demonstrate how Twitter may be used to give information about housing as a relating to public health issue, in the case study #characterbuildings campaign that launched in New Zealand in 2014.	HashTag
Woo et al. [30]	Twitter	To investigate tweets posted in response to a natural disaster in Korea and to determine changes in user attitude	Posts
Cao et al. [31]	Twitter	To find time and space-related behaviour patterns that contribute to user well-being and enjoyment.	Posts

2.2. Textual Analysis

Textual analysis is a sort of qualitative approach in which a text is analysed in order to arrange the concepts under consideration into “nodes”. This method can be used to analyse an event, a business, or any other object of study. NVivo (QSR International, Melbourne, Australia) is the most often used programme for this analysis [25]. The strategy results in the grouping of concepts into nodes, which is quite beneficial for performing exploratory analysis. This strategy is more descriptive than if no software was utilised.

The nodes are organised hierarchically. At the most fundamental level, there are notions that are conceptually distinct from one another. At the second level, there are branches that originate from each of the nodes and are organised hierarchically. Finally, there are indicators that can be derived from research findings and are relevant to the study’s objective. Saura et al. [25] conducted one of the important research studies, in which textual analysis was utilised to discover environmental influences. The authors identify three nodes that correlate to negative, positive, and neutral influences.

How Does Society Respond to Plastic Pollution?

Over 7.8 billion people in the world [32] account for generating a large amount of plastic waste in a single month, which is enough to fill three and a half times the weight of Malaysia’s Petronas Twin Towers [33]. As part of the effort towards sustainability, many activities such as bring your bottle (BYOB) and reusable utensils, metal and eco-friendly straws are being introduced to reduce the amount of plastic use. Environmental agencies and organisations such as Greenpeace, Heal the Bay, and Friend of Earth, have played a crucial role in reducing the plastic pollution crisis. They take the initiative in leading anti-plastic campaigns, conduct beach clean-up activities in the coastal areas, raise awareness on plastic pollution, and help in efforts to reduce plastic pollution. Ocean Conservancy has partnered with governments, corporations and participated in a programme known as the International Coastal Cleanup to remove plastic debris on beaches [34]. These activities have greatly attracted school students’ interest, and can be used as a vehicle for environmental education. Exploring this topic online can assist the study of how well society is educated and reacts to the issues and the current state of the community’s commitment toward this issue.

Meanwhile, governments have set up laws to fine manufacturers and individuals who do not follow proper methods of plastic waste disposal. For example, the UK government discourages using plastic bags, and shops and supermarkets in the UK charge customers for new plastic bags. The Zimbabwe government has announced a ban on polystyrene usage, as it takes approximately a million years to decompose. Those who are found guilty of breaching the law will be fined 30 dollars to 500 dollars [35]. In Malaysia, the Selangor government issued a ban on the use of plastic bags, and people have to pay twenty cents for each plastic bag that is used in shops [36], while in the Penang state, a new plastic bag is charged one ringgit from 2021 [37]. Moreover, Malaysia’s government introduced the Malaysia Roadmap Towards Zero Single-Use Plastic law in 2019 and now strictly restricts the dumping of plastic bags.

All the strategy mentioned above is to urge society to look into the serious issue of plastic pollution. Every individual has a moral responsibility to recycle used plastics. American and European residents have played an active role in initiating recycling activities, and the national recycling rate of polyethylene film has been increasing.

2.3. Twitter

As one of the most popular social media sites, Twitter serves to be a social connection among people, particularly individuals within their in-group that share the same perspective [38]. Social interaction fulfils the emotional needs that people may not otherwise find elsewhere. If an emotional need were met, the viewer would engage in that similar behaviour intensively [38]. In other words, Twitter is potentially helpful in driving social

behaviour change. Hence, understanding and exploring topic prevalence and correlation could drive solidarity in society.

Why tweets? Twitter is connectable to Facebook and Instagram. Instagrammers are the users of the photo-sharing and social networking site Instagram. Instagram has been used as a marketing tool to discuss a wide range of well-being issues, such as food consumption, obesity, eating disorders and dieting [39]. Mainly to gather popularity, many Instagram accounts are dedicated to documenting their lives concerning these issues. In other words, compared to the role of Instagram and Facebook, Twitter has an even larger community of users. It is also larger than both of these competitors on their own. Hence, our data exploration starts with Twitter.

2.4. Topic Modelling

In this age of information, it is challenging for individuals to process and understand a vast amount of unstructured data; hence, topic modelling is commonly applied to declare, determine and define an unknown semantic framework in a text collection [40]. A variety of techniques have been proposed and implemented for this purpose.

Generally speaking, a topic model is a type of mathematical model which discovers and analyses abstract topics from a group of documents. It is also defined as a probabilistic semantic model, because mathematical algorithms are implemented to reveal the underlying semantic framework in a text collection [41]. Over the past years, topic modelling has been used across different fields, concentrating on text mining and retrieving information [42]. The model has also obtained much attraction among researchers from other areas of study. It has been successfully applied in the fields of brand management [43], personal branding [44], and social networks [45].

There are a variety of topic modelling techniques. In this paper, we explore five methods, including (a) Latent Dirichlet Allocation (LDA), (b) Hierarchical Dirichlet Process (HDP), (c) Latent Semantic Indexing (LSI), (d) Non-Negative Matrix Factorisation (NMF), and (e) extension of LDA—Structural Topic Model (STM) for further analysing topics of the Twitter data.

LDA is an unsupervised method for identifying key topics within a set of documents [46]. LDA contains a generative statistical model which assumes that the observed documents are produced from a mix of latent topics. The unobserved latent topics whose number is determined by the researcher beforehand will then generate words that associate with each other based on the probability distributions. Suppose LDA is given a number of topics and a set of documents. In that case, the LDA algorithm will identify what combination of the unknown issues can be generated by those documents [47].

Another topic modelling approach extended from LDA is the Hierarchical Dirichlet Process (HDP). It is a Bayesian nonparametric model used to model groups of data with a potentially infinite number of components [48]. This approach assumes that the components are word distributions representing recurrent topics in the collection. The number of topics and the characteristics of the distribution are defined by posterior inference automatically. Latent Semantic Indexing (LSI) is a statistical model used to enhance information indexing and retrieval. It uses the singular value decomposition (SVD) technique to find the latent relationship between words in given documents and classify them into topics.

Non-negative matrix factorisation (NMF) is a dimensionality reduction technique for multivariate data. It was introduced by Paatero and Tapper in 1994 under the name positive matrix factorization [49]. It is easy to implement and became known as non-negative matrix factorisation after Lee and Seung (2001) published algorithms of two types of factorisation [50]. The basic idea of NMF is that the negative numbers are meaningless in the data process. Hence, it overcomes the issue of negativity by placing non-negativity constraints on the data model. Many researchers have used it for topic modelling, such as [51].

The Structural Topic Model (STM) is an extension of the LDA, which provides many valuable features, including rich ways to explore topics, estimate uncertainty, and visualise interest quantities [52]. It was shown to be useful in analysing trends and understanding

topic prevalence in tweets [53], studying the topic prevalence and how topic evolves across international news in social studies [54]. Lobbying reports [55]. Compared to standard topic models, STM adds two extra functionalities that allow researchers to determine the topic prevalence and topic contents [47]. It facilitates the study of topic prevalence, topic evolution, and topic correlation. Several studies have adopted the STM approach to analyse social media [56], newspapers [57], and research papers [58].

The STM has been growing in popularity, as it is fully implemented as a publically available R package. The package offers multiple features such as topic discovery, text preprocessing, model validation and model search, and topic visualisation [59]. The STM is a seminal word count model document level covariate information used for topic modelling. By influencing topical prevalence, topical content, or both of them, qualitative and assumption interpretability can be enhanced [54]. A data creation process is illustrated for every document so that the data can be further applied to search for the possible values for the framework within the model [52]. The model starts with both topic word and document topic distributions, which produces metadata documents related to them. With STM, researchers can conveniently determine the relationship between topics and documents via metadata, and the measurement of hypothesis among the relationships can be conducted [52].

3. Data and Methodology

Text analytics are applied to extract meaningful and valuable information from unstructured texts [60]. Text analytics usually involve data collection, data preprocessing, categorisation, and analysis [61]. Meaningful information is extracted from the corpus, a collection of textual contents associated with the topic of interest, where data processing and area of expertise are in charge of organising the genuineness of content [62]. Figure 1 shows the overall four phases involved in the text analytics process.

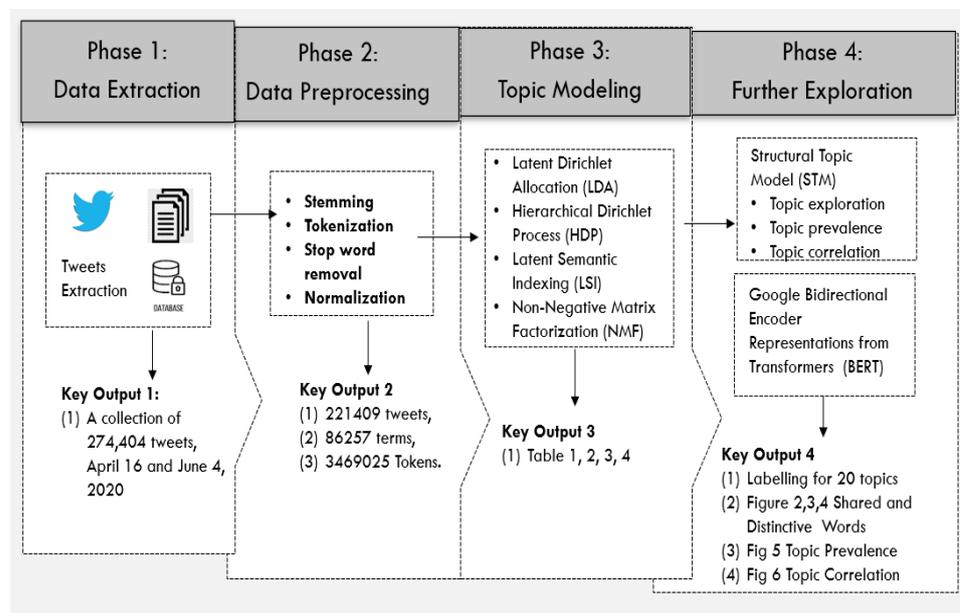


Figure 1. The main phases involved in text analytics.

3.1. Data Extraction

The number of tweets provided for behavioural research studies commonly ranges from 50,000 to 200,000 tweets [63–65]. In this study, 274,404 tweets were crawled from Twitter using keywords between 16 April and 4 June 2020. These tweets were extracted based on predefined keywords of interest, and are related to plastic pollution, such as “plastic-free”, “sustainable”, “sustainability”, “waste”, “plastic”, and “pollution”. The corpus comprises 21 variables, including full name, tweet text, tweet ID, followers, location,

bio, and many more. It was further pre-processed using R programming, and a clean data-set was stored in an Excel file.

3.2. *Methods and Tools for Topic Analysis*

In this paper, we explore a number of mainstream methods and tools for assisting the extraction and analysis of public opinions on hot topics related to plastic pollution from social media on a large scale. For this purpose, we selected a set of standard topic modelling models and tools to process our data. Our aim was not to compare the accuracy of these different methods; rather, we wanted to explore the feasibility of combining the different methods to achieve a wider coverage of topics contained in the data.

As shown in our literature review, there exist numerous algorithms, models and tools for topic detection. For our study, we selected four topic modelling methods, including LDA, LSI, HDP and NMF, as they are among the most commonly used models. In addition, we also used the LDA-based STM (Structural Topic Model) package, as it facilitates various analysis and visualisation of analysis results, in addition to the core topic modelling functionality.

In addition to the detection of topics themselves, we also analysed the key words under the topics, in order to examine the semantic affinity and relations of the key words that collectively represent individual topics. Such analysis reveals how reliably these topic modelling methods can identify main topics from textual data, and how the topics can be linked via shared keywords. Moreover, the STM package was used to analyse the relevance and correlation of the extracted topics illustrated using its visualising functionality.

3.3. *Data Pre-Processing*

Data pre-processing consists of stemming, tokenisation, stop word removal, and normalisation to prepare the corpus and improve the accuracy of further analysis. In this study, text from tweets is first loaded as a corpus. Data cleaning of unnecessary information is required, as the type of text data applied in tweets is generally noisy. Removal of extra white space, numbers, and punctuations from the text is performed. Besides that, arbitrary characters are replaced with space, and words are converted to lowercase. By applying normal expressions, text normalisation is performed too. The normalisation of text in any natural language processing task is a must, as the text is usually represented in an informal and ambiguous method. Text normalisations involve text stemming, where the texts are prepared to be further processed. Words are usually represented in different forms according to the purpose of grammar; for example, the verb “to be” can be represented in “is”, “are”, “were”, and many more. Therefore, texts which are in their basic “stem” forms are reduced in stemming. Then, stop words which are unimportant but common words, such as “for”, “to”, “the”, “an”, “and”, “a”, “but”, and “or”, are removed. Next, the list of top words related to each topic is identified for better topic exploration. By examining the top words from each topic, labelling can be carried out more effectively and efficiently. After performing data pre-processing, the final dataset consisted of 221,409 tweets, 86,257 terms, and 3,469,025 tokens.

3.4. *Topic Labelling*

This study applied search intent to create a suitable and relevant label for automatically generated topics. Search intent illustrates the users’ query on search engines, which represents a goal that users are attempting to achieve. By performing keyword searches through Google, results shown based on the keywords searched could provide a clearer idea of creating relevant labels. For instance, the top results shown by searching for keywords such as “zero”, “waste”, “time”, “just”, and “people” in Google are mostly about zero waste. Hence, the topic with these keywords is labelled as “zero waste”.

Google is one of the most used web search engines that could guess users’ search intent, and provides the most related results for their search. To be more capable of determining users’ search intent, Google has started to improve its algorithm over the years, which

focused on searching, ranking, and returning the most related results for every search term. According to the latest Google Bidirectional Encoder Representations from Transformers (BERT) in 2019 [66], it provides a machine learning algorithm where the framework and connections of all the search terms are considered, instead of just identifying the terms individually based on their order. Therefore, this method is implemented through Google, as it provides the most highly recommended search outcomes, which assist in better labelling for further text analysis.

4. Topic Extraction and Analysis

4.1. Topic Extraction with LDA, LSI, HDP, NMF

One of the two main parts of our experiment is the topic extraction using four popular topic modelling algorithms, namely LDA, LSI, HDP and NMF. The tweet data were passed to each of the tools, and extracted the top twenty topics. Using the topic labelling methods explained earlier, we tagged extracted topics to find meaningful labels. Table 2 shows the labelled topics selected from the results.

As shown in Table 2, altogether, we could manually label forty of the total eighty topics, many of which are related to environmental issues (see highlighted topics), including the plastic pollution issue. In the “Topic types” column, the three digits denote the number of environment-related, manually labelled, and a total number of extracted topics. For example, “4/14/20” for the NMF model indicates that this model extracted twenty topics, of which fourteen could be manually labelled, of which four are relevant to the environment. Some of the topics are semantically close to each other and fall under the same labels. For example, for the HDP model, three topics are about plastic waste. Hence, all of them are labelled as “Plastic Waste”, to which the numbers are attached solely to distinguish different topic items.

Because the Twitter data were collected using keywords related to plastic pollution, we expected to extract many environment-related topics. In this regard, the four algorithms performed differently. The HDP and LSI produced more such topics (nine each) compared to the other models. It demonstrates that they are more effective in picking up general topics of the data. However, as reflected by their lower coherence scores, the topics’ quality (in terms of semantic similarity of words) is poorer than those produced by NMF and LDA.

Tables 3 and 4 show two sets of words of corresponding topics (Climate Change and Plastic Pollution) for a comparison. Table 3 lists ten words under each of the two topics extracted by NMF with a coherence score of 0.4964, which is the highest among the four models. On the other hand, Table 4 lists ten words under the corresponding topics extracted by LSI and HDP, with coherence scores of 0.3503 and 0.3440, respectively. When we closely examine the words, we can observe that, in Table 3, the sets of ten words are semantically closely related, whereas, in Table 4, the sets of words have a loose semantic connection between most of the words.

Table 2. Automatically extracted topics that could be labelled manually, with environment-related topics highlighted.

Topic Model	Coheren	Meaningful Topics Labelled Manually	Topic Types
NFP	0.4956	3R Concept, Plastic Pollution, Climate Change, Sustainable, Audiobook, Romance, Bear Market, Free Stuff, Online Learning, Game, Music, stock, writing, Free books	4/14/20
LDA	0.4014	3R Plastic Pollution, Environment, Sustainable, Poems, Present, Experience, Peace and violent	3/7/20
HDP	0.3503	Waste Management, Plastic Pollution, Plastic Pollution2, Plastic Pollution3, Plastic Waste, Plastic Waste2, Plastic Waste3, Environment, Environment2	9/9/20
LSI	0.3440	3R Concept, 3R Concept2, Recycling, Recycling2, Climate Change, Renewable Energy, Sustainable, Sustainability, Sustainability2, Giveaways	9/10/20

Such different performance of different topic modelling algorithms has implications with regard to their practical applications. In this case, the NMF and LDA are more effective in clustering-related text (tweets in this case) under highly representative topics. On the other hand, the LSI and HDP are more helpful in identifying the general topic of the entire data (environment issue in this case). Our primary purpose in this paper is not to compare the performance of individual models. Instead, we focus on analysing the contents of the Tweet data by combining the topics extracted by these four algorithms.

As we mentioned earlier, our primary focus of this paper is not the comparison of different topic modelling models. Instead, we explore the feasibility of identifying social media hot topics by combining a set of models, namely LDA, LSI, HDP, and NMF, in this particular case. When we combine all the topics extracted using the four models, we found that the most frequent topics of the data include “Plastic Pollution” and “Sustainable/Sustainability”, which accurately reflect our data’s main contents. Table 5 shows the frequency of the environment-related topics. As shown in the table, in the social media conversation, the public are highly aware of environmental issues, including issues related to plastic pollution, reflected by Plastic Pollution, Plastic Waste, Recycling, Waste Management, and its effect to the environment, reflected by Sustainable/Sustainability, Environment, 3R Concept, Climate Change, Renewable Energy.

All the topics above reflect a widespread concern in society about plastic pollution and related environmental issues. Such issues have become popular topics of chatting and discussions on social media. Although a further investigation is needed, our experimental results demonstrate that an appropriate combination of a set of topic modelling algorithms can potentially provide an effective means of tracking and analysing public issues and opinions regarding plastic pollution and more general environmental issues. Furthermore, we examined the semantic network of the topics by observing the collection of the top twenty words extracted by the topic modelling tools for each topic. Figure 2 shows a word cloud generated using all the words under eighty topics extracted in the first step of our experiment. In this word cloud, the frequency of each word determines its font size: the more frequent, the bigger. As shown in Figure 2, the most frequent words include (f = frequency):

“Waste” (f = 22), “Plastic” (f = 20),
 “Sustainability” (f = 12), “environment” (f = 10),
 “pollution” (f = 9), “climate” (f = 9),
 “nature” (f = 9), “time” (f = 9).

Except for the word “time”, which does not have explicit relevance to the environment issue, the rest of the words correctly reflect the core theme of the data collectively, and connect some topics. The wordcloud also reveals some interesting sub-topics related to the main theme, such as “books”, “games”, “romance”, “music”, etc. Such sub-topics may potentially show connections between some media or activities and environmental issues.

We also examined the wordcloud of each of the four topic models, as shown by Figures 3–6. Each word cloud is generated using the words from twenty topics extracted by each of the topic models. As shown in Figure 3, HDP extracted quite a few environment-related words for the topics, including “waste” (f = 14), “plastic” (f = 10), “nature” (f = 8), “environment” (f = 7). Similar case for LSI, whose topics contain frequent words, including “sustainability” (f = 11), “plastic” (f = 7), “climate” (f = 7), “recycle” (f = 6), “pollution” (f = 6), etc., which are closely related to the environment issue. On the other hand, NMF and LDA extracted a wider range of different words, which form more diverse topics, but with a good coherence level, although many are not related to environment issues. These two models should be more useful for identifying a wider range of diverse topics that are potentially related to environmental topics in indirect ways. Note that all of the results are extracted from the same data set.

Given that individual topic models have different features and performances for specific requirements, a practical approach can combine them to produce a complete result. Our experiment demonstrates that this is feasible, as illustrated by Figure 2, with which we

can identify deeper and wider thematic and semantic, and structure and network within social media data.

Table 3. Two samples of topics extracted by NMF.

Climate Change	Plastic Pollution
pollution	plastic
air	use
climate	bag
change	bottle
environment	ocean
lockdown	recycle
clean	recycling
level	single
see	plasticfree
water	packaging

Table 4. Two samples of topics extracted by LSI and HDP.

Climate Change (LSI)	Plastic Pollution (HDP)
climate	swear
month	trigger
change	well
trial	handle
free	become
student	doll
click	plastic
faster	waste
ship	idk
purchase	rank

4.2. Topic Exploration with STM Package

Another major experiment is the topic extraction using STM. The STM package, a high-quality R package that allows data analysis to be carried out effectively and efficiently, was used to extract topics, topic prevalence, and topic correlation. This package provides various functions for data pre-processing [59].

4.2.1. Topic Extraction

Two methods implemented in STM are employed in this study to carry out topic exploration. The first method is for identifying the list of words that are related to each of the topics. Meanwhile, the second method determines the actual documents, which are tweets, that are highly related to each topic. Functions that are used for topic exploration in the STM package are the labels Topics function and Final thoughts function. In the STM package, the label Topics function is applied to explore each topic's top words. The Final thoughts function in the STM package is then used to search for tweets related to a specific topic. This function is beneficial as it provides a better understanding of each topic's contents to illustrate its meaning. Table 2 shows the top words for each topic and the original tweets associated with each of the topics. By applying the plot.STM,

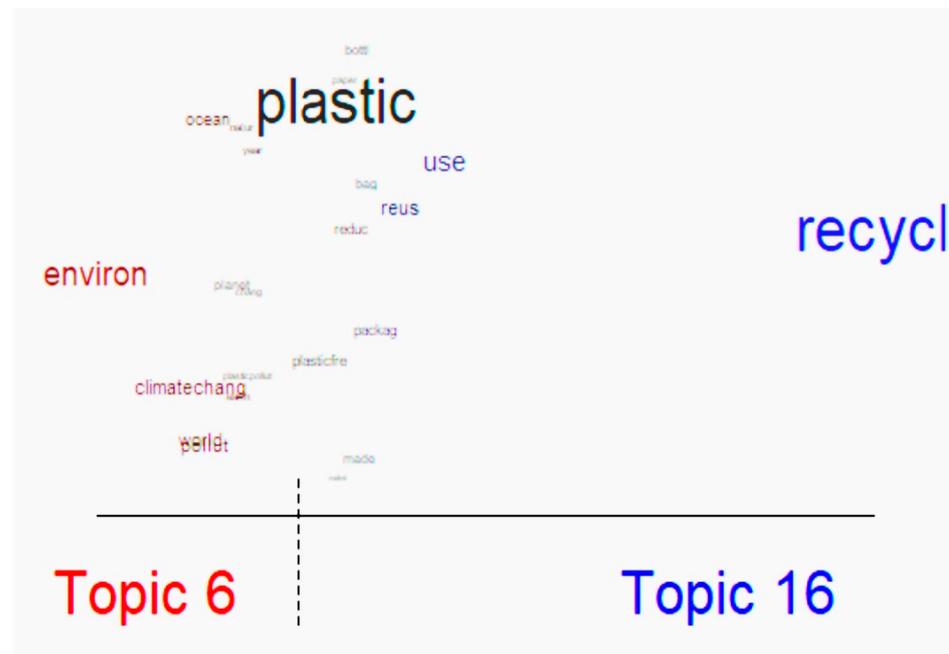


Figure 8. Shared and distinctive words between Topic 6 and Topic 16.

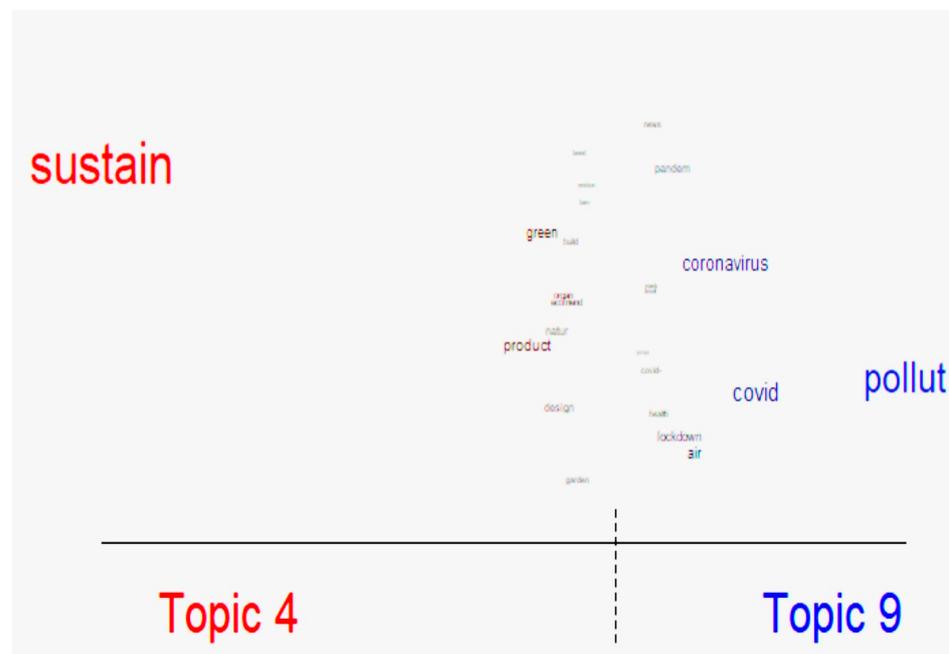


Figure 9. Shared and distinctive words between Topic 4 and Topic 9.

Referring to Figure 10, the top 5 well-known topics made in Twitter conversations between April 16 to June 4 in 2020, in descending proportion order, range from zero waste (Topic 8), sustainable energy (Topic 5), help needed (Topic 20), reading (Topic 7), to online learning (Topic 12). This conversation depicts that society is aware of the environment and recognises the importance of dealing with plastic pollution. They converse in ways that indicate the need for immediate actions to protect the environment, including discussing how to solve problems like reducing waste and sustainable issues. Within two months during the pandemic, while everybody was working from home, online shopping seemed to have high potential to create plastic pollution issues due to the intensifying packaging and delivery services. Furthermore, the topic of reading and online learning also comes

into the discussion. Meanwhile, individuals on Twitter also showed interest in learning more about sustainability, especially the discussion of wandering animals around the city and how society's concern about environmental issues was discussed (see Appendix A for sample Tweets for Topics).

Surprisingly, some topics of discussion that are not related to environmental issues also appear within the batch of tweets in the tweets data collection pile; this topic has an element that linked to the discussion about videos (Topic 11), free games (Topic 18), free music (Topic 10), giveaways (Topic 14), and fashion (Topic 1). However, these are comparatively less popular. Society is more indulging in video and free games during those times instead of talking about it online. Users are still attracted to giveaway information that is posted on Twitter.

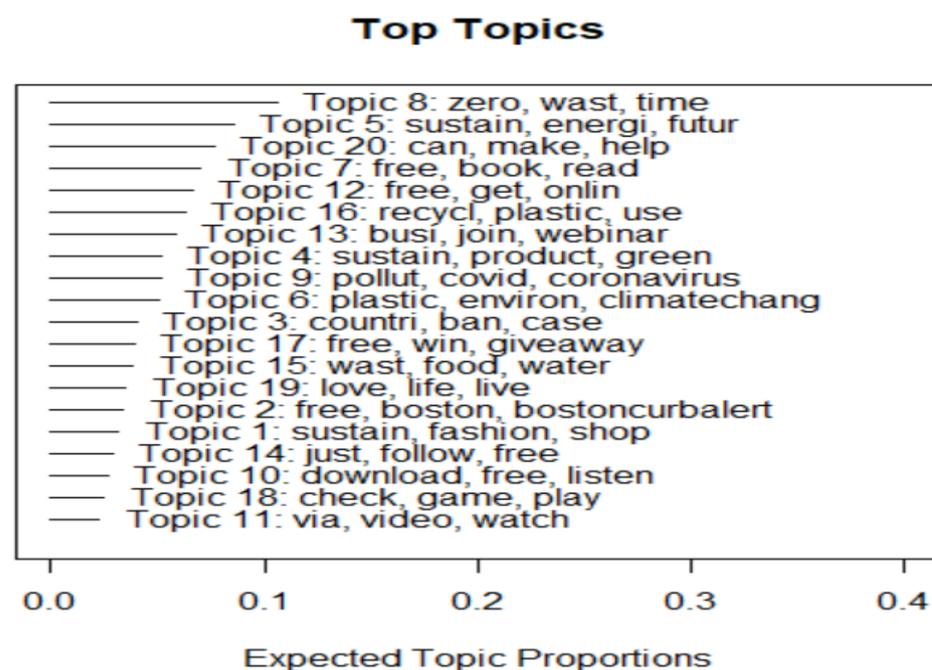


Figure 10. Topic Prevalence.

4.2.3. Topic Correlation

A network graph is generated to illustrate the correlation between the set of 20 topics; Directed and undirected graph frameworks are also presented with the line connected to each topic. Each node represented a topic, and a link connected referred to a relationship between the two nodes. Figure 11 contains a collection of all nodes that are connected.

The topic correlation network presented 20 topics that were connected to one or even more topics. The 20 topics have established newsworthy communities, where the largest community is defined by 10 of those 20 topics. The community is formed by topics that were mainly about sustainability and the environment, which are Topic 4 (going green), Topic 5 (sustainable energy), Topic 6 (plastic pollution), Topic 15 (wastage), Topic 16 (3R concept), Topic 20 (help needed), and others. On the other hand, another community is represented by topics that were more related to leisure activities and free kinds of stuff, instead of the environment, which includes Topic 10 (free music), Topic 11 (watch video), Topic 14 (giveaways), Topic 18 (free games) and so on.

Meanwhile, Topic 2 and Topic 17 have the least connection, as they are only connected to one node. Topic 2 (Boston free stuff) is connected to Topic 1 (fashion), while Topic 17 (gift cards giveaways) is linked to Topic 14 (giveaways). These topics are not commonly discussed on Twitter, based on their expected topic proportions in 4.2 topic prevalence.

Topic 20 (help needed) is the most correlated topic that consists of the most nodes connected to it. It is connected to seven topics, such as Topic 5 (sustainable energy), Topic 8 (zero waste), Topic 9 (covid-19), Topic 13 (webinar), Topic 15 (wastage), Topic 16 (3R

concept), and Topic 19 (happy life). Most of the topics connected to topic 20 contained conversations about environmental issues faced and actions that should be applied to solve the problems and save the environment.

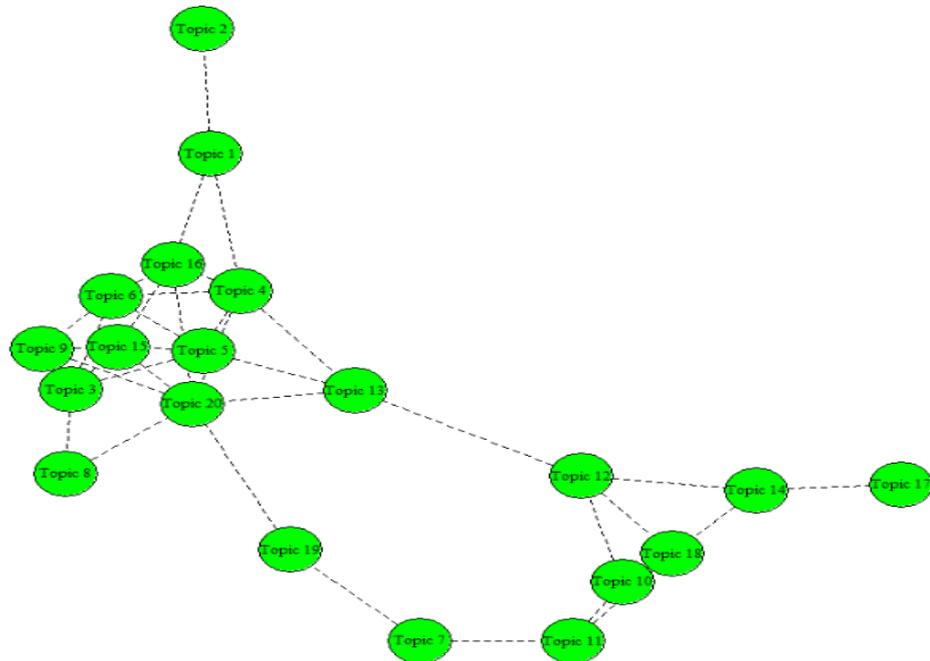


Figure 11. Topic Correlation.

5. Discussions

In this modern information age, text data can be collected from various online sources easily, such as social media, traditional media, surveys, and many more. Nowadays, the public tend to share their opinions and feelings on social media platforms like Twitter, Facebook, and Instagram [67]. Therefore, topics related to the environment, plastic pollution in our particular case, can be identified by analysing data from social media platforms.

Our experiments demonstrate that, by applying various topic modelling algorithms and tools, such as LSI, LDA, HDP, NMF, and STM, on social media data, we can identify topics related to plastic pollution discussions, the popularity of each topic, and correlation among different topics of interest. Furthermore, this approach implemented is proven to be capable of handling a large amount of online textual data effectively and efficiently. Hence, it can distill important information regarding plastic pollution and the environment in general from social media data for stakeholders, such as the government and organisations.

For example, with STM analysis, motivation related to plastic pollution on Twitter consists of various topics, such as sustainability, zero waste, 3R concept, reading, learning, etc. Approximately 55% of the topics discussed are related to the environment, while the rest is about leisure activities. Governments and organisations should be able to get a deeper understanding of individuals' motivation by observing detailed information on every topic of interest, and take timely actions to address the issues of reducing the negative impact of plastic pollution.

According to our STM analysis, the most popular topic discussed between April and June 2020 on Twitter is Topic 8 (zero waste). Creating a zero-waste future where every product and good are treated with value, the disaster caused by plastic pollution can be tackled. Businesses need to develop products and services which avoid single-use plastic to begin a zero-waste supply chain. Governments and environmental organisations also play an important role in spreading awareness and implementing zero-waste acquisition guidance for private companies. In addition, topic correlation showed that Topic 8 (zero waste) is linked to Topic 3 (ban on countries) and Topic 20 (help needed). The interrelated

connections between these three topics emphasise the conversations involving how and what to do to achieve zero waste via international collaboration and help from each other.

Moreover, Topic 6 (plastic pollution) is worth further attention, and is closely related to topics such as Topic 3 (ban on countries), Topic 4 (going green), Topic 5 (sustainable energy), Topic 9 (covid-19), and Topic 16 (3R concept). Due to the COVID-19 pandemic, the amount of single-use plastics such as gloves and masks has rapidly increased, contributing to global plastic pollution [68]. One of the well-known methods in solving plastic waste is that individuals in their daily lives apply the 3R concept to reduce, reuse, and recycle. Topic 7 (reading) and Topic 12 (online learning) are also frequently discussed on Twitter, where individuals can read and learn about the information associated with plastic pollution and the environment through online platforms. Compared to traditional learning methods, free courses and programs provided online are more accessible and convenient. Therefore, governments and organisations could promote online courses to reduce the number of plastics used to avoid plastic pollution. Online learning can also minimise several adverse environmental effects that can reduce wastage and protect natural resources.

Some of the topics like STM Topic 11 (watch video), Topic 14 (giveaways), and Topic 17 (gift cards giveaways) that are not associated with plastic pollution are also worth attention. Video tends to reach out to more people and attract individuals' attention better compared to normal blog posts. Individuals nowadays prefer video content more than written content, as they are more interesting and engaging. To attract and interact with individuals, organisations or governments can create video content related to plastic pollution and post it online so that it is accessible to the public. Knowledge related to sustainability and methods to protect the environment can be promoted more quickly to the public by producing and posting videos online. By providing giveaways and freebies, individuals' attention can also be generated to organisations' platforms.

6. Conclusions

Plastic pollution, along with correlated terms such as zero waste, sustainability, and 3R concept, are important topics that should be further discussed and studied by academic researchers. Big data can be gathered from social media platforms such as Twitter, where individuals are more willing to voice their opinions and feelings. Useful information related to the environment can be extracted from text data like tweets. Previous studies tend to concentrate on applying social media analysis tools that only involve keyword and sentiment analysis, which makes it hard to determine suggestions, desires, and questions from the public. The goal of this study is to explore topics that are discussed in the Twitter-sphere, the popularity of topics, and connections between topics.

Most previous studies aim to extract meaningful insights from tweets to solve research problems that could not be explained by traditional approaches [69]. Twitter has been used as a powerful platform for organisations to interact with individuals, as it is more sentimental and less formal. This study introduces the use of big data and computational methods for future research related to plastic pollution, as there is huge potential in big data. Researchers need to require knowledge related to computational data collection methods, so that studies can be carried out more effectively and efficiently. Data collection methods, such as web crawling and APIs, allow a large amount of data to be gathered rapidly. This study recommends the use of text analytics instead of traditional analysis techniques like manual coding, which are not practical for big data. Besides, structural topic modelling is suggested, as it can quantify topic exploration, topic popularity, and topic correlation automatically.

We must solve the problem of plastic pollution which leads to negative impacts on human beings, animals, and plants, such as causing illness [70] and killing wildlife [71]. Everyone should implement immediate actions to get closer to a future where individuals are capable of adjusting the expenditure with the needs of the natural world. Instead of just reducing plastic waste and pollution, governments from all around the world should encourage sustainable growth and promote both blue and green economies to

their citizens. Meanwhile, organisations can collaborate with other businesses to stimulate plastic compilation and education. By reducing, reusing, and recycling packaging made from plastics across businesses supply chains, plastic waste can also be avoided. Therefore, organisations, businesses, politicians, and society should work together by modifying their consumption habits and public policy to make a difference for the environment and future.

There is a limitation to our current research. The Twitter data we used only cover one and a half months. In order to develop a more authentic and in-depth understanding of plastic pollution, future research should collect more data for a longer period of time, and from various data sources. In addition, future research should also apply more techniques, such as Twitter user metrics and sentiment analysis.

Author Contributions: All the authors contributed to the work and jointly wrote the article. P.L.T. proposed the idea, designs, and writing. S.P., M.A., H.F.O. and A.A. suggested directions for the detailed designs as well as coordinated the work. All authors have read and agreed to the published version of the manuscript.

Funding: Sunway University Research Support.

Institutional Review Board Statement: The study was conducted according to the guidelines of Sunway University, Malaysia.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Sample Tweets for Topics.

Label	Top Words	A Sample of Original Tweet	Expected Topic Proportions
T1 Fashion	Sustain, fashion, shop, cloth, size	Vintage 1980s does 1950s full skirt 80s novelty print black, red and white skirt size medium #vintageclothing #sustainable #onlineshopping #fashion #vintagefashion #retroverbiz #Vintagelifestyle	0.035
T2 Boston free stuff	Free, Boston, Boston curb alert, freebies, wood	Wood #BostonCurbAlert #Boston #Free	0.04
T3 Ban on countries	Countries, ban, case, government, state	The Turkish national #government just declared a 48-hour complete #ban for citizens of most cities around the country, including #Istanbul and #Ankara effective tonight at midnight because of #coronavirus. The ban was announced ONE HOUR BEFORE it took effect.	0.044
T4 Going green	Sustain, product, green, nature, design	Would you prefer a contractor that used green products? #yqr #regina #sask #sk #saskatchewan #construction #conservation #renovation #custom #interiordesign #design #nature #greenproducts #eco #sustainable #ecofriendly #earth #green	0.05
T5 Sustainable energy	Sustain, energy, future, develop, industries	Global partnerships to accelerate a #CleanEnergy future continue to grow. 7 companies based in #Singapore and #Japan have joined forces to help Singapore develop a #sustainable energy system with #hydrogen as a key element. #HydrogenNow	0.08

Table A1. Cont.

Label	Top Words	A Sample of Original Tweet	Expected Topic Proportions
T6 Plastic pollution	Plastic, environment, climate change, world, pollution	Wake up before it's too late, save the environment, save nature, save forests, save water Air pollution, climate change, global warming, plastic pollution, water pollution are the biggest threats to the environment #ForNature #biodiversity #environment #WorldEnvironmentDay #india #Earth #Lockdown #stayathome #staysafe #Read #FREE #eBooks on #Kindle	0.05
T7 Reading	Free, book, read, kindle, ebook	Also available for free download, two of my titles #AnArrangedMatch #Madeinheavendotcom Download the Kindle app on your phones/tablets or use a Kindle eReader and start reading! #KindleIndia	0.07
T8 Zero waste	Zero, waste, time, just, people	The environment is just nurturing us since time immemorial, It is time to give back. Let's pledge to protect it. here are some zero waste tips that u can implement in your daily life. let's try to be more sustainable! #WorldEnvironmentDay #foharmalai #saveocean @	0.11
T9 Covid-19	Pollution, covid, coronavirus, air, lockdown	The positive impact of the world's biggest coronavirus lockdown in an India recent report of SAFAR the graph of PM2.5 and NO ₂ fell during this lockdown. Source: #airpollution #pollution #covid19 #lockdowneffect #lockdown #coronavirus	0.05
T10 Free music	Download, free, listen, art, new	Want to download free (No Copyright Music) then click the link to see a new Youtube Channel called Sound Studio and start listening to some music. #YouTube #free #music #new #sound #YouTubers #youtubechannel #StayHome #StayAtHome #staysafe #soundtrack	0.025
T11 Watch video	Via, watch, share, video, full,	Free Udacity nano degree course FREE Subscribe the channel plzz and watch video step by step carefully #share #udacity #free #freecourses #YouTube #Twitter #StayHomeStaySafe #QuarantineTime #studywithflow #Saturday	0.015
T12 Online learning	Free, get, online, learn, course	Learn about #sustainability and take action with our #online course in#environmental natural resources and sustainable development. 100% #scholarship offered by Government of #India on #iLearn- #StudyiLearn #eVBAB @IndiainDRC @indiainzambia @IndiainUganda	0.065
T13 Webinar	Business, webinar, support, join, may,	Join to support this webinar for an overview of the sustainable plastics marketplace and businesses that have developed sustainable solutions. #EENcanhelp RT @KTN-Creative: Now Live Designing Sustainable Plastic Solutions comp. @innovateuk will invest up to "800k to fund early-stage, human-centered design projects to reduce the harm that plastics have on our environment and increase productivity and amp; growth of the UK economy.	0.055
T14 Giveaways	Just, follow, free, giveaway, post	Don't forget !! CashApp Giveaway tomorrow!! #CashAppFriday #cashapp #stocks #robinhood #webull #m1 #stash #freemoney #moneygiveaway #Stimuluscheck #free #freecash RT @bdippr: @Bdippr @Bdippr @bdippr Weekly Cash App Giveaway (x3) CashApp Prizes FREE money! Invest it ,It's super easy to enter!! 1. Follow @bdippr 2. Tag a friend 3. Retweet this post.	0.03
T15 Wastage	Waste, food, water, time, service	Want to produce delivered to your door AND save money or time? Imperfect provides produce too ugly for the grocery store but still perfectly edible! Less food waste means less water waste #climatechange, #pollution	0.044
T16 3R concept	Recycle, plastic, use, reuse, package	From villain to the hero during #COVID19. Demand for #plastic masks, gloves and packaged goods are soaring. Plastics seem now indispensable: cheap and amp; disposable. A backlash to the #recycling and amp; ban #singleuse plastic movements? Getting back to the #circulareconomy will take some effort	0.06

Table A1. Cont.

Label	Top Words	A Sample of Original Tweet	Expected Topic Proportions
T17 Gift cards giveaways	Free, win, giveaway, gift, card	I am running in a #free #giveaway by @ameel-rahimi for a chance to win a 2GAME Gift card, Make sure you enter as well! #free #Fortnite #vbucks #steam #game #lockdown #paypal #entertowin #Giveaway #Giveaway-Alert #COVID19	0.044
T18 Free games	Check, game, play, free, fun	Give your kids something super fun to do with Seussville! They can watch videos, play games, and explore the wonderful world of Dr. Seuss here! > . #DrSeuss #reading #imagination #fun #learning #activities #games #kids #free	0.018
T19 Happy life	Love, life, live, happiness, families	IDENTIFY and REMOVE TOXINS: REJUVENATE YOUR #BODY #health #love #life #live #money #travel #vegan #wisdom #wellness #motivation #inspiration #nature #family #Covid-19 #fitness #FamiliesFirst #children #art #free #ThursdayThoughts #ThursdayMotivation #style	0.04
T20 Help needed	Can, make, help, need, will	This coming week we want to help solve the single-use plastic problem that is facing our planet. What are some of the ways we can reuse single-use plastic? #Reuse #Plastic #Pollution #CleanBeaches #Trash #Garbage #Environment #LoveEarth #Green #Recycle #SturdyWings #StaySturdy	0.075

References

- Koelmans, A.A.; Nor, N.H.M.; Hermesen, E.; Kooi, M.; Mintenig, S.M.; De France, J. Microplastics in freshwaters and drinking water: Critical review and assessment of data quality. *Water Res.* **2019**, *155*, 410–422. [CrossRef] [PubMed]
- McGoran, A.R.; Cowie, P.R.; Clark, P.F.; McEvoy, J.P.; Morrill, D. Ingestion of plastic by fish: A comparison of Thames Estuary and Firth of Clyde populations. *Mar. Pollut. Bull.* **2018**, *137*, 12–23. [CrossRef]
- Kaza, S.; Yao, L.; Bhada-Tata, P.; Van Woerden, F. *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*; Urban Development, World Bank Publications. 2018. Available online: <https://openknowledge.worldbank.org/handle/10986/30317> (accessed on 1 November 2021).
- Parker, L. Here's How Much Plastic Trash Is Littering the Earth. 2018. Available online: <https://www.nationalgeographic.com/science/article/plastic-produced-recycling-waste-ocean-trash-debris-environment> (accessed on 19 March 2021).
- Bashir, N.H. Plastic problem in Africa. *Jpn. J. Vet. Res.* **2013**, *61*, S1–S11.
- Brandt, T.; Bendler, J.; Neumann, D. Social media analytics and value creation in urban smart tourism ecosystems. *Inf. Manag.* **2017**, *54*, 703–713. [CrossRef]
- Banu, S.H.; Chitrakala, S. Trending Topic Analysis using novel sub topic detection model. In Proceedings of the 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, India, 27–28 February 2016; pp. 157–161.
- Kusiak, A.; Zhang, Z.; Verma, A. Prediction, operations, and condition monitoring in wind energy. *Energy* **2013**, *60*, 1–12. [CrossRef]
- Del Vecchio, P.; Mele, G.; Ndou, V.; Secundo, G. Open innovation and social big data for sustainability: Evidence from the tourism industry. *Sustainability* **2018**, *10*, 3215. [CrossRef]
- Tumasjan, A.; Sprenger, T.; Sandner, P.; Welpe, I. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the International AAAI Conference on Web and Social Media, Washington, DC, USA, 23–26 May 2010; Volume 4.
- Gu, Y.; Qian, Z.S.; Chen, F. From Twitter to detector: Real-time traffic incident detection using social media data. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 321–342. [CrossRef]
- Kumar, S.; Kumar, N.; Vivekadhish, S. Millennium development goals (MDGS) to sustainable development goals (SDGS): Addressing unfinished agenda and strengthening sustainable development and partnership. *Indian J. Community Med. Off. Publ. Indian Assoc. Prev. Soc. Med.* **2016**, *41*, 1. [CrossRef]
- Mihi-Ramírez, A.; García-Rodríguez, Y.; Cuenca-García, E. Innovation and international high skilled migration. *Eng. Econ.* **2016**, *27*, 452–461. [CrossRef]
- Pedercini, M.; Zuellich, G.; Dianati, K.; Arquitt, S. Toward achieving sustainable development goals in Ivory Coast: Simulating pathways to sustainable development. *Sustain. Dev.* **2018**, *26*, 588–595. [CrossRef]
- Vladimirova, K.; Le Blanc, D. Exploring links between education and sustainable development goals through the lens of UN flagship reports. *Sustain. Dev.* **2016**, *24*, 254–271. [CrossRef]
- Fuster Morell, M.; Espelt, R. A framework for assessing democratic qualities in collaborative economy platforms: Analysis of 10 cases in Barcelona. *Urban Sci.* **2018**, *2*, 61. [CrossRef]

17. Hubert, M.; Blut, M.; Brock, C.; Backhaus, C.; Eberhardt, T. Acceptance of smartphone-based mobile shopping: Mobile benefits, customer characteristics, perceived risks, and the impact of application context. *Psychol. Mark.* **2017**, *34*, 175–194. [CrossRef]
18. Porcher, S.; Renault, T. Social distancing beliefs and human mobility: Evidence from Twitter. *PLoS ONE* **2021**, *16*, e0246949. [CrossRef]
19. Mellon, J.; Prosser, C. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Res. Politics* **2017**, *4*. [CrossRef]
20. Pavalanathan, U.; Eisenstein, J. Confounds and consequences in geotagged Twitter data. *arXiv* **2015**, arXiv:1506.02275.
21. Malik, M.M.; Lamba, H.; Nakos, C.; Pfeffer, J. Population bias in geotagged tweets. In Proceedings of the Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015.
22. Palomino, M.; Taylor, T.; Göker, A.; Isaacs, J.; Warber, S. The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to “nature-deficit disorder”. *Int. J. Environ. Res. Public Health* **2016**, *13*, 142. [CrossRef] [PubMed]
23. Bennett, D.; Yábar, D.P.B.; Saura, J.R. University incubators may be socially valuable, but how effective are they? A case study on business incubators at universities. In *Entrepreneurial Universities*; Springer: Berlin, Germany, 2017; pp. 165–177.
24. John, L.K.; Emrich, O.; Gupta, S.; Norton, M.I. Does “liking” lead to loving? The impact of joining a brand’s social network on marketing outcomes. *J. Mark. Res.* **2017**, *54*, 144–155. [CrossRef]
25. Saura, J.R.; Palos-Sanchez, P.; Rios Martin, M.A. Attitudes expressed in online comments about environmental factors in the tourism sector: An exploratory study. *Int. J. Environ. Res. Public Health* **2018**, *15*, 553. [CrossRef] [PubMed]
26. Ekenga, C.C.; McElwain, C.A.; Sprague, N. Examining public perceptions about lead in school drinking water: A mixed-methods analysis of Twitter response to an environmental health hazard. *Int. J. Environ. Res. Public Health* **2018**, *15*, 162. [CrossRef]
27. Zhou, X.; Chen, L. Event detection over twitter social media streams. *VLDB J.* **2014**, *23*, 381–400. [CrossRef]
28. Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the International Conference on Language Resources and Evaluation, Valletta, Malta, 17–23 May 2010.
29. Chisholm, E.; O’Sullivan, K. Using Twitter to explore (un) healthy housing: Learning from the# Characterbuildings campaign in New Zealand. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1424.
30. Tlebere, T.; Scholtz, B.; Calitz, A.P. Using social media to improve environmental awareness in higher education institutions. In *Information Technology in Environmental Engineering*; Springer: Berlin, Germany, 2016; pp. 101–111.
31. Cao, X.; MacNaughton, P.; Deng, Z.; Yin, J.; Zhang, X.; Allen, J.G. Using Twitter to better understand the spatiotemporal patterns of public sentiment: A case study in Massachusetts, USA. *Int. J. Environ. Res. Public Health* **2018**, *15*, 250. [CrossRef]
32. Worldmeter. 2021. Available online: <https://www.worldometers.info/world-population/> (accessed on 1 March 2021).
33. Plastic Statistics. 2018. Available online: <https://oceancrusaders.org/tag/plastic-statistics/> (accessed on 1 March 2021).
34. Burgess, R.M.; Ho, K.T.; Mallos, N.J.; Leonard, G.H.; Hidalgo-Ruz, V.; Cook, A.M.; Christman, K. Microplastics in the aquatic environment—Perspectives on the scope of the problem. *Environ. Toxicol. Chem.* **2017**, *36*, 2259–2265. [CrossRef]
35. FE Online. 2018. Available online: <https://www.financialexpress.com/india-news/which-countries-are-way-ahead-of-india-in-curbng-plastic-pollution/1201608/> (accessed on 1 March 2021).
36. Kamaruddin, R.; Yusuf, M.M. Selangor Government’s “No plastic Bag Day” Campaign: Motivation and acceptance level. *Procedia-Soc. Behav. Sci.* **2012**, *42*, 205–211. [CrossRef]
37. Plastic Bags Will Be Charged RM1 Each In Penang Starting January 2021—Penang Foodie. 2021. Available online: <https://penangfoodie.com/> (accessed on 1 February 2021).
38. Smith, L.R.; Pegoraro, A.; Cruikshank, S.A. Tweet, retweet, favorite: The impact of Twitter use on enjoyment and sports viewing. *J. Broadcast. Electron. Media* **2019**, *63*, 94–110. [CrossRef]
39. Phua, J.; Jin, S.V.; Kim, J. The roles of celebrity endorsers’ and consumers’ vegan identity in marketing communication about veganism. *J. Mark. Commun.* **2020**, *26*, 813–835. [CrossRef]
40. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
41. Cho, H.W. Topic Modeling. *Osong Public Health Res. Perspect.* **2019**, *10*, 115. [CrossRef] [PubMed]
42. Liu, L.; Tang, L.; Dong, W.; Yao, S.; Zhou, W. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus* **2016**, *5*, 1–22. [CrossRef]
43. Liu, X.; Burns, A.C.; Hou, Y. An investigation of brand-related user-generated content on Twitter. *J. Advert.* **2017**, *46*, 236–247. [CrossRef]
44. Geva, H.; Oestreicher-Singer, G.; Saar-Tsechansky, M. Using retweets when shaping our online persona: Topic modeling approach. *MIS Q.* **2019**, *43*, 501–524. [CrossRef]
45. Jiang, S.; Qian, X.; Shen, J.; Mei, T. Travel recommendation via author topic model based collaborative filtering. In *International Conference on Multimedia Modeling*; Springer: Berlin, Germany, 2015; pp. 392–402.
46. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [CrossRef]
47. Lindstedt, N.C. Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature, 2005–2017. *Soc. Curr.* **2019**, *6*, 307–318. [CrossRef]
48. Teh, Y.W.; Jordan, M.I.; Beal, M.J.; Blei, D.M. Sharing clusters among related groups: Hierarchical Dirichlet processes. In Proceedings of the Advances in Neural Information Processing Systems Conference, Vancouver, BC, Canada, 13–18 December 2004.

49. Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126. [[CrossRef](#)]
50. Févotte, C.; Idier, J. Algorithms for nonnegative matrix factorization with the β -divergence. *Neural Comput.* **2011**, *23*, 2421–2456. [[CrossRef](#)]
51. Shi, T.; Kang, K.; Choo, J.; Reddy, C.K. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 1105–1114.
52. Roberts, M.E.; Stewart, B.M.; Tingley, D. Stm: An R package for structural topic models. *J. Stat. Softw.* **2019**, *91*, 1–40. [[CrossRef](#)]
53. Chae, B.K.; Park, E.O. Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling. *Sustainability* **2018**, *10*, 2231. [[CrossRef](#)]
54. Roberts, M.E.; Stewart, B.M.; Airoldi, E.M. A model of text for experimentation in the social sciences. *J. Am. Stat. Assoc.* **2016**, *111*, 988–1003. [[CrossRef](#)]
55. Milner, H.V.; Tingley, D. *Sailing the Water's Edge*; Princeton University Press: Princeton, NJ, USA, 2015.
56. Davidson, T.; Bhattacharya, D. Examining racial bias in an online abuse corpus with structural topic modeling. *arXiv* **2020**, arXiv:2005.13041.
57. Chandelier, M.; Steuckardt, A.; Mathevet, R.; Diwersy, S.; Gimenez, O. Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling. *Biol. Conserv.* **2018**, *220*, 254–261. [[CrossRef](#)]
58. Vanhala, M.; Lu, C.; Peltonen, J.; Sundqvist, S.; Nummenmaa, J.; Järvelin, K. The usage of large data sets in online consumer behaviour: A bibliometric and computational text-mining-driven analysis of previous research. *J. Bus. Res.* **2020**, *106*, 46–59. [[CrossRef](#)]
59. Roberts, M.E.; Stewart, B.M.; Tingley, D.; Lucas, C.; Leder-Luis, J.; Gadarian, S.K.; Albertson, B.; Rand, D.G. Structural topic models for open-ended survey responses. *Am. J. Political Sci.* **2014**, *58*, 1064–1082. [[CrossRef](#)]
60. Netzer, O.; Feldman, R.; Goldenberg, J.; Fresko, M. Mine your own business: Market-structure surveillance through text mining. *Mark. Sci.* **2012**, *31*, 521–543. [[CrossRef](#)]
61. Fan, W.; Wallace, L.; Rich, S.; Zhang, Z. Tapping the power of text mining. *Commun. ACM* **2006**, *49*, 76–82. [[CrossRef](#)]
62. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*; Sage Publications: Thousand Oaks, CA, USA, 2018.
63. Teh, P.L.; Yap, W.L. GoVegan: Exploring motives and opinions from tweets. In *World Conference on Information Systems and Technologies*; Springer: Berlin, Germany, 2021; pp. 3–12.
64. Kenski, K.; Filer, C.R.; Conway-Silva, B.A. Lying, liars, and lies: Incivility in 2016 presidential candidate and campaign tweets during the invisible primary. *Am. Behav. Sci.* **2018**, *62*, 286–299. [[CrossRef](#)]
65. Wong, S.C.; Teh, P.L.; Cheng, C.B. How different genders use profanity on Twitter? In Proceedings of the 4th International Conference on Compute and Data Analysis (ICDDA 2020), Silicon Valley, CA, USA, 9–12 March 2020; pp. 1–9.
66. Zhang, M.; Geng, G.; Zeng, S.; Jia, H. Knowledge Graph Completion for the Chinese Text of Cultural Relics Based on Bidirectional Encoder Representations from Transformers with Entity-Type Information. *Entropy* **2020**, *22*, 1168. [[CrossRef](#)] [[PubMed](#)]
67. Xiang, Z.; Schwartz, Z.; Gerdes, J.H., Jr.; Uysal, M. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int. J. Hosp. Manag.* **2015**, *44*, 120–130. [[CrossRef](#)]
68. Silva, A.L.P.; Prata, J.C.; Walker, T.R.; Duarte, A.C.; Ouyang, W.; Barcelò, D.; Rocha-Santos, T. Increased plastic pollution due to COVID-19 pandemic: Challenges and recommendations. *Chem. Eng. J.* **2021**, *405*, 126683.
69. Yang, Y.; Pan, B.; Song, H. Predicting hotel demand using destination marketing organization's web traffic data. *J. Travel Res.* **2014**, *53*, 433–447. [[CrossRef](#)]
70. Calafat, A.M.; Weuve, J.; Ye, X.; Jia, L.T.; Hu, H.; Ringer, S.; Huttner, K.; Hauser, R. Exposure to bisphenol A and other phenols in neonatal intensive care unit premature infants. *Environ. Health Perspect.* **2009**, *117*, 639–644. [[CrossRef](#)] [[PubMed](#)]
71. Gregory, M.R. Environmental implications of plastic debris in marine settings—Entanglement, ingestion, smothering, hangers-on, hitch-hiking and alien invasions. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 2013–2025. [[CrossRef](#)] [[PubMed](#)]