Using social media and machine learning to understand sentiments towards Brazilian National Parks

Carolina Neves Souza[a*,+]; Javier Martínez-Arribas[b+]; Ricardo A. Correia[c,d,e]; João A. G. R. Almeida[f]; Richard Ladle[b]; Ana Sofia Vaz[b,g,h]; Ana Cláudia M. Malhado[a]

Affiliations:

[a] Institute of Biological and Health Sciences, Federal University of Alagoas, Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, 57072-90, Maceió, AL, Brazil.

[b] CIBIO-InBIO, Research Centre in Biodiversity and Genetic Resources, University of Porto, Campus de Vairão, 4485-661 Vairão, Portugal

[c] Biodiversity Unit, University of Turku, 20014 Turku, Finland

[d] Helsinki Lab of interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland

[e] Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, 00014 Helsinki, Finland

[f] Institute of Computing, Federal University of Alagoas, Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, 57072-90, Maceió, AL, Brazil

[g] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal

[h] NBI, Natural Business Intelligence, R´egia Douro Park, Andraes, Vila Real, Portugal

***Corresponding author:**

Carolina Neves Souza.
E-mail: carolina.souza@icbs.ufal.br
Address: Institute of Biological and Health Sciences, Federal University of Alagoas, Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, 57072-90, Maceió, AL, Brazil.
+55 82 99111 5699

[+]**Shared first authorship**

Acknowledgements

We are grateful to all the researchers at the 21st Century Conservation Lab (Lacos21), at the Federal University of Alagoas and the CIBIO and TROPBIO team at the University of Porto for the insights related to this research.

Funding

The authors would like to acknowledge the support of Fundo Brasileiro para a Biodiversidade (FUNBIO) and Instituto Humanize (#026/2021). RJL and ACMM are funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (#309879/2019-1; #309980/2018-6). RJL and ACMM are also supported via the European Union's Horizon 2020 research and innovation programme under grant agreement No. 854248. CNS acknowledges personal funding from CNPq doctoral grant (#141872/2020-9). RAC acknowledges personal funding from the Research Council of Finland (Grant agreement #348352) and the KONE Foundation (Grant agreement #202101976).

Abstract

Protected areas (PAs) play a vital role in the conservation of natural and cultural heritage while supporting local livelihoods. However, in Brazil, where limited resources and poor effectiveness lead to negative sentiments and are leveraged as criticism towards PAs, it is necessary to better comprehend public perceptions of Brazilian PAs and identify the key factors contributing to negative sentiments. Here, we use data from online discussions about Brazilian national parks (NPs) on Twitter and sentiment analysis to explore this question. We classified the sentiment of ~100,000 tweets collected over a twelve-year period (2011-2022) using the BERTimbau Base model. We also performed a topic modelling with the BERTopic model to identify prevalent subjects concerning Brazilian NPs. We identified 18,388 (17.30%) posts expressing negative sentiment towards NPs, mostly associated with wildfires occurring between 2011 to 2017 and concerning government decisions impacting conservation efforts after 2019. The results revealed six prominent topics: (1) Wildfires; (2) Security; (3) Regulations; (4) Wildlife roadkill; (5) Privatization; (6) Lack of financial resources, reflecting a diverse range of negative sentiments regarding the parks, surpassing isolated events. Furthermore, examining specific topics on a per-park basis proved beneficial in identifying distinct issues and conflicts in the five most tweeted NPs, facilitating targeted conservation actions. Using social media data to better understand public perceptions of NPs can strengthen their management and governance by reinforcing their conservation initiatives and enhancing visitor experiences. Our findings underscore the value of sentiment analysis in identifying gaps and driving improvements in the management of protected areas.

Keywords

Culturomics; Protected areas; Sentiment analysis, Wildfires; Twitter.

1. Introduction

Protected areas (PAs) are a key strategy for promoting the preservation of biological resources and the sustainable use of natural benefits, including ecosystem services and cultural practices (Maretti, et al., 2012; Watson et al., 2014). As a megadiverse country, Brazil has a significant responsibility to protect its biodiversity (Rylands and Brandon, 2005). According to the National Register of Conservation Units (CNUC) of 2023, Brazil is home to a total of 2,859 protected areas (PAs), which collectively cover an area of approximately 2,583,237 km^2. Considering the Brazilian continental area, 19.01% is protected by PAs; while considering the Brazilian exclusive economic zone, 26.49% is being protected by PAs (CNUC, 2023). Brazilian PAs face numerous biophysical and political challenges, from climate change (Soares-Filho et al., 2010) to limited resources for management and monitoring (Silva et al., 2021). Furthermore, they are often viewed as opportunity costs (obstacles to economic development) by politicians and decision-makers (Ferreira et al, 2014). Indeed, the downsizing, downgrading, and degazettement of PA's (PADDD) have affected 72,892km2 of Brazilian PAs between 1981 and 2012 (Bernard et al., 2014). Protecting Brazilian PAs from PADDD requires that negative attitudes are countered by fully demonstrating their value to society (e.g., Jepson et al., 2017) and by showing politicians that they have broad public support (Guedes-Santos et al. 2021).

Of all categories, national parks perhaps have the most potential to draw society's attention to the importance of PAs because they are often popular (Correia et al., 2018) and hold an iconic status by reconciling conservation goals with opportunities to engage with natural outdoor scenic attractions that can inspire and captivate visitors (Dudley, 2008). Moreover, national parks (category II by the IUCN), being the oldest and most visited protected area category, hold a strong connection to recreational activities and a broad range of assets that are important to visitors and generate broad societal value (Bragagnolo et al., 2021), which are closely tied to public interest and support for PAs. However, national parks can also elicit opposing views among visitors, representing either conservation spaces with management restrictions (Hausmann et al., 2020), or positive places for nature-human interactions with psychological and physical benefits for well-being (De Haan et al., 2014). It is important to recognize that the importance of individual sentiments is not only limited to transitory expressions of emotions such as joy, anger, interest, sadness, and gratitude. Rather, people's feelings towards national parks may lead to long-term impact on people's behaviour over the environment (Fredrickson, 2001).

According to Lemberg (2010), perception encompasses how people sense, mentally process, and respond to information derived from their surroundings. It is shaped by sociodemographic characteristics, attitudes, and values, which have the potential to directly impact the experience, satisfaction, and behaviours associated with protected areas (Hoeffel et al., 2008; Rossi et al., 2015). Within this context, it is important to acknowledge that negative sentiments towards protected areas can unfavourably influence perceptions, thereby leading to directed behaviours and dissatisfaction. The connection between how people feel (sentiments) and their perceptions is very significant for conservation, as adverse sentiments can fundamentally shape how people interpret and interact with the natural environment. A concrete example of this is the concession of national parks, which can have negative consequences on local communities, such as impacting service provision as observed in Tijuca national Park (Maciel, 2015), and the distancing of neighbouring communities and low-income visitors from national parks described in Queensland, Australia (Rossi et al., 2016).

Developing a better understanding of how people perceive national parks, evaluate their experiences, and identifying what motivates or hinders visitation is therefore crucial for more informed decision-making (Agyeman et al., 2019; Bragagnolo et al., 2016; Griggs and Lacey, 2022; Rossi et al., 2016). For example, understanding the mobility patterns of visitors to protected areas is essential for formulating conservation strategies (Kim et al. 2023). This proves crucial not only for local management, but also for global marketing initiatives, given that visitors' activities can result in direct or indirect impacts on the environment (Toivonen et al. 2019). This knowledge can help guide the efforts of managers and policymakers towards a more effective management of PAs identifying which aspects and events lead to a negative public attitude towards PAs and addressing more effective strategies to manage tensions and promote changes in favour of conservation (Hausmann et al., 2020; Hockings et al., 2006; Instituto Semeia, 2022).

One of the major challenges in building this knowledge is accessing comprehensive sources of information that can indicate people's feelings/perceptions towards the national parks. In this context,

social media has emerged as a significant source of information. Social media has become an integral part of modern daily lives, offering unprecedented opportunities for communication, debate, and information sharing, and thus presents a powerful platform to understand public perceptions and feelings (Sudhir and Suresh, 2021). A key advantage of analysing digital data lies in the vast volume of information generated by people on a wide range of topics, including political preferences (Ceron et al., 2014), customer satisfaction (Ahani et al., 2019), and nature conservation (Souza et al., 2023; Ladle et al. 2021, Di Minin et al., 2015). Compared to other methods such as survey-based questionnaires, social media data can be collected quickly, at a lower financial cost, and on a larger geographical scale (Becken et al., 2017), and can therefore complement other more targeted approaches. The increasing availability of big online social media data, including text, images, and videos, represents valuable and complementary information for researchers, conservation practitioners, and policymakers to explore citizens' opinions about PAs and biodiversity conservation (Correia et al., 2021). Indeed, the use of online data from social media platforms such as Twitter, Facebook or Instagram has already been applied to measure public interest and to understand users' perceptions about a broad range of conservation topics (Almeida et al., 2022; Fink et al., 2020; Papworth et al., 2015; Tenkanen et al., 2017).

One exciting tool for analysing digital data about human-nature interactions is sentiment analysis, which leverages the use of social media data to comprehend people's values and emotions towards the natural environment (Drijfhout et al., 2016). Sentiment analysis employs computational techniques to extract and evaluate opinions and emotions related to a specific entity or topic (Serrano-Guerrero et al., 2015). It has been extensively applied in marketing and other domains, and more recently it has been used by conservation scientists to assess people's sentiments on a range of topics including environmental management (Bhatt and Pickering, 2021), the impacts of tourism on wildlife (Otsuka et al., 2020), and tourists' preferences, experiences, and opinions (Hausmann et al., 2020). Sentiment analysis employs natural language processing (NLP) through machine learning algorithms that classify the content of textual data based on positive, negative, or neutral perceptions (Liu, 2012). It can provide a mechanistic understanding of how people perceive and feel about nature and conservation (Drijfhout et al., 2016), including PAs. For instance, it can detect negative emotions, such as frustration or disappointment, and can indicate the need for improvements in the visitation areas of protected areas (Agyeman et al., 2019) and in management practices or communication strategies. Controversial management practices, such as restrictions on the use of space and action against illegal hunting (Lubbe et. al. 2019) have the potential to generate dissatisfaction, lack of support and generate conflicts. Thus, sentiment analysis applied to social media data related to PAs can help identify problems such as those mentioned above, and can support decision-making and improve management strategies.

Despite the great potential of sentiment analysis to generate insights for conservation, one current barrier to its wider application is the limited availability of tools and methods for languages other than English (Kaity and Balakrisnhan, 2020). To our knowledge, no previous studies have leveraged social media data to analyse sentiments about Brazilian PAs and this is at least partly due to methodological challenges of working in Portuguese. However, such information is crucial to assist in the management of the vast and diverse network of Brazilian PAs, whose vast biodiversity and ecosystems have increasingly come under attack due to an attempt to dismantle environmental and conservation policies over the past decade (Bernard et al., 2014, Fearnside, 2019; Vale et al 2021). Furthermore, improving the application of sentiment analysis to other languages, such as Portuguese, can ensure the broader application of these methods. Here, our main objective is to understand the people's perception about the Brazilian national parks through social media and sentiment analysis in the Portuguese language. Specifically, we aim to identify the key topics that contribute to the public perceptions (positive, neutral or negative) of Brazilian national parks. In this context, negative sentiments are especially important as the factors that promote them are likely to represent the main challenges for PA management. To achieve this goal, we used natural language processing approaches to classify the sentiments related to Brazilian national parks. In doing so, we hope to demonstrate how sentiment analysis can assist in identifying opportunities to improve the management of Brazilian PAs.

2. Material and Methods

The methodological development of this study is divided into four distinct parts. The first section contextualises the research area, focusing on Brazilian national parks. Subsequently, we outline the

data collection process carried out on the social media platform Twitter, as well as the procedures adopted for data cleaning and filtering. In the third part, we elucidate the sentiment analysis method. This section addresses the underlying concepts of sentiment analysis, the model employed for sentiment classification, and the datasets used for model training. Finally, we describe the analyses conducted based on the results obtained from the sentiment classification of tweets related to Brazilian national parks.

2.1 Study area

Our study focussed on the category of Brazilian protected area with the highest number of tweets in our dataset, the national parks. Brazil has a vast and diversified network of protected areas that safeguard its natural heritage, biodiversity, and cultural resources (BRASIL, 2000). Brazilian national parks fall within the group of strictly protected areas and are designed to protect natural areas of exceptional beauty, diversity, and ecological significance (BRASIL, 2000). National parks are managed by the Chico Mendes Institute for Biodiversity Conservation (ICMBio), with the aim of preserving natural ecosystems, protecting threatened species, and promoting scientific research and environmental education. National parks also provide different cultural and social services to human populations (Nabout et. al 2022), through contemplation, religious rituals and recreation, for example, which gives visitors the opportunity to get to know Brazil's stunning landscapes and wildlife, promoting the protection and conservation of these areas.

In 2022, when this study was carried out, there were 74 officially recognized national parks in Brazil, covering an area of over 268.037 km^2. National parks represent 2.59% of all conservation areas in Brazil, covering 3.11% of the country's land area and contributing only 0.09% of the protected areas in the marine environment (CNUC, 2023). In addition to marine areas, these parks are distributed across all of Brazil's different biomes, including the Caatinga, Cerrado, Atlantic Forest, Pampa and Pantanal (Fig. 1).

2.2 Data Collection

Digital data for conservation culturomics analysis can be applied to different sources of digital data, such as texts, videos and images (Correia et al., 2021). The data mining techniques involved data collection, cleaning, processing and analysis (see Fig. 2). In this study, we collected textual content from Twitter posts, using the Twitter v2 API associated with a Twitter's Academic access which allowed it to collect up to 10 million tweets monthly, access older conversation histories, and apply more search filters than the basic API. It's worth noting that the restrictions and availability of APIs for free usage have been volatile recently, as the platform has undergone a policy change in data access and is now referred to as “X”. The data mining code was developed using the Python language program v.3.9 (http://www.python.org) and was based on the Full-Archive-Search API node. A query composed of 18 keywords in Portuguese language (See supplementary material) was selected to extract tweets related to 11 categories of protected areas in Brazil, including national parks (Souza et al., 2023). We collected tweets posted between 01 January 2011 to 31 December 2022, as well as information on: (i) users (author_id, name, username); (ii) date of publication; (iii) geographic data of tweets; (iv) publication data (number of likes, retweets, replies and whether it is a reply to another user); and (v) text, of each tweet whenever available.

The tweets were downloaded by year and compiled into a single CSV file for data cleaning and filtering. First, a filter was applied to the geographical metadata provided by Twitter, specifically in the 'country' and 'country_code' columns, to identify tweets originating from countries other than Brazil. Next, a manual review of each originally foreign tweet was carried out to exclude those whose textual content was not related to Brazilian protected areas (Souza et al., 2023). Finally, we selected from the dataset the tweets that fell into the category of national parks - the core of our study. The final validated list contained a total of 106,240 tweets about Brazilian national parks. (See https://github.com/CIBIO-TropiBIO/Sentiment-Analysis-Brazilian-National-Parks).

2.3 Sentiment analysis

Sentiment analysis, also known as opinion mining, is a field of study that analyses people's expressed opinions, sentiments, appraisals, attitudes, and emotions towards entities and their attributes in written text (Birjali et al., 2021). The entities can take the form of products, services,

organisations, individuals, events, issues, or topics (Liu, 2012). In this study, we focus on Brazilian national parks as our entity. Sentiment is typically categorised as either "positive", "neutral", or "negative" and is assessed based on the literal meaning of the text. For instance, the sentence "Itatiaia national park: the centre of the problems of the universe is there!" expresses a negative sentiment, while "I liked it and I recommend Iguaçu national park to all my friends! =)" expresses a positive sentiment, and "I'm at Tijuca national park." expresses a neutral sentiment.

Different models, such as Naive Bayes, Maximum Entropy, Support Vector Machine (SVM), and Bidirectional Encoder Representations for Transformers (BERT), have been used to perform sentiment classification on large Portuguese-language datasets (Pereira, 2021). However, the use of BERT is one of the most widely used models in research (Souza et al. 2020) and although it already presents a multilingual natural language processing model, the development of a monolingual approach such as BERTimbau can be effective in training pre-trained language models, especially in languages with few annotated resources, as in the case of Brazilian portuguese. These models replicate the perception of human operators who manually classify the data, based on statistical and structural patterns in the texts. In the case of BERT, this classification can be better developed, since the model was created to have a greater sense of context and language flow than one-way language models because it is trained bidirectionally. Bidirectional training refers to an approach in which a language model is trained to understand the context of a word or token not only by considering previous words, but also subsequent words in a text. This approach favours the understanding of perceptions written in text because they tend to better understand the context of texts and their possible ambiguities (Souza et. al 2020). In this sense, we analysed the polarity of people's expressions of opinions, sentiments, appraisals, attitudes, and emotions towards the Brazilian national parks, based on a sentiment analysis of "positive", "neutral", or "negative" meanings of the collected tweets. To do so, we adapted a pre-trained BERTimbau Base model, due to its performance compared to other natural language processing models and its focus on Brazilian portuguese language (Souza et al., 2020; see Supplementary Material for details).

Our study mined over 100,000 tweets about Brazilian national parks. Although sentiment analysis models perform better with data that is similar in terms of size, style, and text type (Mozetič et al., 2016), good results can be achieved by using data of different nature through appropriate preprocessing tailored to the form of the target data. In a first attempt to classify the sentiments of the texts in our dataset, we used a corpus of tweets with all three sentiments (Portuguese tweets for sentiment analysis, 2018). However, due to limited information, the training of this first model led to most predictions being classified as "neutral", resulting in errors when classifying "positive" and "negative" tweets about the national parks. Therefore, in our second attempt, we decided to train a model using another corpus that had also already been categorised into different types of sentiment (positive, negative, or neutral) - the dataset containing over 200,000 user opinions in Brazilian portuguese about products sold online to accomplish this task (B2W-Reviews01, 2018; Corpus Buscapé, 2013). We used data similar to another study by Avanço and Nunes (2014), properly preprocessed to better adapt these data to the target tweets, as it produced the most accurate predictions for negative tweets which are the main focus of this work. This opinion corpus has a "classification" field with scores ranging from 1 to 5, with 1 being the most negative and 5 being the most positive. For our classification task, we transformed scores 1 and 2 into "negative", 3 into "neutral", and 4 and 5 into "positive". To avoid possible biases arising from different frequencies of occurrence between the categories, we configured the algorithm to take this divergence into account. During the preprocess of training, we assigned a lower weight to errors occurring in the more predominant categories and a higher weight to errors in the rarer categories - what we can call weighting between the classes.

To assess the accuracy of our automated classification, we manually and independently annotated a random sample of 2,000 tweets from our dataset into three classes of sentiment polarity. We then compared the manual annotation with the predicted classes, which confirmed the second model's classification as the most reliable for sentiment classification in our Brazilian national park database. The performance evaluations that we employed to assess our model's performance encompassed accuracy and F1-score. Accuracy means the proportion of true positives and true negatives divided by the total number of predictions, or in other words, how correctly it predicts the categories. This metric is applicable when all the classes are of equal importance, but if there are a different number of observations for each category, it is possible to achieve great accuracy by making all our predictions from the majority class, which is somewhat illusory. For this reason, we also use the F1 score, which is

a harmonic mean between precision and recovery, offering a robust assessment for instances of misclassification, i.e., false positives and false negatives (Capellaro, 2021), or a better metric of incorrectly classified cases.

2.4 Time series analysis

We carried out a time series analysis of the number of daily negative and non-negative tweets in order to identify the main events that generated the peaks in tweets about Brazilian national parks. To do this, we summarised the number of tweets by grouping them by date and year (2011-2022). We then generated a line graph describing the changes in the number of tweets over the years. Using the plotly package (Sievert, 2020) in the R programming language (Team R, 2017), we interactively identified with the graph which days had the highest number of tweets published, and the corresponding event that potentially triggered these publication peaks was determined based on a search of the textual content in our dataset. All analyses were carried out in the R programming language (Team R, 2017), using the dplyr package (Wickham et al., 2021) for data processing and the ggplot2 package (Wickham, 2008) for visual representations of the data.

2.5 Topic modelling analysis

BERTopic is a Python library for natural language processing topic modelling that combines transformer embeddings with clustering algorithms to identify topics in a corpus of texts (Grootendorst, 2022). The BERTopic model supports over 50 languages and has been compared to other models, such as LDA, for performing topic modelling on short texts from social media platforms and has shown exceptional performance in extracting topic representations (Egger and Yu, 2022).

Following the identification of the values associated with each term within the topics, a comprehensive evaluation and inspection of the topics was conducted to detect any potential content that might be misconstrued as a singular topic (See Table 1 in results). This consideration, as noted by Egger and Yu (2022), highlights a potential limitation of the model, particularly when dealing with extensive amounts of data for analysis. Although BERTopic offers the advantage of leveraging domain-specific knowledge to search for specific topics, as done in this study, this process can still be considered exhaustive.

For the purpose of our study, two main steps were undertaken: (i) identification of potential negative topics within our corpus, encompassing all Brazilian national parks, and (ii) segregation of tweets specifically related to the five most tweeted parks in our dataset, followed by clustering to discern the prominent negative topics associated with each individual park. The reason for adopting this filter was to understand whether sentiment analysis has the potential to identify specific topics that are particular to each park. To achieve this, we performed the BERTopic model with the following hyperparameters:

- For the Uniform Manifold Approximation and Projection (UMAP) algorithm we set n_neighbors or the number of samples used during the manifold approximation to 15, n_components or the dimensionality that holds the most information possible to 5, min_dist to 0, in order to get more clustered embeddings and selected the cosine metric to compute distances in high dimensional space.
- For Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) we set the metric to Euclidean in order to compute distances in an array and prediction_data to True to be able to apply to our dataset later, not just to fit the model, for all the datasets, no matter what park the tweets are from. And we set the min_cluster_size parameter or minimum size of the clusters depending on the number of observations we have. The purpose is to reach a reasonable number of topics and also that they contain coherent information to know what they are talking about.
- We set the parameter nr_topics to auto in order to focus on the interpretation of the topics. Besides we use the function CountVectorizer with a list of portuguese stop words and ngram_range between 1 and 2 n-gram words to be extracted, and the function ClassTfidfTransformer in order to reduce the impact of the most frequent words, also the MaximalMarginalRelevance function in order to limit the number of duplicate words that we can find in each topic, and finally, the function SentenceTransformer with the BERT-base-

portuguese-cased model in order to use the same embedding model for the negative tweets selected as in the previous prediction step.

Due to the randomness of some parts of the BERTopic model, it is possible to obtain a slight variation in the topics obtained in each execution of it. After several tests, we identified, based on our knowledge, the topics that remained consistent across all generated models.

3. Results

3.1 Public perceptions about Brazilian national parks

We collected and analysed a total of 106,240 valid tweets about 74 Brazilian national parks generated by 38,432 Twitter users between 2011 and 2022. Of these, 18,388 (17.3%) were categorised as depicting negative sentiment, and 87,852 (82.7%) were categorised as non-negative. The categorisation of positive and neutral tweets into non-negative was mainly due to the model's limitation in differentiating between these types of tweets.

Our two-epoch training process achieved an accuracy score of 0.81 and an F1 score of 0.82 in the validation set. Specifically focussing on the classification of each sentiment group, our final model achieved an accuracy score of 0.23 for neutral tweets, 0.44 for positive tweets and 0.83 for negative tweets in the test set. In general terms, we can assess the accuracy metrics and F1 score as highly satisfactory (above 0.9), good (between 0.8 and 0.9), adequate (between 0.5 and 0.8) or unsatisfactory (below 0.5).

3.2 Trends of non-negative perceptions over time

We identified 4,381 non-negative peaks about the Brazilian national parks between 2011 and 2022. Of these, we selected the day with the highest number of tweets per year to identify the events that caused the greatest public interest. Our temporal analysis of non-negative tweets about Brazilian national parks revealed the presence of several distinct peaks (Fig. 3), with an average of 179 (83 - 377 posts/peak; SD: 93.37) non-negative posts per annual peak. The peaks of similar events ranged from annual celebrations of the anniversaries of certain parks, to large numbers of visitors and government authorisations for concessions in national parks. The highest volume of non-negative tweets (n=377) was related to a higher visitation in Iguaçu national Park on 17th February 2015, followed by the news of the death of an old volunteer from the Itatiaia national Park (n=284) on 28th October 2016. Of the 12 peaks of interest, 3 were related to events about the authorisation of concessions for private companies in national parks, such as (i) Concession of the Chapada dos Veadeiros national park on 20th December 2019 [88 tweets posted]; (ii) Government authorisation to privatize two national parks on 10th august 2020 [83 tweets posted]; and, (iii) concession made in the Iguaçu national park for 375 million reais on 22nd march 2022 [86 tweets posted].

3.3 Trends of negative perceptions over time

Regarding negative perceptions, we identified 3,002 peaks in Brazilian national parks between 2011 and 2022. The peaks indicated some similar events that stimulated the publication of negative tweets about Brazilian national parks (Fig. 4), with an average of 171 (24 - 535 posts/peak; SD: 135.13) negative posts per annual peak. From 2011 to 2017 of the 7 events (1 per year) that caused spikes in negative tweets, 6 were related to opinions about large-scale wildfires in national parks. The highest volume of negative tweets (n=535) was related to a forest fire event in Chapada dos Guimarães national park on 5th September 2015. However, from 2018 onwards, the events that caused the highest negative perceptions diversified and included: (i) the oil disaster off the Brazilian coast on 2 November 2019 [108 tweets posted]; (ii) the increase in mining exploration requests in the Amazon on 11 August 2020 [121 tweets posted]; (iii) the political attempts to reopen a highway inside a national park on 1 June 2021 [147 tweets posted]; and (iv) the seizures of illegal timber on 26 October 2022 [70 tweets posted].

3.4 Main topics associated negative perceptions

Our analysis of negative topics identified six thematic topics, based on the clustering of the most frequent tweets within each cluster. The posts were categorized according to their content across the dataset (18,388 negative tweets): (1) Wildfires (arson and non-arson); (2) Security; (3) Regulations; (4) Wildlife roadkill; (5) Privatization; and, (6) Lack of financial resources. The topic related to wildfires consistently appeared in all the analyses conducted by the BERTopic model. To see the main topics performed by BERTopic, refer to (Table 1).

To gain a specific understanding of what generated negative sentiments among the public for each park, we also explored the five most tweeted parks between 2011-2022 and conducted a topic analysis. As the parks had different numbers of tweets and independent subjects, they had different topic numbers. Here, they are presented in the ranking of the most tweeted national Park. The identified topics for each park (Fig. 5) were as follows: Iguaçu national park: (1) Regulations; (2) Political flexibility; (3) Wildlife roadkill. Tijuca national park: (1) Security; (2) Expropriation. Itatiaia national park: (1) Wildfires; (2) Security. Chapada dos Veadeiros national park: (1) Wildfires; (2) Downsizing. Lençóis Maranhenses national park: (1) Downsizing; (2) Regulations; (3) Political flexibility.

4. Discussion

This study provides the first assessment of perceptions related to Brazilian national parks, employing natural language processing techniques to analyse textual content shared on the social media platform Twitter.

4.1 Non-negative perceptions of Brazilian national parks

Our results revealed that users of the social media Twitter expressed a non-negative (neutral or positive) sentiment towards Brazilian national parks, which is in line with the results of previous studies by Hausmann et al. (2018) and Cao et al. (2022). The non-negative sentiment results are probably driven by many factors, including that the Brazilian national park system has experienced a number of celebratory milestones in recent years. The anniversary celebrations of the creation of the Itatiaia and Iguaçu national parks, as well as high visitor rates on commemorative dates in the Iguaçu national park, generated great public interest among Twitter users. An interesting result is almost 50% of the peaks of public interest related to Iguaçu national park (Fig.3), corroborating recent studies that have shown that it is the most mentioned Brazilian park on the national and global internet (Correia et al., 2018) and that it is also the most tweeted protected area (among all Brazilian categories) in the period from 2011 to 2020 (Souza et al. 2023). This may be due to its exceptional natural attributes and high annual number of visitors (ICMBio, 2022). In addition, Iguaçu national park was the first Brazilian park to be granted to a private company, leading to higher investment and greater media exposure.

However, in addition to the clearly positive findings, we also identified events that provoke public interest, but which are problematic for the classification of sentiments. The first of these is the news related to the death of a volunteer in Itatiaia national park. Such an event can evoke feelings of sadness and grief, but also be written about in positive terms celebrating the importance of the park to the deceased person. Another event that provoked negative sentiments was the attempt to reopen a road in Iguaçu national park and the authorisation of concessions for national parks to private initiatives, revealing possible conflicts between development aspirations and conservation needs.

In situations where the political and social context is mixed analysing sentiment becomes more challenging (Avanço and Nunes, 2014). As mentioned (section 3.1) previously, differentiating between positive and neutral sentiments was exceedingly challenging. Generally, news stories are evaluated by algorithms as neutral, because they don't contain words that can be classified as positive and/or negative. For example, national park concessions to private initiatives are widely reported by the media, often generating debate about environmental impacts and nature conservation.

Analysing sentiment in such a diverse context requires a more sophisticated approach, capable of capturing the duality of emotions present in different topics, offering a more balanced view of public reactions. Therefore, given the weaker accuracy in the classification of non-negative tweets (0.23 for neutral tweets; 0.44 for positive tweets), and considering that negative tweets express negative feelings more strongly, often related to issues and conflicts perceived as directly relevant to improving

biodiversity conservation and the management of these areas, we chose to discuss the results related to negative feelings in more detail.

4.2 Negative perceptions of Brazilian national parks

As mentioned by Fink et al., (2020), negative events tend to elicit stronger reactions from the public. However, it is worth noting that, in general, negative events associated with PAs not only tend to trigger an immediate response, as in the case of reaction on social media platforms, but also tend to have lasting negative impacts on attitudes towards these PAs (Bragagnolo et al. 2015). For instance, displacements of people from protected lands and conflicts between natural resource users and PA managers have long-term negative impacts on individuals, even when these actions result in positive conservation outcomes (Brumatti and Rozendo, 2021; Maciel, 2015; Rossi et al., 2016). From 2011 to 2017, wildfires were the main driver of negative sentiments about the national parks, representing on average 22.51% of the negative perceptions of all four models (Table 1). The media played a critical role in disseminating these events, with a substantial proportion of the tweets originating from news agencies. It is well-known that the media (traditional or online) exert influence over people's perception regarding various environmental and political issues (Shah et al., 2007). When combined with popular social media discussion platforms, these agencies become ideal means to promote public engagement and can also be effective in shaping public perceptions and mobilising real-world actions (Almeida et al., 2022; Stanley, 2020). This further supports the observation that Twitter data can be employed to measure the public response to environmental risks and hazards, such as wildfires (Shook and Turner, 2016).

Our results revealed that the majority of news regarding fires were related to national parks of the Cerrado region and were posted during the dry season (Fig.4). Even in the Brazilian Cerrado, an ecosystem in which fire plays a fundamental role in evolutionary terms and the maintenance of crucial ecological processes, fire is often related to human activities (including climate change) and driven by agricultural practices generating a strong external pressure derived from land use changes in the national park’s vicinities (Pivello et al., 2021). In addition, high-impact wildfire events are subjects of great media interest, and the intensive dissemination of information about the dangers during a short period of imminent disaster threat can sensitise people to the impending event (Perry et al., 1982). Thus, the way news is conveyed, taking into account the factual aspects, plays an important role in raising awareness among both the public and policymakers regarding this issue.

Over time, our findings suggest that while wildfires were the main factor of negative sentiment towards national parks from 2011 to 2017, other issues related to environmental and political concerns have emerged as important factors in recent years. These include oil spills on the Brazilian coast (Almeida, et al 2022), mining requests in protected areas (Siqueira-Gay et. al 2022), political attempts to reopen a road within a national park (Prasniewski et al. 2020), and the seizure of illegal timber in an Amazonian national park (Alencar et al. 2022). Brazilian environmental policy has been subject to criticism and controversy in recent years (Dobrovolski et. al. 2018), especially from 2019 onwards, when the federal government began to relax environmental laws and reduce surveillance and protection of protected areas (Barbosa et al., 2021; Fearnside, 2019). This has generated high levels of concern and reactions from civil society, academia, and national and international organisations, as the government considered environmental restrictions and procedures as obstacles to progress (Abessa et al., 2019). This may help explain the peaks of negative tweets related to Brazilian national parks in the analysed period. It is worth noting that wildfires have not ceased to occur; however, the emphasis on environmental disregard of the federal government of 2019 drew attention to the decisions being made in this area. For example, the oil spill in northeast Brazil led to many Instagram users expressing despair over the government's inaction and lack of response (Almeida et al. 2022).

In addition to our temporal analysis, our aim was also to identify the most discussed topics over the past 12 years. Our topic modelling analysis identified r topics such as wildfires and wildlife roadkill, which are typically isolated events generating heightened engagement. Topic analysis is a powerful tool for attaining a comprehensive and in-depth understanding of the content within a dataset (Grootendorst, 2022). Notably, the presence of cattle grazing and the insufficiency of financial resources (Table 1) emerged as intriguing findings in the analysis, highlighting conflicts stemming from the agricultural sector's push to introduce cattle into protected areas, an activity prohibited in national parks, coupled with the challenges in fostering collaboration between the management body of these areas and cattle breeders to establish inclusive and participatory management agreements (Borges et

al., 2014). It also draws attention to the daunting challenge of resource scarcity (Gerhardinger et al., 2011) in protected area management. Brazilian national parks, on average, have just one staff member per 11,000 hectares (Instituto Semeia, 2021). This broader perspective also underscores the potency of social media and sentiment analysis in monitoring trends, patterns, and reactions pertaining to management and conservation-related matters (Fink et al., 2020; Soriano-Redondo et al., 2017) and its ability to provide fine scale context dependent information.

In examining the management of the most tweeted parks, we identified specific topics that yielded distinctive results for each park. Negative issues were identified that converge between the parks, and distinct topics were also identified for each park (Fig. 5), probably due to their unique characteristics and management factors. Among the topics uncovered, the most prominent intersection revolved around regulations. Several regulations were found to have adverse effects (Refer to Table 2 in the supplementary material for details), from the prohibition of taxis in Iguaçu national park to the prohibition of quad bikes in Lençóis Maranhenses national park. It is widely recognised that management plan restrictions can promote negative perceptions regarding the use of space and can lead to a disconnection between society and protected areas, thus undermining appreciation of the importance of nature (Hausmann et al., 2020). On the other hand, it is also important to acknowledge that practices such as quad biking can both generate value for participants while diminishing value for other park users (e.g. birdwatchers). In addition to regulations, problems related to expropriation in the Tijuca national park, the possibility of reducing the size of the Lençóis Maranhenses national park, and political flexibilisation for the creation of a new category of protected area (Estrada-Parque) in the Iguaçu national park are external pressures that can be related to PADDD events. Bernard and colleagues (2014) identified 41 PADDD events that occurred in Brazil from 1979 to 2012 as a result of chronic deficiencies in financial, personnel and enforcement resources. Government agencies often implement PADDD without consulting civil society, jeopardising the integrity of Protected Areas. These areas exemplify common public goods that require a set of robust governance practices that take into account the complexity of socio-environmental systems, transcend conflicts of interest regarding ecosystems and safeguard the legitimacy of decision-making processes and spaces. In this sense, (Macedo and Medeiros, 2018) have already proposed that participatory incentives (in a non-top-down approach) and knowledge incentives serve as the main drivers of cooperation and effectiveness in protected areas. The adoption of transparent and communicative measures between individuals and institutions is imperative to increase society's interest in the management of conservation territories and promote the inclusion of all social actors in decision-making spaces (Souza et al., 2022).

4.3 Potential, limitations and future research

Overall, sentiment analysis proved to be a valuable methodology for discerning conflicts, such as cases of expropriation, cattle grazing, disagreements linked to restrictions, and safety concerns within national parks (Fig. 5), thus furnishing a potent analytical instrument for protected area management. Monitoring public sentiment through social media could be a way to: (i) monitor events that cause more public interest and negative engagement, enabling rapid action by protected area management to crises and adverse events; (ii) identify gaps through topic analysis and improve strategies to maximise positive results and minimise negative impacts, such as creating fire brigades and environmental education and interpretation strategies on fire in protected areas; (iii) understand the public's perception of and response to any park use rules, promoting more effective communication, addressing concerns, clarifying misunderstandings and fostering trust and collaboration between society and protected area management. Still, the application of sentiment analysis within this context introduces several intertwined challenges and limitations that merit careful consideration. These challenges encompass issues concerning data quality (Mozetič et al., 2016), the presence of subjectivity and irony within textual content (Ravi and Ravi, 2015), along with concerns relating to representativeness (Di Minin et al., 2015) and ethical considerations tied to data collection (for a comprehensive discussion, refer to (Di Minin et al., 2021). Moreover, during our analysis, we noted that the datasets employed to train the model, encompassing both the categorised Twitter dataset and the dataset of opinions about online products, inadequately distinguished between positive and neutral sentiments, resulting in their aggregated categorization as non-negative. It is pivotal to emphasise that the efficacy of classification models is more reliant on the quality, representativeness, and extent of training data than on the particular model type employed (Di Minin et al., 2015; Mozetič et al., 2016). Beyond this limitation, we also observed that sentiments expressed by Twitter users concerning Brazilian national parks, even when perceived as neutral, frequently mirrored a favourable perception and general appreciation of experiences within protected areas. However, it is important to underscore

that the dataset employed (B2W-Reviews01, 2018; Corpus Buscapé, 2013) exhibited satisfactory performance in identifying tweets containing negative sentiments, which are the primary focus of this study.

Despite these challenges and limitations, sentiment analysis can be a valuable tool in the field of conservation management, if approached carefully, and has the potential to contribute to the sustainable management and effective governance of protected areas. By extending the use of sentiment analysis in conservation to the Portuguese language, our research outlines new avenues for this research domain that can focus on developing customised approaches for languages other than English. Given the widespread use of the Portuguese language, which is spoken by more than 200 million people worldwide (Instituto Camões, 2022), our study offers an avenue for gaining insights into the opinions and attitudes of the population by analysing the sentiments expressed in Portuguese about the protected areas present in Portuguese-speaking countries.

For future work, we recommend the use of a pre-categorised sentiment database focused on environmental themes, which can help increase the accuracy of machine learning models and obtain better results in the classification of positive and neutral sentiments. In addition, we propose incorporating field studies into future research, thus bringing together and validating online results with personal perceptions of protected areas. This multi-faceted approach would produce a more holistic understanding of human-nature interactions in protected areas.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT 3.5 in order to improve legibility and translation of some sentences into English. These sentences were then reviewed and rewritten by one of the coauthors (RJL), who is a native speaker with over 20 years' experience of scientific writing.

References

[dataset] B2W-Reviews01., 2021. Open corpus of product reviews. GitHub. https://github.com/b2wdigital/b2w-reviews01

[dataset] Corpus Buscapé, 2013. Portuguese product reviews https://drive.google.com/file/d/1IZJuvt1uxQ4oPGAvGQQxQ_h_ZiV-Be72/view

[dataset] Portuguese tweets for Sentiment Analysis | Kaggle, 2018. GitHub. https://www.kaggle.com/datasets/augustop/portuguese-tweets-for-sentiment-analysis/download?datasetVersionNumber=2

Abessa, D., Famá, A., Buruaem, L., 2019. The systematic dismantling of Brazilian environmental laws risks losses on all fronts. Nat Ecol Evol. https://doi.org/10.1038/s41559-019-0855-9

Agyeman, Y.B., Aboagye, O.K., Ashie, E., 2019. Visitor satisfaction at Kakum National Park in Ghana. Tourism Recreation Research 44, 178–189. https://doi.org/10.1080/02508281.2019.1566048

Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A.R., Knox, K., Samad, S., Ibrahim, O., 2019. Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. Journal of Retailing and Consumer Services 51, 331–343. https://doi.org/10.1016/j.jretconser.2019.06.014

Alencar, A., Silvestrini, R., Gomes, J., & Savian, G. (2022). Amazônia em chamas: o novo e alarmante patamar do desmatamento na Amazônia. Nota Técnica, (9).

Almeida, J.A.G.R., Guedes-Santos, J., Vieira, F.A.S., Azevedo, A.K., Souza, C.N., Pinheiro, B.R., Correia, R.A., Malhado, A.C.M., Ladle, R.J., 2022. Public awareness and engagement in relation to

the coastal oil spill in northeast Brazil. An Acad Bras Cienc 94, 1–10. https://doi.org/10.37002/biobrasil.v12i2.2177

Avanço L.M. and Nunes, M. D. G. V. Lexicon-Based Sentiment Analysis for Reviews of Products in Brazilian Portuguese. 2014. Brazilian Conference on Intelligent Systems, São Paulo, Brazil, 2014, pp. 277-28. https://doi.org/10.1109/BRACIS.2014.57.

Barbosa, L.G., Alves, M.A.S., Grelle, C.E.V., 2021. Actions against sustainability: Dismantling of the environmental policies in Brazil. Land use policy 104. https://doi.org/10.1016/j.landusepol.2021.105384

Becken, S., Stantic, B., Chen, J., Alaei, A.R., Connolly, R.M., 2017. Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing. J Environ Manage 203, 87–97. https://doi.org/10.1016/j.jenvman.2017.07.007

Bernard, E., Penna, L.A.O., Araújo, E., 2014. Downgrading, downsizing, degazettement, and reclassification of protected areas in Brazil. Conservation Biology 28, 939–950. https://doi.org/10.1111/cobi.12298

Bhatt, P., Pickering, C.M., 2021. Public Perceptions about Nepalese National Parks: A Global Twitter Discourse Analysis. Soc Nat Resour 34, 683–700. https://doi.org/10.1080/08941920.2021.1876193

Birjali, M., Kasri, M., Beni-Hssane, A., 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowl Based Syst 226. https://doi.org/10.1016/j.knosys.2021.107134

Borges, S. L., Eloy, L., Ludewigs, T., 2014. O gado que circulava: desafios da gestão participativa e impactos da proibição do uso do fogo aos criadores de gado de solta da Reserva de Desenvolvimento Sustentável Veredas do Acari. Biodiversidade Brasileira. 4 (1), pp.130-156.

Bragagnolo, C., Correia, R.A., Gamarra, N.C., Lessa, T., Jepson, P., Malhado, A.C.M., Ladle, R.J., 2021. Uncovering assets in Brazilian national parks. J Environ Manage 287. https://doi.org/10.1016/j.jenvman.2021.112289

Bragagnolo, C., Costa Gamarra, N., Claudia Mendes Malhado, A., James Ladle, R., 2016. Proposta Metodológica para Padronização dos Estudos de Atitudes em Comunidades Adjacentes às Unidades de Conservação de Proteção Integral no Brasil. Biodiversidade Brasileira 6(1), 190–208.

BRASIL, 2000. Lei No 9.985, de 18 de julho de 2000. Institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providências. Law.

Brumatti, P.N.M., Rozendo, C., 2021. Parques Nacionais, turismo e governança. Revista Brasileira de Pesquisa em Turismo 15, 2119. https://doi.org/10.7784/rbtur.v15i3.2119

Cao, H., Wang, M., Su, S., Kang, M., 2022. Explicit quantification of coastal cultural ecosystem services: A novel approach based on the content and sentimental analysis of social media. Ecol Indic 137. https://doi.org/10.1016/j.ecolind.2022.108756

Capellaro, L., 2021. Análise de polaridade e de tópicos em tweets no domínio da política no Brasil. São Carlos, São Paulo.

Ceron, A., Curini, L., Iacus, S.M., Porro, G., 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. New Media Soc 16, 340–358. https://doi.org/10.1177/1461444813480466

CNUC, 2023. Painel de Unidades de Conservação. https://cnuc.mma.gov.br/powerbi (accessed 06 November 2023).

Correia, R.A., Jepson, P., Malhado, A.C.M., Ladle, R.J., 2018. Culturomic assessment of Brazilian protected areas: Exploring a novel index of protected area visibility. Ecol Indic 85, 165–171. https://doi.org/10.1016/j.ecolind.2017.10.033

Correia, R.A., Ladle, R., Jarić, I., Malhado, A.C.M., Mittermeier, J.C., Roll, U., Soriano-Redondo, A., Veríssimo, D., Fink, C., Hausmann, A., Guedes-Santos, J., Vardi, R., Di Minin, E., 2021. Digital data sources and methods for conservation culturomics. Conservation Biology 35, 398–411. https://doi.org/10.1111/cobi.13706

De Haan, F.J., Ferguson, B.C., Adamowicz, R.C., Johnstone, P., Brown, R.R., Wong, T.H.F., 2014. The needs of society: A new understanding of transitions, sustainability and liveability. Technol Forecast Soc Change 85, 121–132. https://doi.org/10.1016/j.techfore.2013.09.005

Di Minin, E. Di, Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. Front Environ Sci 3, 1–6. https://doi.org/10.3389/fenvs.2015.00063

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., Kulkarni, R., 2021. How to address data privacy concerns when using social media data in conservation science. Conservation Biology 35, 437–446. https://doi.org/10.1111/cobi.13708

Drijfhout, M., Kendal, D., Vohl, D., Green, P.T., 2016. Sentiment Analysis: ready for conservation. Front Ecol Environ 14, 525–526. https://doi.org/10.1002/fee.1435

Dobrovolski, R., Loyola, R., Rattis, L., Gouveia, S.F., Cardoso, D., Santos-Silva, R., Gonçalves-Souza, D., Bini, L.M. and Diniz-Filho, J.A.F. (2018). Science and democracy must orientate Brazil's path to sustainability. Perspectives in Ecology and Conservation, 16(3) 121-124.

Dudley, N., 2008. Guidelines for applying Protected Area Management Categories, IUCN, Gland, Switzerland. IUCN. https://doi.org/10.1103/PhysRevB.38.10724

Egger, R., Yu, J., 2022. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Frontiers in Sociology 7. https://doi.org/10.3389/fsoc.2022.886498

Fearnside, P.M., 2019. Setbacks under President Bolsonaro: A Challenge to Sustainability in the Amazon. Sustentabilidade International Science Journal 1, 38–52.

Fink, C., Hausmann, A., Di Minin, E., 2020. Online sentiment towards iconic species. Biol Conserv 241. https://doi.org/10.1016/j.biocon.2019.108289

Fredrickson, B.L., 2001. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American Psychologist 56, 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Gamarra, N.C., Correia, R.A., Bragagnolo, C., Campos-Silva, J.V., Jepson, P.R., Ladle, R.J., Mendes Malhado, A.C., 2019. Are Protected Areas undervalued? An asset-based analysis of Brazilian Protected Area Management Plans. J Environ Manage 249, 109347. https://doi.org/10.1016/j.jenvman.2019.109347

Gerhardinger, L.C., Godoy, E.A.S., Jones, P.J.S., Sales, G., Ferreira, B.P., 2011. Marine protected dramas: The flaws of the Brazilian national system of marine protected areas. Environ Manage 47, 630–643. https://doi.org/10.1007/s00267-010-9554-7

Griggs, N., Lacey, G.T., 2022. Barriers and limitations to national park visitation by millennials: Perceptions from second-generation Australians. Annals of Tourism Research Empirical Insights 3. https://doi.org/10.1016/j.annale.2022.100074

Grootendorst, M., 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure.

Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Kulkarni, R., Tenkanen, H., Di Minin, E., 2020. Understanding sentiment of national park visitors from social media data. People and Nature pan3.10130. https://doi.org/10.1002/pan3.10130

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., Di Minin, E., 2018. Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. Conserv Lett. https://doi.org/10.1111/conl.12343

Hockings, M., Stolton, S., Leverington, F., 2006. Evaluating effectiveness : a framework for assessing management effectiveness of protected areas, 2nd edition, Evaluating effectiveness : a framework for assessing management effectiveness of protected areas, 2nd edition. https://doi.org/10.2305/iucn.ch.2006.pag.14.en

Hoeffel, J.L., Fadini, A.A.B., Machado, M.K., Reis, J.C., 2008. Trajetórias do Jaguary - Unidades de conservação, percepção ambiental e turismo: Um estudo na apa do sistema Cantareira, São Paulo. Ambiente e Sociedade 11, 131–148. https://doi.org/10.1590/s1414-753x2008000100010

ICMBio, 2022. Unidades de conservação federais recebem mais de 21 milhões de visitas em 2022. https://www.gov.br/icmbio/pt-br/assuntos/noticias/ultimas-noticias/unidades-de-conservacao-federais-recebem-mais-de-21-milhoes-de-visitas-em-2022 (accessed 20 November 2023)

Instituto Camões, 2022. Dados sobre a Língua Portuguesa. https://www.instituto-camoes.pt/images/pdf_noticias/Dados_sobre_a_l%C3%ADngua_portuguesa_de_2022.pdf (accessed 20 April 2023)

Instituto Semeia, 2021. Diagnóstico do uso público em parques brasileiros: a perspectiva da gestão.

Instituto Semeia, 2022. Parques do Brasil: percepções da População 2022.

J. Ferreira et al., Brazil's environmental leadership at risk. Science 346,706-707(2014). DOI:10.1126/science.1260194

Jayalaxshmi Mistry, Marcelo Bizerril, 2011. Por Que é Importante Entender as Inter-Relações entre Pessoas, Fogo e Áreas Protegidas? Biodiversidade Brasileira I, 40–49. https://doi.org/https://doi.org/10.37002/biobrasil.v%25vi%25i.137

Jepson, P.R., Caldecott, B., Schmitt, S.F., Carvalho, S.H.C., Correia, R.A., Gamarra, N., Bragagnolo, C., Malhado, A.C.M., Ladle, R.J., 2017. Protected area asset stewardship. Biol Conserv 212, 183–190. https://doi.org/10.1016/j.biocon.2017.03.032

Kaity, M., & Balakrishnan, V. (2020). Sentiment lexicons and non-English languages: a survey. Knowledge and Information Systems, 62(12), 4445-4480.

Kim, H., Shoji, Y., Mameno, K., Kubo, T., & Aikoh, T. (2023). Changes in visits to green spaces due to the COVID-19 pandemic: Focusing on the proportion of repeat visitors and the distances between green spaces and visitors' places of residences. Urban forestry & urban greening, 80, 127828.

Ladle, Richard J.; Souza, Carolina N.; Correia, R., 2021. Culturomics for (not against!) protected areas In. Biol Conserv 256, 109197. https://doi.org/10.1016/j.biocon.2021.109015

Lemberg D., “Environmental Perception” in Warf, B., Encyclopedia of Geography, Sage, 2010.

Liu, B., 2012. Sentiment analysis and opinion mining. Morgan & Claypool.

Lubbe, B.A; du Preez, E.A; Douglas, A; Fairer-Wessels, F. 2017. The impact of rhino poaching on tourist experiences and future visitation to National Parks in South Africa, Current Issues in Tourism. 10.1080/13683500.2017.1343807

Macedo, H.S., Medeiros, R.P., 2018. Rethinking governance in a Brazilian multiple-use marine protected area. Mar Policy 0–1. https://doi.org/10.1016/j.marpol.2018.08.019

Maciel, G.G., 2015. Mercantilização da cidade do Rio de Janeiro e suas implicações na gestão de unidades de conservação: um estudo sobre a concessão do Setor Paineras/Corcovado (Parque Nacional da Tijuca - RJ) e os efeitos sobre os moradores das favelas do Cerro Corá e do Guararapes. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Maretti, C.C., Catapan, M., Abreu, M. J., Oliveira, J.E.D., 2012. Áreas protegidas: definições, tipos e conjuntos – reflexões conceituais e diretrizes para a gestão.

Mozetič, I., Grčar, M., Smailović, J., 2016. Multilingual twitter sentiment classification: The role of human annotators. PLoS One 11. https://doi.org/10.1371/journal.pone.0155036

Nabout, J. C., Tessarolo, G., Pinheiro, G. H. B., Marquez, L. A. M., & de Carvalho, R. A. (2022). Unraveling the paths of water as aquatic cultural services for the ecotourism in Brazilian Protected Areas. Global Ecology and Conservation, 33, e01958.

Otsuka, R., Yamakoshi, G., Id, R.O., Yamakoshi, G., 2020. Analyzing the popularity of YouTube videos that violate mountain gorilla tourism regulations. PLoS One 15, 1–20. https://doi.org/10.1371/journal.pone.0232085

Papworth, S.K., Nghiem, T.P.L., Chimalakonda, D., Posa, M.R.C., Wijedasa, L.S., Bickford, D., Carrasco, L.R., 2015. Quantifying the role of online news in linking conservation research to Facebook and Twitter. Conservation Biology 29, 825–833. https://doi.org/10.1111/cobi.12455

Pereira, D.A., 2021. A survey of sentiment analysis in the Portuguese language. Artif Intell Rev 54, 1087–1115. https://doi.org/10.1007/s10462-020-09870-1

Perry, R.W., Lindell, M.K., Greene, M.R., 1982. Threat Perception and Public Response to Volcano Hazard, The Journal of Social Psychology, 116:2, 199-204. 10.1080/00224545.1982.9922771

Pivello, V.R. 2006. Fire management for biological conservation in the Brazilian cerrado. p. 129-154. In Mistry, J. & Berardi, A. (eds.). Savannas and dry forests: linking people with nature. Ashgate Publications.

Prasniewski, V. M., Szinwelski, N., Sobral-Souza, T., Kuczach, A. M., Brocardo, C. R., Sperber, C. F., & Fearnside, P. M. 2020. Parks under attack: Brazil’s Iguaçu National Park illustrates a global threat to biodiversity. Ambio. doi:10.1007/s13280-020-01353-5.

Ravi, K., Ravi, V., 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowl Based Syst 89, 14–46. https://doi.org/10.1016/j.knosys.2015.06.015

Rossi, S.D., Byrne, J.A., Pickering, C.M., Reser, J., 2015. “Seeing red” in national parks: How visitors’ values affect perceptions and park experiences. Geoforum 66, 41–52. https://doi.org/10.1016/j.geoforum.2015.09.009

Rossi, S.D., Pickering, C.M., Byrne, J.A., 2016. Not in our park! Local community perceptions of recreational activities in peri-urban national parks. Australasian Journal of Environmental Management 23, 245–264. https://doi.org/10.1080/14486563.2015.1132397

Rylands, A.B., Brandon, K., 2005. Brazilian protected areas. Conservation Biology. https://doi.org/10.1111/j.1523-1739.2005.00711.x

Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E., 2015. Sentiment analysis: A review and comparative analysis of web services. Inf Sci (N Y) 311, 18–38. https://doi.org/10.1016/j.ins.2015.03.040

Shah, D. V., McLeod, D.M., Kim, E., Sun Young Lee, Gotlieb, M.R., Ho, S.S., Breivik, H., 2007. Political consumerism: How communication and consumption orientations drive lifestyle politics. Annals of the American Academy of Political and Social Science 611, 217–235. https://doi.org/10.1177/0002716206298714

Shook, E., Turner, V.K., 2016. The socio-environmental data explorer (SEDE): a social media–enhanced decision support system to explore risk perception to hazard events. Cartogr Geogr Inf Sci 43, 427–441. https://doi.org/10.1080/15230406.2015.1131627

Sievert, C., 2020. Interactive Web-Based Data Visualization with R, plotly, and shiny. https://doi.org/https://doi.org/10.1201/9780429447273

Silva, J.M.C. da, Dias, T.C.A. de C., Cunha, A.C. da, Cunha, H.F.A., 2021. Funding deficits of protected areas in Brazil. Land use policy 100. https://doi.org/10.1016/j.landusepol.2020.104926

Siqueira-Gay, J., Metzger, J.P., Sánchez, L.E.; Sonter, L.J. 2022. Strategic planning to mitigate mining impacts on protected areas in the Brazilian Amazon. Nat Sustain 5, 853–860. https://doi.org/10.1038/s41893-022-00921-9

Soares-Filho, B., Moutinho, P., Nepstad, D., Anderson, A., Rodrigues, H., Garcia, R., Dietzsch, L., Merry, F., Bowman, M., Hissa, L., Silvestrini, R., Maretti, C., 2010. Role of Brazilian Amazon protected areas in climate change mitigation. Proc Natl Acad Sci U S A 107, 10821–10826. https://doi.org/10.1073/pnas.0913048107

Soriano-Redondo, A., Bearhop, S., Lock, L., Votier, S.C., Hilton, G.M., 2017. Internet-based monitoring of public perception of conservation. Biol Conserv 206, 304–309. https://doi.org/10.1016/j.biocon.2016.11.031

Souza, C.N., de Barros, E.L.S.F.C., Dantas, I.F. V., Bragagnolo, C., Malhado, A.C.M., Selva, V.F., 2022. Inclusion and governance in the managing Council of the Costa dos Corais Environmental Protection Area. Ambiente e Sociedade. 25. https://doi.org/10.1590/1809-4422ASOC20210074R1VU2022L3AO

Souza, C.N., Almeida, J.A.G.R, Correia R.A, Ladle R.J, Carvalho A.R, Malhado A.C.M. 2023. Assessing Brazilian protected areas through social media: Insights from 10 years of public interest and engagement. PLoS ONE 18(10): e0293581. https://doi.org/10.1371/journal.pone.0293581

Souza, F., Nogueira, R., Lotufo, roberto, 2020. BERTimbau: Pretrained BERT Models for Brazilian Portuguese, in: 9th Brazilian Conference. pp. 403–417. https://doi.org/https://doi.org/10.1007/978-3-030-61377-8

Stanley, P., 2020. Unlikely hikers? Activism, Instagram, and the queer mobilities of fat hikers, women hiking alone, and hikers of colour. Mobilities 15, 241–256. https://doi.org/10.1080/17450101.2019.1696038

Sudhir, P., Suresh, V.D., 2021. Comparative study of various approaches, applications and classifiers for sentiment analysis. Global Transitions Proceedings 2, 205–211. https://doi.org/10.1016/j.gltp.2021.08.004

Team, R.C., 2017. R: A Language and Environment for Statistical Computing.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci Rep 7. https://doi.org/10.1038/s41598-017-18007-4

Vale M M, Berenguer E, Argollo de Menezes M , Viveiros de Castro E B, Pugliese de Siqueira L and Portela R Q 2021 The COVID-19 pandemic as an opportunity to weaken environmental protection in Brazil Biol. Conserv. 255 108994

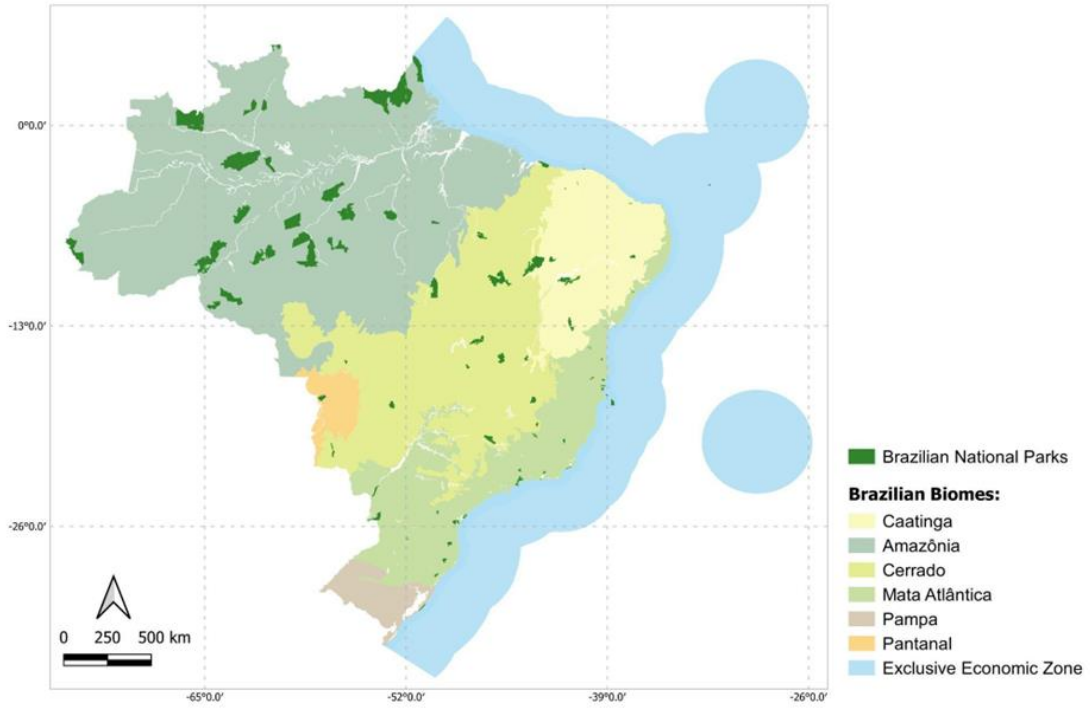
Watson, J.E.M., Dudley, N., Segan, D.B., Hockings, M., 2014. The performance and potential of protected areas. Nature 515, 67–73. https://doi.org/10.1038/nature13947

Wickham, H., 2008. Elegant Graphics for Data Analysis: ggplot2, Applied Spatial Data Analysis with R.

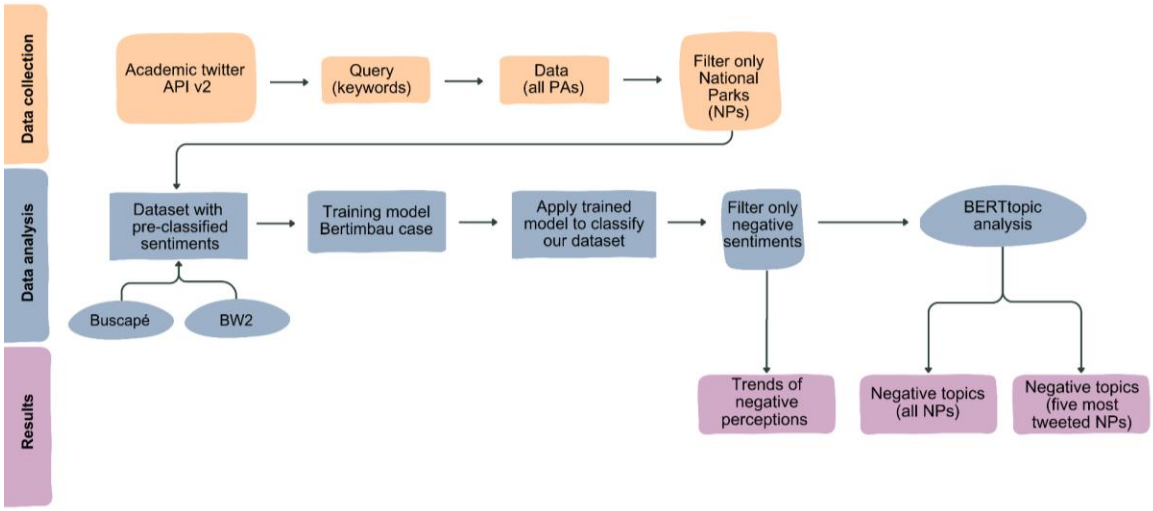
Wickham, H.; François, R.; Henry, L.; Müller, K., 2021. dplyr: A Grammar of Data Manipulation.

Wilkins, L., 1985. Television and newspaper coverage of blizzard: Is the message helplessness? Newsp Res J 51–65.

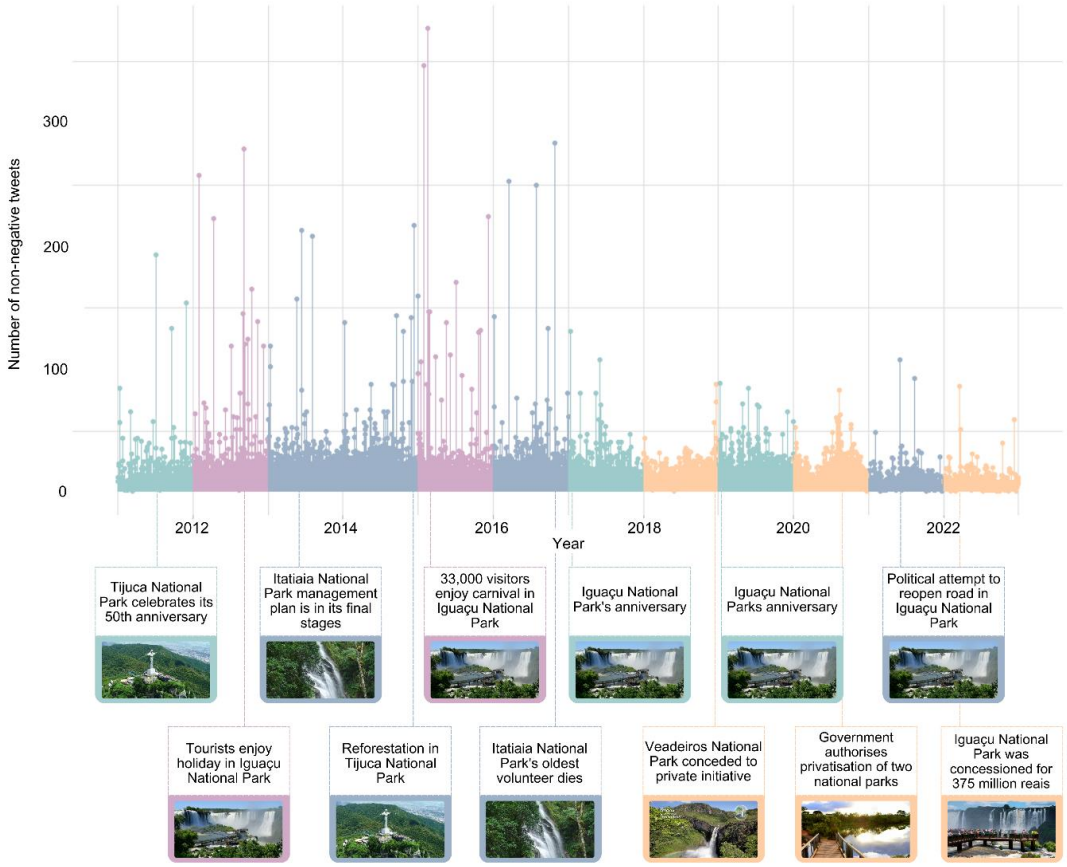
Figure (all colored) and Table Legends



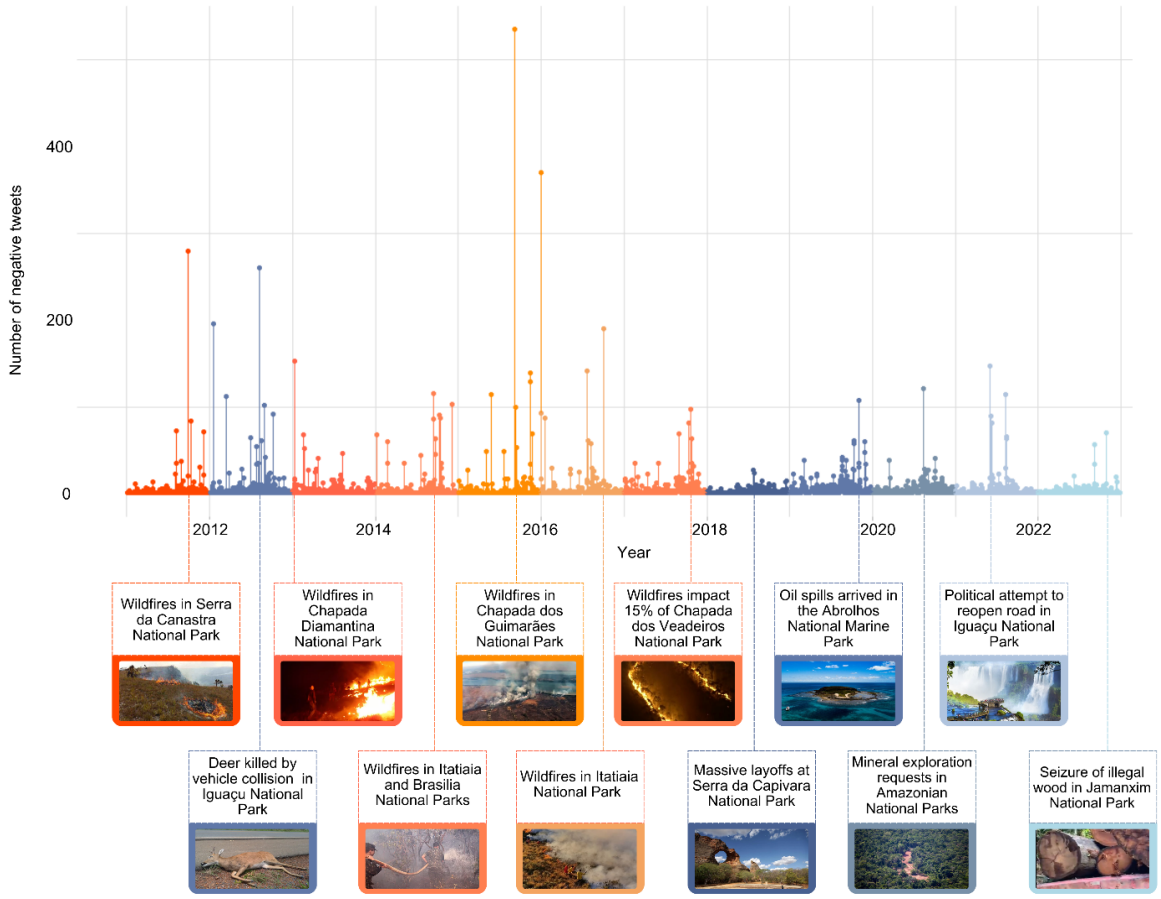
[1.5 column] **Fig. 1.** Map of distribution of 74 Brazilian national parks. Spatial distribution of the 74 Brazilian national parks located in their respective biomes.



[2-column] **Fig. 2.** Methodological flowchart. Methodological flowchart from data collection to results. The flowchart showed all the steps used during the research: data collection, data analysis, and results.



[2-column] **Fig. 3.** The daily counts related to non-negative sentiments of Twitter posts regarding Brazilian national parks from January (2011) to December 2022. The data were obtained from Twitter's application programming interfaces (API). The green colour represents events involving national parks anniversary; the purple represents the visitation in the parks; the orange represents events related with the concessions; and, the blue colour represents others events with peaks of public interest on Twitter.



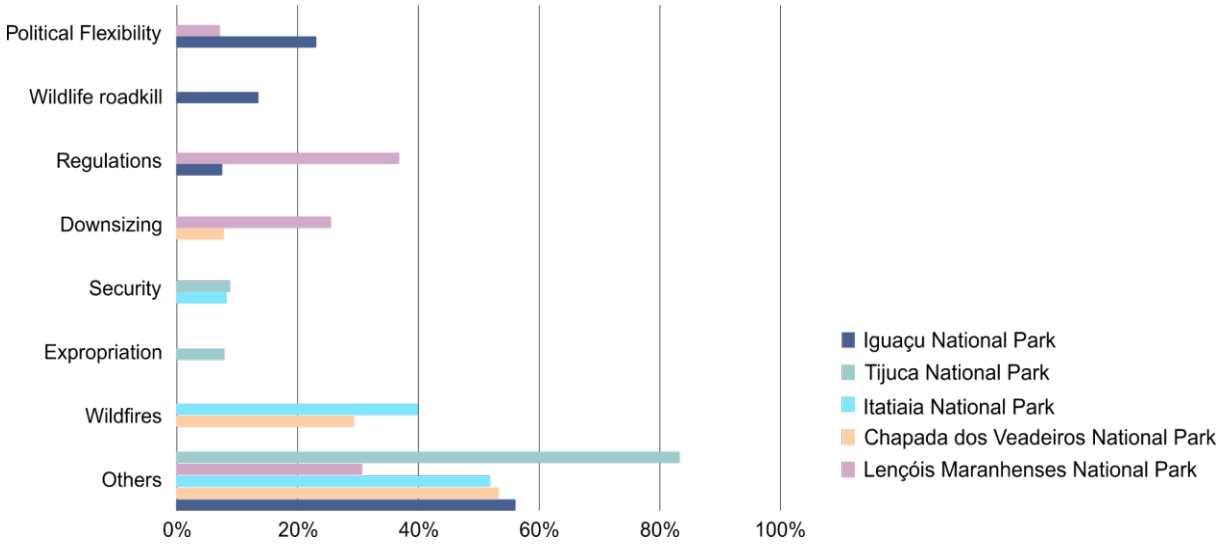
[2-column] **Fig. 4.** The daily counts related to negative sentiments of Twitter posts regarding Brazilian national parks from January (2011) to December 2022. The data were obtained from Twitter's application programming interfaces (API). The orange colours represent events involving wildfires and the blue colours other events with peaks of public interest on Twitter.

Table 1: Topics identified by the BERTopic models during the analysis of the negative tweet dataset. Models 1 to 4 were performed with the nr_topics to auto and different min cluster sizes for the HDBSCAN parameter model.

N of models performed using BERTopic	Topics	#N of tweets	Topic named by BERTopic model	Percentage	Categorised topics
1	1	2929	1_next_christ_national forest_to do.	15.93%	Wildfires
	2	887	2_criminals fire_arson.	4.82%	Wildfires
	3	508	3_canastra_50_50 firefighters_flames ap_fire	2.76%	Lack of financial resources
	4	338	4_cipó minas_gerais fires_minas gerais_wildfires	1.84%	Wildfires
	5	336	5_df_while doing_bite_was walking	1.83%	Security
	6	301	6_dead inside_deer found_found dead	1.64%	Wildlife roadkill
	7	296	7_photo park_publish photo_finished publishing	1.61%	Photo publishing*
	8	254	8_iguaçu police_environmental police_seizes	1.38%	Regulations
	-	12539	Others and Outliers	68.19%	-
2	1	2896	1_christ_station_smoke_points	15.75%	Wildfires
	2	913	2_criminals fire_arson.	4.97%	Wildfires
	3	416	3_flames ap_fire ap_50 firefighters_ap	2.26%	Lack of financial resources
	4	394	4_woman raped_veadeiros park_trail_sp	2.14%	Security
	5	313	5_alerts_km park_inpe_spacial	1.70%	Wildfires
	6	302	6_photo park_publish photo_finished publishing	1.64%	Photo publishing*
	7	296	7_dead inside_deer found_found dead	1.61%	Wildlife roadkill
	8	276	8_environmental police_remove cattle_environmental	1.50%	Regulations
	-	12582	Others and Outliers	68.43%	-

3	-1	2960	1_station_smoke_christ_larger	16.10%	Wildfires
	2	912	2_criminals fire_arson_wildfires	4.96%	Wildfires
	3	326	3_protest_Iguaçu decision_taxis_federal prohibit 2_	1.77%	Regulations
	4	310	4_dead inside_deer found_found dead	1.69%	Wildlife roadkill
	5	293	5_ iguaçu police_environmental police_environmental	1.59%	Regulations
	6	289	6_destroys area_hit 12_12 thousand_main hotspots	1.57%	Wildfires
	7	277	7_ photo park_publish photo_finished publishing	1.51%	Photo publishing*
	8	238	8_privatization_privatization park_privatize	1.29%	Privatization
	-	12783	Others and Outliers	69.52%	-
4	1	2896	1_christ_station_smoke_points	15.75%	Wildfires
	2	913	2_criminals fire_arson_wildfires	4.97%	Wildfires
	3	416	3_flames ap_fire ap_50 firefighters_ap	2.26%	Lack of financial resources
	4	394	4_woman raped_veadeiros park_trail_sp	2.14%	Security
	5	313	5_alerts_km park_inpe_spacial	1.70%	Wildfires
	6	302	6_ photo park_publish photo_finished publishing	1.64%	Photo publishing*
	7	296	7_dead inside_deer found_found dead	1.61%	Wildlife roadkill
	8	276	8_environmental police_remove cattle_environmental	1.50%	Regulations
	-	12582	Others and Outliers	68.43%	-

* The topic "Publishing photos" appears as a negative topic in our analysis, but was not considered in our results because it can be considered a non-negative topic.



[2-column] **Fig. 5.** A bar chart representing the dominant negative topics per park. Each topic is identified using BERTopic analysis and then subjectively classified based on the model output. The parks included are the five most tweeted Brazilian Parks (2011-2022).

Supplementary Materials

Supplementary data associated with this article can be found in the GitHub page (https://github.com/CIBIO-TropiBIO/Sentiment-Analysis-Brazilian-National-Parks).