

**UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
Programa de Pós-Graduação em Diversidade Biológica e Conservação nos
Trópicos**

CAROLINA NEVES SOUZA

**A CONTRIBUIÇÃO DAS REDES SOCIAIS PARA A CONSERVAÇÃO:
compreendendo os interesses e percepções digitais sobre as áreas protegidas
brasileiras.**

**MACEIÓ - ALAGOAS
Março/2024**

CAROLINA NEVES SOUZA

**A CONTRIBUIÇÃO DAS REDES SOCIAIS PARA A CONSERVAÇÃO:
compreendendo os interesses e percepções digitais sobre as áreas protegidas
brasileiras.**

Tese apresentada ao Programa de Pós-Graduação em
Diversidade Biológica e Conservação nos Trópicos,
Instituto de Ciências Biológicas e da Saúde.
Universidade Federal de Alagoas, como requisito para
obtenção do título de Doutora em CIÊNCIAS
BIOLÓGICAS, área de concentração em Conservação
da Biodiversidade Tropical.

Orientadora: Profa. Dra. Ana Cláudia M. Malhado
Coorientador: Prof. Dr. Ricardo A. Correia
Coorientadora: Profa. Dra. Adriana R. Carvalho

**MACEIÓ - ALAGOAS
Março/2024**

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

S729c Souza, Carolina Neves.

A contribuição das redes sociais para a conservação : compreendendo os interesses e percepções digitais sobre as áreas protegidas brasileiras / Carolina Neves Souza. – 2024.

114 f. : il.

Orientadora: Ana Cláudia M. Malhado.

Co-orientador: Ricardo A. Correia.

Co-orientadora: Adriana R. Carvalho.

Tese (Doutorado em Ciências Biológicas) – Universidade Federal de Alagoas. Instituto de Ciências Biológicas e da Saúde. Programa de Pós-Graduação em Diversidade Biológica e Conservação nos Trópicos. Maceió, 2024.

Inclui bibliografias.

Apêndices: f.107-114.

1. Twitter (Rede social on-line) - Engajamento. 2. Análise de sentimento. 3. Monitoramento ambiental. 4. Dados digitais. 5. Culturômica. I. Título.

CDU: 504(81)

*À minha filha Clarice, que me inspira todos os dias
a conservar e ser mais curiosa sobre nossa
relação com a natureza.*

AGRADECIMENTOS

Gostaria de expressar minha gratidão à minha orientadora, Profa. Dra. Cláudia Mendes Malhado pela sua orientação, amizade e apoio ao longo desta jornada acadêmica. Seu comprometimento e paciência foram fundamentais para o desenvolvimento desta tese de doutorado. Além disso, agradeço aos meus coorientadores, Prof. Dr. Ricardo Correia e Profa. Dra. Adriana Carvalho por suas valiosas contribuições e ideias que enriqueceram significativamente a qualidade deste trabalho. Gostaria ainda de agradecer ao Dr. Ricardo por toda a sua generosidade e paciência ao ensinar e compartilhar seu conhecimento e ideias.

Estendo meus agradecimentos aos pesquisadores que me acolheram durante o período de doutorado-sanduíche na Universidade do Porto em Portugal, Msc. Javier Martinez e Dra. Ana Sofia Vaz. Sua hospitalidade foi muito importante para o meu crescimento acadêmico. Um abraço especial aos colegas do LACOS21, cuja colaboração e amizade foram essenciais para enfrentar os desafios e celebrar os sucessos, especialmente ao Prof. Dr. Richard Ladle, João Arthur Almeida, Thainá Lessa, Evelynne Cardoso, Ludmila Costa-Pinto, Ana Carla Rodrigues, Jhonatan Guedes-Santos e Felipe Vieira. Finalmente, reconheço o trabalho dedicado e as contribuições construtivas dos revisores dos artigos científicos originados deste trabalho e dos membros das minhas bancas de avaliação do doutorado, cujas sugestões valiosas contribuíram para a qualidade e rigor deste estudo. A todos, meu sincero agradecimento por fazerem parte desta jornada científica.

Agradeço ao CNPq, ao FUNBIO e ao Instituto Humanize pelas bolsas concedidas. Os apoios financeiros destas instituições foram vitais para realização desta pesquisa, e para que eu, enquanto pesquisadora, pudesse me dedicar integralmente a este trabalho.

Por fim, para além do ambiente acadêmico, desejo expressar minha profunda gratidão à minha família, que sempre me apoiou, incentivou meus estudos e acreditou nos meus sonhos. Um agradecimento especial é destinado ao meu marido, Iran, e a minha filha, Clarice, pelo amor e compreensão que tornaram esta jornada possível.

RESUMO

As áreas protegidas (APs) desempenham um papel vital na proteção do patrimônio natural e cultural, ao mesmo tempo que apoiam os meios de subsistência locais. No entanto, apesar dessa importância, no Brasil, elas enfrentam desafios ligados à falta de recursos financeiros e a percepção de baixa eficácia da gestão, o que pode resultar na falta de apoio da sociedade com relação a essas áreas. Diante desse cenário, é necessário compreender melhor as percepções e interesses do público pelas APs brasileiras. Nesta tese, foram utilizados dados da rede social Twitter (renomeada X) para explorar tendências de interesse e percepções sobre as áreas protegidas no Brasil. Para isso, na primeira fase, foram coletados tweets em português sobre todas as categorias de APs no período de 2011 a 2020. Em seguida, uma metodologia inovadora de análise de sentimentos foi aplicada, focando especificamente nos parques nacionais (PARNAs) brasileiros, e ampliando o período de coleta de dados até 2022. Os conteúdos textuais dos tweets foram analisados com base em métricas de postagem e engajamento dos usuários (curtidas e retuïtes), classificação dos sentimentos expressos no texto e modelagem de tópicos. Os resultados indicam que o número de usuários/tweets que postam sobre as APs brasileiras permaneceu estável ao longo do período amostral, no entanto, o engajamento cresceu consideravelmente a partir de 2018, coincidindo com mudanças no governo federal. Embora os parques nacionais tenham recebido mais menções, especialmente relacionadas às atividades turísticas, os tweets relacionados aos conflitos nas APs atraíram mais discussões. A análise de sentimentos identificou 18.388 (17,30%) de postagens expressando sentimentos negativos em relação aos PARNAs, sendo a maioria relacionada aos incêndios florestais ocorridos entre 2011 e 2017 e ao impacto das decisões governamentais que afetaram os esforços de conservação pós-2019. Foram identificados seis tópicos de discussão proeminentes: (1) Incêndios florestais; (2) Segurança; (3) Regramentos; (4) Vida selvagem morta por atropelamento; (5) Privatização (Concessões); e, (6) Falta de recursos financeiros, refletindo a variedade de sentimentos negativos em relação aos parques. Além disso, a modelagem de tópicos por parques revelou-se benéfica na identificação de diferentes problemas e conflitos nos cinco PARNAs mais tuitados, facilitando assim ações de conservação direcionadas. Neste sentido, o estudo destaca a importância da análise de dados das redes sociais para compreender o interesse público e promover uma gestão mais eficaz das áreas protegidas. Isso pode subsidiar ações de conservação, melhorar a experiência dos visitantes e comunicar a importância das APs para a sociedade. Em última análise os resultados ressaltam o valor da culturômica para identificar lacunas e promover melhorias que alcancem o apoio público às áreas protegidas brasileira.

Palavras-chave: Engajamento. Análise de sentimento. Twitter. Monitoramento ambiental. Dados digitais. Culturômica.

ABSTRACT

Protected areas (PAs) play a vital role in protecting natural and cultural heritage while supporting local livelihoods. However, despite this importance, in Brazil they face challenges linked to a lack of financial resources and the perceived low effectiveness of management, which can result in a lack of support from society for these areas. Given this scenario, it is necessary to better understand the public's perceptions and interests in Brazilian PAs. In this thesis, data from the social media Twitter (renamed X) was used to explore trends in interest and perceptions of protected areas in Brazil. To do this, in the first phase, tweets were collected in Portuguese about all categories of PAs from 2011 to 2020. Next, an innovative sentiment analysis methodology was applied, focussing specifically on Brazilian national parks (PARNAs) and extending the data collection period to 2022. The textual content of the tweets was analysed based on posting metrics and user engagement (likes and retweets), classification of the sentiments expressed in the text and topic modelling. The results indicate that the number of users/tweets posting about Brazilian PAs remained stable throughout the sample period, however, engagement grew considerably from 2018 onwards, coinciding with changes in the federal government. Although national parks received more mentions, especially related to tourism activities, tweets related to conflicts in PAs attracted more discussion. Sentiment analysis identified 18,388 (17.30%) posts expressing negative sentiment towards PARNAs, with the majority related to the forest fires that occurred between 2011 and 2017 and the impact of government decisions affecting post-2019 conservation efforts. Six prominent discussion topics were identified: (1) Forest fires; (2) Security; (3) Regulations; (4) Wildlife killed by trampling; (5) Privatisation (Concessions); and, (6) Lack of financial resources, reflecting the variety of negative sentiments towards the parks. In addition, topic modelling by park proved beneficial in identifying different problems and conflicts in the five most tweeted PARNAs, thus facilitating targeted conservation actions. In this sense, the study highlights the importance of analysing social media data to understand public interest and promote more effective management of protected areas. This can subsidise conservation actions, improve the visitor experience and communicate the importance of PAs to society. Ultimately, the results emphasise the value of culturomics in identifying gaps and promoting improvements that achieve public support for Brazilian protected areas.

Keywords: Engagement. Sentiment analysis. Twitter. Environmental monitoring. Digital data. Culturomics.

LISTA DE FIGURAS

Revisão de literatura

Figura 1 - Boxplot representando a distribuição do volume de pesquisa relativa global 25

Capítulo 1

Figura 1 - Methodological flowchart.....	41
Figura 2 - Volume of posts, users and engagement of Brazilian protected areas	46
Figura 3 - Bootstrapped estimates of mean post engagement per year	47
Figura 4 - Public interest related to the Brazilian protected areas over the years.....	48
Figura 5 - Bootstrapped estimates of mean post engagement per year and per user type	49
Figura 6 - Geographical distribution of the tweets posted about Brazilian PAs	50
Figura 7 - Brazilian PAs that have generated the most public interest and engagement	51
Figura 8 - Most published content related to Brazilian protected areas	54

Capítulo 2

Figura 1 - Map of distribution of 74 Brazilian national parks.....	76
Figura 2 - Methodological flowchart.....	77
Figura 3 - The daily counts related to non-negative sentiments of Twitter posts regarding Brazilian national parks from January (2011) to December 2022.....	84
Figura 4 - The daily counts related to negative sentiments of Twitter posts regarding Brazilian national parks from January (2011) to December 2022.....	85
Figura 5 - A bar chart representing the dominant negative topics per park	86

Apêndice B

Figura 1 - Word frequency about product reviews on Buscapé	110
Figura 2 - Boxplot illustrating the sentiment per word in the reviews from the Buscapé website	110

LISTA DE TABELAS

Capítulo 2

Tabela 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.87

Apêndice A

Tabela 1 - Keywords used in the query for tweets about Brazilian protected areas107

SUMÁRIO¹

1 Apresentação.....	13
Referências	15
2 Revisão da literatura	17
2.1 Áreas protegidas brasileiras	17
2.2 Percepções, atitudes e sentimentos relacionados as áreas protegidas	20
2.3 Redes sociais e conservação ambiental	22
2.3.1 Limitações das redes sociais como ferramenta investigativa	26
Referências	27
3 Objetivos	34
3.1 Objetivo geral	34
3.2 Objetivos específicos.....	34
4 Capítulo 1 - Assessing Brazilian protected areas through social media: insights from 10 years of public interest and engagement.....	35
4.1 Abstract.....	36
4.2 Introduction	36
4.3 Material and Methods.....	39
4.3.1 Brazilian protected areas.....	39
4.3.2 Data collection.....	40
4.3.3 Data analysis.....	43
4.4 Results	46
4.4.1 Volume of tweets about Brazilian PAs.....	46
4.4.2 Users characteristics	48
4.4.3 Content analysis of tweets.....	50
4.4.4 Which protected areas generate most public interest and engagement?	51

¹ Este sumário está organizado conforme a estrutura da tese, a qual foi dividida em dois capítulos elaborados no formato de artigos científicos. Os artigos e suas respectivas seções foram escritos na língua inglesa, de acordo com as exigências de submissão às revistas científicas pertinentes.

4.4.5 Geographic focus of tweets	53
4.5 Discussion	55
4.6 Conclusions.....	60
4.7 Acknowledgments.....	61
References	61
5 Capítulo 2 - Using social media and machine learning to understand negative sentiments towards Brazilian National Parks	69
5.1 Abstract.....	70
5.2 Introduction	70
5.3 Material and Methods.....	74
5.3.1 Study area.....	75
5.3.2 Data collection.....	76
5.3.3 Sentiment analysis	78
5.3.4 Time series analysis.....	80
5.3.5 Topic modelling analysis	81
5.4 Results	82
5.4.1 Public perceptions about Brazilian National Parks	82
5.4.2 Trends of non-negative perceptions over time	83
5.4.3 Trends of negative perceptions over time.....	84
5.4.4 Main topics associated negative perceptions	85
5.5 Discussion	89
5.5.1 Non-negative perceptions of Brazilian National parks	89
5.5.2 Negative perceptions of Brazilian National parks	90
5.5.3 Potential, limitations and future research.....	93
References	95
6 Considerações finais	103
Apêndice A - Palavras-chave utilizadas na coleta de dados no Twitter	107
Apêndice B - Informações metodológicas sobre a análise de sentimento com dados de tweets em português	108

1 APRESENTAÇÃO

As áreas protegidas (APs) são essenciais para a conservação da biodiversidade e possuem significativo valor cultural e socioeconômico no Brasil (Maretti et al., 2012). O sistema brasileiro de APs é um dos maiores do mundo, mas enfrenta uma série de desafios que ameaçam sua integridade e sustentabilidade em longo prazo. Entre os desafios estão a dependência de sistemas de governança ultrapassados; crise institucional e falta de transparência (Gerhardinger et al., 2011; Bragagnolo et al. 2016). Além disso, as APs podem ser enxergadas por parte dos cidadãos e formuladores de políticas como obstáculos ao desenvolvimento socioeconômico (Bernard et al., 2014).

Estudos na área destacam a falta de apoio da sociedade como uma das principais barreiras para a efetividade das APs (Souza, 2017; Cozzolino e Irving, 2016; McClanahan et al., 2005), ressaltando a necessidade de uma abordagem inclusiva que promova uma boa comunicação entre o órgão gestor das áreas protegidas e seus usuários (Macedo e Medeiros, 2018).

As interações humanas com as áreas protegidas variam amplamente, oferecendo benefícios físicos e psicológicos. No entanto, as restrições de uso do espaço e a falta de conhecimento sobre a importância dessas áreas, podem contribuir para o desinteresse público e gerar um sentimento de distanciamento e desconexão das pessoas em relação ao meio ambiente natural (de Hann et al., 2014). O apoio público desempenha um papel fundamental na legitimação das APs, por isso é importante compreender essas interações, sentimentos e interesses. Neste contexto, a coleta e análise de percepções online através das redes sociais ou outras plataformas surge como uma ferramenta promissora para alcançar esse entendimento (Ladle et al., 2017). Estes métodos têm sido utilizados com sucesso para avaliar o ecoturismo em ambientes conflituosos, aumentar a sensibilização pública sobre a conservação e compreender as percepções e sentimentos do público sobre as áreas protegidas (Toivonen et al., 2019). No Brasil, dados do Google Trends e Wikipedia já foram utilizados anteriormente para avaliar o interesse público e a importância da Internet em relação às APs brasileiras (Guedes-Santos et al., 2021; Correia et. al 2018). No entanto, pesquisas adicionais são

imprescindíveis para uma análise mais aprofundada da interação entre os usuários das redes sociais e as áreas protegidas. É importante investigar não apenas o conteúdo publicado pelos usuários, mas também sua origem geográfica, os tópicos mais debatidos e compartilhados, além de uma análise mais detalhada dos sentimentos do público em relação a essas áreas.

Nesta tese, foram utilizados dados da plataforma da rede social Twitter (recentemente renomeada X) para investigar o interesse e os sentimentos do público sobre as APs brasileiras. O Twitter é uma plataforma de rede social e microblogging bastante popular, com mais de 666 milhões de usuários ativos em 2023 em todo o mundo (Statista, 2023a). No Brasil, o Twitter tem em torno de 24,3 milhões de usuários ativos (Statista, 2023b), que publicam milhões de comentários (os chamados "tweets") todos os dias contendo pensamentos e opiniões. O Twitter é muito usado por jornalistas, cientistas, políticos, gestores e pela sociedade em geral (Collins et al. 2016; Mohammadi et al. 2018) para divulgar informações, promover o discurso público e, portanto, serve como um barômetro potencialmente sensível da opinião pública.

Desta maneira, esta tese representa uma oportunidade significativa para ampliar a discussão sobre o interesse da sociedade em relação às APs brasileiras. Isso envolve não apenas a expansão do diálogo prático, mas também do debate teórico, relacionado as abordagens metodológicas inovadoras, como a aplicação da análise de dados de redes sociais para compreender as percepções e sentimentos online sobre as áreas protegidas no Brasil.

A fim de abordar os diversos aspectos desta discussão, esta tese foi dividida em dois capítulos elaborados em formato de artigo científico. O primeiro capítulo intitulado: *Assessing Brazilian protected areas through social media: insights from 10 years of public interest and engagement*, aborda a análise das métricas de publicação e engajamento dos usuários do Twitter (curtidas e retuítes), buscando identificar padrões e fatores relacionados ao interesse público em todas as áreas protegidas brasileiras. O segundo capítulo intitulado: *Using social media and machine learning to understand negative sentiments towards Brazilian National Parks*, concentra-se na análise dos sentimentos da

sociedade em relação aos parques nacionais brasileiros e explora como essa ferramenta analítica pode contribuir para a identificação de problemas e conflitos nestas áreas.

REFERÊNCIAS

- Bernard, E., Penna, L. A. O., & Araújo, E. (2014). Downgrading, downsizing, degazettement, and reclassification of protected areas in Brazil. *Conservation Biology*, 28(4), 939–950. <https://doi.org/10.1111/cobi.12298>
- 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.
- Collins, K., Shiffman, D., & Rock, J. (2016). How are scientists using social media in the workplace? *PLoS ONE*, 11(10). <https://doi.org/10.1371/journal.pone.0162680>
- 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. *Ecological Indicators*, 85, 165–171. <https://doi.org/10.1016/j.ecolind.2017.10.033>
- Cozzolino, L. F. F., & Irving, M. de A. (2016). Por Uma Concepção Democrática De Governança Para a Esfera Pública. *Revista Políticas Públicas*, 19(2), 497. <https://doi.org/10.18764/2178-2865.v19n2p497-508>
- 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. *Technological Forecasting and Social Change*, 85, 121–132. <https://doi.org/10.1016/j.techfore.2013.09.005>
- Gerhardinger, L. C., Godoy, E. A. S., Jones, P. J. S., Sales, G., & Ferreira, B. P. (2011). Marine protected areas: The flaws of the Brazilian national system of marine protected areas. *Environmental Management*, 47(4), 630–643. <https://doi.org/10.1007/s00267-010-9554-7>
- Ladle, R. J., Jepson, P., Correia, R. A., & Malhado, A. C. M. (2017). The power and the promise of culturomics. *Frontiers in Ecology and the Environment*, 15(6), 290–291. <https://doi.org/10.1002/fee.1506>
- Lockwood, M. (2010). Good governance for terrestrial protected areas: A framework, principles and performance outcomes. *Journal of Environmental Management*, 91(3), 754–766. <https://doi.org/10.1016/j.jenvman.2009.10.005>

Macedo, H. S., & Medeiros, R. P. (2018). Rethinking governance in a Brazilian multiple-use marine protected area. *Marine Policy*, July, 0–1. <https://doi.org/10.1016/j.marpol.2018.08.019>

Maretti, Cláudio C; Catapan, Marisete; Abreu, Maria Jasylene; Oliveira, J. E. Dantas. (2012). Áreas protegidas: definições, tipos e conjuntos – reflexões conceituais e diretrizes para a gestão.

McClanahan, T., Davies, J., & Maina, J. (2005). Factors influencing resource users and managers' perceptions towards marine protected area management in Kenya. *Environmental Conservation*, 32(1), 42–49. <https://doi.org/10.1017/S0376892904001791>

Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. In *PLoS ONE* (Vol. 13, Issue 5). Public Library of Science. <https://doi.org/10.1371/journal.pone.0197265>

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>

Statista. Most popular social networks worldwide as of October 2023, ranked by number of monthly active users. 2023a. Disponível em: <<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>> Acessado em: 15 de janeiro de 2024.

Statista. Leading countries based on number of X (formerly Twitter) users as of January 2023b. Disponível em: <<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>> Acessado em: 15 de janeiro de 2024.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, 233(January), 298–315. <https://doi.org/10.1016/j.biocon.2019.01.023>

2 REVISÃO DA LITERATURA

A presente revisão de literatura está dividida em três partes, as quais fundamentam a pesquisa desta tese. A primeira parte introduz o conceito e os referenciais teóricos, proporcionando uma visão abrangente dos principais desafios para a efetividade das áreas protegidas brasileiras. A segunda parte destaca os conceitos relacionados à percepção ambiental subjacentes a este estudo e explora como as teorias comportamentais podem ser aplicadas ao ambiente das redes sociais. Por fim, a terceira parte apresenta tanto o potencial quanto as limitações do uso das redes sociais como ferramenta de investigação das percepções e sentimentos da sociedade em relação à conservação da natureza.

2.1 Áreas protegidas brasileiras

As Áreas Protegidas (APs) desempenham um papel crucial como estratégia para fomentar a preservação dos recursos biológicos e a utilização sustentável dos benefícios naturais, abrangendo serviços ecossistêmicos e práticas culturais em escala global (Maretti, et al., 2012; Watson et al., 2014). No Brasil, as áreas protegidas são legalmente instituídas como unidades de conservação². De acordo com a legislação que institui o Sistema Nacional de Unidades de Conservação (SNUC), uma unidade de conservação é definida como um “espaço territorial e seus recursos ambientais, incluindo as águas jurisdicionais, com características naturais relevantes, legalmente instituído pelo Poder Público, com objetivos de conservação e limites definidos, sob regime especial de administração, ao qual se aplicam garantias adequadas de proteção” (BRASIL, 2000, art. 2º, inciso I). As terras indígenas também contribuem significativamente para a conservação da biodiversidade, embora não sejam formalmente reconhecidas como unidades de conservação pelo SNUC.

² Vale ressaltar que, a presente pesquisa concentrou-se exclusivamente nas unidades de conservação. No entanto, a fim de manter a consistência terminológica em toda a tese, incluindo os artigos publicados nas revistas apresentadas nos próximos capítulos, as unidades de conservação são consideradas, neste trabalho, como áreas protegidas.

Dada a condição de país megadiverso, o Brasil possui uma grande responsabilidade na conservação de sua biodiversidade (Rylands e Brandon, 2005). Segundo o Cadastro Nacional de Unidades de Conservação (CNUC) de 2023, o Brasil abriga um total de 2.859 áreas protegidas, que cobrem coletivamente uma área de aproximadamente 2.583.237 km². Considerando a área continental brasileira, 19,01% estão protegidos por APs; enquanto na zona econômica exclusiva brasileira, 26,49% estão sob proteção por APs (MMA, 2023).

O estabelecimento de um sistema nacional de áreas protegidas e a ampliação da criação destas áreas, está em consonância com as metas globais de conservação delineadas pela Convenção sobre Diversidade Biológica (CDB, 1992; BRASIL, 1994; BRASIL, 1998). Entretanto, permanece a incerteza quanto à efetividade dessas áreas na conservação da biodiversidade e em seus impactos socioeconômicos no Brasil. As áreas protegidas tiveram origem no modelo americano de Yellowstone, o primeiro Parque do Mundo criado em 1872. A adoção desse modelo, estimulou a exclusão das populações tradicionais residentes dentro ou nos arredores dessas áreas as quais foram estabelecidas primordialmente para a preservação dos recursos naturais e para servirem como locais para a apreciação da natureza (DIEGUES, 1996). Essa abordagem preservacionista, desde o início, restringiu a capacidade das populações tradicionais de se identificarem com as áreas protegidas e de desenvolverem um senso de pertencimento e responsabilidade sobre elas.

Em decorrência disso, diversos episódios de degradação ambiental foram desencadeados pela emergência de conflitos resultantes da marginalização das comunidades tradicionais presentes nessas áreas. Paralelamente, a adoção de práticas não sustentáveis e o avanço urbano, aliado aos padrões crescentes de consumo, contribuíram para a proliferação de novas categorias de áreas protegidas em todas as esferas governamentais (SELVA et al., 2016). Os desafios enfrentados pelo sistema brasileiro de áreas protegidas, além daqueles decorrentes do modelo top-down de sua criação, são notáveis. Este sistema, um dos mais vastos do mundo, confronta ameaças que comprometem sua integridade e sustentabilidade em longo prazo. Dentre tais desafios, destacam-se: i) a dependência de sistemas de governança desatualizados e centralizados, prejudicando a participação direta da sociedade (Engen et al., 2021); ii)

uma crise institucional persistente e de longa duração na agência federal responsável pela conservação da biodiversidade (Gerhardinger et al., 2011); iii) a falta de transparência nas práticas de gestão e na comunicação sobre a importância dessas áreas, intensificando a distância entre a sociedade e a natureza e podendo resultar no aumento dos crimes ambientais (Bragagnolo et al., 2016); e, iv) um déficit crescente de financiamento, enfraquecendo as áreas protegidas que não conseguem cobrir integralmente seus custos de gestão (Silva et al., 2021).

Ademais, as áreas protegidas são frequentemente interpretadas como custos de oportunidade, representando obstáculos ao desenvolvimento econômico segundo perspectivas de alguns políticos e tomadores de decisão (Ferreira et al., 2014). De fato, eventos de recategorização, redução e extinção de áreas protegidas (PADDD) impactaram uma extensão de 72.892 km² de áreas protegidas no Brasil entre os anos de 1981 e 2012 (Bernard et al., 2014). Salvar as áreas protegidas brasileiras contra os eventos PADDD exige que as atitudes negativas sejam combatidas por meio da demonstração de seu valor para a sociedade, conforme exemplificado por Jepson et al. (2017), e da evidenciação junto aos políticos de que tais áreas contam com um amplo respaldo público (Guedes-Santos et al., 2021). Em síntese, é de suma importância fomentar uma conexão mais robusta entre as áreas protegidas brasileiras e a sociedade, com vistas a evitar que sejam percebidas como custos de oportunidade pelos cidadãos e líderes políticos (Bernard et al., 2014).

A compreensão da percepção das pessoas sobre o ambiente protegido não apenas é crucial para avaliar o valor das áreas protegidas para a sociedade, mas também desempenha um papel fundamental na eficácia dessas áreas. A análise conduzida por Macedo e Medeiros (2018) destaca que a colaboração da sociedade na promoção da eficácia das áreas protegidas está intrinsecamente vinculada à interação entre os incentivos participativos e os incentivos ao conhecimento. Nesse sentido, a promoção de ações transparentes e comunicativas entre indivíduos e instituições torna-se não apenas essencial, mas imperativa, com o intuito de ampliar o interesse da sociedade na gestão do território das áreas protegidas e assegurar a inclusão de todos os atores sociais nos processos decisórios (Souza et al., 2022).

2.2 Percepções, atitudes e sentimentos relacionados as áreas protegidas

As distintas maneiras como as pessoas percebem e avaliam o meio ambiente são sentidas individualmente ou no contexto de grupos sociais, e podem ser influenciadas por diferentes experiências, aspirações e contextos socioeconômicos (McNeill, 2002). Além disso, a percepção ambiental pode ser vista como um processo que vai além das respostas sensoriais, incorporando impressões e sentimentos que refletem respostas mentais resultantes de experiências individuais. Essas experiências estão intrinsecamente associadas a processos culturais, como destacado por Hoeffel e Fadini (2007), e são enriquecidas pelas memórias e emoções que geram uma conexão significativa das pessoas com os lugares (Tuan, 1980).

As áreas protegidas podem evocar tanto sentimentos positivos quanto negativos nas pessoas. A interação com a natureza não apenas desencadeia emoções positivas, contribuindo para a melhoria da saúde física, mental e psicológica (Velarde et al., 2007), mas também atende a uma necessidade fundamental da sociedade. A percepção dos atributos naturais dessas áreas pode significativamente influenciar a qualidade de vida das pessoas (De Haan et al., 2014). Esses sentimentos positivos podem fomentar um vínculo favorável com a natureza, promovendo um senso de pertencimento ao local (Raffestin, 1993; Tuan, 1980) e estimulando comportamentos pró-conservação ambiental (Hausmann et al., 2016).

Por outro lado, a insatisfação associada a experiências desfavoráveis na visita de áreas protegidas (Hausmann et al., 2020) ou questões relacionadas aos objetivos de criação e gestão dessas áreas, como tamanho, restrições de uso do espaço e regulamentações relacionadas à pesca, caça e proteção de espécies (Bragagnolo et al., 2016; Souza et al., 2022), além da percepção de que as áreas protegidas prejudicam o crescimento econômico, podem resultar na falta de apoio à sua criação, desinteresse e alienação em relação à natureza (Hausmann et al., 2020).

É importante reconhecer que os sentimentos individuais não se limitam apenas a expressões transitórias de emoções como alegria, raiva, interesse, tristeza e gratidão. Em vez disso, os sentimentos das pessoas em relação as áreas protegidas podem causar

um impacto de longo prazo no comportamento das pessoas em relação ao meio ambiente (Fredrickson, 2001).

A compreensão da percepção, conforme destacado por Lemberg (2010), engloba a forma como as pessoas sentem, processam mentalmente e respondem às informações derivadas do ambiente, sendo influenciada por características sociodemográficas, atitudes e valores. Esses elementos têm o potencial de impactar diretamente a experiência, a satisfação e os comportamentos associados às áreas protegidas (Rossi et al., 2015). No contexto da conservação, a conexão entre os sentimentos das pessoas e suas percepções desempenha um papel decisivo, uma vez que sentimentos adversos podem fundamentalmente moldar a interpretação e interação das pessoas com o ambiente natural. Um exemplo ilustrativo dessa dinâmica pôde ser observado por Maciel (2015) no Parque Nacional da Tijuca, onde a concessão do Parque acarretou impactos socioeconômicos negativos para a comunidade local, influenciando consequentemente na percepção sobre a área protegida.

Associativamente, as percepções ambientais não são imparciais, refletindo, entre outros aspectos, os interesses de grupos sociais distintos, atitudes, valores e visões de mundo (Tuan, 1980). Pesquisas voltadas para compreender as relações homem-natureza, incluindo percepções, atitudes e comportamentos, são essenciais para uma aplicação mais eficaz e adaptada das medidas de conservação em áreas protegidas (Badola et al., 2012; Bragagnolo et al., 2016; Broad e Sanchirico, 2008; Souza et al., 2022). Esses esforços de pesquisa estão sendo complementados por estudos de percepções e opiniões sobre o meio ambiente nas redes sociais (Fink et al., 2020; Hausmann et al., 2018), proporcionando uma compreensão mais abrangente das dinâmicas contemporâneas nas relações homem-natureza.

No âmbito teórico, duas importantes bases dos estudos sobre percepções, atitudes e comportamentos são a Teoria da Ação Racional (TAR) de Fishbein & Ajzen (1975) e a Teoria do Comportamento Planejado (TCP) de Ajzen (1991). A TCP é uma extensão da teoria da TAR. A TAR pressupõe que a intenção comportamental está relacionada a dois fatores: (i) a atitude sobre o comportamento baseada nas crenças comportamentais pessoais e (ii) normas subjetivas relacionadas às pressões sociais. No

entanto, a teoria recebeu muitas críticas por pressupor que a intenção comportamental estava relacionada basicamente a fatores internos (julgamento pessoal favorável ou não sobre determinado ato) e externos (como o ato será percebido pelo outro) (Moutinho e Roazzi, 2010). Ajzen (1991) sugeriu então na TCP que além destes fatores, a intenção comportamental era influenciada por um terceiro fator - o controle comportamental presumido, onde se avalia os obstáculos ou a simplicidade em agir de determinada forma. Essas teorias têm sido valiosas para compreender e prever as motivações por trás das ações humanas relacionadas ao meio ambiente.

No contexto das redes sociais, onde as interações online são proeminentes, as teorias comportamentais podem ser aplicadas de maneira pertinente. Considerando a TAR e a TCP, é possível destacar que a atitude positiva ou negativa em relação a esses comportamentos digitais é moldada por crenças pessoais de cada indivíduo. Além disso, as expectativas sociais, que abrangem as normas subjetivas relacionadas às pressões sociais, influenciam diretamente as decisões das pessoas sobre o que é aceitável socialmente escrever, curtir e compartilhar. No entanto, para além dos fatores internos (pessoais) e externos (sociais), a facilidade ou obstáculo percebido para realizar determinados compartilhamentos de opiniões nas redes sociais desempenha um papel crucial nas intenções e, por conseguinte, no comportamento real.

Dessa forma, ao compreender como as teorias comportamentais se aplicam ao ambiente das redes sociais, podemos estabelecer uma ligação relevante com as percepções sobre as APs. A forma como as pessoas interagem e compartilham opiniões online pode influenciar não apenas seus comportamentos digitais, mas também se estender para suas percepções e atitudes presenciais sobre questões ambientais, incluindo as áreas protegidas.

2.3 Redes sociais e conservação ambiental

Tradicionalmente, as interações humanas com a natureza têm sido investigadas por meio de pesquisas sociais. Recentemente, sugeriu-se que os enormes volumes de dados gerados pelas pessoas nas redes sociais online e outras plataformas digitais poderiam ser uma abordagem complementar utilizada para quantificar essas interações

em uma escala maior de população e acessibilidade geográfica (Di Minin et al., 2015) - esse campo de estudo é chamado de culturômica da conservação (Ladle et al., 2016). A culturômica refere-se à aplicação de métodos computacionais e análise de dados em grande escala para estudar padrões culturais e mudanças ao longo do tempo, utilizando grandes conjuntos de dados: textuais, vídeos, fotos e dados compartilhados em meio digital (Ladle et al., 2017, 2016).

As redes sociais tornaram-se parte integrante do dia a dia moderno, proporcionando oportunidades sem precedentes de comunicação, debate e partilha de informação, para a compreensão das percepções e sentimentos do público (Sudhir and Suresh, 2021). Uma das principais vantagens da análise de dados digitais é a grande quantidade de informações geradas pelas pessoas sobre uma ampla gama de tópicos, incluindo preferências políticas (Ceron et al., 2014), satisfação do cliente (Ahani et al., 2019) e proteção natureza (Ladle et al., 2021; Di Minin et al., 2015; Souza et al., 2023). Os dados das redes sociais podem ser recolhidos rapidamente, com custos financeiros mais baixos do que outros métodos (por exemplo, questionários) (Becken et al., 2017) e podem, portanto, ser usados como complementares – não invalidando o uso de questionários e entrevistas face a face.

Em âmbito comercial, foi publicado um estudo sobre interesse e engajamento na rede social Instagram que mostrou como as métricas de curtidas e comentários influenciam na compra de produtos online (Aragão et al., 2016); outra pesquisa utilizando as redes sociais como ferramenta investigativa mostrou por meio de análise de sentimento com dados do Twitter, que existe uma capacidade notável dos meios de comunicação social em prever os resultados eleitorais, bem como uma correlação significativa entre os resultados das redes sociais e os resultados dos questionários tradicionais (Ceron et al., 2014). Assim sendo, o crescente uso das redes sociais por um público mais vasto de cidadãos vem aumentando fortemente a utilização dos dados da internet como um dispositivo para investigar os interesses públicos e sentimentos da sociedade com relação a natureza.

De fato, a utilização de dados online como do Google Trends e Wikipedia, e de redes sociais como Twitter, Facebook e Instagram já demonstrou a sua aplicabilidade

para medir o interesse público e compreender as percepções dos utilizadores sobre diversas questões de conservação (Almeida et al., 2022; Correia et al., 2021, 2017; Guedes-Santos et al., 2021; Tenkanen et al., 2017). Um estudo específico destinado a compreender o interesse público online em espécies ameaçadas analisou vídeos do YouTube relacionados com o turismo de gorilas das montanhas (*Gorilla beringei beringei*). O estudo de Otsuka e Yamakoshi (2020) mostra que vídeos que apresentam contato físico ou interação com gorilas recebem mais visualizações e curtidas, destacando a importância do monitoramento do interesse das pessoas pela natureza e enfatizando a necessidade de melhorar a comunicação das ações voltadas à proteção de espécies ameaçadas.

Em 2020, outro estudo analisou o interesse público da sociedade com relação aos parques nacionais (PARNAs) no período da pandemia de COVID-19 em todo o mundo (Souza et al., 2021). A pesquisa se concentrou, especificamente, nos impactos da limitação da mobilidade das pessoas (através do isolamento social e de medidas de lockdowns), considerando a forte relação causal do interesse público e a visitação nos PARNAs; avaliando-os por meio do volume de pesquisas globais na Internet para cada parque, utilizando a plataforma Google Trends. Os resultados revelaram que apesar da redução global do interesse público nos PARNAs durante a pandemia, foi encontrada uma diferenciação no comportamento tanto em nível local de cada parque como também quando considerados em nível de país no qual está inserido (Figura 1). Esse entendimento se mostra crucial não apenas para a gestão local, mas também para iniciativas de marketing global, uma vez que as atividades dos visitantes podem resultar em impactos diretos ou indiretos no meio ambiente (Toivonen et al., 2019). Esse conhecimento pode ajudar a orientar os esforços de gestores e formuladores de políticas para uma gestão mais eficaz das APs, identificando quais aspectos e eventos levam a uma atitude pública negativa em relação às APs e abordando estratégias mais eficazes para gerenciar tensões e promover mudanças em favor da conservação (Hausmann et al., 2020; Hockings et al., 2006).

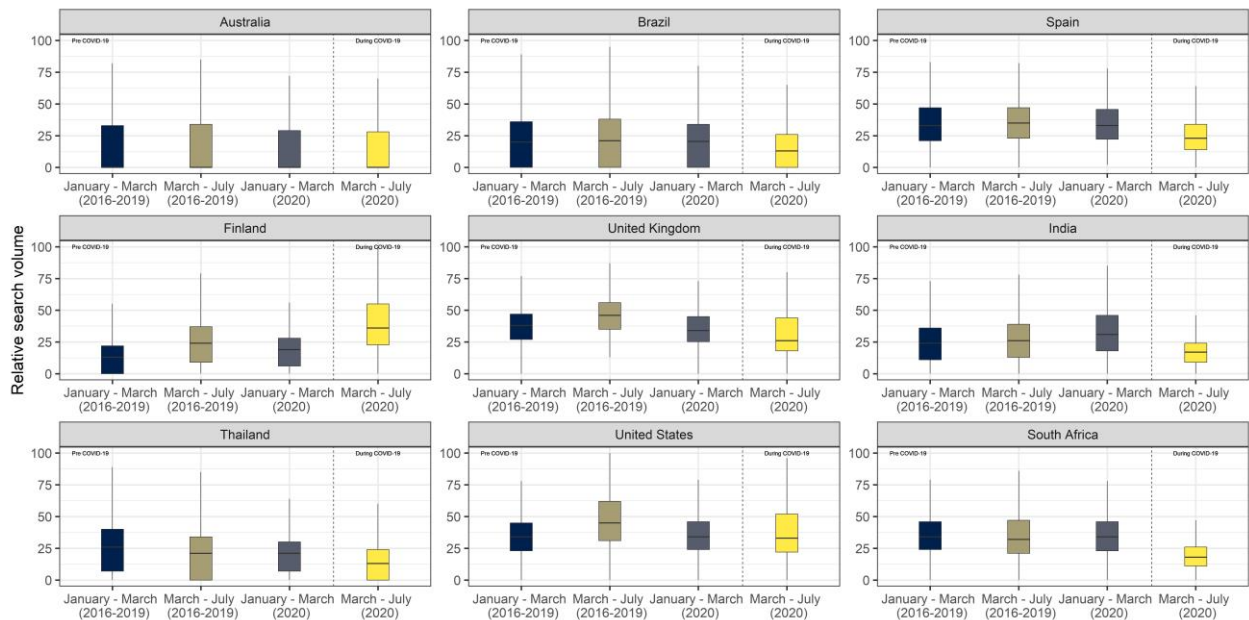


Figura 1: Boxplot representando a distribuição do volume de pesquisa relativa global para parques nacionais na Austrália, Brasil, Espanha, Finlândia, Reino Unido, Índia, Tailândia, Estados Unidos e África do Sul durante quatro períodos de tempo diferentes: janeiro a março de 2016-2019, março a julho de 2016-2020, janeiro a março de 2020 (antes da COVID-19) e março a julho de 2020 (durante a Covid-19). Fonte: Souza et. al. (2021).

Além dos padrões de interesse, outra ferramenta interessante para analisar dados digitais sobre interações entre humanos e natureza é a análise de sentimentos, que aproveita os dados textuais de rede social para compreender os valores e as emoções das pessoas em relação ao ambiente natural (Drijfhout et al., 2016). A análise de sentimento emprega técnicas de processamento de linguagem natural (PLN) para “analisar as opiniões das pessoas, sentimentos, avaliações, apreciações, atitudes e emoções em relação a entidades como produtos, serviços, organizações, indivíduos, questões, eventos, tópicos e seus atributos” (Liu, 2012, p.1), classificando o conteúdo de dados textuais com base em percepções positivas, negativas ou neutras.

Essa abordagem tem uma ampla aplicação em marketing e outros campos de estudo, e, mais recentemente, cientistas da conservação têm empregado essa metodologia para avaliar os sentimentos das pessoas em relação à natureza, incluindo gestão ambiental (Bhatt e Pickering, 2021), impactos do turismo na vida selvagem (Otsuka et al., 2020) e as preferências, experiências e opiniões dos turistas (Hausmann et al., 2020). Ao identificar emoções negativas, como frustração ou decepção, essa

abordagem pode indicar a necessidade de melhorias nas áreas de visitação de áreas protegidas (Agyeman et al., 2019), bem como em práticas de gestão ou estratégias de comunicação. Práticas polêmicas de gestão, como restrições ao uso do espaço e medidas contra a caça ilegal (Lubbe et al., 2019), têm o potencial de gerar insatisfação, falta de apoio e conflitos. Nesse sentido, a aplicação da análise de sentimento em dados de rede social relacionados a áreas protegidas pode ser instrumental para identificar problemas como os mencionados acima, contribuindo para a tomada de decisões informadas e aprimoramento de estratégias de gestão (Agyeman et al., 2019).

2.3.1 Limitações das redes sociais como ferramenta investigativa

Mesmo considerando o potencial significativo das redes sociais para fornecer percepções valiosas para a conservação, existem algumas limitações significativas importantes de serem discutidas. Apesar de aproximadamente 90% dos brasileiros terem acesso à Internet e ao menos 83,6% fazer uso de alguma rede social (IBGE, 2023), é vital reconhecer que essa representação demográfica não abrange uniformemente todo o país. No entanto, essa limitação não invalida o uso dessa abordagem investigativa, uma vez que os resultados obtidos na pesquisa podem ter relevância e extrapolar o escopo dos usuários de redes sociais para a sociedade (Ceron et al., 2014)

Outra limitação referente ao uso de dados online, envolve a presença significativa de gírias, coloquialismos, acrônimos e emoticons em seus conteúdos textuais, o que representa um desafio para a coleta e análise de dados. Além disso, a ausência da geolocalização em dispositivos de alguns usuários pode limitar análises espaciais. Portanto, todos os dados de rede social requerem limpeza extensiva e análise crítica (Toivonen et al., 2019). Apesar destas limitações, os vastos volumes de dados gerados por plataformas de redes sociais como o Twitter por exemplo, podem certamente fornecer insights sobre as percepções de pessoas que não utilizam redes sociais (Ceron et al., 2014) e, de forma crítica, podem avaliar padrões a nível macroscópico entre as interações homem e natureza (Ladle, et al., 2021).

Partindo para a abordagem da análise de sentimento, apesar de seu grande potencial para gerar percepções sobre a conservação da biodiversidade, uma limitação

significativa é a escassa disponibilidade de ferramentas e métodos para outros idiomas além do inglês (Kaity and Balakrishnan, 2020). No contexto brasileiro, a lacuna de estudos voltados a exploração dos sentimentos das pessoas sobre as áreas protegidas, pode se dá pelos desafios metodológicos de aplicar essa ferramenta na língua portuguesa. No entanto, quanto mais pesquisas nesta área forem desenvolvidas, mais dados e resultados serão gerados a fim de fortalecer o uso desta importante ferramenta metodológica.

Mesmo diante das limitações do uso de dados digitais para investigação, é de conhecimento que os dados de culturomics são excepcionais em sua capacidade de capturar, identificar e mapear sistematicamente interações homem-natureza em amplas escalas espaciais e temporais. Diversos estudos já empregaram com sucesso dados de redes sociais para informar comunicação científica (Papworth et al., 2015), explorar o ecoturismo (Wibowo et al., 2019), avaliar o sentimento online em relação a espécies ameaçadas (Fink et al., 2020) e compreender as percepções públicas sobre APs (Hausmann et al., 2018; Souza et al., 2023). Entretanto, mais pesquisas são necessárias para avaliar a relação específica entre o conteúdo das redes sociais e as APs, bem como para investigar os sentimentos do público em relação a essas áreas. Tais pesquisas são fundamentais para auxiliar na gestão da vasta e diversificada rede de APs brasileiras, que têm sido cada vez mais atacados devido a uma tentativa de dismantelar as políticas ambientais e de conservação na última década (Bernard et al., 2014; Fearnside, 2019; Ferrante e Fearnside, 2019).

REFERÊNCIAS

- 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>
- Ajzen, I., 1991. *The Theory of Planned Behavior*.

Almeida, J.A.G.R., et al., 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>

Aragão, F., Farias, F., Mota, M., Freitas, A., 2016. Curtiu, comentou, comprou. A mídia social digital Instagram e o consumo. *Revista Ciências Administrativas* 22, 130–161.
<https://doi.org/10.5020/2318-0722.2016.v22n1p130>

Badola, R., Barthwal, S., Hussain, S.A., 2012. Attitudes of local communities towards conservation of mangrove forests: A case study from the east coast of India. *Estuar Coast Shelf Sci* 96, 188–196. <https://doi.org/10.1016/j.ecss.2011.11.016>

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>

Bragagnolo, C., Gamarra, N.C., Malhado, A.C.M., Ladle, R.J., 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 Federal Nº 9.985, de 18 de julho de 2000. Institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providências. Disponível em: < https://www.planalto.gov.br/ccivil_03/leis/l9985.htm >. Acesso em: 10 de jan. de 2024.

BRASIL, 1998. Decreto legislativo Nº 2519, de 16 de março de 1998. Promulga a Convenção sobre Diversidade Biológica, assinada no Rio de Janeiro, em 05 de junho de 1992. Disponível em: < https://www.planalto.gov.br/ccivil_03/decreto/d2519.htm >. Acesso em: 01 de fev. de 2024.

BRASIL, 1994. Decreto legislativo Nº 02, de 1994. Aprova o texto do Convenção sobre Diversidade Biológica, assinada durante a Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento, realizada na Cidade do Rio de Janeiro, no período de 5 a 14 de junho de 1992. Disponível em: < <https://www2.camara.leg.br/legin/fed/decleg/1994/decretolegislativo-2-3-fevereiro-1994-358280-publicacaooriginal-1-pl.html> >. Acesso em: 01 de fev. de 2024.

Broad, K., Sanchirico, J.N., 2008. Local perspectives on marine reserve creation in the Bahamas. *Ocean Coast Manag* 51, 763–771.
<https://doi.org/10.1016/j.ocecoaman.2008.07.006>

CBD. Convenção da Diversidade Biológica. 1992. Disponível em: <
https://www.planalto.gov.br/ccivil_03/decreto/1998/anexos/and2519-98.pdf>. Acesso em: 01 de fev. de 2024.

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>

Correia RA, Ladle R, Jarić I, Malhado ACM, Mittermeier JC, Roll U, et al. Digital data sources and methods for conservation culturomics. *Conservation Biology*. 2021;35: 398–411. doi:10.1111/cobi.13706

Correia, R.A., Jepson, P., Malhado, A.C.M., Ladle, R.J., 2017. Internet scientific name frequency as an indicator of cultural salience of biodiversity. *Ecol Indic* 78, 549–555.
<https://doi.org/10.1016/j.ecolind.2017.03.052>

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>

Diegues, A.C. O mito moderno da natureza intocada / Antonio Carlos Santana Diegues. — 3.a ed. — São Paulo : Hucitec Núcleo de Apoio à Pesquisa sobre Populações Humanas e Áreas Úmidas Brasileiras, USP, 2000.

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>

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>

Engen, S., Hausner, V.H., Gurney, G.G., Broderstad, E.G., Keller, R., Lundberg, A.K., Murguzur, F.J.A., Salminen, E., Raymond, C.M., Falk-Andersson, J., Fauchald, P., 2021. Blue justice: A survey for eliciting perceptions of environmental justice among coastal planners' and small-scale fishers in Northern-Norway. *PLoS One* 16, 1–20.
<https://doi.org/10.1371/journal.pone.0251467>

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

Ferrante, L., Fearnside, P.M., 2019. Brazil's new president and "ruralists" threaten Amazonia's environment, traditional peoples and the global climate. *Environ Conserv.* <https://doi.org/10.1017/S0376892919000213>

Ferreira, J., Aragão, L.E.O.C., Barlow, J., Barreto, P., Berenguer, E., Bustamante, M., Gardner, T.A., Lees, A.C., Lima, A., Louzada, J., Pardini, R., Parry, L., Peres, C.A., Pompeu, P.S., Tabarelli, M., Zuanon, J., 2014. Brazil's environmental leadership at risk: Mining and dams threaten protected areas. *Science* (1979). <https://doi.org/10.1126/science.1260194>

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>

Fishbein, M. e Ajzen, I. 1975. Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley. 578p

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>

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>

Guedes-Santos, J., Correia, R.A., Jepson, P., Ladle, R.J., 2021. Evaluating public interest in protected areas using Wikipedia page views. *J Nat Conserv* 63. <https://doi.org/10.1016/j.jnc.2021.126040>

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. Percepção ambiental. In: Ferraro-Jr., L.A (org). Encontros e Caminhos: Formação de Educadoras(es) Ambientais e Coletivos Educadores. Brasília: MMA, Departamento de Educação Ambiental, 2007. p. 253-263.

IBGE. Painel PNAD contínua. 2022. Disponível em: < <https://painel.ibge.gov.br/pnadc/> > Acesso em: 15 de janeiro de 2024.

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. *Knowl Inf Syst* 62, 4445–4480. <https://doi.org/10.1007/s10115-020-01497-6>

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>

Ladle, R.J., Correia, R.A., Do, Y., Joo, G.J., Malhado, A.C.M., Proulx, R., Roberge, J.M., Jepson, P., 2016. Conservation culturomics. *Front Ecol Environ* 14, 269–275. <https://doi.org/10.1002/fee.1260>

Ladle, R.J., Jepson, P., Correia, R.A., Malhado, A.C.M., 2017. The power and the promise of culturomics. *Front Ecol Environ* 15, 290–291. <https://doi.org/10.1002/fee.1506>

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., 2019. The impact of rhino poaching on tourist experiences and future visitation to National Parks in South Africa. *Current Issues in Tourism*. <https://doi.org/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 onservcação: 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.

Maharani Wibowo, J., Dra Sri Muljaningsih, S., Dias Satria, Ms., Wibowo, J.M., Muljaningsih, S., Satria, D., 2019. Tripadvisor sentiment analysis: the policy of ecotourism competitiveness from bromo, tengger, and semeru national park. *International Journal of Business, Economics and Law* 20, 18–24.

Maretti, Cláudio C; Catapan, Marisete; Abreu, Maria Jasylene;Oliveira, J.E.Dantas., 2012. Áreas protegidas: definições, tipos e conjuntos – reflexões conceituais e diretrizes para a gestão.

McNeill, J., 2002. Something New Under the Sun: An Environmental History of the Twentieth-Century World. Oxford University Press 36, 183–185.
<https://doi.org/10.1353/jsh.2002.0109>

MMA, 2023. Painel de Unidades de Conservação. Disponível em:
<https://cnuc.mma.gov.br/powerbi> Acesso em: 06 de novembro de 2023.

Moutinho, K., Roazzi, A., 2010. As teorias da ação racional e da ação planejada: relações entre intenções e comportamentos. *Aval. psicol* 9, 279–287.

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>

Raffestin, C., 1993. O que é Território?, in: *Por Uma Geografia Do Poder*. pp. 143–163.

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>

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

Selva, V. S. F.; Souza, C. N.; Gouveia, R. L.; Santos, E. C. S. **Práticas turísticas em áreas protegidas: um olhar sobre a Área de Proteção Ambiental - APA Costa dos Corais, Brasil**. In: Giovanni Seabra. (Org.). *Terra - paisagens, solos, biodiversidade e os desafios para um bom viver*. 1ed.Ituiutaba: Barlavento, 2016, v. 1, p. 1024-1034

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>

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.
<https://doi.org/10.1371/journal.pone.0293581>

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., Rodrigues, A.C., Correia, R.A., Normande, I.C., Costa, H.C.M., Guedes-Santos, J., Malhado, A.C.M., Carvalho, A.R., Ladle, R.J., 2021. No visit, no interest: How COVID-19 has affected public interest in world's national parks. *Biol Conserv* 256. <https://doi.org/10.1016/j.biocon.2021.109015>
- 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>
- 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>
- Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E., 2019. Social media data for conservation science: A methodological overview. *Biol Conserv* 233, 298–315. <https://doi.org/10.1016/j.biocon.2019.01.023>
- Tuan. Y. *Topofilia: um estudo da percepção, atitudes e valores do meio ambiente*. Trad.: Lívia de Oliveira. Londrina: Eduel, 2012.342p.1980-2012.
- Velarde, M.D., Fry, G., Tveit, M., 2007. Health effects of viewing landscapes - Landscape types in environmental psychology. *Urban For Urban Green* 6, 199–212. <https://doi.org/10.1016/j.ufug.2007.07.001>
- 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>

3 OBJETIVOS

3.1 Objetivo geral

Compreender o interesse e percepções em relação às áreas protegidas brasileiras por meio das redes sociais e da análise de sentimentos.

3.2 Objetivos específicos

- i. Quantificar o volume de postagens no Twitter relacionadas às áreas protegidas brasileiras;
- ii. Mapear geograficamente a origem das postagens realizadas no Twitter sobre as áreas protegidas brasileiras;
- iii. Analisar e identificar quais áreas protegidas brasileiras geram maior volume de publicações e engajamento no Twitter;
- iv. Investigar a relação entre o número de publicações e o engajamento dos usuários que postam sobre áreas protegidas no Twitter;
- v. Identificar os tópicos mais discutidos nas postagens sobre áreas protegidas brasileiras no Twitter;
- vi. Classificar os sentimentos públicos (positivos, neutros e negativos) expressos no conteúdo textual do Twitter sobre os parques nacionais brasileiros;
- vii. Identificar os principais tópicos que contribuem para as percepções negativas relacionadas aos parques nacionais brasileiros no Twitter.

4 ASSESSING BRAZILIAN PROTECTED AREAS THROUGH SOCIAL MEDIA: INSIGHTS FROM 10 YEARS OF PUBLIC INTEREST AND ENGAGEMENT ³

Carolina Neves Souza^{1*}, João A. G. R. Almeida², Ricardo A. Correia^{3,4,5}, Richard J. Ladle^{1,6,#a}, Adriana R. Carvalho⁷, Ana C. M. Malhado¹

¹ Programa de pós-graduação em Diversidade Biológica e Conservação nos Trópicos. Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, Alagoas, Brasil

² Instituto de Computação, Universidade Federal de Alagoas, Maceió, Alagoas, Brasil

³ Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

⁴ Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland

⁵ Biodiversity Unit, University of Turku, Turku, Finland

⁶ Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Universidade do Porto, Vairão, Portugal

^{#a} Current Address: BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal

⁷ Departamento de Ecologia, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brasil

³ Artigo publicado na revista PLOS ONE. 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>

4.1 Abstract

Social media platforms are a valuable source of data for investigating cultural and political trends related to public interest in nature and conservation. Here, we use the micro-blogging social network Twitter to explore trends in public interest in Brazilian protected areas (PAs). We identified ~400,000 Portuguese language tweets pertaining to all categories of Brazilian PAs over a ten-year period (1 January 2011 - 31 December 2020). We analysed the content of these tweets and calculated metrics of user engagement (likes and retweets) to uncover patterns and drivers of public interest in Brazilian PAs. Our results indicate that users / tweets mentioning PAs remained stable throughout the sample period. However, engagement with tweets grew steeply, particularly from 2018 onward and coinciding with a change in the Brazilian federal government. Furthermore, public interest was not evenly distributed across PAs; while national parks were the subject of the most tweets, mainly related to tourism activities, tweets related to conflicts among park users and managers were more likely to engage Twitter users. Our study highlights that automatic or semi-automatic monitoring of social media content and engagement has great potential as an early warning system to identify emerging conflicts and to generate data and metrics to support PA policy, governance and management.

4.2 Introduction

Protected areas (PAs) are a key tool for biodiversity conservation [1]. In Brazil, these areas are not only responsible for protecting different ecosystems, habitats and endangered species, they also safeguard important cultural and socioeconomic values [2]. To align social and economic aims, in addition to conservation and recreation, the Brazilian system includes two broad categories of PAs: strictly protected and sustainable use. Indigenous lands also provide an important contribution to biodiversity conservation, although they are not formally recognized as PAs in the National Protected Areas System (SNUC) [3] and are regulated by different legislation [4]. The Brazilian PA system is one of the largest in the world, but is facing a range of challenges that threaten its integrity and long-term sustainability. Among the most pressing of these challenges are: i) its reliance

on outdated top-down governance systems that do not sufficiently allow for the direct participation of society [5]; ii) the long term and persistent institutional crisis facing the federal biodiversity conservation agency [6]; iii) a lack of transparency in management actions and in communicating the importance of these areas, which tends to result in increasing environmental crimes and corresponding diminishment of management effectiveness and monitoring of these areas [7], and; iv) a growing funding deficit, which weakens PAs that are not able to cover their management costs [8]. In summary, fostering a stronger connection between Brazilian PAs and society is crucial to avoid them being perceived as opportunity costs by citizens and politicians [9]. In addition, understanding how people interact with these PAs can provide important insights on how to increase society's support for conservation efforts.

Individuals interact with PAs in a wide range of ways, generating diverse values and evoking different interests and feelings [10]. In Brazil, depending on the PA category, citizens and visitors can engage in a wide range of activities including recreation, research, developing environmental education activities, or simply visiting and enjoying the iconic landscapes and biological spectacles. Such interactions have demonstrable psychological and physical benefits to humans and can promote well-being [11]. Nevertheless, Brazilian PAs are primarily configured for environmental conservation, with varying use restrictions that can contribute to a lack of interest from wider society and a general alienation from nature [12]. Given that public support is critical for the legitimacy of PAs [7], understanding human interactions, sentiment and public interest in PAs is essential for developing effective strategies to attract societal support and for supporting decision-makers and researchers in conservation planning, financing and public communication activities [13–15].

Traditionally, human interactions with nature have been investigated through social surveys which are necessarily costly and limited in scale. It has recently been suggested that the huge volumes of data generated by social media and other digital platforms could be a complementary approach utilised to quantify these interactions at a larger scale of population and geographic accessibility [16] - this field of study is called conservation culturomics [14]. In comparison to questionnaires, social media analysis has the potential to generate large amounts of data, at a lower financial cost, and on a larger geographical

scale [13]. This in no way invalidates the continued use of questionnaires as a valuable methodological tool; data generated from analysis of social networks have many intrinsic biases [17] and, critically, does not capture the attitudes and behaviours of communities/individuals with limited access to the internet and/or those that do not use social networks [12,18]. Another concern is the reliability of how the data is recorded and made available by the system. According to [19], inconsistent data measurements by the system can also undermine internal validity, making it difficult to infer causality from the responses. The authors therefore suggest that researchers familiarise themselves with the system used in the survey in order to validate the results. Although about 90% of Brazilians have access to the Internet, this demographic representation does not cover the entire country. Nevertheless, this limitation does not invalidate the usefulness of this investigative approach, since the results obtained in the research may often have relevance beyond the scope of social network users [18]. Culturomic data is unrivalled in its potential for systematically capturing, identifying and mapping human-nature interactions at large spatial and temporal scales [17]. Indeed, social media data has already been used successfully to inform science communication [20,21], investigate ecotourism in high conflict environments [22], assess online sentiment towards threatened species [23], enhance public awareness of wildlife conservation [24], and to better understand public perceptions and feelings related to PAs [12,13,25]. In Brazil, Google Trends data has previously been used to assess public interest and internet salience in relation to Brazilian PAs [26]. However, more studies are needed to assess the relationship between social media content and PAs, as well as to investigate public sentiment towards these areas.

Here, we use data from the social media platform Twitter (recently renamed X) to investigate public attitudes and interest in Brazilian PAs. Twitter is one of the most popular social media and microblogging platforms with over 436 million active users worldwide in 2021 [27]. In Brazil, Twitter has up to 14.1 million active users, who posted millions of comments (so-called “tweets”) every day containing thoughts and opinions of up to 280 characters [28]. Despite the character limit, the inclusion of links significantly enhances the informational content and potential impact of tweets, as it enables individuals to access additional resources, broaden their knowledge, and engage in more enriching and

informed discussions. The company has promoted itself as the right place to learn more about “what’s going on” and “what people are talking about right now”. Twitter is heavily used by journalists, scientists, politicians, managers, and wider society [29,30] to spread information, promote public discourse, and thus serves as a potentially sensitive barometer of public opinion [31]. However, it is important to note that due to its open nature Twitter can also be responsible for the dissemination of misinformation and the spread of fake news.

In this study we analysed ten years (2011-2020) of public tweets in Portuguese language that contained content related to Brazilian PAs with the objective of answering the following questions: (i) what is the volume of posts on Twitter about Brazilian PAs? (ii) where are people communicating about Brazilian PAs? (iii) which types of Brazilian PAs generate more posts and engagement (e.g., other users reacting to the original posts through ‘likes’ and ‘retweets’)? (iv) what is the relationship between the number of posts and the engagement of users who post about PAs? (v) what are the most discussed topics in the posts? Answering these questions through digital data analytics can inform targeted communication strategies, improve public engagement and foster a deeper understanding of biodiversity conservation, leading to more effective and impactful conservation initiatives.

4.3 Material and Methods

4.3.1 Brazilian protected areas

Brazil has enormous biodiversity and an extensive system of conservation units, protected spaces that are part of the Brazilian territory and that are managed to conserve its ecological, historical, geological, and cultural heritage [32]. Since Brazil committed to international programmes such as the Convention on Biological Diversity (CBD) and national targets aimed at conservation, the country’s PAs system has rapidly expanded. Although the system of PAs in Brazil was not directly created in response to the CBD, the Convention served as a backdrop for the establishment of the National System of PAs (SNUC). (This occurred because the initial versions of the law that instituted the SNUC

predated the CBD). During its passage through the National Congress, which took approximately 12 years, some of the guidelines from the Convention were incorporated into the text of this legal framework, thereby making it the primary instrument focused on biodiversity conservation in Brazil [33]. The consolidation of the various norms regarding Conservation Units in Brazil was not straightforward due to frequent disagreements between conservationist and preservationist perceptions of these areas [34]. However, after a long process of discussions among technicians, researchers, and public bodies, the National System of Conservation Units (Sistema Nacional de Unidades de Conservação, commonly referred to by its acronym SNUC) was unified under a single law (see [3]).

The SNUC recognizes 12 categories of conservation units, separated according to their management objectives and types of use. These categories fall into two major groups: strict protection and sustainable use PAs [3]. Most, but not all, of Brazil's PAs are documented in the National Registry of Protected Areas (Cadastro Nacional de Unidades de Conservação - CNUC). The CNUC currently includes 2,659 PAs, including marine and private PAs, managed at federal, state, and municipal levels. These areas cover about 18.80% of the continental area and 26.48% of the marine area of Brazil [35]. Between 2003 and 2009, Brazil was solely responsible for 74% of the global increase in PA coverage (km²), mainly due to several large PAs created in Amazonia [36]. Despite Brazil's leading role in global conservation and the immense success of its PA programme, Brazilian politicians and decision-makers seem to be increasingly viewing PAs as opportunity costs that limit economic development, leaving many PAs vulnerable to downgrading, downsizing or degazettement (PADDD) [37]. In this context, demonstrating public support for PAs and revealing their true value to society is an essential step to ensure their long-term sustainability [9].

4.3.2 Data collection

Digital data for conservation culturomics analysis can be collected from different sources (e.g., texts, videos, images, songs) [38]. Based on the framework suggested by [17], we analysed the content of, engagement with and author characteristics of publicly available Twitter posts about Brazilian PAs in Portuguese language. The data mining

techniques involved data collection, cleaning, processing and analysis (see Fig 1). In this study, we utilised the Twitter v2 API to collect all the data. Twitter's Academic access to its v2 API has provided us with the opportunity to gather up to 10 million tweets per month, which is a significant increase of 20 times compared to what was previously possible with the standard v1.1 API [39]. Moreover, this access allows us to retrieve older conversation histories, making Twitter a rich and accessible source of data for textual content analysis. As such, it becomes a valuable tool for gaining insights and a better understanding of public discussions and perceptions about Brazilian PAs. The code for data mining was developed using the Python language program v.3.9 (<http://www.python.org>) and was based on the Full-Archive-Search API node example from Twitter's official repository.

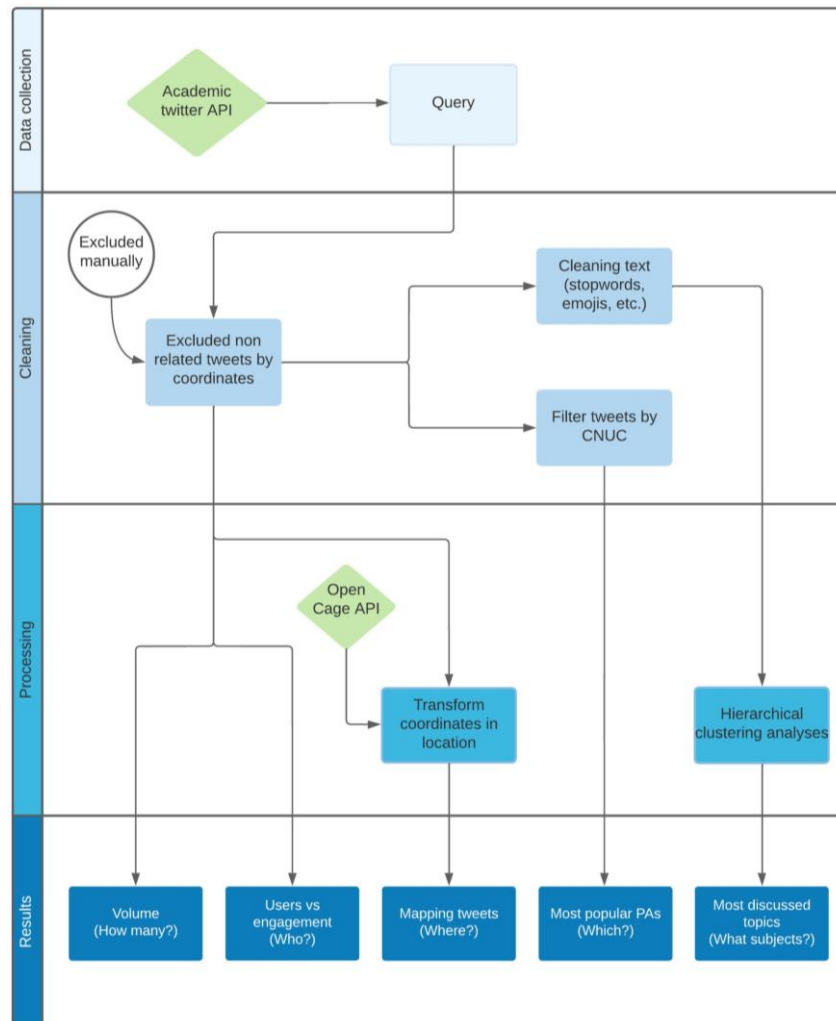


Fig 1. Methodological flowchart. Methodological flowchart from data collection to results. The flowchart shows all the steps used during the research: data collection, cleaning, processing, and analysis.

The second step was to define which content we would like to collect from Twitter. A query of 18 keywords was defined in the Portuguese language to extract the tweets (not the retweets). : 1- Parque nacional (National park); 2- Parque Estadual (State park); 3- Parque natural municipal (Natural municipal park); 4- Parque municipal (Municipal park); 5- Estação Ecológica (Ecological station); 6- Reserva Biológica (Biological reserve); 7- Monumento Natural (Natural monument); 8- Refúgio da Vida Silvestre (Wildlife refuge); 9- Reserva Extrativista (Extractivism reserve); 10- Área de proteção ambiental (Environmental protected area); 11- Floresta nacional (National forest); 12- Floresta estadual (State forest); 13- Floresta municipal (Municipal forest); 14- Reserva de desenvolvimento sustentável (Sustainable development reserve); 15- Área de interesse relevante (Area of relevant interest); 16- Reserva particular do patrimônio natural (Private natural reserve); 17- Unidade de conservação (Conservation unit); 18- Área protegida (Protected area).

All 12 categories of Brazilian PAs were included in the query. PAs with management levels in their name were also added to the query, for example, “national park”, “state park”, and “municipal park”. Furthermore, to collect the tweets that do not mention the names of PAs in their textual content, the query also included the two keywords: conservation unit and protected area (unidade de conservação e área protegida). We also considered adding to the set of terms used in the query the acronyms referring to each PA category, for example APA for Área de Proteção Ambiental (Environmental Protection Area). However, a preliminary analysis of data collected using the acronyms revealed these mostly relate to other topics such as slang, celebrity names and other words that did not result in tweets discussing PAs. Based on this, even though it resulted in reducing our sample size, we decided not to include the acronyms in the query to reduce bias and noise.

The collection of the textual content of retweets was also avoided to prevent possible biases in the results of the multiple counting, and in the salience of topics, since they generally represent repetitive copies of original tweets. Instead of collecting the textual content of retweets, we collected quantitative data from the original comments that contained the number of times the message was retweeted, received likes or comments.

In addition, the use of keywords in our query made it possible to identify and collect all messages, including retweets that had comments related to the original tweet, that mentioned any Brazilian PAs. The tweets were retrieved from the API between April 15th and April 30th, 2021. The sampling period encompassed tweets posted from January 1st, 2011, to December 31st, 2020. In total, before the data cleaning and filtering process, 421,254 tweets were collected.

The following information was collected from each tweet: (i) author (name, username, and whether it is verified); (ii) date of publication; (iii) geographical data (latitude, longitude, city, country); (iv) publication data (number of likes, retweets, replies and whether it is a reply to another user); and (v) the text of the tweet. The tweets were downloaded in JSON format, by year, with a maximum of 500 tweets per page (as per the limit set by the Twitter API). After converting the JSON pages to a single CSV file, the data cleaning process was performed. Initially, a filter was applied to the geographical metadata provided by Twitter (country and country_code columns) to identify tweets from countries other than Brazil. Subsequently, a manual verification of each foreign tweet was conducted to eliminate those that did not correspond to Brazilian PAs. Our final validated list contained a total of 402,508 tweets about Brazilian PAs (see all IDs collected from Twitter in <https://github.com/jagra26/Brazilian-PAs-on-twitter>). Tweets usually contain URLs, emojis, and emoticons, so the dataset needs to be cleaned before analysis. Text cleaning was performed using the R language program (R Core Team, 2017), with the tm package [40]. Specifically, we used the function 'tm_map' to convert all text to lower case and remove any hashtags, URLs, symbols (like □, ■, ◆), numbers, and Portuguese stop words present in the text. In terms of data collection, it's worth noting that the availability of the APIs for academic usage has been volatile recently, as the platform is now referred to as X and has undergone a policy change in data access which has restricted access.

4.3.3 Data analysis

We aimed to understand the public interest in Brazilian PAs based on the number of tweets and public engagement with those posts. First, we compared the volume of tweets about Brazilian PAs on a temporal scale of 2011 to 2020 with the number of users

who post about PAs. To do this, we counted the number of tweets, number of unique users, and the total number of 'likes' and retweets by year to create a line graph. To explore the relationship between the volume of tweets and engagement, we summarised the number of tweets and the average number of 'likes' and retweets per post at the user level. We calculated the bootstrapped mean engagement over time using 1000 samples and a confidence interval of 95% with function 'smean.cl.boot' from R package Hmisc [41], and generated a line plot depicting the disparity between user types (verified and not verified). We also bootstrapped the Spearman's correlation between number of tweets and mean engagement per tweet using 1000 samples and 95% confidence intervals with function 'spearmanRho' from R package rcompanion [42].

We also mapped the number of Tweets about Brazilian PAs based on location information provided by Twitter's users. To identify the geolocation of tweets related to Brazilian PAs, we employed a keyword-based (see query used in the data collection section) sampling strategy that increased the likelihood of capturing tweets about the targeted PAs originating from various locations [25]. The metadata of the tweets provided geographic data such as coordinates, countries, and cities. However, the consistency of this information was limited, with many tweets containing only a single geographic coordinate (e.g., the absence of country-level data despite the availability of coordinates). To overcome this limitation, we adopted a similar approach to [43], where we utilised the OpenCage Geocoding API search engine (<https://opencagedata.com/>) for reverse geocoding. This process allowed us to obtain comprehensive state, and country information, based on the coordinate metadata provided by Twitter. Ultimately, our geospatial dataset consisted of 62,924 tweets (15.63%) with location data (geographic coordinates and states and countries data). Using the folium library [44] in Python version 3.9 (<http://www.python.org/>), we created a choropleth map where the colour intensity of each territory corresponded to the number of tweets posted.

It was necessary to implement a filtering process to assess which Brazilian PAs had the most posts, as the majority of tweets did not explicitly mention the full names of the PAs in their text content. To accomplish this, we filtered our dataset using the proper names of the PAs from the National Register of Protected Areas (MMA/CNUC 2021), without considering the specific category of the PA in the search, using the VLOOKUP

function through Microsoft Excel software (Microsoft 365 2020): =VLOOKUP (search_value, SEARCH (search_text, no_text, [start_num]), search_text). Exceptions to the filtering process using the full name with the category type were made for those areas that had the same name with different categories, such as "Tamoios Ecological Station" and "Tamoios Environmental Protection Area." The list of names used for filtering can be accessed on the repository (<https://github.com/jagra26/Brazilian-PAs-on-twitter>). The final dataset used in this analysis has 189,294 tweets containing PAs in CNUC. Using the new dataset, we generated a lollipop chart using summarised data on the number of posts and average engagement per PA to compare the differences between the public interest of each area.

Finally, to group tweets based on similarity in content and to identify the main topics Twitter users discuss concerning Brazilian PAs, we applied the agglomerative clustering technique, known as AGNES (Agglomerative Nesting). The clustering method used to calculate the distances was the "complete" method, based on the word composition of posts in the dataset [45]. The complete method uses the "largest dissimilarity between a point in the first cluster and a point in the second cluster (furthest neighbour method)" [46]. According to the study conducted by [47], in order to ensure that the final clusters became neither excessively broad (involving few clusters covering PAs that share the most common words) nor excessively specific (resulting in numerous clusters addressing PAs that share less frequent words), the authors carried out several simulations with a range of 2 to 10 clusters (k).

During these simulations, the most frequent words in each cluster were examined in each scenario in order to identify the most prominent characteristics of each cluster in relation to the Brazilian PAs. The cluster analysis produced a dendrogram in which the distances between the branches reflect the similarity between each tweet. This procedure made it possible to define the theme of each cluster (5) based on the similarity of the words grouped in the dendrogram. With exception of the choropleth map, all analyses were performed in the R programming language. The following packages were used in the analysis: dplyr package was used for the data manipulation [48]; to create the graphics we used the packages ggplot2 [49] and plotly [50]; and, finally, for the hierarchical clustering analysis, we used the packages cluster [45] and factoextra [51]. All R and

Python scripts used for data collection and analysis are available on our repository page (<https://github.com/jagra26/Brazilian-PAs-on-twitter>).

4.4. Results

4.4.1 Volume of tweets about Brazilian PAs

We mined a total of 402,508 valid tweets posts about Brazilian PAs between 2011 and 2020, which were those that successfully passed through the data cleaning process. The number of posts and the number of users tweeting about the topic was relatively similar throughout the decade analysed, with only minor decreases in the number of posts after 2016 (Fig 2).

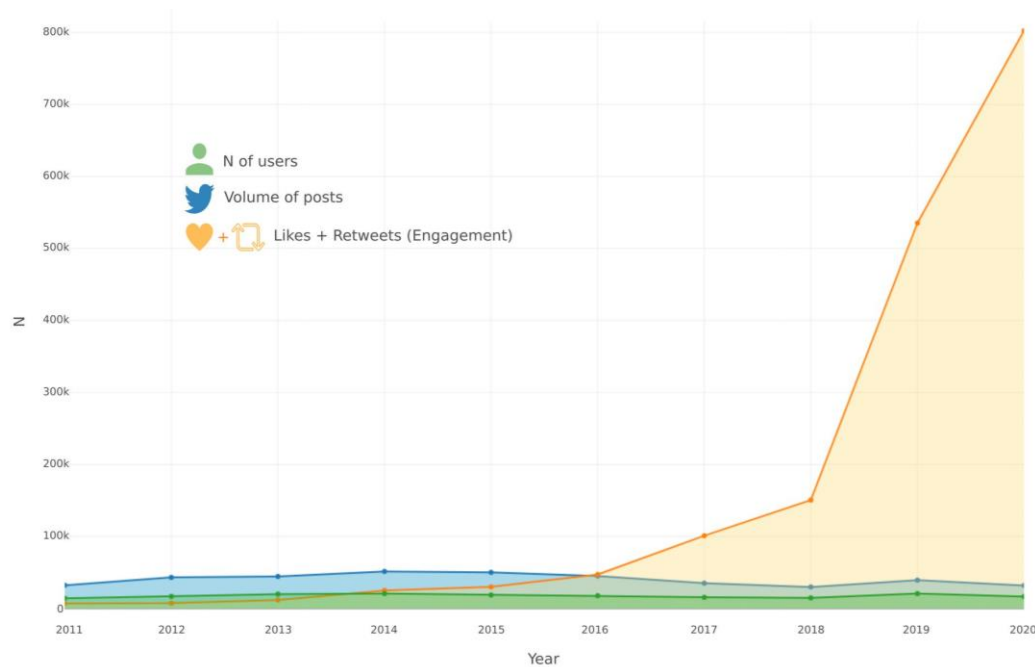


Fig 2. Volume of posts, users and engagement of Brazilian protected areas. Line graph representing the volume of users, tweets, and public engagement (likes + retweets) about Brazilian protected areas during the period from 01 January 2011 to 31 December 2020.

The average number of posts and users posting about PAs in Brazil was around 40,000 posts and around 17,000 users over the 10 years analysed. Metrics of engagement (likes + retweets) with posts about PAs started increasing in 2016, but grew

steeply after 2018. Estimates of the mean engagement per tweet increased from 0.182 (bootstrapped 95% C.I.: 0.165-0.199) in 2011 to 15.3 (bootstrapped 95% C.I.: 11.9-19.6) (Fig 3).

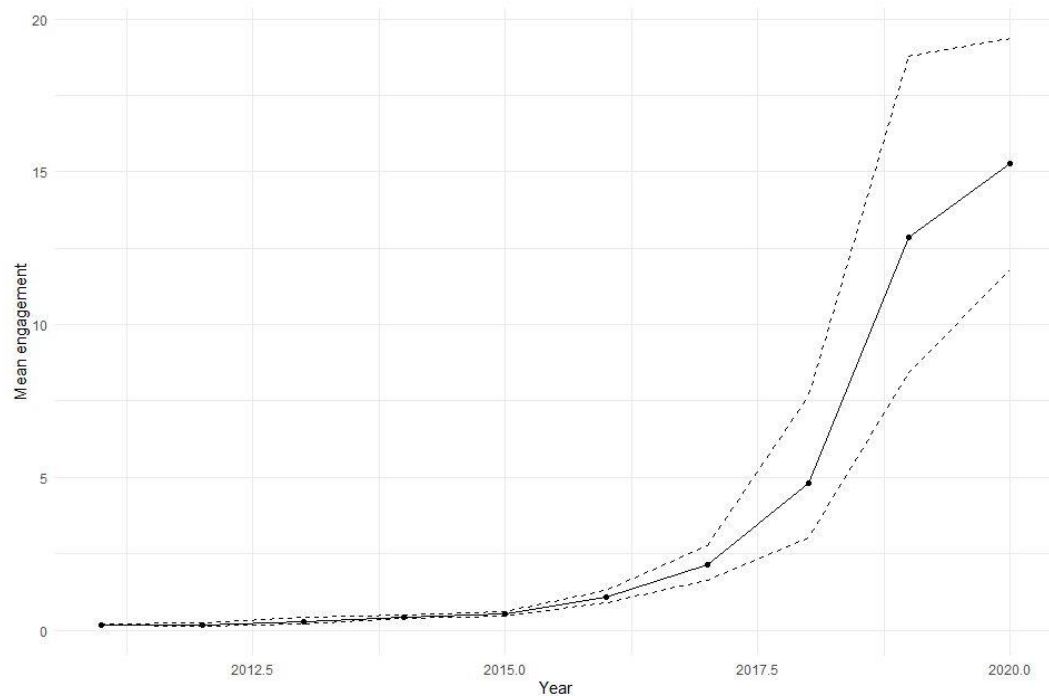


Fig 3. Bootstrapped estimates of mean post engagement per year. Figure shows the results of a Bootstrap analysis representing the mean engagement (likes and retweets) with posts about Brazilian protected areas published between 2011 and 2020.

Compared to 2018, the number of people actively posting about PAs in Brazil in 2019 increased by 40%, posts about what is happening in these areas increased by 31%, and public participation increased by 255% (Fig 4).

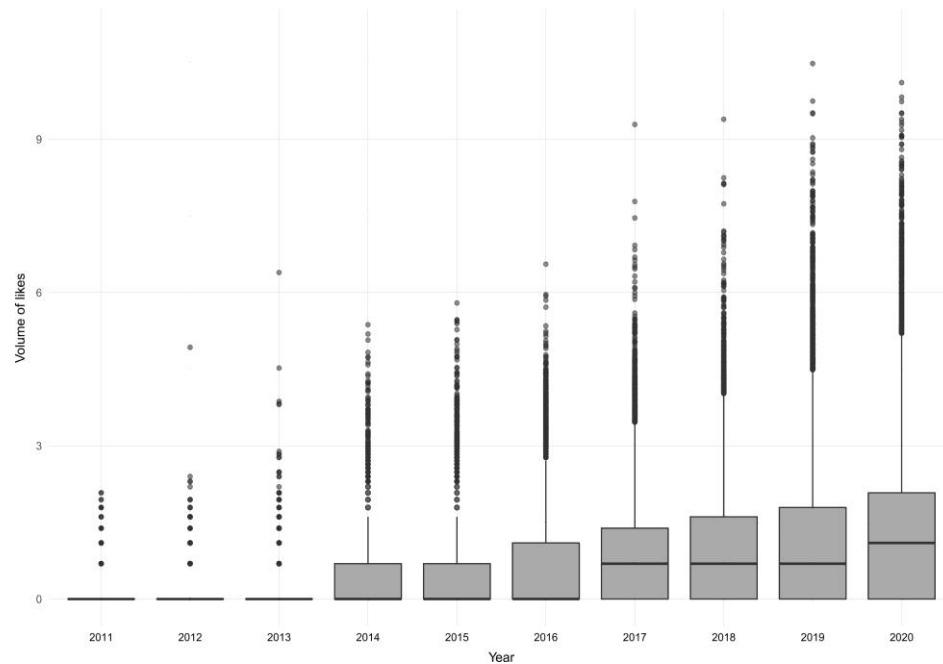


Fig 4. Public interest related to the Brazilian protected areas over the years. Boxplot represents the distribution of the volume of likes for Brazilian protected areas during the period from 01 January 2011 to 31 December 2020. The lower and upper box limits correspond to the first and third quartiles (the 25th and 75th percentiles); a black bar inside the box indicates the median. The volume of tweets was log-transformed (natural logarithm).

4.4.2 Users characteristics

We identified 130,742 Twitter users who have posted content about Brazilian PAs. Some of these Twitter users have many followers, including news media, politicians, celebrities, travel agencies, and a range of international people and organisations such as WWF and UNESCO. These are frequently classified by Twitter as ‘users of public interest’ and, in most cases, received a verification seal. We explored the relationship between user type (verified or not) and the number of tweets about Brazilian PAs and the public interest in those posts (Fig 5). In general, users who post more about Brazilian PAs receive more engagement (likes + retweets). Indeed, there was a small but positive correlation (Spearman’s $\rho = 0.107$; bootstrapped 95% C.I.: 0.102-0.112) between the number of tweets per user and mean engagement per tweet. However, it is worth noting that when we compared engagement between verified and not verified users, our

bootstrap analysis showed that the estimates of the average engagement per tweet of an not verified user increased from 0.354 (bootstrapped 95% CI: 0. 305-0.409) in 2011 to 22.3 (bootstrapped 95% CI: 12.9-34.5) in 2020, while the average engagement per tweet of verified users increased from 7.00 (bootstrapped 95% CI: 4.48-10.1) in 2011 to 651.8 (bootstrapped 95% CI: 388.6-1003.9) (Fig. 5). For example, of the 10 users who have posts with an average engagement of more than 5,000 likes and retweets, seven are verified profiles and have a considerable number of followers (>15,000 followers). Interestingly, the most of these users, despite their large following, posted only one or two tweets related to Brazilian PAs during the analysed ten-year period.

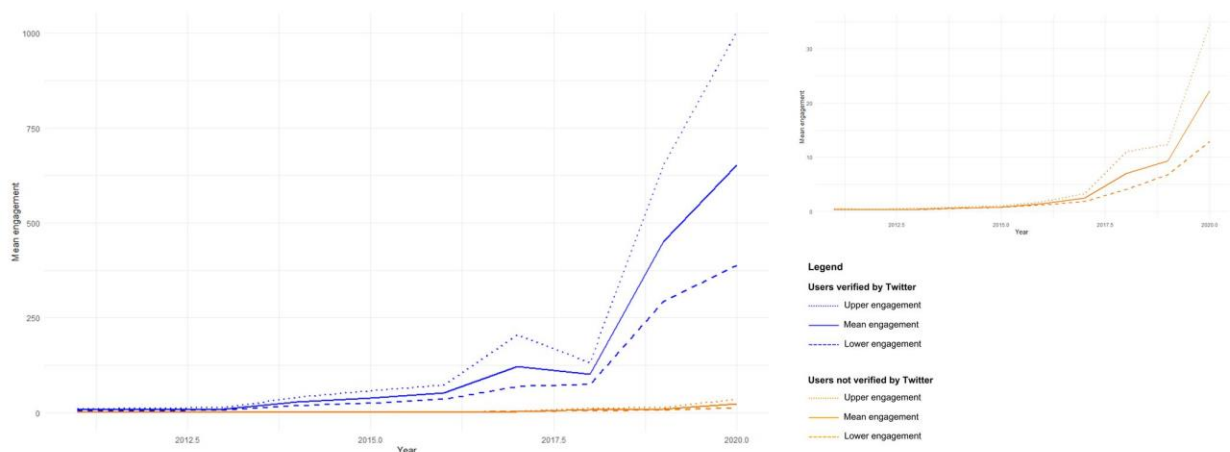


Fig 5. Bootstrapped estimates of mean post engagement per year and per user type. Figure shows the results of a Bootstrap analysis representing the mean engagement (likes and retweets) with posts about Brazilian protected areas, per type of user (verified and not verified) published between 2011 and 2020.

4.4.3 Geographic focus of tweets

The large majority of tweets (provided by the Twitter coordinates metadata) about Brazilian PAs originate in Brazil, as expected, but we also found mentions from countries bordering Brazil, Costa Rica, Portugal, and the United States (Fig 6). This result is similar to the spatial distribution found by lusophone and associated countries (<https://www.cplp.org/>). In Brazil, the state of Rio de Janeiro received the highest number of georeferenced tweets about PAs (12,585), followed by Minas Gerais (11,339), São

Paulo (8,837), and Paraná (5,434). Nevertheless, when we examine which PAs were most georeferenced by Twitter users, we can see that out of the top 100 geo-referenced areas, 64% are urban parks. The most geotagged National Park was Iguaçu National Park (Paraná state) with 2,747 tweets, followed by Tijuca National Park (Rio de Janeiro state) with 1,300 tweets and Serra dos Órgãos National Park (Rio de Janeiro state) with 752 tweets. The most geotagged urban park was Américo Renné Giannetti Municipal Park (Minas Gerais state) with 2,615 tweets, followed by Ponte dos Bilhares municipal Park (Manaus state) with 1,101 tweets and Flamboyant Municipal Park (Goiás state) with 1,033 tweets.

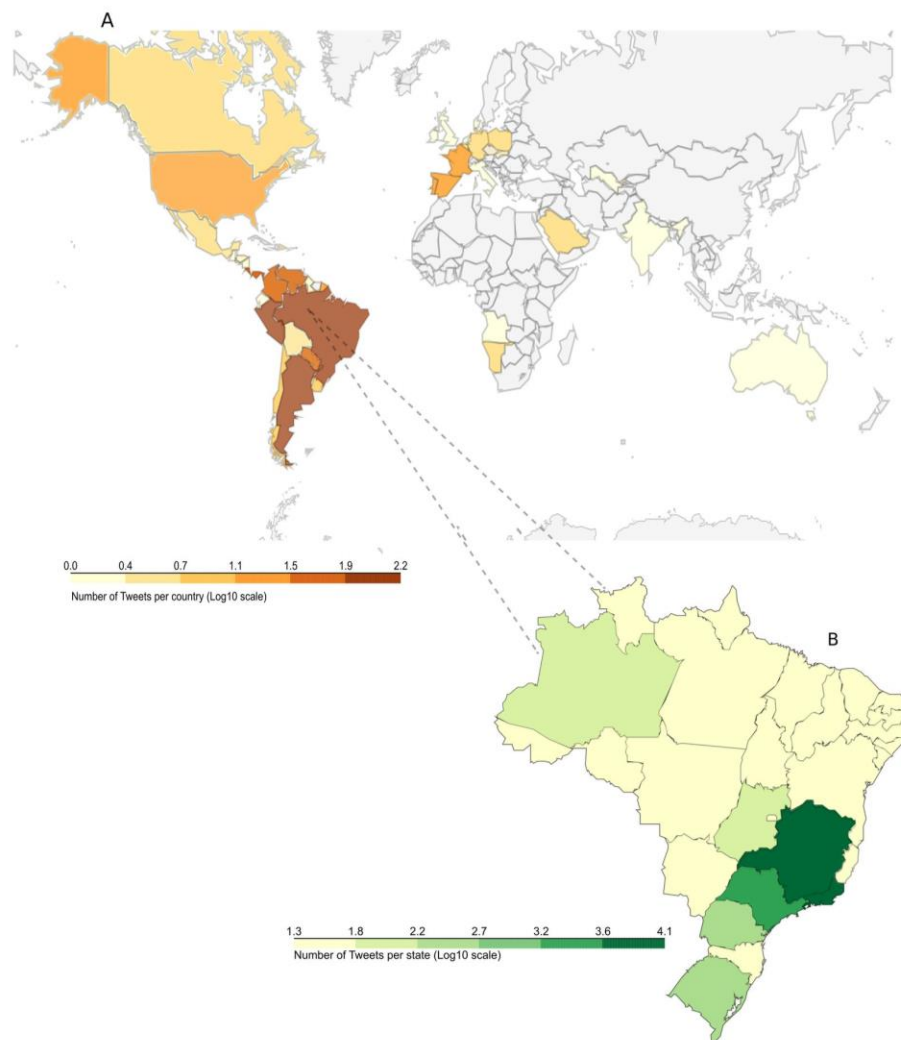
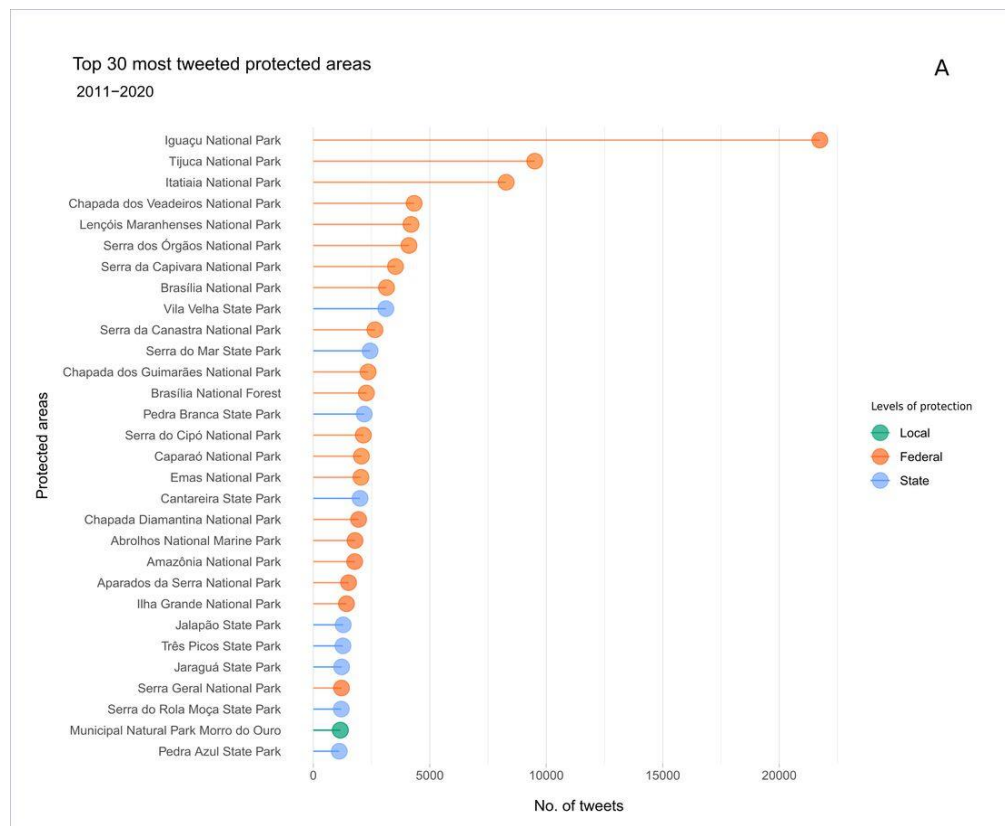


Fig 6. Geographical distribution of the tweets posted about Brazilian protected areas. Choropleth map representing the number of tweets posted about Brazilian protected areas in

Brazil and worldwide from 2011-2020. Locations are extracted from geotagged tweets and Twitter user profiles. The map was cropped above Alaska to enhance the visibility of the other countries.

4.4.4 Which protected areas generate most public interest and engagement?

Based on the analysis of tweet content, national parks were the most posted category of PAs on Twitter, followed by state and municipal parks (Fig 7A). The iconic Iguaçu National Park received over 21,700 tweets, and Tijuca National Park received 9,507 tweets. Tourism emerged as the primary topic tweeted about among the top 30 most-tweeted parks; seven out of the ten most visited national parks in Brazil [52] were among the top 12 most-tweeted about parks. We hypothesized that PAs that generated high public interest, with numerous posts about them, also garnered high engagement (likes + retweets). When evaluating the average engagement associated with PAs, besides national parks, other PA categories surfaced, including environmental protection areas, national forests, and biological reserves (Fig 7B).



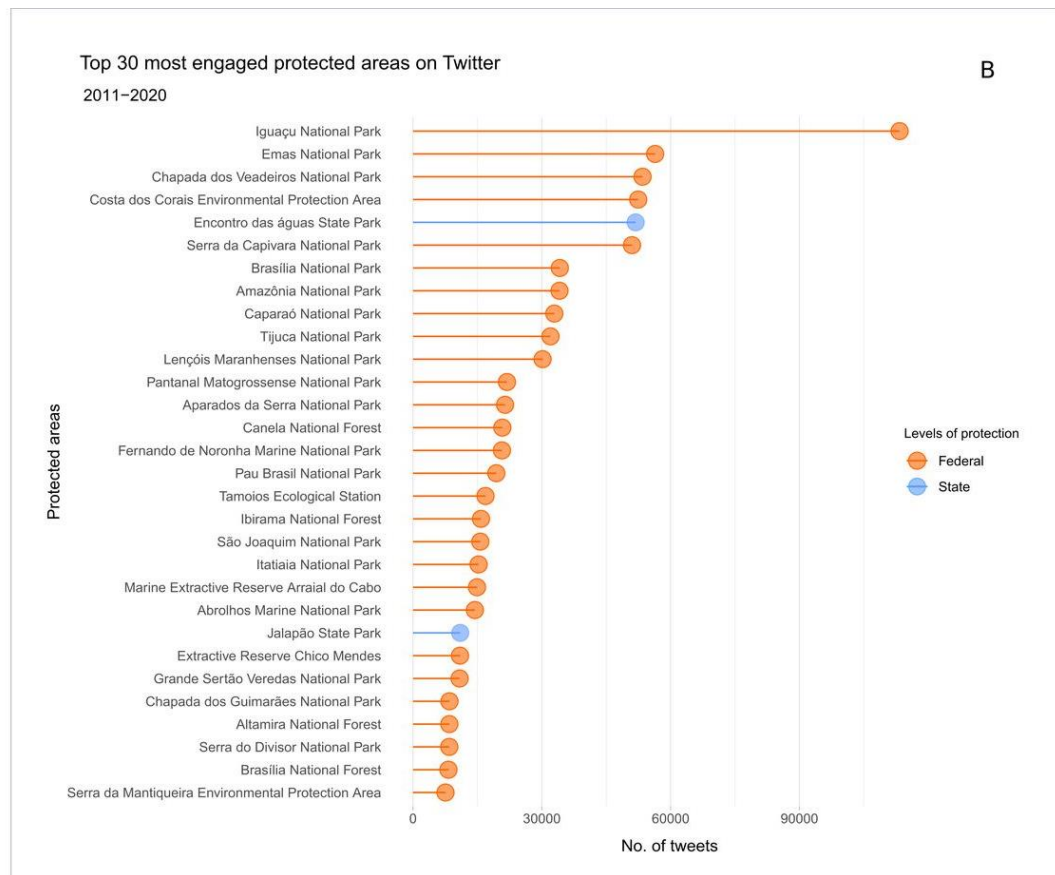


Fig 7. Brazilian protected areas that have generated the most public interest and engagement. (A) Lollipop plot illustrating the volume of tweets about Brazilian protected areas during the period from 01 January 2011 to 31 December 2020. (B) Lollipop plot illustrating the average of engagement in the tweets related to Brazilian protected areas. The top 30 tweets were illustrated in ascending order and grouped by level of protection: federal, state, and local.

In addition to Iguaçu National Park, our results indicate that the most tweeted PAs did not achieve the highest levels of engagement (see comparison in Figs 7A and 7B). Instead, tweets about conflicts, rather than tourism activities, generated the highest engagement, and these tweets were more frequently associated with other PA designations. Upon analysing the five most engaged PAs on Twitter, the tweets with the highest engagement included discussions about fires, jaguar deaths, and administrative abuses, among the most frequent topics.

4.4.5 Content analysis of tweets

A broader look at the topics discussed in the full dataset revealed somewhat similar patterns. Our cluster analysis of the most commonly featured keywords in all tweets identified five thematic categories. Based on the most frequent words in each cluster represented in the dendrogram (Fig 8), we characterised the main topics of discussion as follows:

- (i) Fires:** this cluster included words associated with fires in PAs;
- (ii) Management of protected areas:** themes related to environmental crimes; PADD (downgrading, downsizing, and degazettement) events, mainly focused on the size of PAs, and educational campaigns;
- (iii) Nature protection:** with regard to the creation of new PAs;
- (iv) Nature appreciation:** this cluster included words associated with sharing on social media and natural monuments;
- (v) Visitation:** where the names of national parks, and urban parks were mentioned the most.

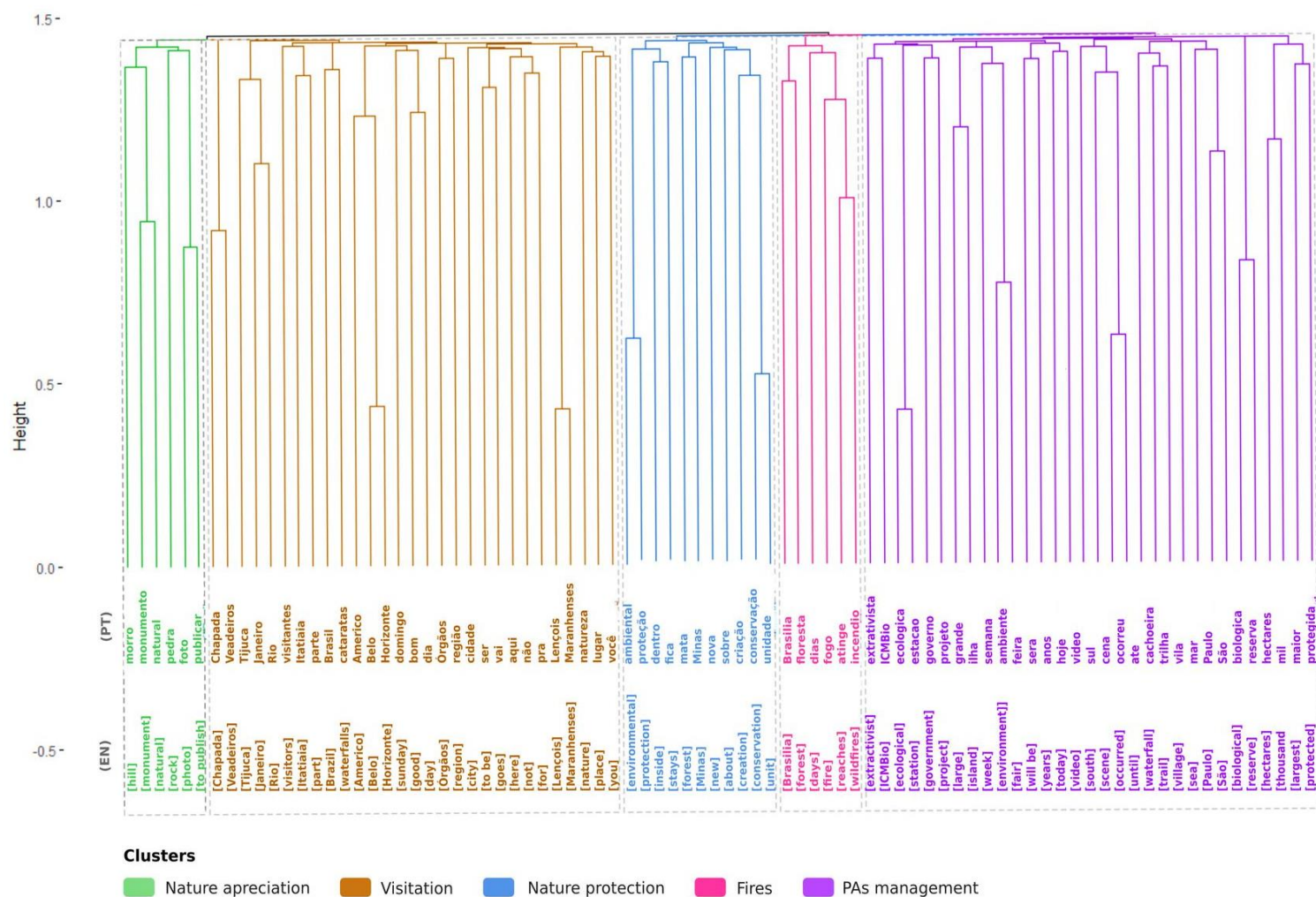


Fig 8. Most published content related to Brazilian protected areas. The contents were published in the period 01 January 2011 to 31 December 2020. Each branch in the dendrogram represents a different cluster, grouped according to the similarity of words and themes.

4.5 Discussion

We assessed Twitter users' interest in Brazilian PAs over a ten-year time scale (2011-2020) using metrics of content and engagement, including likes and retweets in the Portuguese language. Our results suggest that social media content related to PAs remained broadly stable throughout the studied period, though engagement with this content was low until 2016, at which time engagement began to increase. Engagement grew particularly steeply after 2018, following a remarkable increase in likes and retweets related to Brazilian PAs, even though the number of users posting on this theme remained relatively stable over time. There are at least two factors that could be driving this pattern: First, in 2016 Twitter introduced an algorithm to customise the content that users can access on their profile's timeline [53]. Based on accounts users have chosen to follow and posts interacted with, the algorithm presents users with a series of recommendations. This new personalization algorithm amplifies certain content while reducing the visibility of other posts [54]. In this way, people sympathetic to environmental themes will be more likely to be presented with content related to nature (an “echo chamber” effect), such as PAs, and may more readily engage with posts (tweets and retweets) that in the past may not have been easily encountered. In addition, users who carried the verification seal on Twitter, identifying them as figures of public interest, may have experienced a significant increase in visibility and influence within the platform after 2016 (Fig 5). This change can be attributed to the way Twitter's algorithm has optimised the dissemination of tweets from these verified users, especially by making their tweets become viral.

A second reason for the observed engagement trend may be related to the change in Brazilian federal government that took place at the end of the same year, and the increase in public discourse about the environment and environmental policy that started during the political campaign that preceded the election and continued after it took place. In 2019, the recently elected Bolsonaro government made a number of highly controversial decisions to backtrack on environmental policies [55]. Specifically, the incoming government made pledges to halt the expansion of the PA system and to make environmental licensing more flexible [56] leading to a wide scale mobilisation of the environmental movement in Brazil. The resulting conflicts of perspectives and attitudes

generated considerable discussion on Twitter and beyond. This, combined with Twitter's new algorithm, may have prompted concerned citizens to express their opinions (agreeing or disagreeing with government policy) by engaging with content related to what was happening to Brazilian PAs.

Notwithstanding the highly polarised debate around Brazilian PAs and improvements in personalization of Twitter feeds, our results indicate that the volume of posts made by a single user was positively but weakly correlated with levels of engagement. Also, posts by environmental NGOs, official agencies and celebrities had a much greater level of engagement, as has been found in other studies [25]. As illustrated in Figure 5, the growth trajectory of engagement regarding PAs after 2016 is similar between the types of users (verified and unverified). Nevertheless, it is verified users who play a key role in feeding interest and engagement in this subject, due to the substantial number of followers, running into thousands or even millions, who often express appreciation through likes, comments and retweets - a result which, perhaps unsurprisingly, highlights the ability of celebrities to captivate the Twittersphere's attention. Celebrities have the ability to attract people's attention [57] and have a long and complex history of environmental engagement and activism (reviewed in [58]). There were several artists and politicians who only contributed with one or two posts about Brazilian PAs in our database, but due to their enormous number of followers, generated very high levels of public engagement. This further confirms the importance and power of celebrities in raising awareness of environmental issues, promoting science and environmental conservation [57,58]. More generally, recent research has highlighted the existence of different personas engaged in environmental discourses on social media [59], and our results suggest a similar result for PA discussions. Exploring in greater detail which personas drive and shape PA discussions on social media can help better understand public discussions around this topic.

We found a relatively low number of georeferenced posts, possibly due to the decision to collect only Portuguese language tweets. In Brazil, the most populated states generated the most tweets. However, Iguaçu National Park, despite not being located in a highly populated state, had the highest volume for an individual PA. This supports the findings of a recent study of internet salience that found that Iguaçu National Park was the

most mentioned Brazilian PA on the national and global internet [26]. This may be due to its beauty, iconic status and large annual number of visitors [51]. Iguaçu is the second most visited PA in the country, with excellent visitation structure and is located on the border between Brazil and Argentina - the second ranked country with the most geo-referenced tweets about Brazilian PAs.

Another surprising finding was the high number of urban parks georeferenced in the tweets. Urban parks are open green areas that can perform ecological, landscape and recreational functions [60]. Significantly, they are not classified as PAs according to Brazilian environmental policy. This observation highlights the importance of more effective communication regarding PAs and the need for awareness about their functions and values. Despite not being classified by the Brazilian environmental policy as PAs [3], such parks can provide important physical, social and health benefits for urban residents [61]. They are often the gateway for people's first contact with nature, and provide opportunities to escape from the stressful pace of the city [62]. Being located in urban areas, these parks are more easily accessed when compared to other PAs and tend to have better mobile phone signal coverage. This potentially allows more real-time interaction with social media, contributing to the high volume of georeferenced postings. As an example, the Américo Renné Giannetti Municipal Park in the Minas Gerais state had the second highest volume of georeferenced posts in the entire database.

The use of social media data has recognized potential for improving knowledge and monitoring tourism interest in PAs [12,13,63]. Our study suggests that the most officially visited parks were also the most popular subjects for Twitter posts. Such a result matches with previous research that used different research platforms such as Google Trends [26], OpenStreetMap (OSM), and Wikipedia [64,65] and social media platforms such as Instagram, Flickr, and Twitter [66,67]. However, two highly-visited national parks drew attention for not being among the 30 most posted-about on Twitter: Jericoacoara National Park and the Fernando de Noronha Marine National Park. Similar discrepancies between visitation and tweets were observed in research on Nepalese Parks, due to greater local visitation and Twitter use restrictions [25]; factors that are less likely to account for the discrepancies observed in the current study. These are more likely caused by: (i) intrinsic biases in data collection. A limitation of nearly all textual content studies on

the internet is the problem of synonyms [68] - alternative names or spellings for the represented entity. In the current study we used search strings that contained the full names of PAs, and may therefore have missed many tweets that used colloquial names. In the cases of Jericoacoara National Park, many Brazilians refer to the Park as "Jeri" and Fernando de Noronha Marine National Park as "Noronha"; (ii) longer names tend to be less popular on social media platforms [69] and are more likely to contain spelling errors. Abbreviations and misspellings can restrict the search for PAs, however, currently, this is the only viable and standardised way to attach a name to the PA designation at scale [26] and; (iii) Both Jericoacoara and Fernando de Noronha National Parks may have communication gaps regarding their identity as National Parks. The two parks are located in places (municipality and island, respectively) that have the same names as the Park (Jijoca de Jericoacoara municipality and Fernando de Noronha Island). This could confuse visitors or simply lead them to communicate about the place more broadly rather than the park specifically which, again, could lead to underrepresentation in our database due to our chosen search method.

The Park category was the most tweeted category, especially National Parks. The National Park has the primary objective of preserving natural ecosystems of great ecological significance and scenic beauty, enabling scientific research and the development of environmental education and interpretation activities, nature-based recreation, and ecotourism [3]. It is almost certainly that this high public interest, reflected in the substantial number of Twitter posts, is attributed to the fact that national parks incorporate recreation as a significant component of their objectives [70]. Recreation activities are strongly linked to public interest and support for PAs. Moreover, National Parks are the oldest and most visited PA category [71], and many were protected to preserve the scenic beauty of their unique landscapes along with other relevant biophysical assets [72]. In addition, in Brazil, National Parks have larger use concessions and receive more tourists, raising more resources to invest in the media. However, we noticed that a large part of our dataset contained municipal parks that are not currently present in the CNUC database. The most general explanation for this, according to government environmental analysts, is that the registration of municipal units is at the discretion of municipal environmental agencies and is not mandatory. However,

registration in the CNUC is used to verify the criteria for the allocation of funds from federal public policies, such as those from federal environmental compensation. Such funds are destined exclusively for PAs recognized by CNUC as belonging to the SNUC. Our findings indicate that many parks currently unlisted by the CNUC are generating considerable value for society [73] and for biodiversity [74] and should therefore be a valid target for financial investment and improved environmental management [75].

A diversity of topics was found on tourist experiences and nature appreciation (Fig 8), corroborating the importance of PAs in bringing people closer to nature and the physical and psychological benefits they provide [11]. Unlike other content-based research on PAs, our findings did not identify topics related to iconic animals [12,25]. Although Brazil is renowned for its megadiverse ecosystems, offering the opportunity to witness wildlife in PAs such as west indian manatee (*Trichechus manatus*), river dolphins (*Inia geoffrensis*), jaguar (*Panthera onca*), and the largest terrestrial mammal in Brazil, the tapir (*Tapirus terrestris*) [76–78], it's important to note that many Brazilian National Parks were established with the primary goal of conserving scenic landscapes (as referenced in Article 11. [3]), rather than focusing specifically on iconic species. Furthermore, the type of social media and PA category may influence the results related to the topics that are chosen as symbols and the experiences that are chosen to be shared. Besides the topics related to tourism experiences, most of the topics that generate high levels of public interest were related to management actions and conflicts, such as fires, reduction in the size of PAs, and possible environmental crimes. This result corroborates our findings related to engagement with posts (Fig 7B). These results highlight the potential use of social media to monitor PAs in terms of cultural value generated (e.g., for tourists) or public discontent related to management decision-making [14,79]. In the context of the spread of misinformation on Twitter, it is important to recognise that false messages may have influenced user engagement in relation to PAs, despite our observation that the presence of bots did not have a significant impact in our study (only 2.61% of the tweets were duplicates and could potentially be attributed to bots). Previous research [80] indicates that bots play a role in the dissemination of true and false information, but it is human action that amplifies false news more quickly and extensively, as they tend to arouse a greater sense of novelty and excitement in people.

It is well known that the dissemination of incorrect information can result in the delegitimisation of science, the minimisation of real threats to conservation and the generation of polarisation about the importance of protected areas. In this sense, there is a clear need for additional research that focuses on textual and behavioural content analysis in order to understand whether the issues that lead people to engage with environmental issues are false or true. These studies can shed light on the dynamics behind the spread of misinformation, as well as on effective communication strategies to combat its detrimental effects on discussions about protected areas and other environmental issues. Better understanding of human-PA interactions can also be used to inform strategies to attract societal support for conservation and PAs, to improve conservation planning and management, and to tailor communication strategies [13,15,37].

4.6 Conclusions

Overall, our study adds to the rapidly growing literature on the use of culturomic metrics for monitoring human-nature interactions at large spatio-temporal scales [17]. Nevertheless, Twitter data has a few limitations that limit its suitability as a stand-alone monitoring tool. First, Twitter corpora contain high levels of slang, colloquialisms, acronyms and emoticons that can be challenging for data collection and analysis. Second, social networks are not necessarily representative of PA users or of wider society [25]. Finally, not all users enable the geolocation on their devices, limiting this type of data [42]. Thus, all social media data requires extensive cleaning and critical analysis [15]. Despite such limitations, the enormous volumes of data generated by social media platforms such as Twitter mean that their analysis can certainly generate insights into the perceptions of non-social media users [18] and, critically, allows the evaluation of macroscopic patterns of human-nature interactions [81]. Acknowledging both the limitations and potentials inherent in utilising social media as a means of investigation, we propose the integration of various online data sources, such as Wikipedia, Instagram, Facebook, and offline methodologies like questionnaires. This comprehensive approach aims to reinforce the discourse surrounding the attained outcomes and enhance the depth of analysis. Accepting the intrinsic limitations of our data, our analysis of 10 years of Twitter discourse

about Brazilian PAs indicates several areas of policy that could be improved. For example, it is clear that official communication about PAs could be improved, from providing and disseminating basic information about some parks (Fig 5), to the way government agencies responsible for managing PAs communicate their decisions and campaigns (Fig 7). The lack of public engagement with conservation is often attributable to ineffective communication from scientists and decision-makers [21]. In this sense, government agencies and environmental NGOs could use data from social media to create better and more powerful awareness campaigns, potentially in partnership with celebrities and other social media ‘influencers’ (Fig 5). Furthermore, understanding online perceptions and public interest can lead to identifying topics related to conservation actions (Fig 8), where managers can pinpoint gaps and enhance strategies to maximise positive outcomes and minimise negative impacts. In a broader context, our study confirms that the use of content and engagement metrics on social media, combined with other monitoring tools such as environmental data on water quality, air, biodiversity, along with information from field monitoring and traditional media, has significant potential to enable early warnings to identify emerging conflicts in PAs (Fig. 7B) and to identify public interest in these areas (Fig. 7A).

4.7. Acknowledgments

We are grateful to all the LACOS21 colleagues, who contributed with insights to this work. We also appreciate the comments made by the editor, Joseph Millard and one anonymous reviewer that greatly contributed to improve the original manuscript.

REFERENCES

1. Watson JEM, Dudley N, Segan DB, Hockings M. The performance and potential of protected areas. *Nature*. 2014;515: 67–73. doi:10.1038/nature13947
2. Maretti C, Catapan M, Abreu MJ, Oliveira J. Áreas protegidas: definições, tipos e conjuntos – reflexões conceituais e diretrizes para a gestão. 2012.
3. BRASIL. 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. 2000.

4. BRASIL. Decreto nº. 5758, de 13 de abril de 2006. 2006.
5. Engen S, Hausner VH, Gurney GG, Broderstad EG, Keller R, Lundberg AK, et al. Blue justice: A survey for eliciting perceptions of environmental justice among coastal planners' and small-scale fishers in Northern-Norway. *PLoS One*. 2021;16: 1–20. doi:10.1371/journal.pone.0251467
6. Gerhardinger LC, Godoy EAS, Jones PJS, Sales G, Ferreira BP. Marine protected dramas: The flaws of the Brazilian national system of marine protected areas. *Environ Manage*. 2011;47: 630–643. doi:10.1007/s00267-010-9554-7
7. Bragagnolo C, Gamarra NC, Malhado ACM, Ladle RJ. Proposta Metodológica para Padronização dos Estudos de Atitudes em Comunidades Adjacentes às Unidades de Conservação de Proteção Integral no Brasil. *Biodiversidade Brasileira*. 2016;6(1): 190–208.
8. Silva JMC da, Dias TCA de C, Cunha AC da, Cunha HFA. Funding deficits of protected areas in Brazil. *Land use policy*. 2021;100. doi:10.1016/j.landusepol.2020.104926
9. Bernard E, Penna LAO, Araújo E. Downgrading, downsizing, degazettement, and reclassification of protected areas in Brazil. *Conservation Biology*. 2014;28: 939–950. doi:10.1111/cobi.12298
10. Jepson PR, Caldecott B, Schmitt SF, Carvalho SHC, Correia RA, Gamarra N, et al. Protected area asset stewardship. *Biol Conserv*. 2017;212: 183–190. doi:10.1016/j.biocon.2017.03.032
11. De Haan FJ, Ferguson BC, Adamowicz RC, Johnstone P, Brown RR, Wong THF. The needs of society: A new understanding of transitions, sustainability and liveability. *Technol Forecast Soc Change*. 2014;85: 121–132. doi:10.1016/j.techfore.2013.09.005
12. Hausmann A, Toivonen T, Fink C, Heikinheimo V, Kulkarni R, Tenkanen H, et al. Understanding sentiment of national park visitors from social media data. *People and Nature*. 2020; pan3.10130. doi:10.1002/pan3.10130
13. Becken S, Stantic B, Chen J, Alaei AR, Connolly RM. Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing. *J Environ Manage*. 2017;203: 87–97. doi:10.1016/j.jenvman.2017.07.007
14. Ladle RJ, Correia RA, Do Y, Joo GJ, Malhado ACM, Proulx R, et al. Conservation culturomics. *Front Ecol Environ*. 2016;14: 269–275. doi:10.1002/fee.1260
15. Toivonen T, Heikinheimo V, Fink C, Hausmann A, Hiippala T, Järv O, et al. Social media data for conservation science: A methodological overview. *Biol Conserv*. 2019;233: 298–315. doi:10.1016/j.biocon.2019.01.023

16. Minin E Di, Tenkanen H, Toivonen T. Prospects and challenges for social media data in conservation science. *Front Environ Sci.* 2015;3: 1–6.
doi:10.3389/fenvs.2015.00063
17. Correia RA, Ladle R, Jarić I, Malhado ACM, Mittermeier JC, Roll U, et al. Digital data sources and methods for conservation culturomics. *Conservation Biology.* 2021;35: 398–411. doi:10.1111/cobi.13706
18. Ceron A, Curini L, Iacus SM, Porro G. 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.* 2014;16: 340–358.
doi:10.1177/1461444813480466
19. Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 2.
20. Lamb CT, Gilbert SL, Ford AT. Tweet success? Scientific communication correlates with increased citations in *Ecology and Conservation*. *PeerJ.* 2018;2018.
doi:10.7717/peerj.4564
21. Papworth SK, Nghiem TPL, Chimalakonda D, Posa MRC, Wijedasa LS, Bickford D, et al. Quantifying the role of online news in linking conservation research to Facebook and Twitter. *Conservation Biology.* 2015;29: 825–833. doi:10.1111/cobi.12455
22. Otsuka R, Yamakoshi G, Id RO, Yamakoshi G. Analyzing the popularity of YouTube videos that violate mountain gorilla tourism regulations. *PLoS One.* 2020;15: 1–20.
doi:10.1371/journal.pone.0232085
23. Fink C, Hausmann A, Di Minin E. Online sentiment towards iconic species. *Biol Conserv.* 2020;241. doi:10.1016/j.biocon.2019.108289
24. Wu Y, Xie L, Huang SL, Li P, Yuan Z, Liu W. Using social media to strengthen public awareness of wildlife conservation. *Ocean Coast Manag.* 2018;153: 76–83.
doi:10.1016/j.ocecoaman.2017.12.010
25. Bhatt P, Pickering CM. Public Perceptions about Nepalese National Parks: A Global Twitter Discourse Analysis. *Soc Nat Resour.* 2021;34: 683–700.
doi:10.1080/08941920.2021.1876193
26. Correia RA, Jepson P, Malhado ACM, Ladle RJ. Culturomic assessment of Brazilian protected areas: Exploring a novel index of protected area visibility. *Ecol Indic.* 2018;85: 165–171. doi:10.1016/j.ecolind.2017.10.033
27. Statista. Worldwide digital population as of January 2021. 2021 [cited 9 Mar 2022]. Available: <https://www.statista.com/statistics/617136/digital-population-worldwide/>.

28. Statista. Social media usage in Brazil – statistics & facts. In: 2022 [Internet]. [cited 19 Jan 2022]. Available: https://lb-aps-frontend.statista.com/topics/6949/social-media-usage-in-brazil/#topicHeader__wrapper
29. Collins K, Shiffman D, Rock J. How are scientists using social media in the workplace? PLoS One. 2016;11. doi:10.1371/journal.pone.0162680
30. Mohammadi E, Thelwall M, Kwasny M, Holmes KL. Academic information on Twitter: A user survey. PLoS ONE. Public Library of Science; 2018. doi:10.1371/journal.pone.0197265
31. Kirilenko AP, Stepchenkova SO. Public microblogging on climate change: One year of Twitter worldwide. Global Environmental Change. 2014;26: 171–182. doi:10.1016/j.gloenvcha.2014.02.008
32. Mittermeier RA, Da Fonseca GAB, Rylands AB, And ‡, Brandon K. A Brief History of Biodiversity Conservation in Brazil. Conservation Biology. 2005;19: 601–607.
33. Prates PL, Irving AMA. Conservação da biodiversidade e políticas públicas para as áreas protegidas no Brasil: desafios e tendências da origem da CDB às metas de Aichi. Revista Brasileira de Políticas Públicas. 2015. 10.5102/rbpp.v5i1.3014
34. Medeiros R, Araújo FFDS. Dez anos do sistema nacional de unidades de conservação da natureza: lições do passado, realizações presentes e perspectivas para o futuro. 2011.
35. MMA. Cadastro das Unidades de Conservação – CNUC. 2021 [cited 19 Dec 2021]. Available: https://dados.gov.br/dataset/unidadesdeconservacao/resource/baf25448-5064-4ece-9a0e-d778b0eca542?inner_span=True
36. Jenkins CN, Joppa L. Expansion of the global terrestrial protected area system. Biol Conserv. 2009;142: 2166–2174. doi:10.1016/j.biocon.2009.04.016
37. Mascia MB, Pailler S. Protected area downgrading, downsizing, and degazettement (PADDD) and its conservation implications. Conserv Lett. 2011;4: 9–20. doi:10.1111/j.1755-263X.2010.00147.x
38. Ladle RJ, Jepson P, Correia RA, Malhado ACM. The power and the promise of culturomics. Front Ecol Environ. 2017;15: 290–291. doi:10.1002/fee.1506
39. Barrie C, Ho J. academictwitter: an R package to access the Twitter Academic Research Product Track v2 API endpoint. J Open Source Softw. 2021;6: 3272. doi:10.21105/joss.03272
40. Feinerer I; Hornik K, Meyer D. Text Mining Infrastructure in R. J Stat Softw. 2008;25.
41. Harrell Jr F. Hmisc: Harrell Miscellaneous. 2023.

42. Mangiafico S. rcompanion: Functions to Support Extension Education Program Evaluation. New Jersey; 2023.
43. Otero P, Gago J, Quintas P. Twitter data analysis to assess the interest of citizens on the impact of marine plastic pollution. Mar Pollut Bull. 2021;170: 112620. doi:10.1016/j.marpolbul.2021.112620
44. python-visualization. Folium. 2020. Available: <https://python-visualization.github.io/folium/>
45. Kassambara A. Multivariate Analysis I Practical Guide To Cluster Analysis in R Unsupervised Machine Learning. 2017. Available: <http://www.sthda.com>
46. Maechler, M; Rousseeuw, P; Struyf, A; Hubert, M; Hornik K. cluster: Cluster Analysis Basics and Extensions. 2022. Available: <https://cran.r-project.org/package=cluster>.
47. Vieira, F.A.S., Santos, D.T.V., Bragagnolo, C., Campos-Silva, J.V., Correia, R.A.H., Jepson, P., Malhado, C.M., Ladle, R.J., 2021. Social media data reveals multiple cultural services along the 8.500 kilometers of Brazilian coastline. Ocean & Coastal Management, 214, 105918
48. Wickham, H.; François, R.; Henry, L.; Müller K. dplyr: A Grammar of Data Manipulation. 2021. Available: <https://cran.r-project.org/package=dplyr>
49. Wickham H. Elegant Graphics for Data Analysis: ggplot2. Applied Spatial Data Analysis with R. 2008.
50. Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny. 2020. doi:<https://doi.org/10.1201/9780429447273>
51. Kassambara, A.; Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data AnalysesNo Title. 2020. Available: <https://cran.r-project.org/package=factoextra>
52. Instituto Chico Mendes de Conservacao da Biodiversidade. Paineis Dinâmicos de Informações. Uso Público e Turismo. 2021 [cited 9 Mar 2022]. Available: <http://qv.icmbio.gov.br>
53. @mjahr. Never miss important tweets from people you follow. 2016 [cited 3 Mar 2022]. Available: https://blog.twitter.com/en_us/a/2016/never-miss-important-tweets-frompeople-you-follow
54. Huszár F, Ktena SI, O'Brien C, Belli L, Schlaikjer A, Hardt M. Algorithmic amplification of politics on Twitter. Proceedings of the National Academy of Sciences. 2022;119. doi:10.1073/pnas.2025334119/-/DCSupplemental
55. Fearnside PM. Setbacks under President Bolsonaro: A Challenge to Sustainability in the Amazon. Sustentabilidade International Science Journal. 2019;1: 38–52.

56. Ferrante L, Fearnside PM. Brazil's new president and "ruralists" threaten Amazonia's environment, traditional peoples and the global climate. *Environmental Conservation*. Cambridge University Press; 2019. doi:10.1017/S0376892919000213
57. Craig G. *Celebrities and Environmental Activism. Media, Sustainability and Everyday Life*. Palgrave Macmillan UK; 2019. pp. 135–163. doi:10.1057/978-1-137-53469-9_6
58. Brockington D. *Celebrity and the Environment: Fame, Wealth and Power in Conservation*. London: Zed Books; 2009.
59. Chang CH, Armsworth PR, Masuda YJ. Twitter data reveal six distinct environmental personas. *Front Ecol Environ*. 2022;20: 481–487. doi:10.1002/fee.2510
60. Nielsen AB, van den Bosch M, Maruthaveeran S, van den Bosch CK. Species richness in urban parks and its drivers: A review of empirical evidence. *Urban Ecosyst*. 2014;17: 305–327. doi:10.1007/s11252-013-0316-1
61. Özgüner H. Cultural differences in attitudes towards urban parks and green spaces. *Landsc Res*. 2011;36: 599–620. doi:10.1080/01426397.2011.560474
62. Chiesura A. The role of urban parks for the sustainable city. *Landsc Urban Plan*. 2004;68: 129–138. doi:10.1016/j.landurbplan.2003.08.003
63. Hausmann A, Toivonen T, Slotow R, Tenkanen H, Moilanen A, Heikinheimo V, et al. Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. *Conservation Letters*. Wiley-Blackwell; 2018. doi:10.1111/conl.12343
64. Guedes-Santos J, Correia RA, Jepson P, Ladle RJ. Evaluating public interest in protected areas using Wikipedia page views. *J Nat Conserv*. 2021;63. doi:10.1016/j.jnc.2021.126040
65. Levin N, Lechner AM, Brown G. An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. *Applied Geography*. 2017;79: 115–126. doi:10.1016/j.apgeog.2016.12.009
66. Fisher DM, Wood SA, Roh YH, Kim CK. The geographic spread and preferences of tourists revealed by user-generated information on jeju island, south korea. *Land (Basel)*. 2019;8. doi:10.3390/LAND8050073
67. Tenkanen H, Di Minin E, Heikinheimo V, Hausmann A, Herbst M, Kajala L, et al. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Sci Rep*. 2017;7. doi:10.1038/s41598-017-18007-4
68. Correia RA, Jepson P, Malhado ACM, Ladle RJ. Internet scientific name frequency as an indicator of cultural salience of biodiversity. *Ecol Indic*. 2017;78: 549–555. doi:10.1016/j.ecolind.2017.03.052

69. Zmihorski M, Dziarska-Palac J, Sparks TH, Tryjanowski P. Ecological correlates of the popularity of birds and butterflies in Internet information resources. *Oikos*. 2013;122: 183–190. doi:10.1111/j.1600-0706.2012.20486.x
70. Dudley N. Guidelines for applying Protected Area Management Categories. IUCN, Gland, Switzerland. IUCN; 2008. doi:10.1103/PhysRevB.38.10724
71. Figueiredo CCM. From paper parks to real conservation: case studies of national park management effectiveness in Brazil. Ohio State University. 2007. Available: https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1167587930&disposition=inline
72. Gamarra NC, Correia RA, Bragagnolo C, Campos-Silva JV, Jepson PR, Ladle RJ, et al. Are Protected Areas undervalued? An asset-based analysis of Brazilian Protected Area Management Plans. *J Environ Manage*. 2019;249: 109347. doi:10.1016/j.jenvman.2019.109347
73. Prado DAR. Parque Municipal Flamboyant: apropriação e usos para lazer. Universidade Federal de Goiás. 2012. Available: <https://repositorio.bc.ufg.br/tede/handle/tde/1880>
74. Calheiros AR, Souza MA, Costa JG da, Araújo KD. Espécie invasora de bambu e seus impactos sobre a qualidade do solo. *Revista Ibero-Americana de Ciências Ambientais*. 2023;13: 63–73. doi:10.6008/cbpc2179-6858.2022.006.0006
75. Bragagnolo C, Correia RA, Gamarra NC, Lessa T, Jepson P, Malhado ACM, et al. Uncovering assets in Brazilian national parks. *J Environ Manage*. 2021;287. doi:10.1016/j.jenvman.2021.112289
76. Izidoro FB, Schiavetti A. Associated benefits of manatee watching in the Costa dos Corais Environmental Protection Area. *Front Mar Sci*. 2022;9. doi:10.3389/fmars.2022.1002855
77. Vidal MD, Santos MC, Jesus JS, C Santos PM, Jesus JS, Alves LCPS, Chaves, MPSR. Ordenamento participativo do turismo com botos no Parque Nacional de Anavilhanas, Amazonas, Brasil. *Bol Mus Para Emílio Goeldi Cienc Nat*. 2017.
78. Tortato FR, Ribas C, Concone HVB, Hoogesteijn R. Turismo de observação de mamíferos no Pantanal. *Boletim do Museu Paraense Emílio Goeldi - Ciências Naturais*. 2022;16: 351–370. doi:10.46357/bcnaturais.v16i3.814
79. Almeida JAGR, Guedes-Santos J, Vieira FAS, Azevedo AK, Souza CN, Pinheiro BR, et al. Public awareness and engagement in relation to the coastal oil spill in northeast Brazil. *An Acad Bras Cienc*. 2022;94: 1–10. doi:10.37002/biobrasil.v12i2.2177
80. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*. 2018 Mar 9;359(6380):1146-1151. doi: 10.1126/science.aap9559. PMID: 29590045.

81. Ladle, RJ, Souza, CN, Correia R. Culturomics for (not against!) protected areas In. Biol Conserv. 2021;256: 109197. doi:10.1016/j.biocon.2021.109015

5 USING SOCIAL MEDIA AND MACHINE LEARNING TO UNDERSTAND SENTIMENTS TOWARDS BRAZILIAN NATIONAL PARKS⁴

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; 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

*Corresponding author:

Carolina Neves Souza.

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.

⁴ Artigo aceito para publicação na revista Biological Conservation.
<https://doi.org/10.1016/j.biocon.2024.110557>

⁺ Shared first authors.

5.1 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 and 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.

5.2 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 km². 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 Balakrishnan, 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.

5.3 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.

5.3.1 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 km². 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).

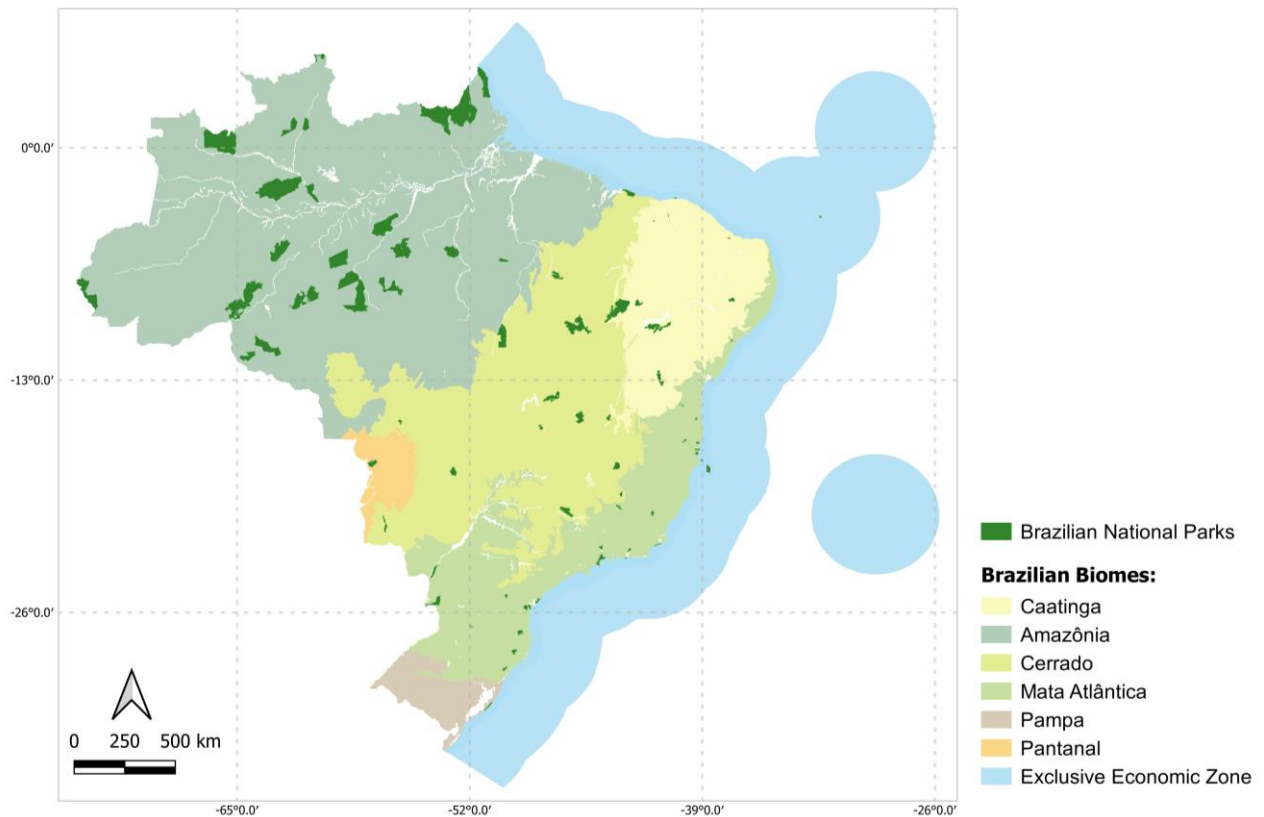


Fig. 1. Map of distribution of 74 Brazilian national parks. Spatial distribution of the 74 Brazilian national parks located in their respective biomes.

5.3.2 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.

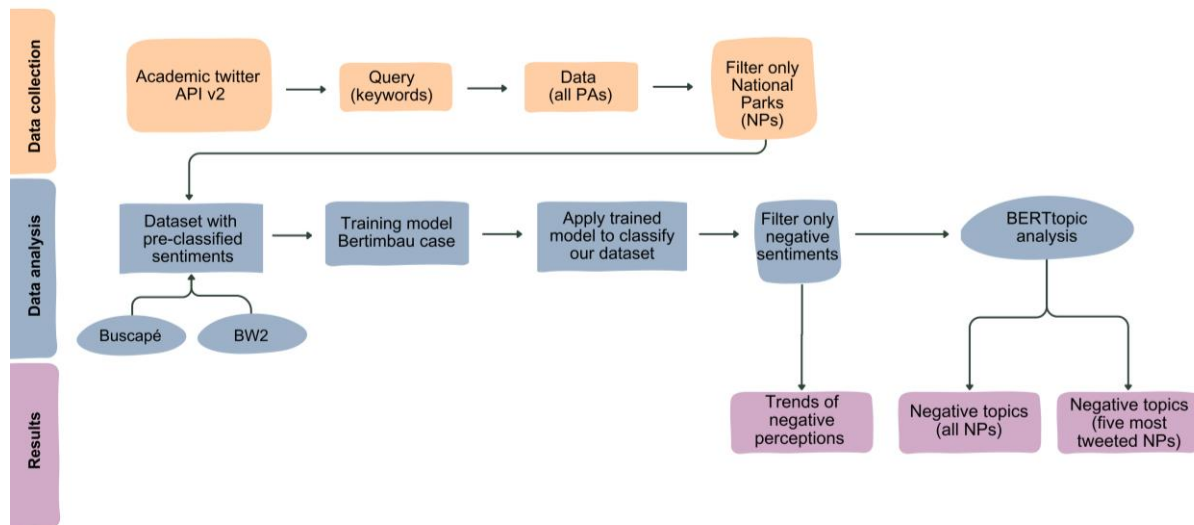


Fig. 2. Methodological flowchart. Methodological flowchart from data collection to results. The flowchart showed all the steps used during the research: data collection, data analysis, and results.

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>).

5.3.3 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.

5.3.4 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.

5.3.5 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.

5.4 Results

5.4.1 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).

5.4.2 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].

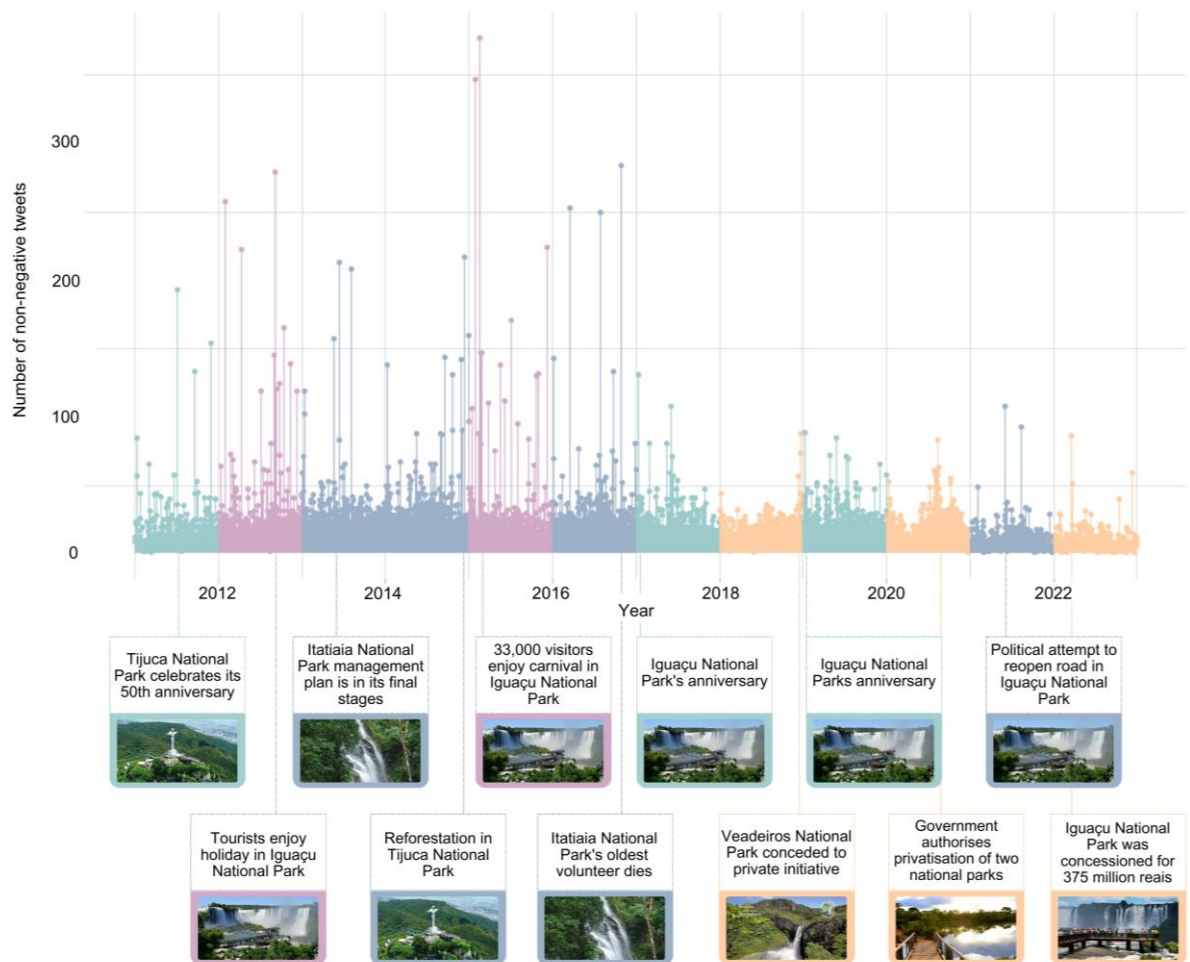


Fig. 3. The daily counts related to non-negative sentiments of Twitter posts regarding Brazilian national parks from January (2011) to December 2022. The data were obtained from Twitter's application programming interfaces (API). The green colour represents events involving National Parks anniversary; 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.

5.4.3 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].

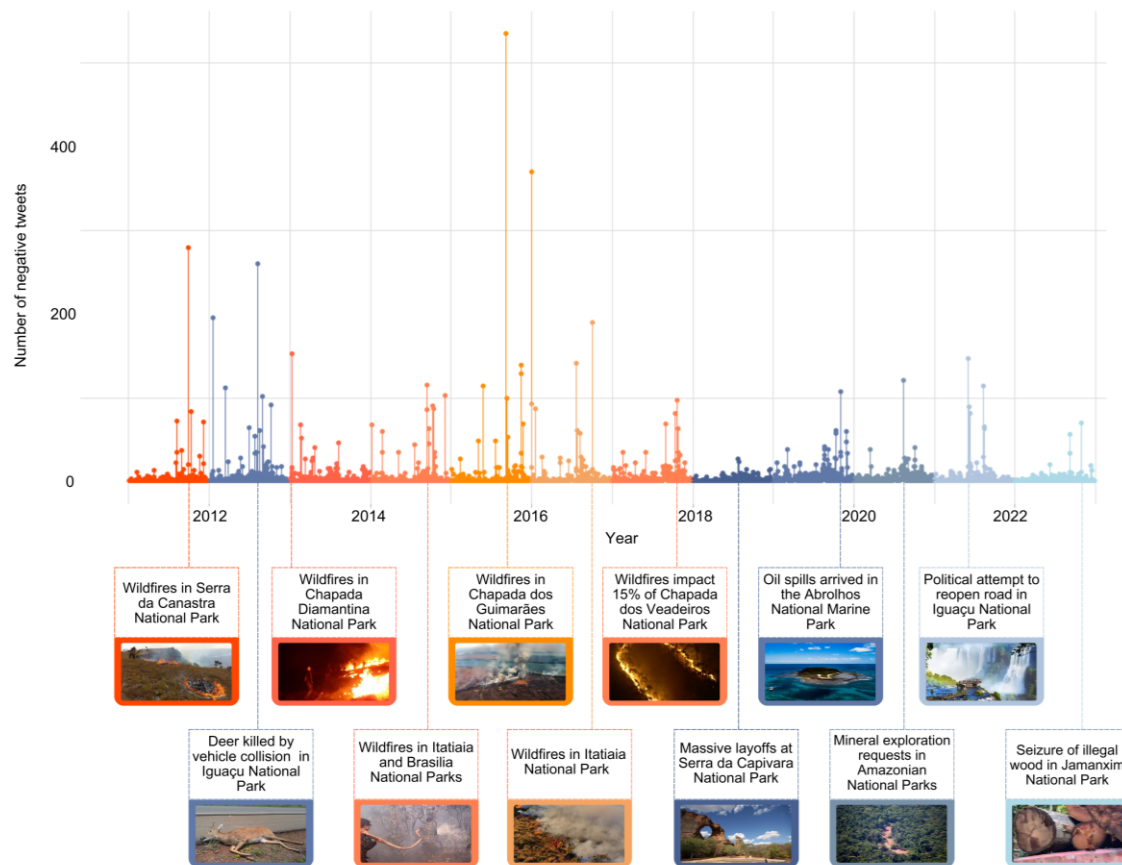


Fig. 4. The daily counts related to negative sentiments of Twitter posts regarding Brazilian national parks from January (2011) to December 2022. The data were obtained from Twitter's application programming interfaces (API). The orange colours represent events involving wildfires and the blue colours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

5.4.4 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.

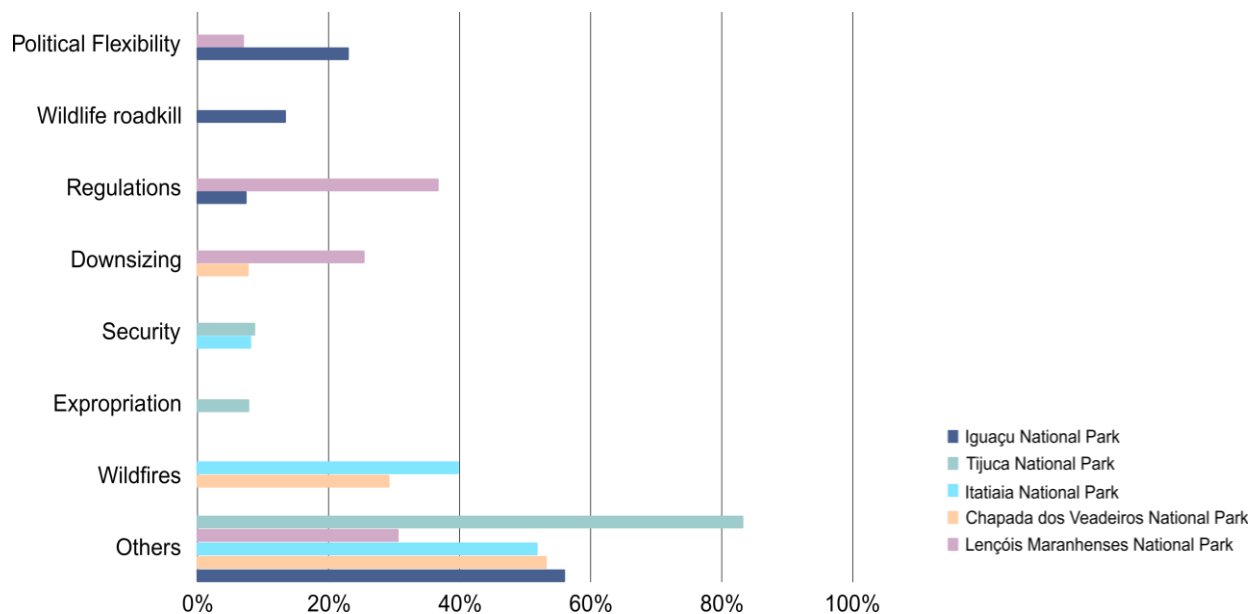


Fig. 5. A bar chart representing the dominant negative topics per park. Each topic is identified using BERTopic analysis and then subjectively classified based on the model output. The parks included are the five most tweeted Brazilian Parks (2011-2022).

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	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	1.77%	Regulations
	4	310	4_dead_inside_deer_found_found_dead	1.69%	Wildlife roadkill
	5	293	5_iguazu_police_environmental_police	1.59%	Regulations
	6	289	6_destroys_area_hit_12_12_thousand_main	1.57%	Wildfires
	7	277	hotspots	1.51%	Photo publishing*
	8	238	7_photo_park_publish_photo_finished	1.29%	Privatization
	-	12783	8_privatization_privatization_park_privatize	69.52%	-
Others and Outliers					
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_spatial	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	1.50%	Regulations
	-	12582	Others and Outliers	68.43%	-

5.5 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.

5.5.1 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.

5.5.2 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).

5.5.3 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.

REFERENCES

- 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. *Tour. Recreat. Res.* 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. *J. Retail. Consum. Serv.* 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. doi:<https://doi.org/10.1109/BRACIS.2014.57>.
- B2W-Reviews01., 2021. Open corpus of product reviews. GitHub. <https://github.com/b2wdigital/b2w-reviews01>.

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. *Conserv. Biol.* 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), 130–156.

Bragagnolo, C., Costa Gamarra, N., Malhado, Claudia Mendes, 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.

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>.

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).

Corpus Buscapé, 2013. Portuguese product reviews https://drive.google.com/file/d/1IZJuvt1_uXQ4oPGAvGQxQ_h_ZiV-Be72/view.

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.

Conserv. Biol. 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., 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. *Conserv. Biol.* 35, 437–446. <https://doi.org/10.1111/cobi.13708>.

Dobrovolski, R., Loyola, R., Rattis, L., Gouveia, S.F., Cardoso, D., Santos-Silva, R., Gonçalves-Souza, D., Bini, L.M., 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.

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>.

Dudley, N., 2008. Guidelines for Applying Protected Area Management Categories. IUCN, IUCN, Gland, Switzerland. <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. *Front. Sociol.* 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.

Ferreira, J., et al., 2014. Brazil's environmental leadership at risk. *Science* 346, 706–707. <https://doi.org/10.1126/science.1260194>.

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. *Am. Psychol.* 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 areas: 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., 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>.

Guedes-Santos, J., Correia, R.A., Jepson, P., Ladle, R.J., 2021. Evaluating public interest in protected areas using Wikipedia page views. *J Nat Conserv* 63. <https://doi.org/10.1016/j.jnc.2021.126040>.

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* 3.10130. <https://doi.org/10.1002/pan3.10130>.

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>

Hoeffel, J.L., Fadini, A.A.B., Machado, M.K., Reis, J.C., 2008. Trajetórias do Jaguar- 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.

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. *Knowl. Inf. Syst.* 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 For. Urban Green.* 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., 2010. Environmental Perception. In: Warf, B., *Encyclopedia of Geography*. Sage.

Liu, B., 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.

Lubbe, B.A., du Preez, E.A., Douglas, A., Fairer-Wessels, F. The impact of rhino poaching on tourist experiences and future visitation to National Parks in South Africa. *Current Issues in Tourism*. <https://doi.org/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.

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

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. Conserv. Biol. 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. J. Soc. Psychol. 116 (2), 199–204. <https://doi.org/10.1080/00224545.1982.9922771>.

Pivello, V.R., 2006. Fire management for biological conservation in the Brazilian cerrado. In: Mistry, J., Berardi, A. (Eds.), Savannas and Dry Forests: Linking People with Nature. Ashgate Publications, pp. 129–154.

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

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. <https://doi.org/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. *Aust. J. Environ. Manag.* 23, 245–264. <https://doi.org/10.1080/14486563.2015.1132397>.

Rylands, A.B., Brandon, K., 2005. Brazilian protected areas. *Conserv. Biol.* doi: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., Lee, Sun Young, 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/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. doi:<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, F., Nogueira, R., Lotufo, Roberto, 2020. BERTimbau: Pretrained BERT Models for Brazilian Portuguese, in: 9th Brazilian Conference. pp. 403–417. doi:<https://doi.org/10.1007/978-3-030-61377-8>

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. doi:<https://doi.org/10.1371/journal.pone.0293581>.

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>.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E., 2019. Social media data for conservation science: A methodological overview. *Biol Conserv* 233, 298–315. <https://doi.org/10.1016/j.biocon.2019.01.023>.

Vale, M.M., Berenguer, E., Argollo de Menezes, M., Viveiros de Castro, E.B., Pugliese de Siqueira, L., 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.

6 CONSIDERAÇÕES FINAIS

Este estudo oferece uma contribuição significativa ao crescente campo da análise de métricas culturômicas para monitorar as interações homem-natureza em grandes escalas. A análise de dados das redes sociais, demonstrou ser uma poderosa ferramenta capaz de promover uma abordagem mais abrangente e participativa para a gestão das áreas protegidas (APs) brasileiras.

No primeiro artigo, a análise de 10 anos de discurso no Twitter sobre as APs destacou o poderoso potencial das redes sociais para a compreensão das percepções sobre essas áreas. O estudo realçou a necessidade de aprimorar a comunicação oficial sobre as APs, destacando a importância de fornecer informações básicas, como o significado dessas áreas, esclarecimento das diferenças entre áreas protegidas e parques urbanos, e a relevância dos regulamentos específicos de cada AP. Adicionalmente, o estudo destacou que os usuários das redes sociais demonstram maior envolvimento com publicações relacionadas à conflitos. O monitoramento das redes sociais pode auxiliar os gestores na compreensão das percepções e interesses do público que utiliza estas plataformas online, possibilitando a identificação de lacunas e a antecipação de possíveis conflitos para aprimorar estratégias visando maximizar os resultados positivos das ações de conservação (Capítulo 1, Fig. 7A e 7B).

No segundo artigo, a análise de sentimentos emergiu como uma ferramenta valiosa para a gestão de APs, fornecendo insights sobre as percepções negativas manifestadas pelos usuários do Twitter com relação aos parques nacionais brasileiros. Foram identificadas áreas críticas abrangendo temas como incêndios nos parques nacionais, segurança dentro do parque, mortalidade da vida selvagem por atropelamentos, regramentos, privatizações (concessões) e a escassez de recursos financeiros para a otimização da gestão dessas áreas (Capítulo 2, Tabela 1). Além disso, diminuimos a escala espacial para compreender o desempenho da ferramenta de análise de sentimentos na obtenção de fatores que provocavam sentimentos negativos no público online considerando os parques nacionais individualmente. A análise, realizada separadamente para cada parque, captou com sucesso, as percepções negativas dos usuários do Twitter sobre cinco diferentes parques nacionais, mesmo diante de suas

características únicas e diferentes contextos de gestão (Capítulo 2, Fig. 5). Destacam-se insatisfações relacionadas a implementação de regimentos, bem como tentativas de flexibilização política, como a abertura de estradas dentro de parques e reduções do tamanho dessas áreas protegidas.

Contudo, a análise de dados de redes sociais, requer cautela devido a desafios relacionados à subjetividade textual como coloquialismos, gírias, ironias e questões éticas relacionadas a coleta de dados. Ademais, é fundamental enfatizar que existe a necessidade de aprimoramento dos conjuntos de dados para classificação dos sentimentos na língua portuguesa. Neste trabalho foram encontradas dificuldades em distinguir tweets positivos e neutros dos tweets negativos. No entanto, ao estender o uso da análise de sentimentos para compreender as percepções das pessoas sobre a conservação ambiental na língua portuguesa, esta pesquisa abre novas possibilidades e fortalece o uso dessa abordagem metodológica para outros idiomas além do inglês. Outro desafio a ser considerado é a representatividade das opiniões online em comparação com as opiniões do mundo real. Mesmo diante do fato de que aproximadamente 83,6% da população brasileira utiliza alguma rede social, é de suma importância avaliar de maneira crítica se os resultados das percepções oriundas das redes sociais são representativos dos usuários das áreas protegidas ou da sociedade em geral.

Reconhecendo estes desafios e o potencial inerente ao uso das redes sociais como ferramenta de investigação, é aconselhável realizar uma limpeza criteriosa e uma análise crítica de todos os dados utilizados, visando minimizar as possíveis subjetividades textuais existentes. Para fins éticos, embora o conteúdo gerado pelos usuários das redes sociais ser disponibilizado gratuitamente por algumas plataformas, é de extrema importância considerar a privacidade e o bem-estar dos usuários. Para tal, o compartilhamento dos dados coletados deve ser realizado utilizando critérios que garantam o anonimato e a proteção à privacidade dos usuários. Com o intuito de aprimorar a precisão dos modelos de aprendizado de máquina voltados à classificação de sentimentos, recomenda-se que futuras pesquisas utilizem bancos de dados com sentimentos previamente classificados, especialmente focados em temas ambientais. Por fim, para mitigar os desafios relacionados à representatividade, é sugerido integrar os resultados das pesquisas em redes sociais a diversas fontes online, como Wikipedia,

Instagram e Facebook. Além disso, validar os resultados online por meio de estudos de campo amplia o potencial de representatividade dos dados, conectando as percepções online às percepções offline relacionadas às áreas protegidas.

Em geral, ao enfrentar esses desafios, a análise de dados das redes sociais, pode se tornar uma ferramenta valiosa na gestão sustentável e eficaz das áreas protegidas. No contexto prático da gestão da biodiversidade brasileira, a utilização das redes sociais pode auxiliar nas seguintes estratégias de gestão:

- i. Complementar os monitoramentos anuais já conduzidos pelos órgãos gestores como o Sistema de Análise e Monitoramento de Gestão (SAMGE), direcionado a avaliação da efetividade das APs, bem como o monitoramento anual da visitação das áreas protegidas. Incluir as percepções e atitudes da população aos sistemas de avaliação e monitoramento podem proporcionar insights holísticos sobre as necessidades de melhorias das áreas protegidas;
- ii. Subsidiar os planos de comunicação das APs, identificando as áreas que exigem maior atenção da equipe gestora. Isso pode envolver uma melhor comunicação sobre a valorização de locais de visitação significativos para os visitantes, bem como o esclarecimento sobre os objetivos de criação, regimentos e até possíveis problemas e conflitos para a sociedade;
- iii. Monitorar eventos que geram mais interesse público e sentimentos negativos, possibilitando uma ação rápida da administração da área protegida em relação a crises e eventos adversos;
- iv. Auxiliar no desenvolvimento ou revisão dos planos de manejo, planos de uso público e de interpretação ambiental, nos quais as percepções e sentimentos do público online podem complementar as informações e opiniões da comunidade local em relação ao uso do espaço público protegido.

Além dos resultados e sugestões apresentados, esta pesquisa suscitou duas questões que merecem uma atenção mais aprofundada em futuras investigações. A primeira diz respeito à representatividade das percepções online sobre áreas protegidas.

Ou seja, o quão representativos são os dados extraídos de redes sociais e até que nível podemos considerar estas percepções como sendo a percepção da sociedade como um todo? E a segunda está relacionada às percepções da sociedade em relação as espécies icônicas protegidas por meio dessas áreas – dado que esta discussão, com exceção do atropelamento de fauna silvestre, não emergiu durante a análise dos mais de 10 anos de dados coletados sobre as áreas protegidas brasileiras no Twitter.

Neste contexto, em meio a ampliação cotidiana de ferramentas online, onde as pessoas compartilham cada vez mais seus pensamentos, opiniões e anseios de maneira pública e abrangente, os dados digitais das redes sociais tornam-se grandes aliados a favor da conservação ambiental, auxiliando na identificação de lacunas e na promoção de melhorias que conquistem o apoio do público às áreas protegidas brasileiras.

APÊNDICE A – MATERIAL SUPLEMENTAR

5 USING SOCIAL MEDIA AND MACHINE LEARNING TO UNDERSTAND SENTIMENTS TOWARDS BRAZILIAN NATIONAL PARKS

Table 1: Palavras-chave utilizadas na coleta de dados no Twitter

N	Keywords used in data collection (PT)	Keywords used in data collection (EN)	Brazilian PA category (BRASIL, Law No 9.985, of 18 July 2000)
1	Parque Nacional	National park	Parque
2	Parque estadual	State park	Parque
3	Parque natural municipal	Municipal natural park	Parque
4	Parque municipal	Municipal park	Parque
5	Estação ecológica	Ecological station	Estação Ecológica
6	Reserva biológica	Biological reserve	Reserva Biológica
7	Monumento natural	Natural monument	Monumento Natural
8	Refúgio da vida silvestre	Wildlife refuge	Refúgio de Vida Silvestre
9	Reserva extrativista	Extractive reserve	Reserva Extrativista
10	Área de proteção ambiental	Environmental protection area	Área de Proteção Ambiental
11	Floresta nacional	National forest	Floresta
12	Floresta estadual	State forest	Floresta
13	Floresta municipal	Municipal forest	Floresta
14	Reserva de desenvolvimento sustentável	Sustainable development reserve	Reserva de Desenvolvimento Sustentável
15	Área de relevante interesse	Area of relevant interest	Área de Relevante Interesse Ecológico
16	Reserva Particular do Patrimônio Natural	Private natural heritage reserve	Reserva Particular do Patrimônio Natural
17	Unidade de conservação	Conservation unit	-
18	Area protegida	Protected area	-

APÊNDICE B – MATERIAL SUPLEMENTAR

5 USING SOCIAL MEDIA AND MACHINE LEARNING TO UNDERSTAND NEGATIVE SENTIMENTS TOWARDS BRAZILIAN NATIONAL PARKS

Material suplementar: Informações metodológicas sobre a análise de sentimento com dados de tweets em português.

In the scope of our study, an extensive dataset comprising over 100,000 tweets concerning the 74 Brazilian National Parks was acquired. The primary objective was to discern the sentiment embedded within these tweets and ascertain the principal themes associated with each sentiment classification. The initial phase encompassed the classification of these tweets based on their sentiment through the utilization of a pre-trained transformer model - BERTimbau Base (a.k.a. "bert-base-portuguese-cased") (Souza et al. 2020). This model was subsequently refined to cater specifically to our task, which involves the categorization of tweets into either positive, negative, or neutral sentiments.

Given the focus on understanding and classifying the sentiment expressed within tweets pertaining to Brazilian national parks, the model necessitated fine-tuning utilizing a pre-structured dataset that had already been categorized across a spectrum of sentiment types. Nonetheless, it is noteworthy that pre-trained models such as BERTimbau Base are primarily designed for three fundamental natural language processing (NLP) tasks: Named Entity Recognition, Sentence Textual Similarity, and Recognizing Textual Entailment, and their direct application to text classification is not inherently straightforward (Tunstall L. et al., 2022). Hence, adaptations were implemented to align the model with our distinct classification objectives.

In this endeavor, we harnessed a comprehensive dataset originating from the B2W e-commerce company, in conjunction with the Corpus Buscapé (B2W-Reviews01, 2018; Corpus Buscapé, 2013). This publicly accessible dataset encompasses in excess of 130,000 user reviews encompassing a diverse range of products. Among its attributes, it

features binary labels denoting whether users would recommend the product to others and ratings scored on a scale from 1 to 5 stars. It is pertinent to mention that our analysis solely engaged with the user rating component.

Opinando, an institution specializing in mining opinions from Portuguese textual content and established under the auspices of the Research Office of the University of São Paulo (USP), has contributed significantly to the domain of natural language processing (NLP). One of its notable contributions is the creation of various robust corpora germane to this field. Among these, the Corpus Buscapé emerges as a substantial compilation of Portuguese product reviews, harvested in the year 2013 from the Buscapé website, renowned for its product and price exploration capabilities. Unlike the previously referenced datasets, this corpus employs a rating spectrum spanning from 0 to 5. As a corollary, comments affiliated with a rating of zero were omitted from the analytical considerations.

The decision to select this particular dataset hinged on its capacity to yield precise predictions for the classification of the target tweets. Notably, this opinion-oriented dataset is stratified based on a "rating" feature, encompassing scores ranging from 1 to 5. A rating of 1 signifies a markedly negative sentiment, while a rating of 5 corresponds to a highly positive sentiment. As part of our classification strategy, ratings of 1 and 2 were grouped as "negative," a rating of 3 as "neutral," and ratings of 4 and 5 as "positive." With the model having undergone training and validation using this dataset, its application was extended to a subset of approximately 2,000 tweets culled from our larger collection. This selected subset had been manually categorized in advance to serve as a dedicated test set for evaluation purposes.

1. How Long Are Our Tweets?

Transformer models possess a defined upper limit for the length of input sequences, commonly denoted as the "maximum context size" or "tokens." Tokens within these models can encompass complete words, subword components, or even singular characters, such as punctuation marks.

For instance, a tweet is constrained to a maximum of 280 characters. Considering the average of approximately 6 characters per word in the Portuguese language, the typical tweet may contain approximately 47 words, contingent on the nature of the content under analysis. The e-commerce sales platform's dataset on product reviews presents the following word frequency:

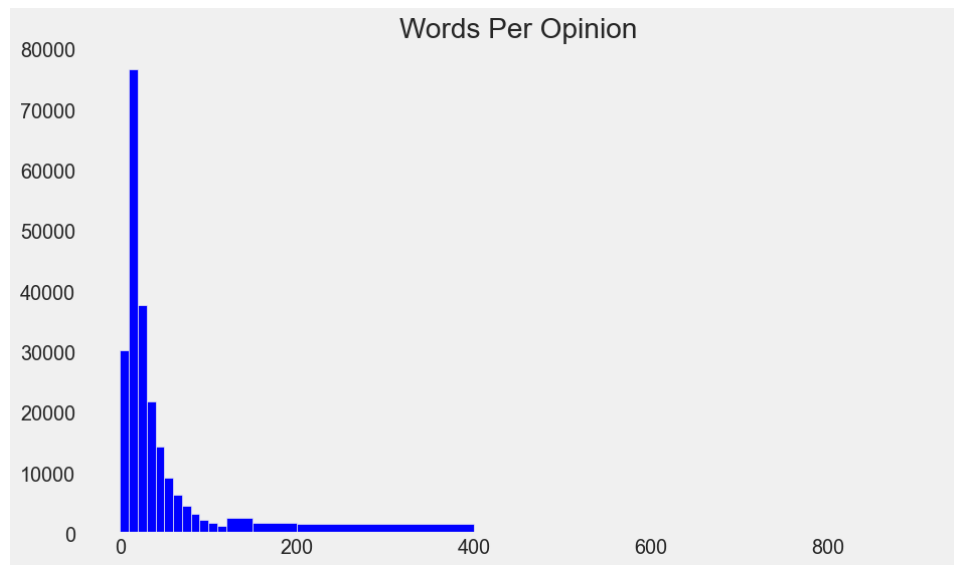


Fig. 1: Word frequency about product reviews on Buscapé.

By type of sentiment:

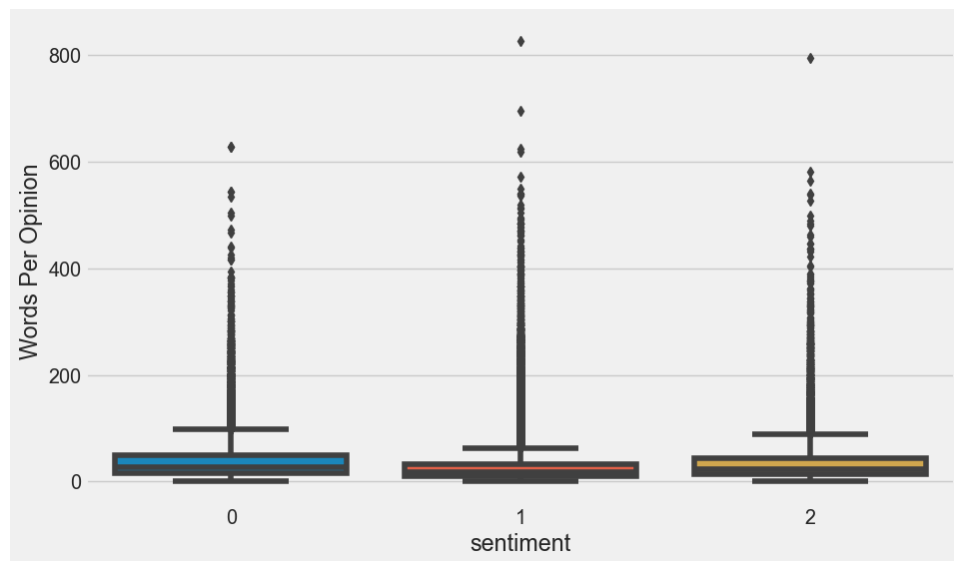


Fig. 2: Boxplot illustrating the sentiment per word in the reviews from the Buscapé website. The number 1 represents positive sentiment, 0 represents negative sentiment, and 2 represents neutral sentiment.

Positive sentiments generally exhibit greater word length in their expressions; however, this trend is subject to exceptions, as evidenced by the elongated tails observed in the boxplot representations. Hence, adaptations were implemented to align the model with our distinct classification objectives through Transfer Learning techniques.

2. BERTopic 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).

In the first step of the BERTopic algorithm, we obtain embeddings for all documents in the corpus, which are numeric vector representations of the documents. The next step is to perform clustering on the embedded documents, for which dimensionality reduction techniques, such as Uniform Manifold Approximation and Projection (UMAP), are employed to reduce the high dimensionality of the embedding vectors (McInnes et al., 2018). The UMAP algorithm is used by default because it preserves both the local and global structure of the data with superior runtime performance, an important factor in representing the semantics of text data. Preprocessing of text data is an optional step in natural language processing. Generally, It is not recommended to remove stop words as a preprocessing step when using the BERTopic model because transformer-based embedding models, which we utilise, require the complete context to generate accurate embeddings.

The default clustering algorithm used by BERTopic is HDBSCAN, which is a density-based model that automatically identifies the number of clusters without requiring a pre-specified number of clusters. HDBSCAN is a hierarchical density-based clustering algorithm proposed by (Campello et al., 2013). In this algorithm, documents with higher similarity are grouped into clusters based on cluster stability. One important characteristic of HDBSCAN is that it does not force the assignment of a data point to a specific cluster.

If the data point does not fit into any similarity-based group, it is considered an outlier (Capellaro, 2021). Once the documents are assigned to clusters, the next step is to obtain the topic representation for each cluster using class-based Term Frequency-Inverse Document Frequency (c-TF-IDF).

This method selects the top words with the highest c-TF-IDF scores within a cluster to represent each topic (Grootendorst, 2022). This means that the higher the value of a term, the more representative it is of its topic. Furthermore, 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. Given that the use of BERTopic also requires significant effort due to the dynamic nature of topic structures, which change when researchers experiment with different numbers of topics, it can be considered a laborious task to access the topics that best represent the database. 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 six most frequently visited parks (ICMBio, 2021), followed by clustering to discern the prominent negative topics associated with each individual park. To achieve this, we performed the BERTopic model with the following hyperparameters:

- For the 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 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.

- For BERTopic 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 stopwords 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.

After evaluating the possible topics generated by the model, we identified, based on our knowledge, the topics that remained consistent across all generated models.

REFERENCES

- Campello, R., Moulavi, D., Sander, J., 2013. Density-Based Clustering Based on Hierarchical Density Estimates. Gold Coast.
- 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>
- Grootendorst, M., 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure.
- ICMBio, 2021. Unidades de conservação federais atingem novo recorde de visitação em 2021. Accessed on 23 April 2023. <<https://www.gov.br/pt-br/noticias/viagens-e-turismo/2022/04/unidades-de-conservacao-federais-atingem-novo-recorde-de-visitacao-em-2021>>
- McInnes, L., Healy, J., Melville J., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. <https://arxiv.org/abs/1802.03426>

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>

Tunstall L., von Werra L., Wolf T., 2022. Natural Language Processing with Transformers. O'Reilly Media, Inc.