Exploring public interest from Twitter in 2021 using natural language processing for post-2020 biodiversity conservation strategies

Shimon Ohtani (s-ohtani@g.ecc.u-tokyo.ac.jp)
The University of Tokyo https://orcid.org/0000-0001-8414-9248

Research Article

Keywords: biodiversity, conservation, public interest, post-2020, Twitter, NLP

Posted Date: March 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1468450/v1

Abstract

The need to develop more effective biodiversity conservation strategies for the post-2020 period is now being called for both globally and in individual countries. As an example of analysis that should serve as a clue for this purpose, this paper presents a methodology for exploring public interest in “biodiversity” by analyzing text posted on Twitter in 2021 using natural language processing, which is one of the big data analyses that have been rapidly advancing in recent years. First, the frequency of idioms was explored by aggregating bigrams in contexts where the word “biodiversity” is used, and then the tweet data set was classified into 40 meaningful topics and defined by LDA topic modeling. In addition, two ways of sentiment analysis by NRC emotion lexicon and VADER were used to visualize the rough emotional trends that can be read from the data set. And I then selected a topic on “Extinction of Species” for intensive discussion, and also picked up and discussed tweets about “30 by 30” as an example of a cross-topic analysis. However, developing measures to compensate for the weaknesses of incompleteness and non-representativeness of big data is also a major challenge. In the future, cultivating interdisciplinary knowledge and promoting collaboration among researchers in different fields and deep social networking literacy to enable a hybrid analysis of multiple data sets will be essential for effective biodiversity conservation strategies.

Introduction

The United Nations Decade on Biodiversity (2010–2020) has been replaced by the United Nations Decade on Ecosystem Restoration (2021–2030), and various measures are already being considered at the global level. Its aim is to halt the degradation of ecosystems and restore them to achieve global goals (UN Decade on Ecosystem Restoration 2022). In October 2021, the first part of the 15th Conference of the Parties (COP15) to the Convention on Biological Diversity (CBD) was held in Kunming, China, after being postponed from the previous year due to the COVID-19 pandemic. The “Kunming Declaration” was adopted during the high-level segment, and a new roadmap for the post-2020 period (i.e., the next 10 years) will be formulated during the second part of COP15, which is scheduled to be held from April to May 2022. Here, reflection on the failure to meet the Aichi Biodiversity Targets established in 2010 calls for multiple targets to be defined as a whole and set at the highest level of ambition (Díaz et al. 2020). Specifically, it is expected that each party will commit to goals that are aware of the Sustainable Development Goals (SDGs) set by the United Nations and that focus on climate change, cities, supply chains, and consumption.

However, no matter how ambitious and precise the goals are set, if they are discussed only among policymakers and a few activists, and are out of touch with the general public, the previous failure will be repeated. While it is known that numerous attempts have been made by policymakers, activists, and researchers to implement the CBD and realize its international goals, it is by no means easy to explore the extent to which these efforts have been shared with an unspecified number of the general public. In fact, how much of the general public is aware of COP15, the set of goals that are being developed, or even the existence of the Convention on Biological Diversity itself? It is undeniable that the term “biodiversity,” which seems to be well established among experts, is very ambiguous, and the general public is not familiar with it and may find it difficult to grasp. Ignoring the existence of such gaps can render noble goals useless. To this end, it is essential to first measure in some way the level of public interest in biodiversity at this stage.

Traditionally, this method of exploring public perceptions has been in the form of questionnaires. However, with the evolution of information technology, it has become known that unexpected findings can be obtained through the analysis of big data. Natural Language Processing (NLP) is the subfield of computer science concerned with using computational techniques to learn, understand, and produce human language content (Hirschberg & Manning 2015). In particular, Social Networking Service (SNS) has been the target of analysis for more than 10 years as a platform where people can write short messages about personal events and thoughts, and its base is gradually expanding with the development of NLP. The purpose of this paper is to comprehensively analyze trends in Twitter posts containing the word “biodiversity” during the year 2021 to understand the recent social networking discourse on “biodiversity” and to provide clues for creating a roadmap for reaching new targets to be achieved by 2030 and a long-term goal by 2050. In particular, due to its ambiguity (Mayer 2006), the term “biodiversity” is currently used in the context of a wide variety of topics, even on Twitter, a platform where an unspecified number of people post. This makes it impossible to ascertain details through mere keyword searches. Therefore, this study focused on what interests of the public can be read from Twitter posts by classifying them by topic using NLP, especially machine learning analysis methods. Many remain skeptical about analyzing SNS, including Twitter (e.g., Alizadeh 2021). However, it is something that can be refined through the accumulation of research. Although this is a developing study, it is worthwhile to provide an example here.

Literature Review

The reality is that the search for public interest in big data has grown rapidly in the last decade, and the methods used have become more sophisticated over the years. Twitter and Google Trends stand out as two important sources of social data. Online data sources provide fresh insights into online sociological research (Kostakos 2018). Although there are still few examples of analyzing social networking and other big data to explore people’s awareness and public interest in biodiversity, Troumbis (2019) provides an example of analyzing public interest in “biodiversity” using Google Trends. In addition, Barrios-O’Neill (2021) selects and analyzes advocacy content on biodiversity threats from major NGOs on Twitter, along with sentiment trends. Another analysis using Twitter and Google Trends together is Cooper et al. (2019). This study, which proposed a way to measure the first of the 20 Aichi biodiversity targets set in 2010 (increasing people’s awareness), was groundbreaking in its use of Twitter and Google Trends together. More recently, a detailed analysis of Twitter discourse on marine plastic pollution has been conducted using topic and sentiment analysis (Otero, et al. 2021). The detailed analysis in which, while overcoming as much as possible the current weaknesses of Twitter analysis, is remarkable and may indicate a direction for similar analysis using natural language processing in the future.

These studies were designed to take advantage of the unique characteristics of SNS and big data, focusing on the unique subject matter and providing unprecedented findings. However, these studies were somewhat limited in their research subjects and did not take full advantage of the benefits of each platform. This paper aims to explore public interest in a more multifaceted, big-picture perspective on “biodiversity. Therefore, all tweets containing “biodiversity” were included, regardless of whether they contained location information or not. In addition, because of lingering doubts about the accuracy of machine translation and to avoid losing the nuance of the posters’ word choice in the sentiment analysis, only tweets with English text were extracted from the collected tweets for the analysis this time. To present as many results as possible, technical descriptions are kept to a minimum to provide a guideline for future policies on biodiversity. Before this study, I obtained an account for the Application Programming Interface (API) V2 provided by Twitter and obtained hundreds of thousands of tweets from the full Twitter archive. There has been no research in this field that has analyzed such a large set of data.

This paper first details the research methodology used in this study in the Methods section presents a visualization of the results in the Results section, reviews the interpretation of the results and the contributions and limitations of this paper in the Discussion section, and summarizes and discusses future directions in the Conclusion section.

Contribution

This paper is intended to be a practical and academic contribution to multiple fields. First, as mentioned at the beginning of this paper, I present results that could help to effectively implement the CBD, which is scheduled to establish new targets for the next decade. Second, to provide an example of analysis of the implementation of NLP into global policy issues, thereby promoting the integration and elaboration of policy issues and computer science. Third, by exploring recent discourses and emotional trends on Twitter on a specific topic, in this case, the conservation of endangered species, in particular, I provide useful information for researchers and activists who are currently engaged in species conservation.

Methods

As noted above, since this study will focus its analysis on Twitter posts, I will first briefly touch on the characteristics, usefulness, and dangers of Twitter. Twitter and other microblogs have become very popular communication tools among Internet users, with millions of messages posted daily. As contributors write about their lives, exchange opinions on a variety of topics, discuss current issues, post about products and services they use, and express their political and religious views, microblogging websites have become a valuable source of people’s opinions and feelings. And these data can be used efficiently for marketing and social research (Pak and Paroubek 2010). Twitter is one of the largest social media platforms on which people express their opinions and feelings on a variety of topics. With 330 million monthly active users, 1.3 billion accounts, and 500 million tweets (240-character microblogs) published daily, Twitter is said to make it easy for users to find specific topics by quickly and on-the-fly creating shared hashtags related to specific events (Morshed et al. 2021).

In exploring public interest, the major difference with Google Trends is that Google Trends requires the analysts to infer public interest from search volume and related keywords, regardless of user intent. Twitter, on the other hand, allows analysts to get a multifaceted and specific read on user sentiment not only by the volume of tweet posts on specific issues but also by more direct text and image posts, as well as retweets and likes. Although the time range available for acquisition for analysis has been limited, its utility has expanded in recent years as certain users have been allowed access to the full archive for the entire period since the service was launched.

However, we should be careful about equating the discourse space on Twitter with the actual discourse space. The bias of user attributes, individual differences in tweet frequency, the existence of bots that automatically repeat certain posts, and a lot of useless noise other than text can hinder the analysis in various ways (Otani 2021). Therefore, when using Twitter as a target for analysis, it is necessary to keep the above limitations and restrictions in mind and develop measures to minimize them from the pre-processing stage of the data set.

Research Design

With the above in mind, the methodology of this study basically follows that of Ohtani (2021), that analyzed the discourse related to “biodiversity” on Twitter from 2010 to 2020. First, I pre-process the acquired data set, count the bigrams, and present the top 20. Next, I constructed optimal topic modeling by Latent Dirichlet Allocation (LDA) (Blei et al. 2003) using machine learning algorithms. In addition, lexicon-based sentiment analysis, the NLP technique, was used to explore the emotional tendencies of each topic. In doing so, two different methods were used to compensate for the uncertainty of the sentiment analysis. Here, in addition to emotion analysis with the NRC emotion lexicon (Mohammad and Turney 2013), the Valence Aware Dictionary for Sentiment Reasoning (VADER) sentiment analysis (Hutto and Gilbert 2014) was also conducted. Subsequently, as examples of analysis based on the information visualized by these methods, a detailed analysis of a specific topic and a cross-topic analysis for a specific policy theme was conducted.

Data Collection

The analysis of this study was based on all tweets with English text containing “biodiversity” posted on Twitter from January 1, 2021, to December 31, 2021. First, I retrieved all tweets containing the English word “biodiversity” using the Python programming language (ver. 3.9.7). This was possible because I already had access to the full Twitter archive through the Academic Research Access of Twitter’s API v2. As a result, a total of 596,731 tweets were obtained. Of these, I extracted 546,288 tweets with purely English text. Here, it was also considered to obtain only tweets containing location information and to conduct an analysis focusing on the region, but in this study, the emphasis was on obtaining public interest from a larger number of tweets and not on the location of the posters. For reference, I present a

16-year trend of the number of tweets from 2006, when Twitter launched its service, to 2021 (Fig. 1). This shows that the number of tweets containing “biodiversity” has increased significantly since 2018. Of course, this alone cannot be taken to indicate an increase in public interest in biodiversity, but it does provide insight.

In this study, this group of tweets in English text in 2021 was pre-processed and narrowed down following Ohtani (2021). The procedure is as follows.

1. Removed tweets with duplicate text.
2. Removed @usernames and links (pasted URLs such as http and www) in the text.
3. Removed special characters and punctuations from the text.
4. Other strings that did not have any particular meaning were excluded by designating them as “stop words”.
5. The texts in the above state were saved for sentiment analysis.
6. Also removed hashtagged words.
7. Tokenized the texts. Deleted tweets with less than 3 tokens.
8. n-grams (bigrams) were counted and saved.
9. Performed lemmatization of the tokens.

After preprocessing, 87,854 tweets, corresponding to 16%, were removed, and 458,434 tweets, corresponding to 84%, were extracted for the final analysis. It is important to note that although there are a wide variety of preprocessing methods, none of them are perfect. The inability to remove all disturbing elements should also be recognized as a current limitation of NLP.

Counting n-gram (bigram)

During the year 2021, I searched for the use of the word “biodiversity” with any word or any idiom in the context of the use of the word “biodiversity”. N-gram is a sequence of *n* contiguous parts from a given sequence of text, and it usually is obtained from a text (Ahmad et al. 2021). That is, a sequence of two words is called a bigram and a sequence of three words is called a trigram. Sometimes simply observing this n-gram is enough to quickly get a rough idea of the discourse on Twitter. In this paper, the top 20 bigrams were obtained and observed using the programming language Python library, NLTK, Gensim.

LDA Topic Modeling

In order to explore the public interest in biodiversity, it is essential to identify not only the words that are used together but also the context in which they are used and the topics in which they are discussed to conduct a more precise analysis. The need for such an approach increases especially in the analysis of Twitter, where all kinds of content are posted indiscriminately. The method used to classify such indiscriminate postings into meaningful topics is topic modeling.

Topic modeling, one of the most popular text mining methods, is an efficient and systematic approach to analyze thousands of documents in a few minutes (Karami et al. 2020). It is one of the unsupervised machine learning methods that propose patterns or clusters of similar expressions without outlining topic tags or prior data learning. It helps place documents into observed themes and determine hidden themes within groups, while presenting a way to automatically organize, knowledgeize, explore, and review large documents (Ahmad et al. 2021). LDA is a type of unsupervised learning algorithm used for topic modeling and one of the most common topic modeling methods. This method aims to detect the topic to which a document belongs based on the words it contains and can discover the semantic content of a document by deconstructing the latent semantic structure inside the document. On the other hand, it can also predict new document topics that are invisible to the eye without the need for labeling or training sets and can be effectively applied to huge text data to discover semantic patterns (Ahmad et al. 2021). In this study, I used the Python library Gensim and the java open software MALLET to search for and visualize the optimal topic model. In addition, each topic was defined based on the keywords and textual content of each topic.

Sentiment Analysis

After categorizing groups of tweets by meaningful topic, exploring their affective tendencies through sentiment analysis is a method that has increased in recent years due to its perceived usefulness in exploring public interest (see e.g., Xue et al. 2020). Lexicon-based approaches rely primarily on the sentiment lexicon, a collection of known and compiled sentiment terms, phrases, and even idioms developed for traditional communication genres, such as the Opinion Finder lexicon (Kharde and Sonawane 2016). Although the detailed analysis methods and visualization techniques vary widely, here again, I basically followed Ohtani (2021) and conducted the analysis using the NRC emotion lexicon (Mohammad and Turney 2013). This is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive), developed by a crowdsourced task for tens of thousands of English words, manually curated, and encoded with emotions (Mohammad 2020).

In addition, another lexicon-based sentiment analysis was conducted in this study to improve the reliability of the sentiment analysis. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media and works well on texts from other domains (Oyewusi et al. 2020). And it is empirically validated by multiple independent human judges, which incorporates a "gold-standard" sentiment lexicon that is especially attuned to microblog-like contexts (Hutto and Gilbert 2014). The specific analysis procedure was to assign Positive, Negative, and Neutral scores between 1 and − 1 according to the VADER Lexicon to all tweets to be analyzed after preprocessing and to calculate a Compound score that combined these scores. Those with a score of 0.05 or higher were assigned a "1" as Positive Sentiment, those with a score of -0.05 or lower were assigned a "-1" as Negative Sentiment, and the intermediate portion that did not fall under any of these categories was assigned a "0" as Neutral. However, the words Positive, Negative, and Neutral are vague and ambiguous. Especially when it comes to emotional tendencies toward deep concepts, as in this study, the information that can be obtained only by a simple assignment of polarity is limited. Therefore, I will analyze the tweets by presenting the degree of polarity score as appropriate, while also including the original tweet text.

Specific research examples

This paper presents further specific research examples based on the results of the overall analysis and visualization. First, as an example of an in-depth analysis of a selected specific topic, I presented the results of a detailed analysis of the topic of “Species Extinction”. Second, as an example of analysis across multiple topics, I presented a detailed analysis of the way “30 by 30” was mentioned in tweets containing “biodiversity”.

Results

Aggregation of bigram

Table 1 lists the top 20 bigrams in English tweets containing “biodiversity” in 2021. The top two categories (‘climate’, ‘change’) and (‘biodiversity’, ‘loss’), as well as (‘biodiversity’, ‘conservation’), have consistently been among the top 20 for the past 12 years since 2010. For reference, these trends since 2010 are shown below (**Fig. 2**). It shows that (‘biodiversity’, ‘conservation’) has been slowly rising, while (‘biodiversity’, ‘loss’) and (‘climate’, ‘change’) have increased rapidly since 2018. This fact means that the frequency with which biodiversity loss is tweeted in relation to climate change has increased in the last three years. Furthermore, in Table 1, (‘biodiversity’, ‘crisis’) and (‘biodiversity’, ‘framework’) have also increased rapidly in recent years, from 2018 onwards. As shown in Fig. 1, the overall number of tweets has shown a clear upward trend since 2018, and these facts together suggest that since 2018, the number of posts expressing a sense of crisis by linking biodiversity loss to climate change issues has continued to increase.

Table 1
Top 20 bigrams used in English tweets containing “biodiversity” in 2021

bigram	count
(‘climate’, ‘change’)	26,906
(‘biodiversity’, ‘loss’)	24,815
(‘climate’, ‘biodiversity’)	12,737
(‘biodiversity’, ‘conservation’)	12,127
(‘biodiversity’, ‘crisis’)	7,830
(‘global’, ‘biodiversity’)	7,209
(‘change’, ‘biodiversity’)	7,021
(‘protect’, ‘biodiversity’)	6,728
(‘nature’, ‘biodiversity’)	6,607
(‘biodiversity’, ‘climate’)	6,533
(‘loss’, ‘biodiversity’)	6,336
(‘biodiversity’, ‘nature’)	4,093
(‘wildlife’, ‘biodiversity’)	3,759
(‘climate’, ‘crisis’)	3,756
(‘biodiversity’, ‘framework’)	3,725
(‘marine’, ‘biodiversity’)	3,700
(‘nature’, ‘wildlife’)	3,646
(‘protecting’, ‘biodiversity’)	3,559
(‘environment’, ‘biodiversity’)	3,436
(‘conservation’, ‘biodiversity’)	3,405

Results of LDA Topic Modeling

When conducting LDA, the number of topics must be determined in advance. In this study, I explored the Coherence Score (Röder et al. 2015), a common measure for determining the number of topics, with varying numbers of topics. After visually checking the distribution of the topics, I determined that the best way to classify the topics in 2021 is to use the Java open software MALLET to classify them into 40 topics. Figure 3 shows the distribution of 40 topics in 2021, visualized in PCoA by pyLDAvis, a Python library. Table 2 also shows the keywords that make up the 40 topics and the number of tweets belonging to each topic. This resulted in a classification of 40 topics consisting of a maximum of 19,181 tweets, a minimum of 5,997 tweets, and an average of 11,461 tweets per topic.

Table 2
Contents of 40 topics in 2021 (Keywords, Tweet count)

Topic	Keywords	Tweet count
1	climate, change, crisis, tackle, address, emergency, loss, solve, challenge, fight	19,181
2	soil, farm, farmer, farming, agriculture, crop, agricultural, practice, organic, grow	17,140
3	food, carbon, system, reduce, emission, energy, production, security, sustainable, waste	12,795
4	stop, destroy, kill, destruction, money, industry, damage, profit, ignore, continue	15,358
5	loss, impact, environmental, pollution, issue, deforestation, problem, effect, lead, degradation	10,217
6	plant, garden, bee, insect, pollinator, grow, flower, native, wild, grass	17,565
7	water, provide, increase, benefit, improve, river, wetland, quality, reduce, clean	11,139
8	forest, tree, plant, native, fire, growth, planting, woodland, forestry, burn	15,330
9	development, sustainable, policy, management, strategy, base, approach, plan, gain, conservation	10,205
10	green, park, city, urban, space, build, area, site, create, design	13,918
11	global, cop, action, post, framework, target, goal, fornature, meet, achieve	13,638
12	join, talk, event, today, week, discuss, free, register, webinar, session	16,541
13	land, biodiversity, high, increase, huge, impact, cost, small, level, scale	7,537
14	specie, extinction, year, lose, threaten, population, decline, habitat, number, endanger	12,671
15	ecosystem, human, biodiversity, health, healthy, essential, maintain, depend, provide, society	8,517
16	project, support, conservation, biodiversity, fund, initiative, launch, work, partner, program	10,983
17	research, study, science, ecology, scientist, conservation, student, paper, show, biology	14,233
18	wildlife, animal, bird, nature, science, wild, trade, elephant, naturelover, mammal	14,560
19	protect, biodiversity, restore, ecosystem, preserve, effort, conserve, restoration, support, step	8,190
20	area, biodiversity, rich, region, hotspot, home, island, unique, large, country	12,065
21	datum, biodiversity, open, team, apply, information, survey, monitor, find, record	12,670
22	people, community, local, biodiversity, indigenous, land, protect, knowledge, amazon, culture	10,539

Topic	Keywords	Tweet count
23	biodiversity, understand, diversity, ecological, include, term, level, process, form, genetic	8,331
24	loss, climate, threat, global, change, face, risk, pandemic, covid, collapse	11,094
25	biodiversity, government, state, national, protection, environment, plan, public, environmental, vote	10,538
26	biodiversity, beautiful, amazing, visit, find, enjoy, photo, wonderful, beauty, love	13,265
27	biodiversity, good, point, question, reason, true, fact, case, word, interest	8,272
28	report, economic, biodiversity, business, economy, risk, finance, company, global, investment	12,393
29	life, planet, earth, save, biodiversity, live, world, human, future, humanity	11,337
30	biodiversity, work, great, share, opportunity, learn, group, youth, idea, bring	7,261
31	biodiversity, part, solution, role, important, today, importance, play, celebrate, happy	9,344
32	biodiversity, natural, resource, habitat, conservation, landscape, important, create, manage, contribute	5,997
33	biodiversity, make, time, year, place, happen, people, move, real, feel	6,416
34	environment, biodiversity, climatechange, sustainability, climateaction, sustainable, conservation, nature, climatecrisis, education	10,379
35	nature, biodiversity, rewilde, bring, recovery, positive, create, based_solution, build, heart	7,476
36	biodiversity, ocean, marine, fish, coastal, plastic, deep, fishery, fishing, blue	14,351
37	biodiversity, good, thing, time, give, love, care, hope, long, gulfkanawut	10,714
38	biodiversity, world, country, protect, action, leader, call, sign, lead, nation	7,287
39	biodiversity, read, article, late, news, check, story, full, link, interesting	11,578
40	biodiversity, learn, watch, start, year, today, video, find, activity, great	7,409

The keywords that make up each topic provide an overview of the discourse on “biodiversity” on Twitter in 2021. In this study, I attempted to define each topic from the tweet text in addition to the keywords. For example, Topic 1 was defined as “Urgent Action” based on the keywords “tackle,” “address,” and “emergency” and the content of the tweet text. Similarly, from the keywords “soil,” “farm,” and “agriculture,” Topic 2 is “Soil and Agriculture”.

Note that the main theme of all tweets was not necessarily “biodiversity”, since I collected all tweets containing the word “biodiversity” in this study. Some of them only mentioned “biodiversity” incidentally among other major themes. However, this approach was desirable for the purpose of exploring public interest in “biodiversity” because it allowed me to cover a wide range of contexts in which the word “biodiversity” is used.

Results of sentiment analysis by NRC emotion lexicon

For the group of tweets classified by topic by LDA, I first explored eight types of emotional tendencies for each topic using the NRC emotion lexicon. Figure 4 is a heat map showing the definitions given to the 40 topics and their emotional tendencies. Here it is simply created by the total number of words used that correspond to the sentiment defined by the lexicon. It is important to note that the NRC emotion lexicon tends to be relatively prone to the emotions “trust” and “anticipation” due to its specifications. Thus, “trust” and “anticipation” here should be viewed as indicating rather neutral emotional tendencies. The heat map shows that “joy” is prominent in Topic 6, named “Plants and Insects,” and Topic 8, named “Forest and Timber. Negative emotions such as “fear,” “sadness,” and “anger” were prominent in topics 1, 4, 5, and 14, which were defined as “Urgent Action,” “Harm,” “Issues and Impacts,” and “Extinction of Species,” respectively. In addition to Topic 4 and Topic 8, “disgust” is prominent in Topic 2, which is named “Soil and Agriculture. This is due to the fact that words such as bacteria and fungus related to soil are assigned to “disgust,” and the lexicon itself may need to be customized in this regard (Ohtani 2021). Similarly, “surprise” is also prominent in Topic 8, and this is also due to the fact that the words “young” and “tree,” which are frequently used in this topic, are counted as “surprise” in the lexicon. Thus, although there are some limitations, it is possible to understand the direction of Twitter posters’ awareness and interests by classifying indiscriminately posted tweets containing “biodiversity” in a meaningful way, defining each topic, and making it possible to list emotional trends.

Results of sentiment analysis by VADER

Figure 5 shows the distribution of VADER Compound scores for all tweets analyzed in 2021. This shows that approximately 60% of tweets containing “biodiversity” in 2021 indicate positive sentiment. The distribution of the Compound score also shows that more tweets with a high positive sentiment of 0.75 or higher are found. Presented below are examples of the original text of tweets with high positive and negative scores, respectively, from the Compound score.

Examples of positive tweets

Wow, what a year! Plagued with global turmoil & challenges, but still we’re thankful for the collective spirit of #creativity, #kindness, & love of #biodiversity in all who engaged w/ us. Enjoy this screen shot of the #Ireland & #Netherlands winners during the award ceremony (0.9913)

It’s a great state. Beautiful. Best State Parks anywhere. Best landscape. Best small towns. Best school systems (except Nashville). Best biodiversity. Best Governor. And Louisville isn’t here. (0.9884)

It’s truly a wonderful act of philanthropic benevolence, direct benefit for the environment, habitat, and biodiversity, I must praise the integrity of those striving selflessly to achieve this wonderful enterprise, and the kind generosity of those who wish to make it a success. (0.9882)

AMAZING people! I agree. Wonderful country. Far more equitable than. Great free education far all. Universal healthcare for all. Very cultured, well educated & kind people. Some of the highest levels of biodiversity on Earth, which they protect very well. Great food and music. (0.9868)

It was only just blooming as we broke up but they love it & we do a whole biodiversity module in Science so it's a fantastic learning resource as well as looking beautiful & doing our bit for nature. It's spectacular. Our students love the beautiful landscaping at our school (0.9865)

Examples of negative tweets

Besides, grouse & pheasant shoots are not about feeding people but numbers killed. Both environmentally damaging activities, negatively impact biodiversity, are associated with criminal activity (illegal killing of wildlife), use harmful lead shot & many killed birds are dumped. (-0.9913)

Exactly far worse in any uk slaughter house but behind closed doors they get away with excruciating cruelty. Wildlife or farm animal or pet same pain in fact farm animals farmfeed also kills wildlife all over the hence leading cause of the climate crisis & biodiversity loss (-0.9874)

These you should know but still i list themWorld Hunger, Poverty, Inequality, War on terror, world debt, nuclear weapon, Climate emergency, Biodiversity loss, waste, consumerism, resource depletion, Crimes, Human Right issues, health, consumerism (-0.9872)

Turning point our present way of living is not sustainable, rather it is self destructive. Evidence of excessiveness, greed, violence, fear, environmental & social disruption, biodiversity loss, climate crisis, human conflict & suffering, danger everywhere at every scale. (-0.9844)

I am telling it slightly differentlyClimate change is a pain. But so is Biodiversity crisis, extinction crisis & consumption crisis. U cant treat one crisis & leave others alone let alone aggravate the others for treating one with Climate UnicornsI ain't selling fossil fuels (-0.9834)

After comparing both polarized groups of tweets, it appears that positive tweets are optimistic, favorable, and friendly, while negative tweets are pessimistic, cynical, and hostile. However, it would not be possible to say that either positives or negatives are generally preferable. Thus, when looking at emotional tendencies in terms of polarity, there is a need to examine the specific tweet text based on the visualized information and with reference to the assigned score, because of the ambiguity involved.

In-depth analysis of the specific topic

It is also possible to perform VADER sentiment analysis on a topic-by-topic basis. In the following, one specific topic is discussed and an example of comprehensive analysis is given. Here, I examined in detail 12,671 tweets belonging to topic 14 named "Extinction of Species". This enabled me to explore what was said about species extinctions in tweets about "biodiversity" posted on Twitter in the year 2021.

First, Fig. 6 shows a radar chart of the percentage of use of the eight types of emotions in the NRC emotion lexicon for the emotional trends of this topic. Especially “fear” and “sadness” are prominent, and “trust” and “anticipation” and “anger” also stand out at a glance. Next, Fig. 7 shows the score distribution of the sentiment analysis of polarity by VADER. According to this, 45% of tweets on this topic are deemed negative, while at the same time 37% of tweets are deemed positive. It is also necessary to determine what differences exist between the two. Below are a few tweets picked out for their respective positive and negative scores.

Examples of positive tweets

Help us measure invasive species awareness in BC and enter to win a $250 prize! The information you provide will help us protect BC’s beautiful natural areas and biodiversity from the effects of non-native invasive species. Visit #BCinvasives (0.9674)

Yesterday we braved the rain in search of the ideal habitat for the elusive and #heathfritillary #butterfly. We hope to help create more of it to try to help bring this beautiful species back from the brink #conservation #biodiversity #wildlife #nature #environment #devon (0.9618)

The best action we can take at this moment, is to promote biodiversity. We can promote food growth and biodiversity at the same time. We just have to be smart about it. For instance, use native species to do their jobs. Or better, promote reservations for endangered species. (0.9595)

Wow - fantastic - thank you kind people! Please RT & donate friends. We can do this. Let’s make a difference! Time is running out for our endangered species; for the land that is losing its biodiversity & vigour & for our planet #ClimateCrisis #COP26 #rewilding #vegan (0.9558)

Today is Endangered Species Day. Help restore and protect the world’s great forests and their wealth of biodiversity to prevent extinctions. $40 sponsors 1/4 acre of rainforest in Sumatra. Visit our website to show some respect for nature. (0.9517)

Human activity occurs on a faster scale than evolution. We’ve radically altered a lot of environments but haven’t gotten fun new species! We’re evolved to like the biodiversity that exists (e.g. “rainforests are pretty”) so preserving and restoring those seems like a better path (0.9391)

Examples of negative tweets

Irresponsible tweet We are in the sixth mass extinction crisis. There is unprecedented ecosystems destruction spreading everywhere by dumb humans. There is a biodiversity crisis going on. There is an insect apocalypse going on. Wild animals are down by 80%, just 20% remain. (-0.959)

Heartbreaking tragedy. This was a very visible extinction. We are also losing species we have not even documented. Think soil biodiversity. “Man’s war against nature is inevitably a war against himself.” Rachel Carson (-0.9571)

Monarch butterflies are beautiful creatures, but we could lose them forever. Their populations are in serious decline from climate change and severe habitat loss. Losing them could destabilize entire biomes and wreak havoc on biodiversity. Spread the word. (-0.9562)

SIGN: Pass New Bill to Stop America’s #Endangered Animals from #Extinction: Over 31,000 threatened & endangered species could face extinction in the US amid an ongoing devastating #biodiversity crisis fueled by reckless human activity & #climatechange. (-0.946)

51 Years the USA has had an endangered species a Puget sound orca in absolute illegal suffering. Shameful the leading driver of the climate crisis is lack of biodiversity lack of habitats what is the gov doing? Looks like nothing won’t recover without wildlife (-0.9442)

Loss of biodiversity is down to the destructive impact of just 1 specie while there are millions of victim species that are threatened and endangered with imminent further extinctions. Reducing the level of threat by limiting and inhibiting the impact of the aggressor is priority (-0.9432)

Looking at the sample of tweets for each, it is clear that the positive ones are marked by words of gratitude and respect, whereas the negative ones are more about raising a sense of crisis, along with somewhat extreme language. Rather than deciding which is more desirable, we should consider that understanding what language users use to express their opinions and feelings on a particular issue can help guide us in exploring the direction of public interest.

Figure 8 shows a co-occurrence network diagram of the top 150 most frequently occurring words in Topic 14, divided into positive and negative tweets. KH Coder (ver. 3.Beta.04a) was used for drawing. The software’s default setting (the Jaccard coefficients calculated) was used here (Higuchi 2016). Degree 1 is a group of words specific to each positive and negative, while Degree 2 is a common group of words. According to this, Degree 1’s positive tweets use words for relief, such as “save,” “help,” and “conservation,” as well as words for space, such as “area”, “forest”, “earth”, and “land”. Conversely, negative tweets used words such as “disappear”, “threat”, and “crisis”, as well as “cause” and “due”, suggesting that they were referring to the cause of the threat. Degree 2 shows that words other than “biodiversity” were also used to refer to objects of protection, such as “species,” “habitat,” “wildlife,” “animals,” and “plants,” while words such as “extinction,” “have,” and “threaten” were commonly used in both positive and negative tweets. From these results, it can be roughly ascertained that positive tweets primarily advocate the protection of extinct species, whereas negative tweets warn of the danger or threat of extinction. In summary, I could conclude that among the topics related to “biodiversity” posted on Twitter in 2021, those related to “Extinction of Species” accounted for more than half of the tweets expressing “fear” and “sadness” over the danger or threat to species, and a significant number of tweets emphasizing the importance of species and advocating their conservation.

Cross-topic analysis

As an example of an analysis that crosses multiple topics for a specific policy keyword, I analyzed what discourse can be extracted from Twitter about “30 by 30”, which has been discussed in recent years in the context of biodiversity conservation policy, and what public interest can be explored.

30 by 30 (or 30x30) is a global initiative by governments to designate 30% of the Earth’s land and sea areas as protected areas by 2030, as proposed by the Science article “*A Global Deal For Nature: Guiding principles, milestones, and targets*” (Dinerstein et al. 2019). Launched by the High Ambition Coalition in 2020, more than 50 countries had agreed by January 2021, expanding to more than 70 by October of the same year (“30 by 30,” 2021). And it is being considered as a major goal of the proposed “Post-2020 Biodiversity Framework” and will be finalized in the second part of COP15 to be held in April-May 2022.

First, I extracted tweets that contained the keywords “30 by 30” and “30 x 30” in the text or hashtags, as well as tweets where it was clear from the text context that this was the case, from the entire data set analyzed in this study. As a result, 93 tweets were found to be applicable. Figure 9 shows the topic distribution of these 93 tweets. This shows that the tweets are distributed among 25 of the total 40 topics. Topic38, entitled “Leaders’ Commitment,” had the most tweets, with 38 tweets, followed by topic19, entitled “Protection and Restoration,” with 11 tweets. More than half of all tweets were on these two topics, which is not surprising given the gist of the “30 by 30”. In addition, 18 of the 25 topics had two tweets or less.

Figure 10 is a network diagram showing the co-occurrence relationship of frequently occurring words from the 93 tweets. KH Coder (ver. 3.Beta.04a) was used for drawing. The software’s default settings (the Jaccard coefficients calculated, and sub-graphs detection based on [Communities: modularity]) were used here and displayed only the minimum spanning tree (Higuchi 2016). By looking at the network from sub-graph 1 to 2, we can get a rough idea of the main content of the tweets. The majority of tweets outlined the “30 by 30” goal set by the High Ambition Coalition (HAC) for Nature and People, an intergovernmental group. In addition, sub-graph 3 revealed several tweets, particularly about the recent “30 by 30” initiative in Canada.

Figure 11 shows a radar chart of the emotional tendencies of the 93 tweets, classified into 8 types by the NRC emotion lexicon, and Fig. 12 shows the distribution of sentiment polarity scores by VADER. According to these results, “trust” and “anticipation” stand out, with the majority of tweets (83 out of 93) being assigned to the positive category. The top three tweets in this category with positive scores are as follows.

Topic 19 (“Protection and Restoration”)

But again, *where* should all these new protected areas be created? We need to make smart decisions about where/how to protect biodiversity & the benefits people get from nature. Now has committed to protecting 25% by 2025 and 30% by 2030! We gotta get this right. (0.9527)

Topic 38 (“Leaders’ Commitment”)

Listening live to some strong international contributions at the one planet summit on global biodiversity Very glad that Ireland will join the countries in High Ambition Coalition & also commit to ensuring 30% protected areas by 2030 (0.9324)

Topic 16 (“Funds and Support”)

This is great! Does it include purchasing land for sale by nonprofits such as #CampWoolsey thatare selling? So many precious places like it to protect for generations to come#Heritage #Climate #biodiversity #30by30 (0.9256)

Conversely, the top three negative scores are as follows.

Topic 5 (“Issues and Impacts”)

The impact of ecosystem destruction remains underestimated. From biodiversity to conflict or supply chain shocks it is not a very long way. #OnePlanetSummit #biodiversity #ForNature #CoP15 #30x30 (-0.8658)

Topic 4 (“Harm”)

Susan Eisenhower reminds us that the US Capitol is not the only sacred American site under attack on this fateful day: #biodiversity #30x30 #50x50 Biden Must Stop Trump’s Reckless Plunder in Alaska (-0.7906)

Topic 5 (“Issues and Impacts”)

yes - agreed- one of the key failings of #30by30 is it fails to tackle the underlying drivers of biodiversity loss linked to the ongoing legacies of colonialism, the nature of capitalist economies and habitat destruction for intensive agriculture etc in fav of Pas (-0.6369)

Thus, although the number of tweets is small, there is a clear tendency for the majority of the content to be spelled out in positive terms. Positive tweets were found to be enlightening, while a small number of negative tweets were skeptical and warning.

The analysis here is limited to the case of English text postings to Twitter in 2021, and other analyses and methodologies could lead to completely opposite results (i.e., more negative results for “30 by 30”). However, by simply looking at the tweets posted, it is unclear what meaning a tweet can have in a large data set. By using NLP techniques to classify tweets by topic and conducting sentiment analysis of different patterns, it is possible to understand the position of a single tweet in the data set and to infer the direction of the poster’s interest. Of course, similar analyses could be performed for other policy themes or specific species by replacing keywords. Furthermore, analyzing a single theme over a period of several years may yield unexpected findings.

Discussion

So far, I have shown that analyzing Twitter posts using NLP has the potential to explore the interests of an unspecified number of people in more detail than Google Trends search volume and related keywords exploration. However, in addition to the caveats to the Twitter analysis mentioned earlier, I must again emphasize its limitations. It is particularly about the uncertainty of big data analysis, especially concerning incompleteness and unrepresentativeness (Liu et al. 2016).

The first issue is about incompleteness. This refers to the case where, even if the amount of data is large, the amount of information is scarce. For example, in the "30 by 30" analysis example, the number of tweets picked up by the author was 93, which means that if Twitter users did not tweet at all, it would be impossible to analyze using the method presented here. Thus, there is always the uncertainty that analysis of only one data set will not always provide the necessary information.

The next issue concerns unrepresentativeness. This is because big data is not scientifically sampled and is limited to a specific demographic. In general, even if we analyze hundreds of millions of tweets in Twitter analysis, they are only tweets, sentiments, and interests of Twitter users. Twitter users cannot be equated with the general public, which means that the results obtained cannot be generalized unconditionally (Boyd and Crawford, 2012). Furthermore, there is always the danger of dismissing a particular user's frequent posting as public interest, so it is important to keep in mind that the frequency of posting varies widely from user to user.

Considering the above, it would be best at this point to reserve the results of a study such as this one, which analyzed a single data set, Twitter, for reference or tentative purposes only, and to reinforce and corroborate the results with analyses of other data sets. In the future, it will be necessary to propose a hybrid type of analysis method that can mutually compensate for the weaknesses of both Twitter and Google Trends. Although there are already examples of studies such as Kostakos (2018) and Cooper et al. (2019), a simple and reliable method that can be applied to all fields has not yet been proposed. For example, a research design that includes multiple big data platforms as well as a quantitative analysis of news coverage while analyzing its correlation with user sentiment on SNS platforms would allow for deeper insights.

In order to overcome the weaknesses of big data research and to enable the kind of sophisticated research described above, it is necessary to cultivate the knowledge of individual researchers and to facilitate collaboration among researchers across multiple disciplines. For example, knowledge of biodiversity and other environmental studies, knowledge of global governance, knowledge of data science such as NLP, and deep social networking literacy will be needed.

Conclusions

In this paper, I have presented examples of analyses that combine multiple methods of NLP while discussing studies that analyze SNS and other big data, including Twitter. However, we need to re-examine why we need to explore public interest in biodiversity. The “First draft of the post-2020 global biodiversity framework” published in July 2021 clearly states “To take urgent action across society to conserve and sustainably use biodiversity and ensure the fair and equitable sharing of benefits from the use of genetics resources, to put biodiversity on a path to recovery by 2030 for the benefit of planet and people” as its mission by 2030 (CBD Secretariat 2021). Coincidentally, the LDA topic modeling of Twitter posts in 2021 conducted in this study yielded 19,181 tweets for topic 1, which I named “Urgent Action.” Even if skepticism about the analysis of Twitter and other social networking sites and other big data itself cannot be completely dispelled, it is undeniable that these results raise a strong inference that public interest is turning to new frameworks.

We failed to achieve the 2010 Biodiversity Targets followed by the Aichi Biodiversity Targets through 2020 as a whole. A more realistic and effective roadmap is needed to ensure that the 2030 Targets, which are already less than a decade away, and the ongoing 2050 Targets do not become a mere skeleton. As a guideline for this purpose, a monitoring system is being considered in which the status of each country’s efforts will be reported, reviewed, and strengthened. To specifically explore the attitudes and interests of an unspecified large number of people toward a certain policy decision, it is not a waste of time and money to focus on the trade-off between public opinion polls and questionnaires, which require time and cost, and big data analysis such as NLP targeting SNS, which is easy and immediate and to refine the analysis. In the future, the results of this study and others like it should attract the interest of more researchers and the attention of more skilled and better analysts. I can only hope that follow-up tests and improvements by such analysts will establish its status as a research method and, eventually, help to make effective environmental policy.

Declarations

Funding

This work was supported by JST SPRING, Grant Number JPMJSP2108.

Competing Interests

The corresponding author states that there is no conflict of interest. This article does not contain any studies involving human participants performed by the author.

Author Contribution

The corresponding author contributed to the study’s conception and design. Material preparation, data collection, and analysis were performed by Shimon Ohtani. The first draft of the manuscript was written by Shimon Ohtani.

Data Availability Statement

This study followed Twitter's Developer's terms and policy (https://developer.twitter.com/en/developer-terms/policy). Other researchers can obtain similar datasets by applying for and being approved for Academic Research Access (https://developer.twitter.com/en/products/twitter-api/academic-research) provided by Twitter.

References

30 by 30. (2021, October 29). In *Wikipedia*. (https://en.wikipedia.org/wiki/30_by_30) Accessed 1 March 2022

Ahmad, W., Wang, B., Xu, H., Xu, M., & Zeng, Z. (2021). Topics, Sentiments, and Emotions Triggered by COVID-19-Related Tweets from IRAN and Turkey Official News Agencies. *SN Computer Science*, 2(5), 1-19. (https://doi.org/10.1007/s42979-021-00789-0)

Alizadeh, Kourosh (2021) Limitations of Twitter Data (https://towardsdatascience.com/limitations-of-twitter-data-94954850cacf) Accessed 4 March 2022

Barrios-O'Neill, D. (2021). Focus and social contagion of environmental organization advocacy on Twitter. *Conservation Biology*, 35(1), 307-315. (https://doi.org/10.1111/cobi.13564)

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679. (https://doi.org/10.1080/1369118X.2012.678878)

Cooper, M. W., Di Minin, E., Hausmann, A., Qin, S., Schwartz, A. J., & Correia, R. A. (2019). Developing a global indicator for Aichi Target 1 by merging online data sources to measure biodiversity awareness and engagement. *Biological Conservation*, 230, 29-36. (https://doi.org/10.1016/j.biocon.2018.12.004)

Díaz, S., Zafra-Calvo, N., Purvis, A.et al. (2020). Set ambitious goals for biodiversity and sustainability. *Science*, 370(6515), 411-413. (https://doi.org/10.1126/science.abe1530)

Dinerstein, E., Vynne, C., Sala, E et al. (2019). A global deal for nature: guiding principles, milestones, and targets. *Science advances*, 5(4), eaaw2869. (https://doi.org/10.1126/sciadv.aaw2869)

Higuchi, K. (2016). KH Coder 3 reference manual. *Kioto (Japan): Ritsumeikan University*.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266. (https://doi.org/10.1126/science.aaa8685)

Karami, A., Lundy, M., Webb, F., Dwivedi, Y.K., (2020). Twitter and research: a systematic literature review through text mining. *IEEE Access* 8, pp.67698–67717. (https://doi.org/ 10.1109/ACCESS.2020.2983656)

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971. (https://doi.org/10.48550/arXiv.1601.06971)

Kostakos, P. (2018). Public perceptions on organised crime, mafia, and terrorism: a big data analysis based on twitter and Google trends. *International Journal of Cyber Criminology*, *12*(1), 282-299. (https://doi.org/10.5281/zenodo.1467919)

Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS journal of photogrammetry and remote sensing*, 115, 134-142 (https://doi.org/10.1016/j.isprsjprs.2015.11.006)

Mayer, P. (2006). Biodiversity—the appreciation of different thought styles and values helps to clarify the term. *Restoration ecology*, 14(1), 105-111. (https://doi.org/10.1111/j.1526-100X.2006.00111.x)

Morshed, S. A., Khan, S. S., Tanvir, R. B., & Nur, S. (2021). Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis. *Journal of Urban Management*, 10(2), 155-165. (https://doi.org/10.1016/j.jum.2021.03.002)

Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council, Canada*, 2. (https://doi.org/10.4224/21270984)

Mohammad, S. M. (2020). Practical and ethical considerations in the effective use of emotion and sentiment lexicons. arXiv preprint arXiv:2011.03492. (https://doi.org/10.48550/arXiv.2011.03492)

Ohtani, S. (2021) How is People's Awareness of "Biodiversity" Measured? Using Sentiment Analysis and LDA Topic Modeling in the Twitter Discourse Space from 2010 to 2020, 05 October 2021, PREPRINT (Version 2) available at Research Square (https://doi.org/10.21203/rs.3.rs-922908/v2)

Otero, P., Gago, J., & Quintas, P. (2021). Twitter data analysis to assess the interest of citizens on the impact of marine plastic pollution. *Marine Pollution Bulletin*, 170, 11262 (https://doi.org/10.1016/j.marpolbul.2021.112620)

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).

The UN Convention on Biological Diversity (CBD) Secretariat (2021). FIRST DRAFT OF THE POST-2020 GLOBAL BIODIVERSITY FRAMEWORK (https://www.cbd.int/doc/c/abb5/591f/2e46096d3f0330b08ce87a45/wg2020-03-03-en.pdf) Accessed 12 March 2022

Troumbis, A. Y. (2019). The time and timing components of conservation culturomics cycles and scenarios of public interest in the Google era. *Biodiversity and Conservation*, 28(7), 1717-1727. (https://doi.org/10.1007/s10531-019-01750-7)

UN Decade on Ecosystem Restoration (https://www.decadeonrestoration.org/) Accessed 3 March 2022

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PloS one, 15*(9), e0239441. (https://doi.org/10.1371/journal.pone.0239441)

Figures

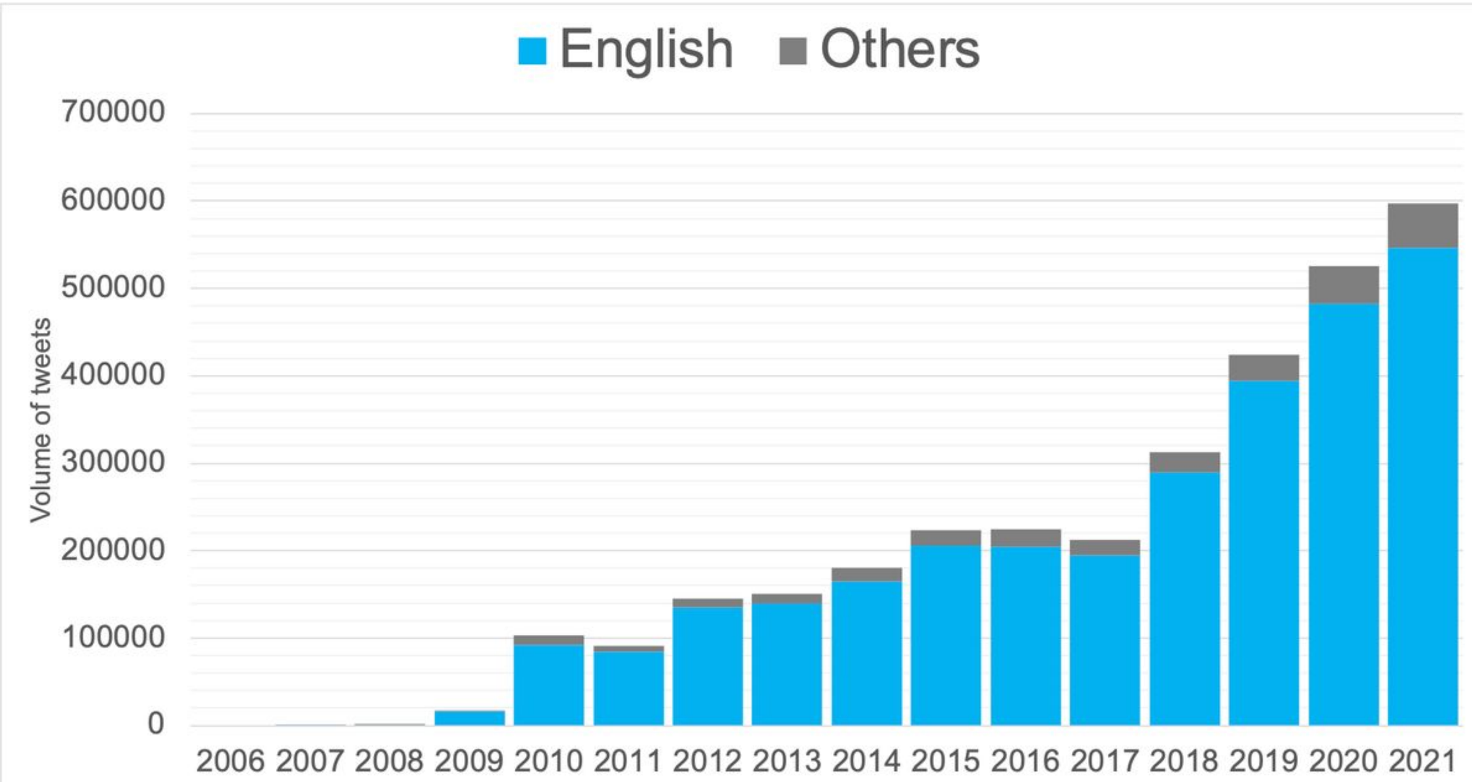

Figure 1

Total number of tweets containing the word “biodiversity” by year from 2006 to 2021 (distinguishing between English text and other language text)

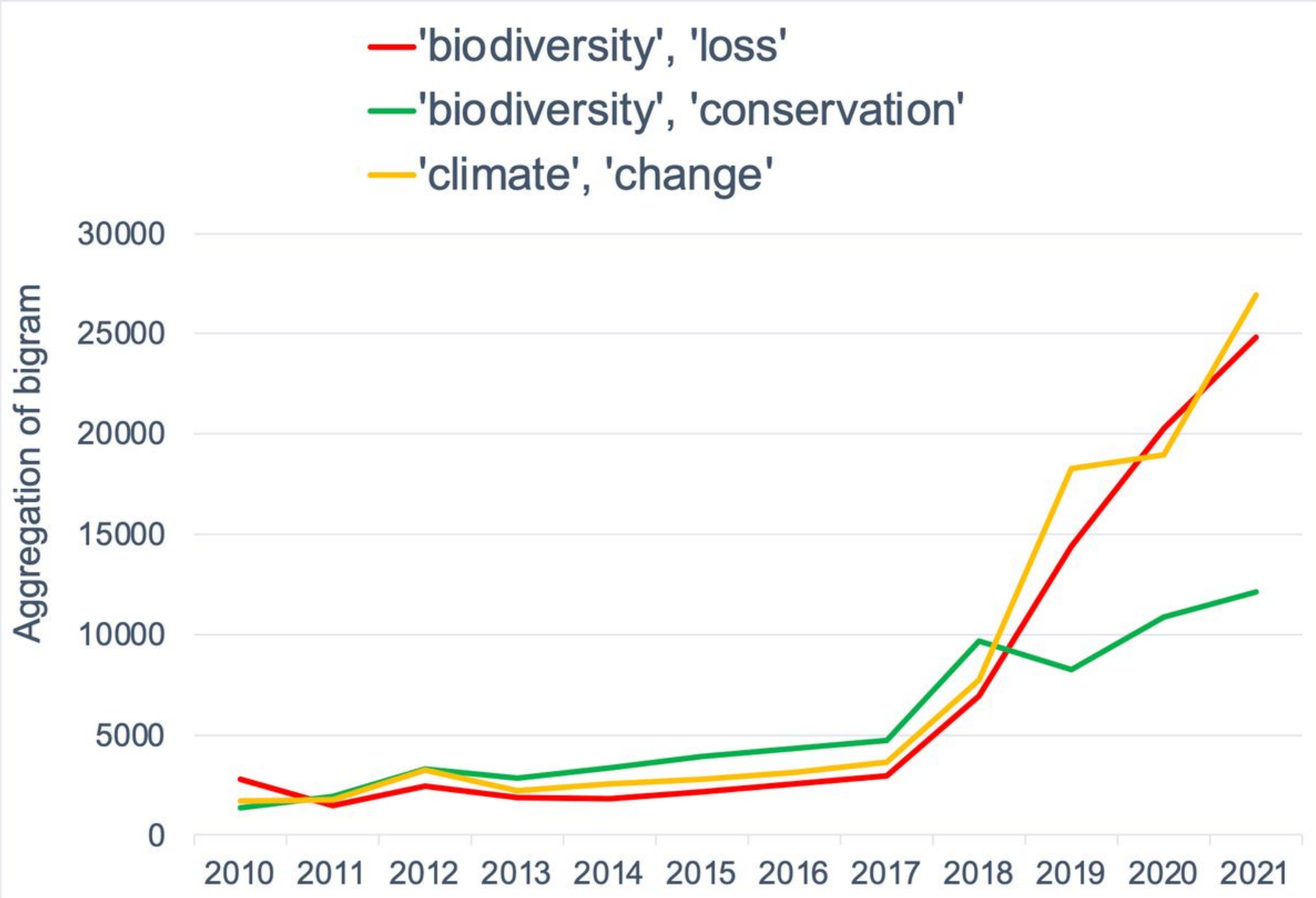

Figure 2

Trends in the three bigrams (‘biodiversity’, ‘loss’), (‘biodiversity’, ‘conservation’), and (‘climate’, ‘change’) from 2010 to 2021

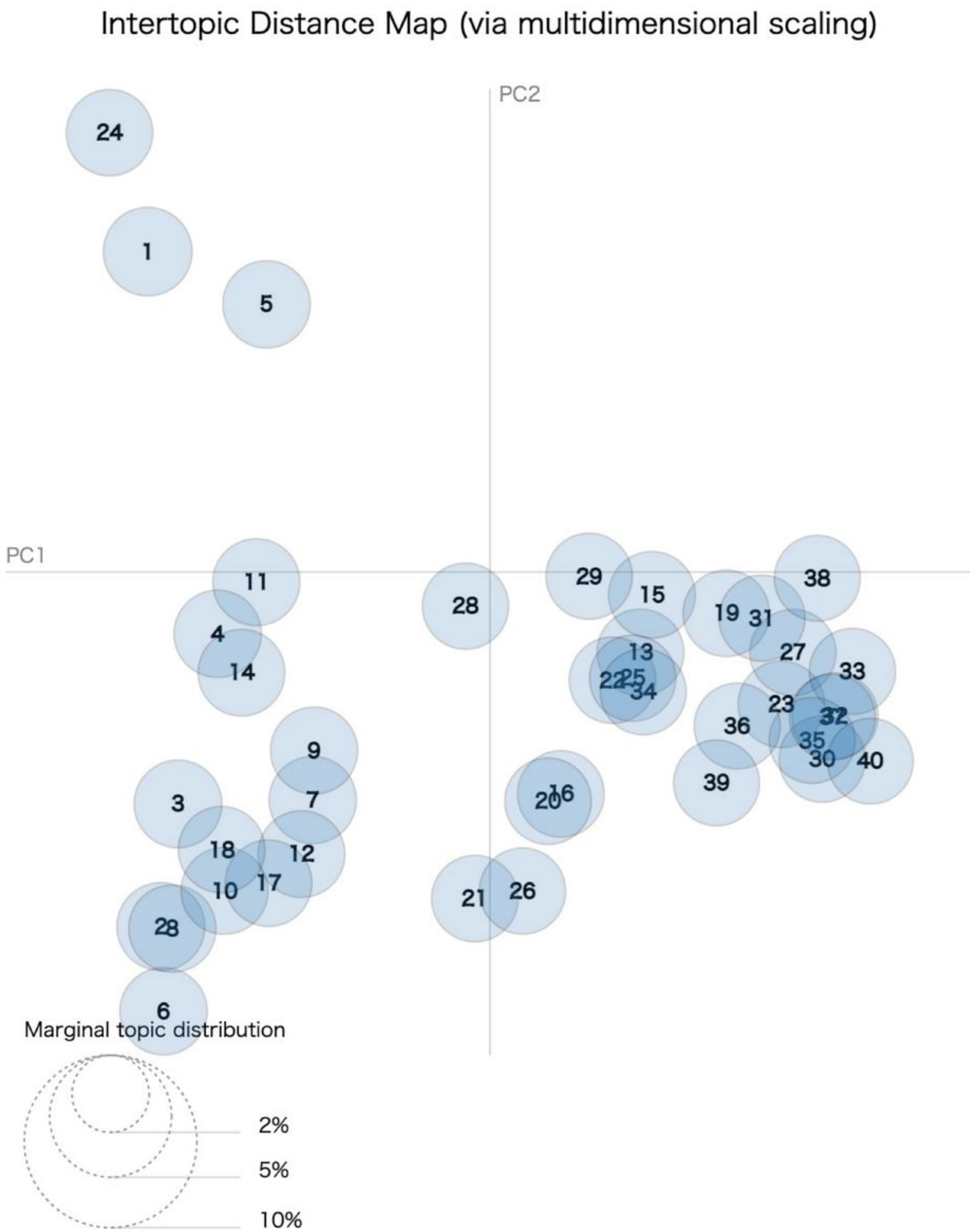

Figure 3

Distribution map of 40 topics in 2021 based on PCoA (visualized by the Python library pyLDAvis)

Figure 4

Heatmap showing 40 topic definitions and 8 types of emotion levels by NRC emotion lexicon (created by the total number of words used in the tweet text that correspond to each emotion defined by lexicon)

Figure 5

Distribution of VADER Compound Scores for tweets containing “biodiversity” in 2021

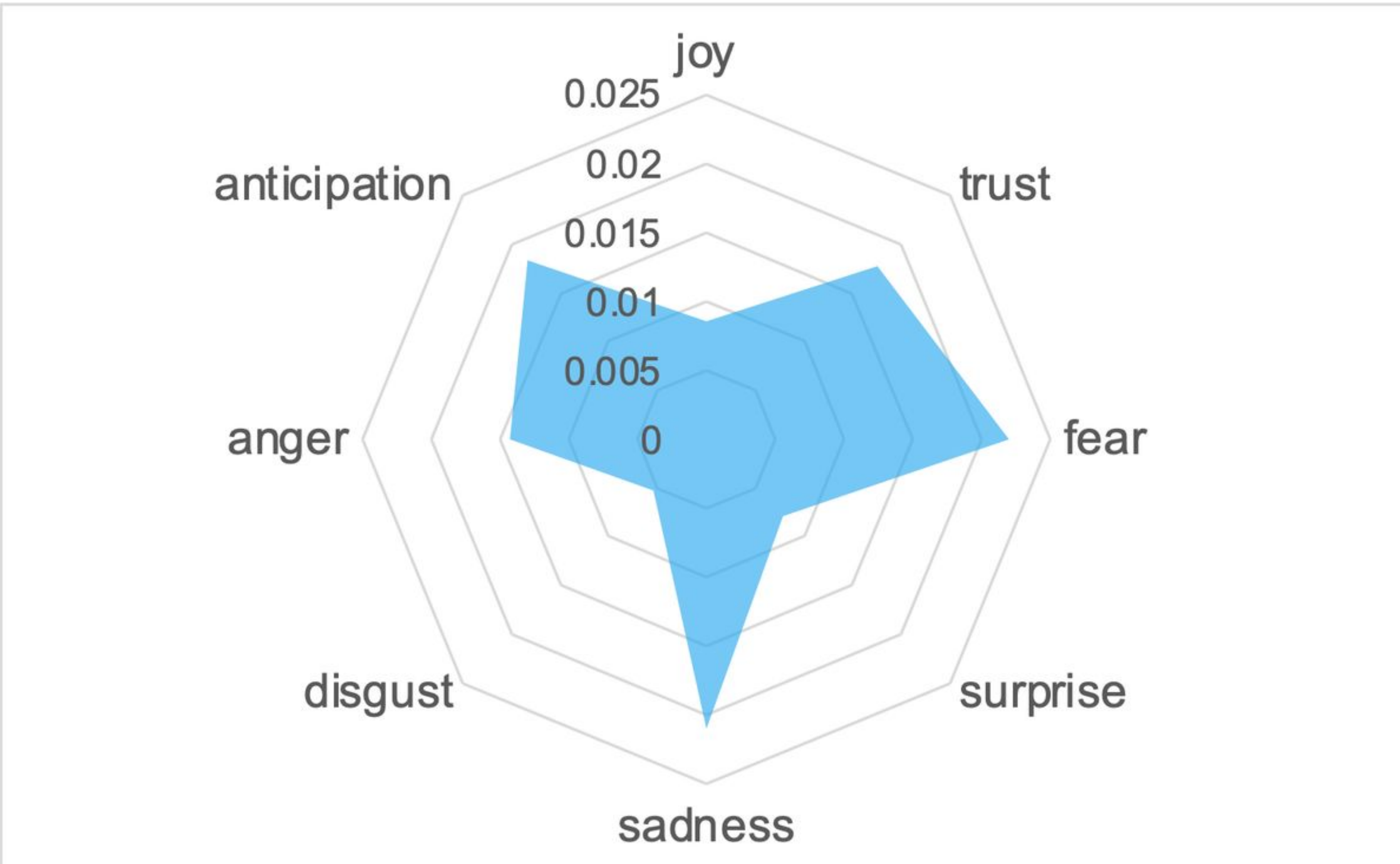

Figure 6

Emotional trends for Topic 14 “Extinction of Species” (chart based on the percentage of words used that correspond to the 8 types of emotions specified by the NRC emotional lexicon)

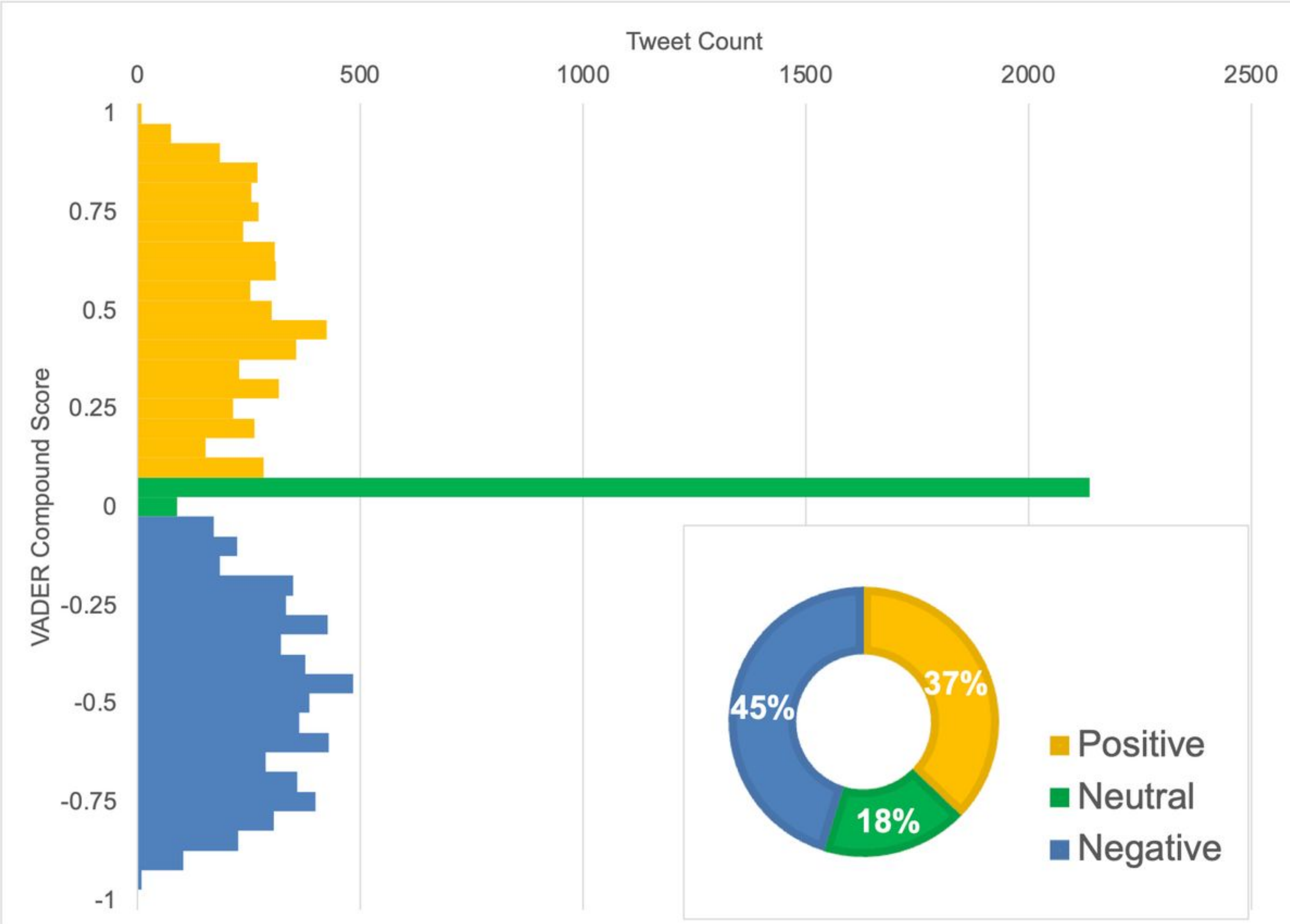

Figure 7

Distribution of VADER Compound Scores for topic 14 “Extinction of Species”

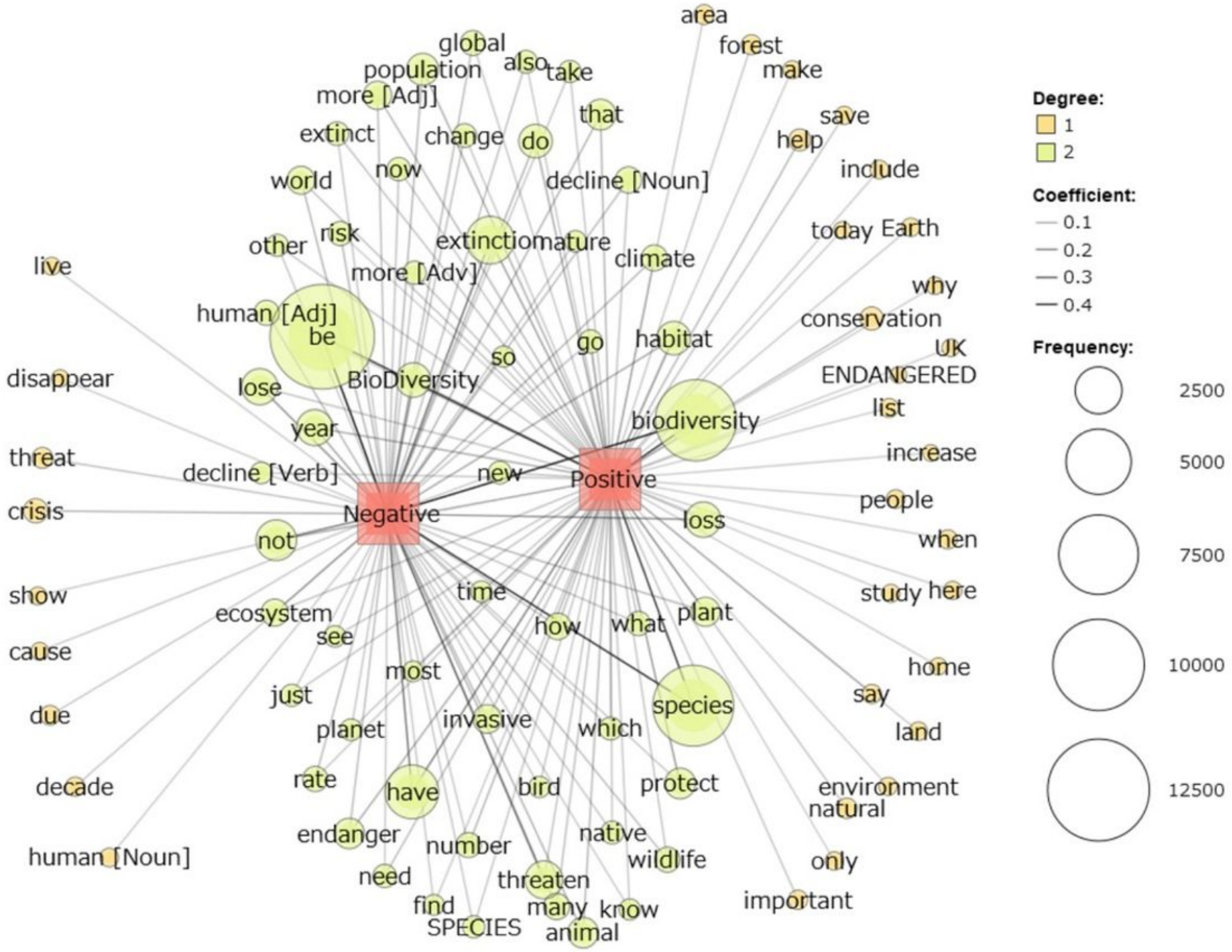

Figure 8

Co-occurrence network diagram of 150 frequently occurring words of Topic 14 (Extinction of Species) by positive and negative (output by KH Coder Version 3.Beta.04a)

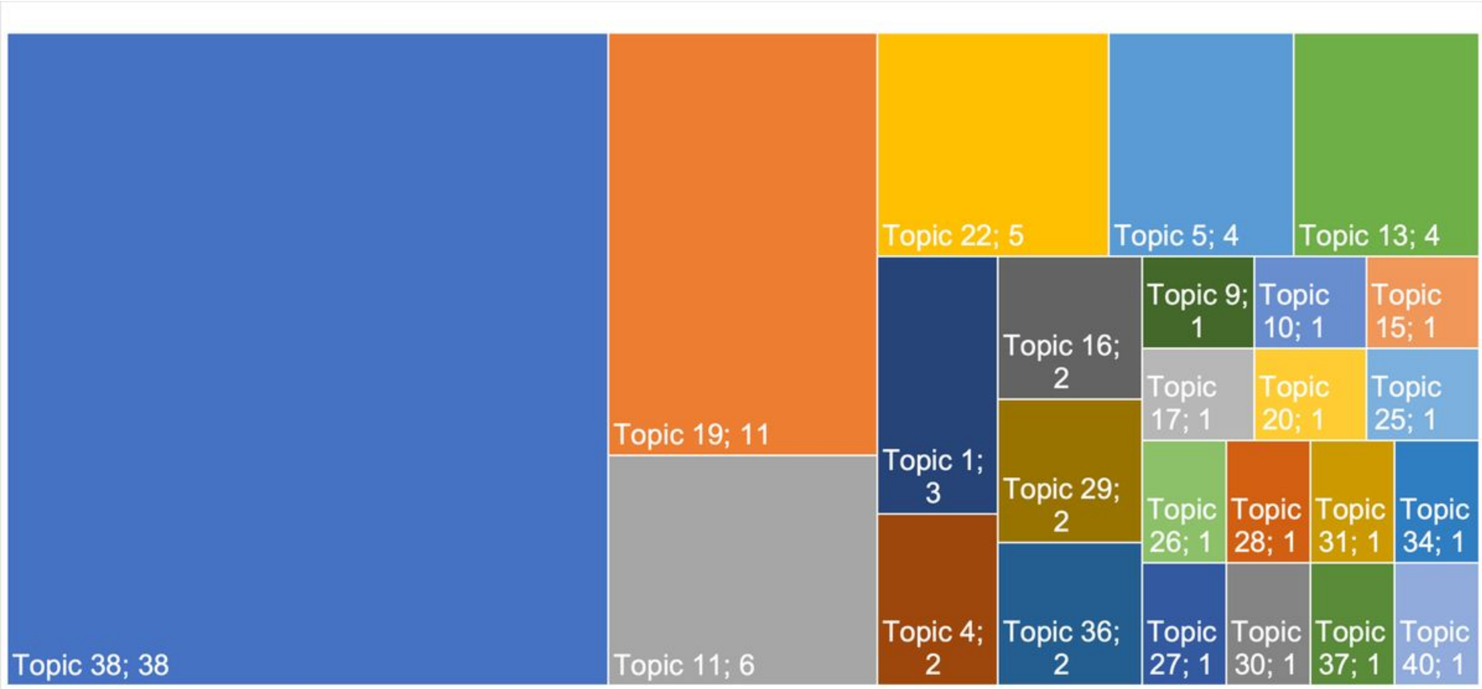

Figure 9

Topic distribution of 93 tweets mentioning “30 by 30” in relation to biodiversity in 2021 (topic number; tweet count)

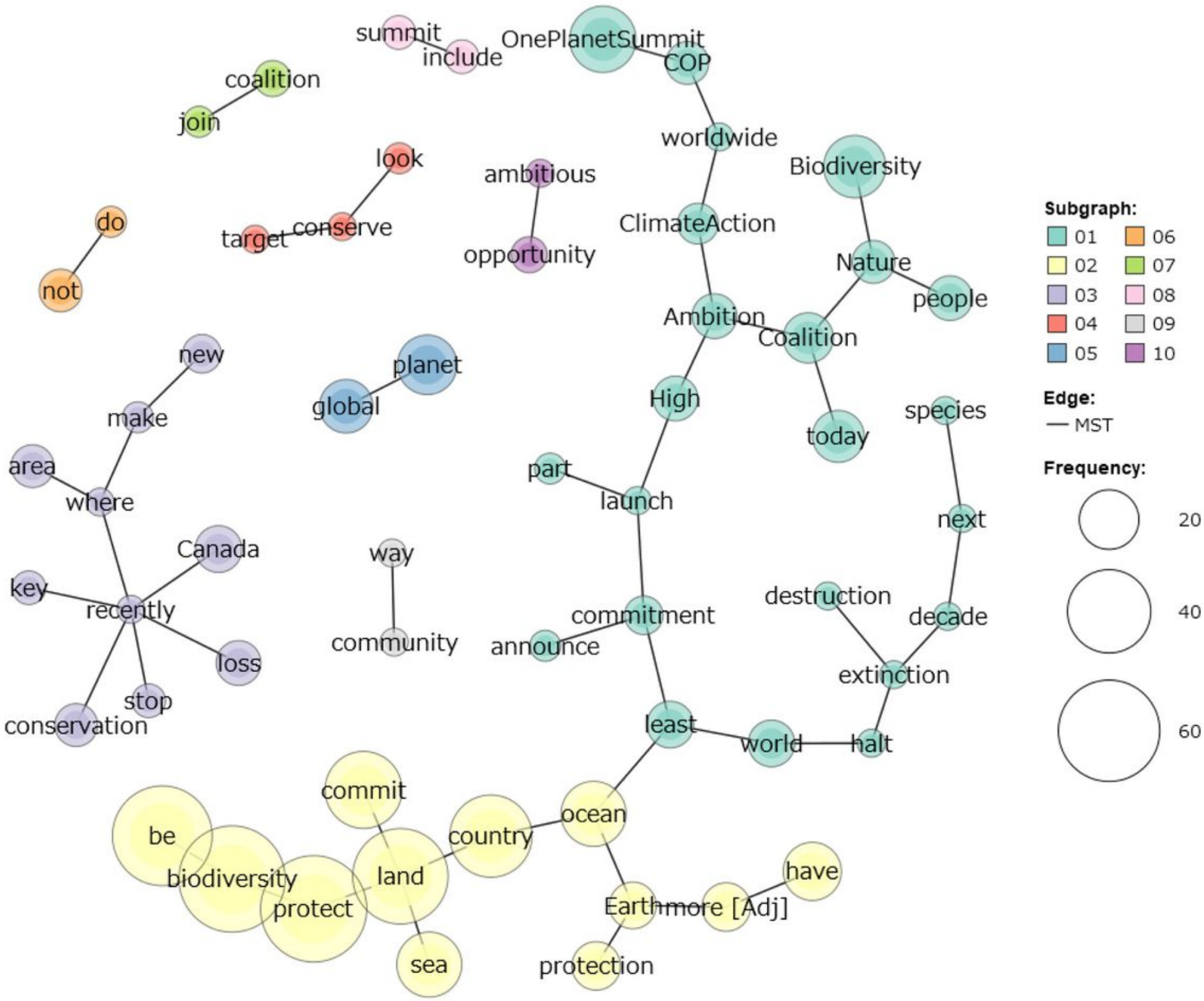

Figure 10

Co-occurrence network diagram of frequently occurring words that mention “30 by 30” in tweets containing “biodiversity” in 2021 (output by KH Coder Version 3.Beta.04a)

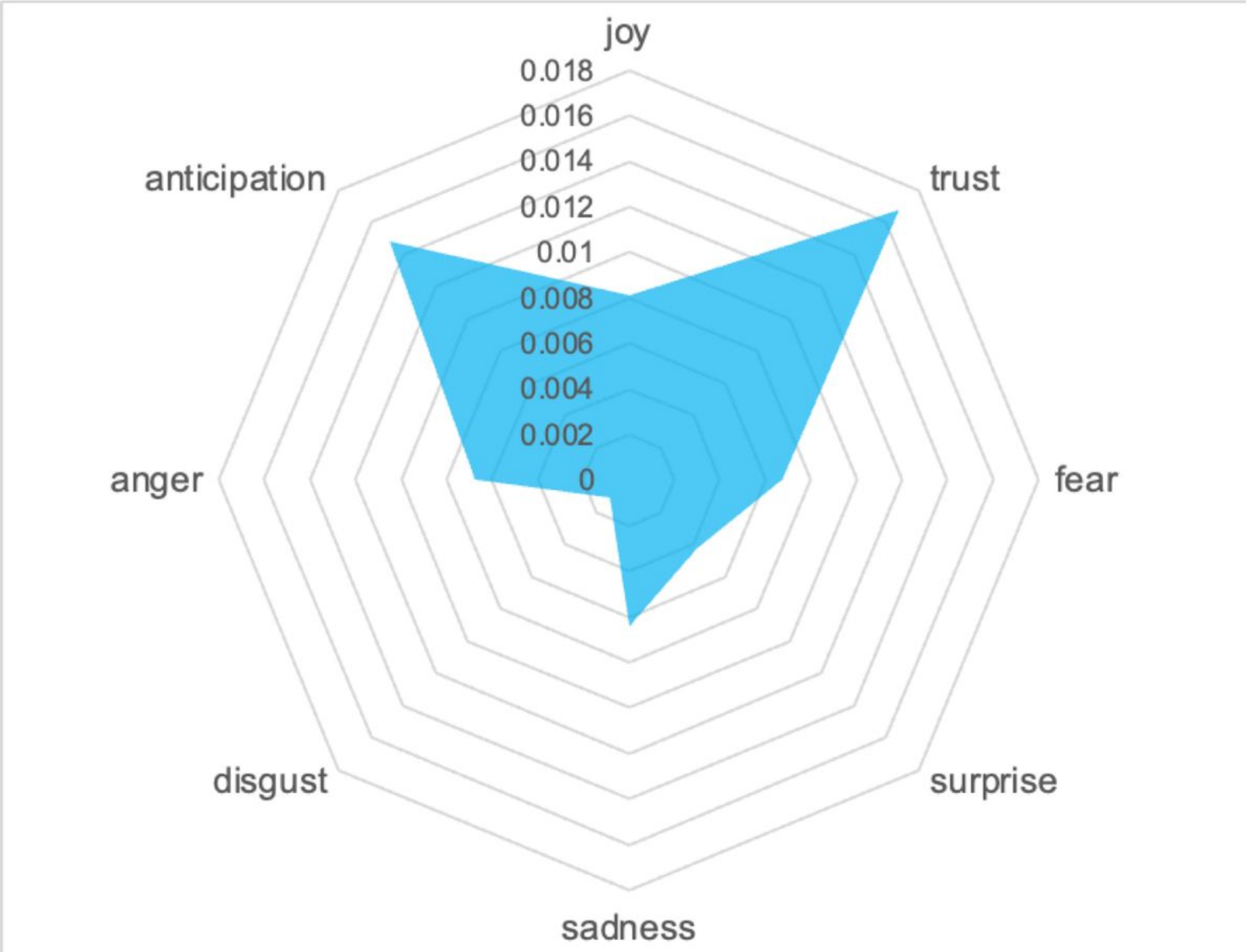

Figure 11

Emotional trend of tweets mentioning “30 by 30” among “biodiversity” related tweets in 2021 (chart showing the percentage of words used that correspond to the 8 types of emotions by NRC emotion lexicon)

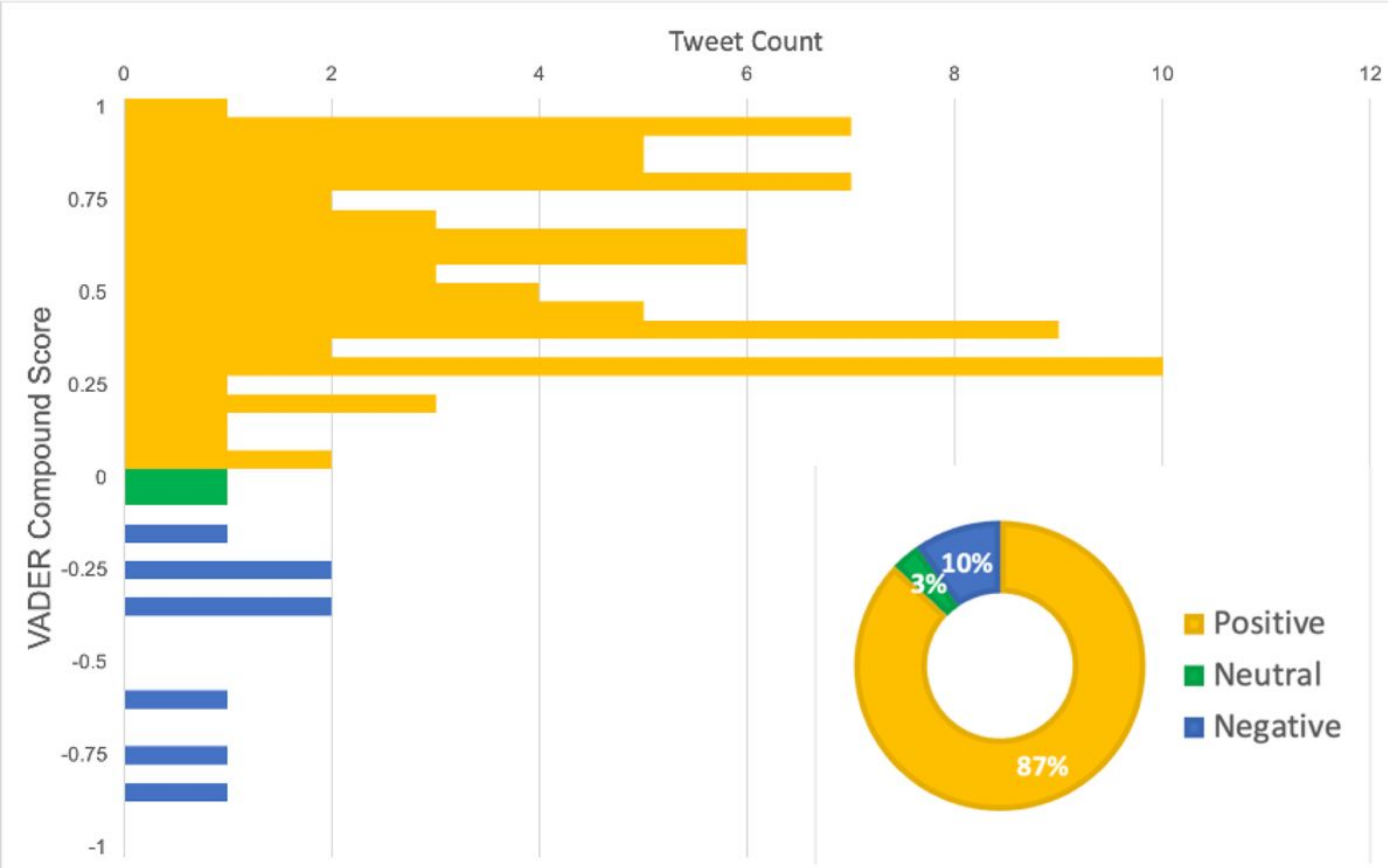

Figure 12

Distribution of VADER Compound Scores of tweets mentioning “30 by 30” among “biodiversity” related tweets in 2021