Understanding sentiment of national park visitors from social media data

Hausmann, Anna

John Wiley & Sons Ltd.
2020-09

Hausmann, A, Toivonen, T, Fink, C, Heikinheimo, V, Kulkarni, R, Tenkanen, H & Minin, E D 2020, 'Understanding sentiment of national park visitors from social media data', People and Nature, vol. 2, no. 3, pp. 750-760. https://doi.org/10.1002/pan3.10130

http://hdl.handle.net/10138/322546
10.1002/pan3.10130

Received: 12 July 2019 | Accepted: 9 June 2020
DOI: 10.1002/pan3.10130

RESEARCH ARTICLE

PEOPLE AND NATURE | BRITISH ECOLOGICAL SOCIETY

Understanding sentiment of national park visitors from social media data

Anna Hausmann[1,2] | **Tuuli Toivonen**[1,2] | **Christoph Fink**[1,2] | **Vuokko Heikinheimo**[1,2] | **Ritwik Kulkarni**[1] | **Henrikki Tenkanen**[1,2] | **Enrico Di Minin**[1,2,3]

[1]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

[2]Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, Finland

[3]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa

Correspondence
Anna Hausmann
Email: anna.hausmann@helsinki.fi

Funding information
Helsinki Institute of Sustainability Science; KONE Foundation; University of Helsinki; Academy of Finland 2016–2019, Grant/Award Number: 296524; European Research Council, Grant/Award Number: 802933

Handling Editor: Maria Puig de la Bellacasa

Abstract

1. National parks are key for conserving biodiversity and supporting people's well-being. However, anthropogenic pressures challenge the existence of national parks and their conservation effectiveness. Therefore, it is crucial to assess how people perceive national parks in order to enhance socio-political support for conservation.
2. User-generated data shared by visitors on social media provide opportunities to understand how people perceive (e.g. preferences, feelings, opinions) national parks during nature-based recreational experiences. In this study, we applied methods from automated natural language processing to assess visitors' sentiment when describing experiences in Instagram posts geolocated inside four national parks in South Africa.
3. We found that visitors' sentiment was positive, and mostly included emotions such as joy, anticipation, trust and surprise, with only a small occurrence of posts with negative feelings. Appreciation of nature, in association with a diverse set of other aspects, such as activities, geographical features and tourist attractions, was used to describe experiences related to nature, wilderness, travelling, holidays and adventures. The type of nature-based experience described by visitors was park specific, revealing different profiles of parks providing wildlife or scenery experiences.
4. Findings support and highlight the societal role of national parks in providing visitors with opportunities to develop positive connections with nature. Social media data may be used to understand visitors' perceptions, and how the image of national parks is constructed by users in the virtual social environment. This may help inform management for promoting a high-quality tourism experience, as well as conservation marketing aimed at fostering socio-political support for national parks and their long-term conservation effectiveness.

KEYWORDS
culturomics, digital conservation, ecotourism, experiences, national parks, natural language processing, opinion mining, sentiment

1 | INTRODUCTION

In the Anthropocene, human activities are dramatically transforming the biosphere, causing an unprecedented loss of biodiversity (Dirzo et al., 2014). Together with a broader system of designated protected areas, national parks are key policy instruments for protecting wide ecological structure and processes necessary to conserve biodiversity (Watson, Dudley, Segan, & Hockings, 2014). However, the needs of a growing human population pose financial, economic and political pressures on national parks, challenging their existence and effective conservation outcomes (Chan et al., 2007). These include, for example, high management costs and lack of resources, pressure to explore other land use options (Di Minin, MacMillan, et al., 2013), poor governance (Eklund & Cabeza, 2017) and pressure from the downgrading, downsizing and degazettement of protected areas (Golden Kroner et al., 2019).

National parks also play a key societal role (Dudley, 2008). Historically, they were firstly designated to conserve a country's natural heritage (Gissibl, Höhler, & Kupper, 2012). Today, this scope was broadened to engage a diverse range of political, cultural, economic and ecological values, including biogeography representativeness and resource management that national parks entail (Gissibl et al., 2012; Watson et al., 2014). In particular, access to recreation, education and other non-material benefits people obtain from cultural ecosystem services (Ament, Moore, Herbst, & Cumming, 2016; Millennium Ecosystem Assessment, 2005) are key aspects defining the primary socio-ecological objectives of national parks (Dudley, 2008; Eagles & McCool, 2002). Therefore, understanding how people perceive national parks during nature-based recreational experiences is important in order to help enhance physical and psychological benefits to visitors, and foster long-term socio-political support for conservation (McCool, 2006).

Visiting national parks may evoke both positive and negative sentiment in people (Terraube, Fernández-Llamazares, & Cabeza, 2017). Interaction with nature elicits positive emotions (Maller et al., 2010), for example, by improving physical, mental and psychological health, reducing stress, increasing physical recovery from illnesses (Velarde, Fry, & Tveit, 2007) and promoting social integration (Abraham, Sommerhalder, & Abel, 2010) and well-being (Puhakka, Pitkänen, & Siikamäki, 2017). These feelings help develop a positive attachment to nature, which can elicit benefits from sense of place, while promoting people's pro-environmental behaviour and support for conservation (Hausmann, Slotow, Burns, & Di Minin, 2016). In contrast, dissatisfaction towards experiences or services and expectations may bring out adverse sentiment towards national parks, disrupting attachment and intention for future visits (Kil, Holland, Stein, & Ko, 2012). For example, controversial management, such as species population control (Gusset et al., 2008) and anti-poaching activities (Lubbe, du Preez, Douglas, & Fairer-Wessels, 2019), may potentially result in lack of support to management, opposition to conservation initiatives and conflict (Chan et al., 2007; Gusset et al., 2008). As a consequence of lack of interest or alienation from nature (e.g. less emotional affinity due to the extinction of experience, Soga et al., 2016), people may also have no emotional response towards national parks, which may result in the lack of support and engagement for biodiversity conservation (Zhang, Goodale, & Chen, 2014).

In order to assess visitors' attitudes towards national parks, managers have been traditionally using surveys, such as feedback questionnaires (e.g. Boshoff, Landman, Kerley, & Bradfield, 2007; Puhakka et al., 2017). However, surveys are generally costly to implement, time-consuming and limited in space and time by providing only a snapshot of the situation, while resources available to managers are limited (McCarthy et al., 2012). On the other hand, social media platforms have become popular means, among national park visitors, for sharing experiences through photos, videos and text (Hausmann et al., 2018). Social media data may provide a cost-effective, widespread and real-time source of information which can be used to understand human–nature interactions (Di Minin, Tenkanen, & Toivonen, 2015; Toivonen et al., 2019), including landscape values (van Zanten et al., 2016), preferences for biodiversity (Hausmann et al., 2018; Willemen, Cottam, Drakou, & Burgess, 2015), and visitation to nature-based destinations (Hausmann et al., 2019; Tenkanen et al., 2017). In addition, compared to surveys, data from social media may reveal how the destination image (i.e. what people know, how do they feel and act in relation to a place; Tasci, Gartner, & Cavusgil, 2007) is represented and constructed in the virtual social environment, which may reflect a different or richer view from what is projected by traditional marketing productions (Hunter, 2016). This is a topic of growing importance in tourism research (Govers & Go, 2008), as digital platforms, including social media, are increasingly playing a significant role in shaping public reputations of places, travellers' behaviour and choices (Zeng & Gerritsen, 2014), and visitors' expectations for a satisfactory experience (Hunter, 2016). However, it is still an unexplored aspect of tourists' visitation in national parks. Novel methods from natural language processing allow to systematically assess the emotional tone and content of digital written language (Barnes, Klinger, Schulte, & Walde, 2017; Ribeiro, Araújo, Gonçalves, André Gonçalves, & Benevenuto, 2016). Among these methods, sentiment analysis has been advocated as a novel way to assess public opinions towards conservation issues (Drijfhout, Kendal, Vohl, & Green, 2016; Ladle et al., 2016), yet its application in conservation science is still widely unexplored. Existing case studies have used Twitter data to investigate human sentiment on the Great Barrier Reef (Becken, Stantic, Chen, Alaei, & Connolly, 2017) or for tracking public opinion towards conservation-related topics over time (Fink, Hausmann, & Di Minin, 2020). However, to our knowledge, no previous study assessed social media users' perceptions and emotions when visiting national parks.

In this study, we assessed the sentiment and the discourse shared by visitors on social media to understand how they perceive national parks, and what they value during recreational experiences. In order to do this, we analysed the content of picture captions in Instagram posts which were geolocated from inside the borders of Kruger, Addo Elephant, Table Mountain and Garden Route National Parks (NPs), in South Africa, between 2013 and 2016. In particular, we used natural language processing and sentiment analysis approaches to assess (a) what is the sentiment and what are the main emotional components attached to social media posts? and (b) how visitors describe experiences in social media posts across and within parks?

2 | METHODS

2.1 | Study areas

South Africa is well known for its environmental and wildlife conservation efforts, and is known for hosting some of the most iconic national parks world-wide (Carruthers, 2017). However, many national parks share the controversial socio-political history of the country, having originated during the colonial time and having experienced alienation from local people (Carruthers, 2017). Today, the management plans recognize the vision for 'a sustainable national parks system, connecting society' in the way that 'national parks will be the pride and joy of all South Africans' (SANParks, 2006, 2017). Understanding how visitors perceive national parks, and what they value during their visit, is therefore a key aspect needed to fulfil this management vision and promote national parks' role into society.

The study focuses on Kruger NP, Addo Elephant NP, Table Mountain NP and Garden Route NP (Figure S1, Appendix A). Table Mountain NP and Kruger NP are the most popular parks in the country having received, between 2016 and 2017, respectively almost 3.5 and 2 million visitors, while Garden Route NP and Addo Elephant NP received almost 500,000 and 270,000 visitors respectively (SANParks, 2017). These parks were chosen as they are the most visited in South Africa and because social media data were found to match official visitations' statistics (Tenkanen et al., 2017). In addition, according to previous survey conducted in Kruger NP and Table Mountain NP, the majority of both national and international tourists visiting the parks actively used social media platforms to share experiences during their visit, with no statistical differences between the groups (Hausmann et al., 2018). Moreover, these parks were chosen as they have different biological and geographical characteristics, allowing to cover different types of nature-based experiences and compare whether visitors' perceptions might change across parks with different characteristics. In particular, Kruger NP (the largest park, covering 19,623 km^2) and Addo Elephant NP (covering 1,642 km^2) have charismatic megafauna, such as lion *Panthera leo*, elephant *Loxodonta Africana*, leopard *Panthera pardus*, white rhinoceros *Ceratotherium simum* and black rhinoceros *Diceros bicornis*. The opportunity to spot these species in the wild make these parks popular destinations for wildlife watching activities, including self or guided game drives, which are part of a 'safari' experience (Boshoff et al., 2007; Grünewald, Schleuning, & Böhning-Gaese, 2016). In addition, the parks are located in different biomes offering a variety of landscape experiences and aesthetic cultural services. While Kruger NP is located in the Savanna and Thickets biomes, Addo Elephant NP is located in the Fynbos, Forest, Nama-Karoo and the Indian Ocean Coastal Belt biomes (Mucina & Rutherford, 2006). Access to both parks is regulated by official gates and borders are fenced. On the other hand, Table Mountain NP (covering 243 km^2) and Garden Route NP (covering 1,570 km^2) are popular destinations for broader nature-based experiences, including for less-charismatic biodiversity and other outdoor activities, such as hiking, nature walks, mountain biking and water activities (Barendse et al., 2016; Hausmann, Slotow, Fraser, & Di Minin, 2017). They are both located in the Fynbos biome (Mucina & Rutherford, 2006) and access to the parks is open as they are mostly unfenced.

2.2 | Social media data collection and processing

Instagram, with up to 1 billion active users (Chaffey, 2018), is among the most popular platforms world-wide, and one of the most used to share nature-based experiences in South Africa's national parks (Hausmann et al., 2018). We used Instagram's public Application Programming Interface (API) for accessing 1 week/month sample of posts, which were geolocated within the national parks borders, between June 2013 and February 2016. Only publicly available posts were accessed and users were de-identified.

In order to extract relevant information from large volume of social media posts, we used methods from natural language processing, which allowed us to perform automated text mining and turn unstructured textual data into structured data, which can be used for quantitative analysis (Nadkarni, Ohno-Machado, & Chapman, 2011). In particular, in R (R Development Core Team, 2013), we used the packages TM (Feinerer & Hornik, 2018) and TIDYTEXT (Silge & Robinson, 2016) for cleaning the text by keeping only alpha-numeric characters and removing irrelevant features, such as links and usernames embedded in the text, all special characters and English stop words based on a pre-defined list of most common words in English, for example, pronouns. An additional set of words (Table S1, Appendix A), which carried obvious information, such as name of the country, regions and parks or that were not related to the content shared, such as highly used hashtags (#instagood, #nofilter), were also removed. Hashtags, a popular way of communicating in short text language of social media, were included in the analysis as they may carry sentiment and emotional meaning (e.g. #love, #happy; Mohammad, Kiritchenko, & Zhu, 2013). Hashtags created from combined words (e.g. #wildlifephotography) were kept as single words. Then, we used the *fastText* framework in Python 3.7.2 (Joulin, Grave, Bojanowski, & Mikolov, 2017) for identifying posts in English language, as this is among the main official languages used in South Africa and the most popular among internet users world-wide (https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/), and discarded the rest from the analysis.

2.3 | Sentiment analysis

Sentiment analysis is a natural language processing method, increasingly popular in several field of research (see e.g. Mäntylä, Graziotin, & Kuutila, 2018), which allows to automatically analyse opinions and subjectivity expressed in online speech, including personal feelings, beliefs and judgement. Several methods, like for example supervised lexicon-based classification or unsupervised machine learning, have been developed to identify positive, negative or, in case no subjectivity is expressed, neutral sentiment polarity in the text. These methods showed varying prediction performances across different research domains, such as computer science, marketing and psychology (see e.g. Alaei, Becken, & Stantic, 2019; Ribeiro et al., 2016). As no specific method has been identified as best for nature-based tourism yet (Alaei et al., 2019), we chose to apply a general lexicon-based approach where the text is classified based on pre-defined dictionary (lexicon) of words associated with sentiment polarity. In particular, we used the NRC Word-Emotion Lexicon (Mohammad & Turney, 2013), which is openly available from the NRC-Canada sentiment system, and accessible in the SYUZHET package in R (Jockers, 2015). The lexicon contains a list of unigrams (i.e. words) in English language that have been manually annotated and validated through an online crowdsourcing system (see more details in Mohammad & Turney, 2013). We chose this dictionary because the NRC Word-Emotion Lexicon includes words annotated both for sentiment polarity (−1, negative, and +1, positive) and in association with eight classes of basic emotions, including anger, fear, anticipation, trust, surprise, sadness, joy and disgust. These classes allow for a more in depth understanding of the emotional components driving the sentiment in the text (Naldi, 2019). Based on the number of words occurring in the text and their respective values annotated in the lexicon, we obtained a summed value for each sentiment and emotion classes for each social media post (see examples in Table S2, Appendix A). Specifically, the 'sentiment value' of each post was calculated as the number of positive words minus the number of negative words, as annotated in the dictionary. Accordingly, the sentiment polarity of a single post was determined as follows: if the score is >0, the post has an overall 'positive' sentiment, if the score is <0, the post has an overall 'negative' sentiment, if the score = 0, the post is considered to be 'neutral'. Score in each emotion class were calculated as the sum of words assigned to each class according to the dictionary. Words that were not present in the lexicon, such as hashtags from combined words (e.g. #hikingDay, #BigFive), or that carried no sentiment or emotional tone, were annotated as 0. We then calculated average and variance values of sentiment and emotion classes overall across parks and within each park. We used ANOVA ($\alpha = 0.05$) and Tukey post hoc tests to assess statistical differences in average sentiment values, and in combined emotional values, across parks. We did this in order to assess whether visitors perceived parks differently and were more likely to express stronger emotional tone (i.e. higher average values of emotion classes) in social media posts.

In order to assess the accuracy of the chosen automated classification, we manually and independently annotated a random sample of 4,500 posts into three classes of sentiment polarity. Then, we compared manual annotation with predicted classes (divided into positive > 1, neutral = 0 and negative < 1), by calculating *F*1-score measures, which is a weighted mean of 'precision' and 'recall' (Ribeiro et al., 2016). Precision is the ratio of the number of correctly predicted posts in a class with respect to the total number of posts predicted in the same class, including false positive. Recall is the ratio of number of correctly predicted posts in a class with respect to the total number of posts which should have been predicted in the same class, including false negative. Finally, we used a Chi-square (χ^2) test on a balanced sample, which was randomly selected within each park ($n = 1{,}000$ posts), in order to assess whether proportions of sentiment polarity classes, and emotion classes, in each park, differed from average values across all parks.

2.4 | Content analysis

To assess the content of the discourse shared on social media text, we firstly extracted the most frequent unigrams, including single words or hashtags, across all parks and within each park. Frequencies were normalized according to the total amount of words in each park in order to be compared. To measure the diversity of the vocabulary used in each park to describe experiences, we used Simpson's λ Index. To calculate the index we considered unigrams as features and their frequencies as abundance measures. In addition, in order to assess whether visitors describe experiences by using a similar language across different types of parks, we assessed the correlation of frequencies of same unigrams between parks by using the Sperman's rank test. To do this, we only considered those unigrams used at least once within all parks. Moreover, we further explored content shared within each park, by extracting the most frequent compounds of two words or hashtags (bi-grams) in order to identify how words are used in combination. By extracting bi-grams, we were also able to detect names generated from two words, such as the iconic location of 'Nature's Valley'. In addition, in order to assess which aspects of the nature experience were frequently mentioned in relation with positive sentiment, we extracted the most frequent words occurring in posts classified with positive sentiment polarity for each park.

Thereafter, in each park, we further explored main topics of discussion on social media by using topic modelling (Hong & Davison, 2010). Specifically, we used Latent Dirichlet Allocation, an unsupervised algorithm which identifies a potential underlying structure in the data, by detecting 'bags' or groups of words without any specific order probabilistically associated with each other. Compared to word frequency analysis, where words stand alone with a single, often literal, meaning, topic modelling allows to capture the broad meaning, or concept, described by the combination of words. For example, the combination of words 'beautiful', 'lion', and 'safari' may capture aspects used to describe a broader concept of 'wildlife experience'. In this sense, results from the Latent Dirichlet Allocation were used to identify the 'type of experience' described by visitors on social media. To do this, topics were labelled through interpreting the non-predefined themes emerging from the data (McAbee, Landis, & Burke, 2017). As Latent Dirichlet Allocation requires a pre-defined number of topics, we identified the optimal number of topics by assessing the rate of perplexity change as a function of numbers of topics (Zhao et al., 2015). Perplexity is a common measurement used to evaluate how well a statistical model describes a dataset, such as the appropriate number of topics that can describe a text (Zhao et al., 2015). The optimal number is defined by the least number of topics maximizing the information covered as close to the original text as possible. We identified this optimum by assessing the perplexity against an increasing number of assumed topics in the model (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30 and 40). It was seen that perplexity was lowest when the model was trained on a single topic, as can happen in certain practical scenarios when using very small documents (Bian et al., 2015). While a measure of perplexity can be a useful guiding principle, the best judgement is often achieved by human interpretations of resultant topics (Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009).

3 | RESULTS

A total of 33,213 Instagram posts in English were used, corresponding to 93% of the total collected posts. Specifically, 16,545 posts were located in Kruger NP, 11,481 posts in Table Mountain NP, 3,066 posts in Garden Route NP and 2,121 posts in Addo Elephant NP. Average post length was 11 ± 9.83 words, with similar values across parks (AENP 10.64 ± 9.65; GRNP 10.75 ± 9.57; KNP 11.62 ± 10.28; TMNP 10.27 ± 9.21).

3.1 | Sentiment analysis

On average, 18% of words per post carried positive or negative sentiment polarity or emotional tone. Sum of sentiment values ranged from −5 to 12 per post, with an average positive value of 0.517. The length of the post was positively correlated with sentiment score (Spearman's $\rho = 0.35$). Compared to manual classification, automatic sentiment classification was highly accurate with F score = 0.81 (precision = 0.88, recall = 0.75). Average sentiment values were significantly different across parks (F value: 20.96, $df = 3$, $p < 0.001$), with higher values in Garden Route NP and Table Mountain NP compared to the other parks (Figure 1; Table S3 in Appendix A). Across all parks, the most frequent sentiment polarity class was neutral (55%), followed by positive (37%) and negative (8%), with same ranking distribution within each park (Figure 2). However, proportional distribution across polarity classes was significantly different across parks ($\chi^2 = 35.56$, $df = 6$, $p > 0.001$). Compared to average values across all parks, proportion of positive posts was lower in Addo Elephant NP ($\chi^2 = 6.62$, $p < 0.01$) and higher in Garden Route NP ($\chi^2 = 5.32$, $p < 0.05$); proportion of neutrals was higher in Addo Elephant NP ($\chi^2 = 4.72$, $p < 0.01$), and lower in Garden Route NP

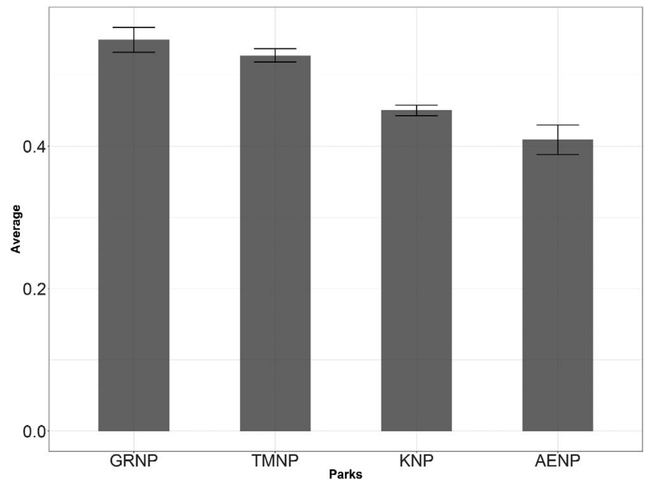

FIGURE 1 Average sentiment values of social media posts within Addo Elephant (AENP), Garden Route (GRNP), Kruger (KNP) and Table Mountain (TMNP) National Parks. Error bars show 95% confidence intervals of the average values

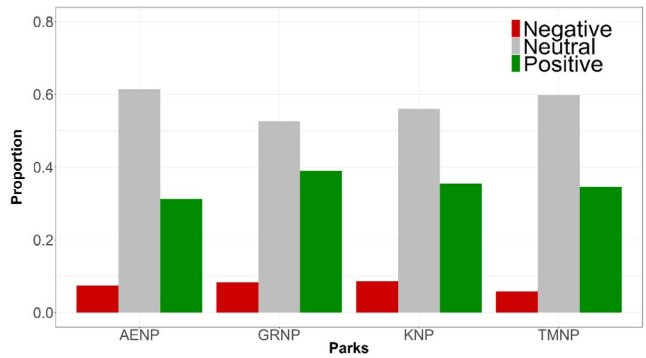

FIGURE 2 Proportions of social media posts in each class of sentiment polarity (negative, neutral, positive) within Addo Elephant (AENP), Garden Route (GRNP), Kruger (KNP) and Table Mountain (TMNP) National Parks

($\chi^2 = 4.03$, $p < 0.05$); and proportion of negative was higher in Kruger NP ($\chi^2 = 6.28$, $p < 0.05$) and lower in Table Mountain NP ($\chi^2 = 5.49$, $p < 0.05$; Figure 2).

Across all parks, posts were mostly expressing joy (43%), followed by anticipation (36%) and surprise (19%) and with same ranking distribution within each park (Figure 3). However, posts in Addo Elephant NP had significantly lower expression of joy ($\chi^2 = 7.92$, $p < 0.01$), anticipation ($\chi^2 = 21.53$, $p < 0.01$) and surprise ($\chi^2 = 4.732$, $p < 0.05$); posts in Garden Route NP had significantly higher expression of joy ($\chi^2 = 13.311$, $p < 0.01$); posts in Kruger NP had significantly higher expression of fear ($\chi^2 = 5.06$, $p < 0.05$) and disgust ($\chi^2 = 6.12$, $p < 0.05$); posts in Table Mountain NP had significantly lower expression of fear ($\chi^2 = 11.97$, $p < 0.01$), and anger ($\chi^2 = 4.95$, $p < 0.05$). Average values of overall emotional tone expressed in the text was also significantly different among parks (F value: 47.44, $df = 3$, $p < 0.001$), and more likely to be higher in Garden Route NP compared to all other parks (Figure 3; Table S3 in Appendix A).

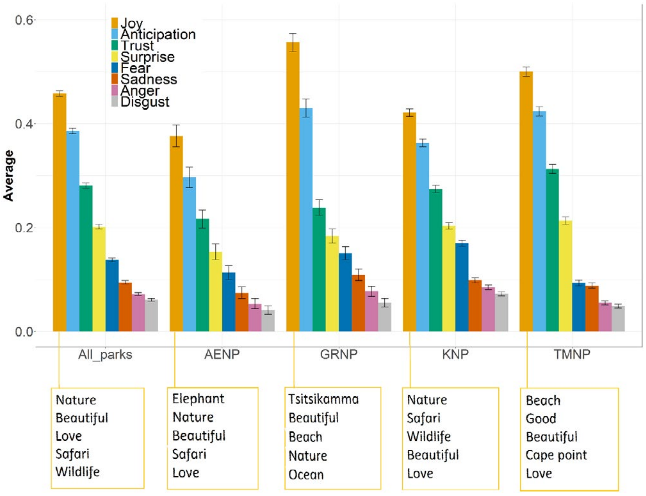

FIGURE 3 Average values of emotional classes expressed in the text of social media posts across all parks and within Addo Elephant (AENP), Garden Route (GRNP), Kruger (KNP) and Table Mountain (TMNP) National Parks. Error bars show 95% confidence intervals of the average values. Boxes show an example of the top five words in social media posts classified as Joy

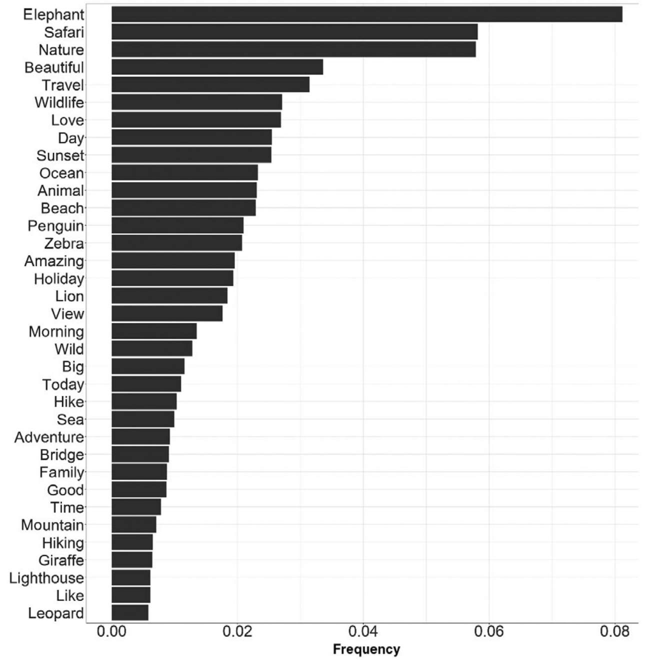

FIGURE 4 Top 35 words/hashtags (excluding place names) used in Instagram post captions between 2013 and 2016 across Addo Elephant (AENP), Garden Route (GRNP), Kruger (KNP) and Table Mountain (TMNP) National Parks

3.2 | Content analysis

Overall, 33,658 different unigrams, including 17,943 words and 15,715 hashtags, were used in social media to describe the parks. Hashtags amounted for 64% of words per post on average. After processing the dataset, the size of the vocabulary for each of the parks was 11,188 in Table Mountain NP, 4,305 in Garden Route NP, 2,969 in Addo Elephant NP and 14,224 in Kruger NP. Across all parks, 'elephant', 'safari' and 'nature' were the most frequent unigrams (Figure 4). The vocabulary used was highly diverse in all parks (Simpson's Index: Addo Eelephant NP $\lambda = 0.952$; Garden Route NP $\lambda = 0.977$; Kruger NP $\lambda = 0.970$; Table Mountain NP $\lambda = 0.973$). However, specific unigrams were used differently between parks (Figure 4). Some unigrams occurred in only one park, such as the word 'penguins' which occurred only in Table Mountain NP. Moreover, some unigrams occurred only in few of the parks, such as 'elephant' and 'safari' occurring only in Addo Elephant NP and Kruger NP, while other unigrams, such as 'nature', occurred in all of the parks. Between parks, vocabulary used was highly correlated across parks (Table S4, Appendix A), especially between Addo Elephant NP and Kruger NP (Spearman's $\rho = 0.69$, $p < 0.001$), and between Garden Route NP and Table Mountain NP (Spearman's $\rho = 0.69$, $p < 0.001$).

Most frequent bi-grams (Table 1) in Table Mountain NP and Garden Route NP were related to activities, geographical attractions and other tourist attractions. For instance, the bi-gram 'otter'—'trail' referred to a popular hiking route in Garden Route NP. Moreover, the bi-gram 'Chapman'—'Peak' referred to the name of an iconic mountain in Table Mountain NP. Furthermore, the bi-gram 'Boulder'—'penguins' referred to the location of a beach hosting a colony of African penguins—*Spheniscus demersus*—in Table Mountain NP. In Addo Elephant NP and Kruger NP, word associations were mostly related to natural features and nature-based activities. For instance, the bi-gram 'elephant'—'wildlife' in Addo Elephant NP referred to natural features in the park, while the bi-gram 'wildlife'—'wildlife-photography' in Kruger NP referred to the activity of photographing wildlife in the park. Moreover, most frequent words used in posts with positive sentiment polarity (Table S5, Appendix A) showed that wildlife-related words, such as 'elephant' or 'safari', and words referring to iconic places, such as 'Tsitsikamma' or 'Cape point', were the most frequent words in Addo Elephant and Kruger NPs, and in Garden Route and Table Mountain NPs respectively. Similarly, most frequent words used in posts expressing joy (Figure 3) highlight the use of words related to broader concepts, including 'nature', 'beautiful' or 'love', together with context-specific features, such as iconic places.

Across all parks, the best number of modelled latent topics of discussion was one (Table S6, Appendix A), showing that 'nature' and its meanings appear to be the predominant topic to describe experiences. Topic modelling revealed a combination of words which were either common to all parks or unique to each park (Table 2). Words common to all parks were related to nature experiences, such as 'nature', 'wilderness' or 'sunset'; activities, including 'travel', 'hiking' or 'safari'; and positive emotions, such as 'beautiful' or 'love'. Topics also included words referring to park-specific features. These were related to species, such as 'penguins', 'lion' or 'elephant'; other nature attractions, including 'Storms river', 'ocean' or 'beach'; and iconic places, such as 'Cape point' or 'Tsitsikamma'. Specifically, Addo Elephant NP and Kruger NP topics were labelled as 'wildlife-experience parks', as words referred mostly to species common names (e.g. elephant, lion) and wildlife-watching

TABLE 1 Top 10 bi-grams occurring in social media post captions across Addo Elephant (AENP), Garden Route (GRNP), Kruger (KNP) and Table Mountain (TMNP) National Parks

AENP	KNP	GRNP	TMNP
Crossing-zebra	Day-safari	Ocean-stormsriver	Ocean-two
Elephant-today	Safari-sunset	Fitfam-fitspo	Capepoint-ilovecapetown
Elephant-wild	Nature-safari	Nature-ocean	Chapman-peak
Animal-wildlife	Nature-sultan	Ocean-sea	Boulder-penguin
Elephant-nature	Wildlife-wildlifephotography	Stormsriver-stormsrivermouth	Mountain-table
Nature-wildlife	Wild-wildlife	Beautiful-nature	Lion-head
Animal-elephant	Animal-wildlife	Otter-trail	Capepoint-lighthouse
Safari-zebra	Safari-wildlife	Mouth-river	Beach-penguin
Elephant-wildlife	Elephant-safari	Nature-valley	Beach-boulder
Elephant-safari	Nature-wildlife	River-storm	Bouldersbeach-penguin

TABLE 2 List of latent topics identified on Instagram within Addo Elephant (AENP), Garden Route (GRNP), Kruger (KNP) and Table Mountain (TMNP) National Parks by using Latent Dirichlet Allocation. Topics were labelled by interpreting themes emerging from the words identified in each park

Wildlife-experience parks		Scenery-experience parks	
AENP	KNP	GRNP	TMNP
Elephant	Safari	Tsitsikamma	Cape point
Safari	Nature	Nature	Penguins
Elephants	Wildlife	Beautiful	Beach
Nature	Elephant	Beach	Ocean
Wildlife	Sunset	Ocean	Travel
Zebra	Animals	Storms river	Good
Animals	Travel	Wilderness	Nature
Travel	Lion	Sunset	View
Beautiful	Wild	River	Beautiful
Eastern Cape	Beautiful	Hiking	Love

activities (e.g. safari). Topics in Garden Route NP and Table Mountain NP, instead, were labelled as 'scenery-experience parks' as words described a broader nature-based experience related to appreciation of sceneries, such as sunset, view or ocean, and iconic places, such as Cape point or Storms river.

4 | DISCUSSION

This study provides an assessment of visitors' attitudes and perceptions of national parks, by using automatic natural language processing of textual content shared on social media during visitation. Overall, we found that the polarity of visitors' sentiment on social media was positive, and was mostly expressing emotions such as joy, anticipation, trust and surprise, with only a small occurrence of posts with negative feelings. In particular, the most frequent aspect used to describe experiences across all parks was appreciation of nature and its components, including species, landscapes, beach, ocean and images and ideas of nature. These findings support and highlight the societal role of national parks in providing visitors with opportunities to develop positive connections with nature (Russell et al., 2013), which can then generate physical, psychological and social benefits (Ament et al., 2016; Hausmann et al., 2016; Hausmann, Slotow, et al., 2017; Puhakka et al., 2017). In addition, our study reveals that user-generated content shared on social media may help understand how visitors share and communicate such experiences in the virtual environment, including the context-specific attributes valued in each park.

While looking at the images shared on social media may help understand visitors' preferences for specific national park features, such as biodiversity groups, landscapes or activities (Hausmann et al., 2018; Heikinheimo et al., 2017), our study suggests that analysing the text content of the posts helps understand what these features may symbolize to visitors. For example, we found that visitors use a diverse set of national park features, including species, activities, geographical landmarks and iconic places, in association with a broader meaning of experiences related to, for example, nature, wilderness, travelling, holidays and adventures. Building on previous research on tourism experience (e.g. Lau, 2011; Lekies & Whitworth, 2011), our study suggests that posts shared in national parks may represent a collection of sights, or attributes of the parks, chosen as symbols and that give meaning to what visitors consider experiences worth sharing. Therefore, content shared on social media may help understand how the image (i.e. people's cognition, affection and attitude; Tasci et al., 2007) of national parks as tourism destinations is reflected by visitors and constructed in the virtual social environment (Hunter, 2016). Since general public perception of the identity and reputation of a tourism destination is shaped by the online appearance of the place (Alaei et al., 2019), social media data may help inform ecotourism and conservation marketing strategies. In particular, the methods applied in this study can be used by managers of national parks to understand the tourism profile of parks, according to physical and psychological dimensions symbolizing a satisfactory experience. In addition, it can facilitate the design of high-quality tourism experiences, which could foster socio-political support for national parks and their long-term conservation effectiveness (McCool, 2006). Analysing visitors' posts may also help detect potential threats to biodiversity, which might be represented and self-reinforced on social media as symbols of a positive experience. These may include taking close-up selfies with wildlife, or identifying emerging tourist hotspots in potentially sensitive areas (Hausmann et al., 2019). An early detection of these threats can help inform the design of targeted interventions, such as awareness campaigns, which may promote positive visitor experiences in line with biodiversity conservation objectives.

The type of nature-based experience described by visitors varies according to park-specific context. We found that, although the language used to describe experiences was highly diverse across all parks, similar words were used in parks with similar characteristics, suggesting that people may tend to perceive parks according to the type of experience they offer. Such experiences referred to watching charismatic wildlife in Kruger NP and Addo Elephant NP, while they focused more on scenery experiences in Garden Route NP and Table Mountain NP. Interestingly, we found that sentiment expressed on social media within Garden Route NP and Table Mountain NP, where topics were mostly related to landscape, iconic places and outdoor activities, was more positive than in Kruger NP and Addo Elephant NP, where topics were mostly related to species and wildlife watching. Charismatic megafauna are considered key attractors driving national and international tourists to national parks and are generally used as flagship for ecotourism and conservation marketing campaigns (Di Minin, Fraser, Slotow, & MacMillan, 2013). We found that charismatic species (e.g. elephant, lion, zebra), in relation to wildlife watching activities (e.g. game driving, safari), were indeed part of most frequent topics shared on social media in parks where the species occur. However, the presence of potentially dangerous animals, and restrictions on some activities for security reasons, such as in the case of independent walking, may explain higher negative emotions, such as fear. On the other hand, parks offering broader nature-based experiences where outdoor activities, such as hiking, are allowed to elicite higher positive feelings in social media users. Nature-based tourism markets in sub-Saharan Africa are not limited to charismatic species (Hausmann, Slotow, et al., 2017; Hausmann, Toivonen, et al., 2017) and these conservation areas can be promoted for the well-being feelings they generate.

While social media data may provide important insights in understanding human–nature interactions (Di Minin et al., 2015), the nature of the data involves limitations, including biases related to geographical, population and visitation representativeness. These include, for instance, that posts may be incorrectly geotagged (Toivonen et al., 2019), that social media is mostly used among younger people (Hausmann et al., 2018) and that social media data are a better proxy of tourists' visitation in more popular parks (Tenkanen et al., 2017). In addition, social media text often includes abbreviations, slang, emojis and hashtags combining multiple words. This short and unconventional language challenges the effectiveness of sentiment predictions when applying automatic computational methods. Future studies may explore the use of machine learning approaches for training sentence-based algorithms, in order to better increase prediction performances (Di Minin, Fink, Hiippala, & Tenkanen, 2019; Di Minin, Fink, Tenkanen, & Hiippala, 2018; Toivonen et al., 2019). Moreover, expressing positive sentiment and emotions is generally perceived as more appropriate than negative feelings when sharing experiences on social media platforms (Waterloo, Baumgartner, Peter, & Valkenburg, 2018), potentially creating an overall positivity bias. Across all social media platforms there is need to quantify this positivity bias, for example, by comparing social media data with other traditional data sources, including surveys, as empirical information is currently lacking. However, by acknowledging this bias, sentiment analysis within the same platform can be implemented to compare relative variations in polarity (e.g. Schwartz, Dodds, O'Neil-Dunne, Danforth, & Ricketts, 2019). Further exploring the effect of national parks in promoting happiness in visitors, may involve comparing sentiment of social media posts from within parks with data shared in different spatio-temporal contexts, such as in urban areas, or before and after visiting parks. In addition, while general purpose sentiment dictionaries built from crowd-sourcing tools may reflect simple general public perceptions, more specific dictionaries may help to better assess opinions of a target population (Alaei et al., 2019), such as national parks visitors. For example, developing a recreation-specific dictionary for national parks, which better considers the emotional meaning of words used to describe experiences, may help improve the accuracy of automated sentiment analysis in such contexts and inform managers for monitoring visitors' experiences over space and time. Finally, the dynamic environment of digital data, with new platforms emerging in popularity or being closed in short times, requires further exploration of the methods used in this study by using alternative sources, including other social media platforms, blogs, and web reviews. This may help to validate our results beyond a single platform and ensure viability of our approach in the future (Fink et al., 2020). Validating social media data with real world information (e.g. survey-based methods, Hausmann et al., 2018), may also help to better understand people's behaviour when sharing experiences on social media while visiting national parks.

In conclusion, our study complements previous research on stakeholders' attitudes towards protected areas (Bragagnolo, Malhado, Jepson, & Ladle, 2016) by revealing that social media data may provide opportunities for understanding how people perceive national parks, including identifying the context-specific aspects sought by visitors. The approach and methods used in this study can be used elsewhere by conservation scientists and managers to understand the online image of national parks constructed by visitors. This could help inform decision-making for enhancing the social value of the parks and to build political support to justify their existence (Chan et al., 2007). However, social media data do not always cover all stakeholders' views, such as people without internet access or not using the platforms. In order to generate a comprehensive understanding of the social impact of national parks, it is important to integrate attitudes of different stakeholders both in the virtual and the real environment, into assessments of management effectiveness. Future studies may help to further investigate the role of social media data to understand the views of various stakeholders, such as both visitors and people living within and in the surroundings of national parks, and which factors may be driving positive sentiment. Moreover, new analytical methods, and the development of computationally efficient algorithms, will provide emerging opportunities to synthesize the growing amount of digital data to the relevant information (Gandomi & Haider, 2015) useful for conservation decision-making (Toivonen et al., 2019). Sentiment analysis, and content of social media, can be further explored to inform conservation science and practice (Drijfhout et al., 2016; Toivonen et al., 2019), including understanding and monitoring people's reactions towards events related to biodiversity conservation (Fink et al., 2020), or controversial topics (e.g. culling, recreational hunting).

ACKNOWLEDGEMENTS

A.H. was supported by Helsinki Institute of Sustainability Science (HELSUS) through a grant allotted to E.D.M. V.H. is indebted to the KONE Foundation for supporting T.T. through a grant. C.F. thanks the University of Helsinki for funding E.D.M., who also acknowledges the assistance provided by the Academy of Finland 2016–2019, via Grant 296524. E.D.M., R.K. and A.H. thank the European Research Council (ERC) for funding under the European Union's Horizon 2020 research and innovation program (grant agreement #802933). Finally, we would like to thank the Editors, P. Jepson and A. Schwartz for comments that helped us improve our manuscript.

CONFLICT OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTIONS

A.H., T.T. and E.D.M. conceived the idea and designed the study; A.H., C.F., V.H. and H.T. collected data; A.H. and R.K. undertook analyses; A.H. and E.D.M. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

R packages and codes used for natural language processing (https://www.tidytextmining.com/index.html) and sentiment analysis (https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html) are openly available online. In order to ensure full protection of users' privacy and compliance with General Data Protection Regulation (GDPR 2016/679), raw data used in the study cannot be made publicly available. Summaries of data are available into a Dryad online repository linked to this manuscript (https://doi.org/10.5061/dryad.cjsxksn3h). Data may be granted by request to the corresponding author, with permission of all parties involved with the research and in compliance with GDPR requirements.

ORCID

Anna Hausmann https://orcid.org/0000-0002-9639-9532
Tuuli Toivonen https://orcid.org/0000-0002-6625-4922
Christoph Fink https://orcid.org/0000-0003-1251-9726
Vuokko Heikinheimo https://orcid.org/0000-0001-5119-0957
Ritwik Kulkarni https://orcid.org/0000-0002-1320-9693
Henrikki Tenkanen https://orcid.org/0000-0002-0918-4710
Enrico Di Minin https://orcid.org/0000-0002-5562-318X

REFERENCES

Abraham, A., Sommerhalder, K., & Abel, T. (2010). Landscape and well-being: A scoping study on the health-promoting impact of outdoor environments. *International Journal of Public Health*, *55*, 59–69. https://doi.org/10.1007/s00038-009-0069-z

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, *58*, 175–191. https://doi.org/10.1177/0047287517747753

Ament, J. M., Moore, C. A., Herbst, M., & Cumming, G. S. (2016). Cultural ecosystem services in protected areas: Understanding bundles, trade-offs and synergies. *Conservation Letters*, *10*, 440–450. https://doi.org/10.1111/conl.12283

Barendse, J., Roux, D., Erfmann, W., Baard, J., Kraaij, T., & Nieuwoudt, C. (2016). Viewshed and sense of place as conservation features: A case study and research agenda for South Africa's national parks. *Koedoe – African Protected Area Conservation and Science*, *58*, 1–16. https://doi.org/10.4102/koedoe.v58i1.1357

Barnes, J., Klinger, R., Schulte, S., & Walde, I. (2017). Assessing state-of-the-art sentiment models on state-of-the-art sentiment datasets. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 2–12). Copenhagen, Denmark, September. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-5202

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. *Journal of Environmental Management*, *203*, 87–97. https://doi.org/10.1016/j.jenvman.2017.07.007

Bian, J., Yoshigoe, K., Hicks, A., Yuan, J., He, Z., Xie, M., … Modave, F. (2016). Mining twitter to assess the public perception of the 'internet of things'. *PLoS ONE*, *11*, 1–14. https://doi.org/10.1371/journal.pone.0158450

Boshoff, A. F., Landman, M., Kerley, G. I. H., & Bradfield, M. (2007). Profiles, views and observations of visitors to the Addo Elephant National Park, Eastern Cape, South Africa. *South African Journal of Wildlife Research*, *37*, 189–196. https://doi.org/10.3957/0379-4369-37.2.189

Bragagnolo, C., Malhado, A. M., Jepson, P., & Ladle, R. (2016). Modelling local attitudes to protected areas in developing countries. *Conservation and Society*, *14*, 163. https://doi.org/10.4103/0972-4923.191161

Carruthers, J. (2017). *National park science: A century of research in South Africa (ecology, biodiversity and conservation)*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781108123471

Chaffey, C. (2018). *Global social media research summary 2018|Smart Insights*. Retrieved from https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Chan, K. A. I. M. A., Pringle, R. M., Ranganathan, J. A. I., Boggs, C. L., Chan, Y. L., Ehrlich, P. R., … Macmynowski, D. P. (2007). When agendas collide: Human welfare and biological conservation. *Conservation Biology*, *21*, 59–68. https://doi.org/10.1111/j.1523-1739.2006.00570.x

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems* (pp. 288–296). Cambridge, MA: The MIT Press.

Di Minin, E., Fink, C., Hiippala, T., & Tenkanen, H. (2019). A framework for investigating illegal wildlife trade on social media with machine learning. *Conservation Biology*, *3*, 210–213. https://doi.org/10.1111/cobi.13104

Di Minin, E., Fink, C., Tenkanen, H., & Hiippala, T. (2018). Machine learning for tracking illegal wildlife trade on social media. *Nature Ecology & Evolution*, *2*, 406–407. https://doi.org/10.1038/s41559-018-0466-x

Di Minin, E., Fraser, I., Slotow, R., & MacMillan, D. C. (2013). Understanding heterogeneous preference of tourists for big game species: Implications for conservation and management. *Animal Conservation*, *16*, 249–258. https://doi.org/10.1111/j.1469-1795.2012.00595.x

Di Minin, E., MacMillan, D. C., Goodman, P. S., Escott, B., Slotow, R., & Moilanen, A. (2013). Conservation businesses and conservation planning in a biological diversity hotspot. *Conservation Biology*, *27*, 808–820. https://doi.org/10.1111/cobi.12048

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*, 1–6. https://doi.org/10.3389/fenvs.2015.00063

Dirzo, R., Young, H., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the anthropocene. *Science*, *345*, 401–406. https://doi.org/10.1126/science.1251817

Drijfhout, M., Kendal, D., Vohl, D., & Green, P. T. (2016). Sentiment analysis: Ready for conservation. *Frontiers in Ecology and the Environment*, *14*, 525–526. https://doi.org/10.1002/fee.1435

Dudley, N. (Ed.) (2008). *Guidelines for applying protected area management categories*. Gland, Switzerland: IUCN.

Eagles, P., & McCool, S. (2002). *Tourism in national parks and protected areas: Planning and management*. Wallingford, UK: CABI Publishing.

Eklund, J., & Cabeza, M. (2017). Quality of governance and effectiveness of protected areas: Crucial concepts for conservation planning. *Annals of the New York Academy of Sciences*, *1399*, 27–41. https://doi.org/10.1111/nyas.13284

Feinerer, I., & Hornik, K. (2018). *tm: Text mining package*. R package version 0.7-6. Retrieved from https://CRAN.R-project.org/package=tm

Fink, C., Hausmann, A., & Di Minin, E. (2020). Online sentiment towards iconic species. *Biological Conservation*, *241*. https://doi.org/10.1016/j.biocon.2019.108289

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gissibl, B., Höhler, S., & Kupper, P. (2012). *Civilizing nature: National parks in global historical perspective*. New York, NY: Berghahn Books.

Golden Kroner, R. E., Qin, S., Cook, C. N., Krithivasan, R., Pack, S. M., Bonilla, O. D., … Mascia, M. B. (2019). The uncertain future of protected lands and waters. *Science*, *364*, 881–886. https://doi.org/10.1126/science.aau5525

Govers, R., & Go, F. M. (2008). Deconstructing destination image in the information age. *Information Technology & Tourism*, *6*, 13–29. https://doi.org/10.3727/109830503108751199

Grünewald, C., Schleuning, M., & Böhning-Gaese, K. (2016). Biodiversity, scenery and infrastructure: Factors driving wildlife tourism in an African savannah national park. *Biological Conservation*, *201*, 60–68. https://doi.org/10.1016/j.biocon.2016.05.036

Gusset, M., Maddock, A. H., Gunther, G. J., Szykman, M., Slotow, R., Walters, M., & Somers, M. J. (2008). Conflicting human interests over the re-introduction of endangered wild dogs in South Africa. *Biodiversity and Conservation*, *17*, 83–101. https://doi.org/10.1007/s10531-007-9232-0

Hausmann, A., Slotow, R., Burns, J. K., & Di Minin, E. (2016). The ecosystem service of sense of place: Benefits for human well-being and biodiversity conservation. *Environmental Conservation*, *43*, 117–127. https://doi.org/10.1017/S0376892915000314

Hausmann, A., Slotow, R., Fraser, I., & Di Minin, E. (2017). Ecotourism marketing alternative to charismatic megafauna can also support biodiversity conservation. *Animal Conservation*, *20*, 91–100. https://doi.org/10.1111/acv.12292

Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Tenkanen, H., Butchart, S. H. M., … Di Minin, E. (2019). Assessing global popularity and threats to important bird and biodiversity areas using social media data. *Science of the Total Environment*, *683*, 617–623. https://doi.org/10.1016/j.scitotenv.2019.05.268

Hausmann, A., Toivonen, T., Heikinheimo, V., Tenkanen, H., Slotow, R., & Di Minin, E. (2017). Social media reveal that charismatic species are not the main attractor of ecotourists to sub-Saharan protected areas. *Scientific Reports*, *7*, 763. https://doi.org/10.1038/s41598-017-00858-6

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. *Conservation Letters*, *11*, e12343. https://doi.org/10.1111/conl.12343

Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, *6*, 85. https://doi.org/10.3390/ijgi6030085

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. In P. Melville, J. Leskovec, & F. Provost (Eds.), *Proceedings of the First Workshop on Social Media Analytics – SOMA '10* (pp. 80–88). New York, NY: ACM Press.

Hunter, W. C. (2016). The social construction of tourism online destination image: A comparative semiotic analysis of the visual representation of Seoul. *Tourism Management*, *54*, 221–229. https://doi.org/10.1016/j.tourman.2015.11.012

Jockers, M. (2015). *Syuzhet: Extract sentiment and plot arcs from text*. Retrieved from https://github.com/mjockers/syuzhet

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Short papers* (pp. 427–431). Valencia, Spain: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/E17-2068

Kil, N., Holland, S. M., Stein, T. V., & Ko, Y. J. (2012). Place attachment as a mediator of the relationship between nature-based recreation benefits and future visit intentions. *Journal of Sustainable Tourism*, *20*, 603–626. https://doi.org/10.1080/09669582.2011.610508

Ladle, R. J., Correia, R. A., Do, Y., Joo, G. J., Malhado, A. C. M., Proulx, R., … Jepson, P. (2016). Conservation culturomics. *Frontiers in Ecology and the Environment*, *14*, 269–275. https://doi.org/10.1002/fee.1260

Lau, R. W. (2011). Tourist sights as semiotic signs: A critical commentary. *Annals of Tourism Research*, *38*, 711–714. https://doi.org/10.1016/j.annals.2010.11.002

Lekies, K. S., & Whitworth, B. (2011). Constructing the nature experience: A semiotic examination of signs on the trail. *American Sociologist*, *42*, 249–260. https://doi.org/10.1007/s12108-011-9129-y

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*, *22*, 8–15. https://doi.org/10.1080/13683500.2017.1343807

Maller, C., Townsend, M., Leger, L. S., Henderson-Wilson, C., Pryor, A., Prosser, L., & Moore, M. (2010). Healthy parks, healthy people: The health benefits of contact with nature in a park context. *The George Wright Forum*, *26*, 51–83.

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis – A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, 16–32. https://doi.org/10.1016/j.cosrev.2017.10.002

McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, *27*, 277–290. https://doi.org/10.1016/j.hrmr.2016.08.005

McCarthy, D. P., Donald, P. F., Scharlemann, J. P. W., Buchanan, G. M., Balmford, A., Green, J. M. H., … Butchart, S. H. M. (2012). Financial

costs of meeting global biodiversity conservation targets: Current spending and unmet needs. *Science*, *338*, 946–949. https://doi.org/10.1126/science.1229803
McCool, S. F. (2006). Managing for visitor experiences in protected areas: Promising opportunities and fundamental challenges. *Parks: The International Journal for Protected Areas Managers*, *16*, 3–9.
Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Washington, DC: Island Press.
Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In S. Manandhar, & D. Yuret (Ed.), *Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises SemEval-2013* (pp. 321–327). Atlanta, GA: Association for Computational Linguistics.
Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *23*, 236–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x
Mucina, L., & Rutherford, M. C. (2006). *The vegetation of South Africa, Lesotho and Swaziland*. Strelitzia 19. Pretoria: South African National Biodiversity Institute. ISBN: 978-1919976-21-1
Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*, *18*, 544–551. https://doi.org/10.1136/amiajnl-2011-000464
Naldi, M. (2019). A review of sentiment computation methods with R packages. arXiv:1901.08319.
Puhakka, R., Pitkänen, K., & Siikamäki, P. (2017). The health and well-being impacts of protected areas in Finland. *Journal of Sustainable Tourism*, *25*, 1830–1847. https://doi.org/10.1080/09669582.2016.1243696
R Development Core Team. (2013). *R: The R Project for Statistical Computing. R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/
Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench – A benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*, 1–29. https://doi.org/10.1140/epjds/s13688-016-0085-1
Russell, R., Guerry, A. D., Balvanera, P., Gould, R. K., Basurto, X., Chan, K. M. A., … Tam, J. (2013). Humans and nature: How knowing and experiencing nature affect well-being further. *Annual Review of Environment and Resources*, *38*, 473–502. https://doi.org/10.1146/annurev-environ-012312-110838
SANParks (South African National Parks). (2006). Policy context: SANParks' mandate and values, Chapter 2. In *A Framework for Developing and Implementing Management Plans for South African National Parks*. Pretoria, South Africa: South African National Park.
SANParks (South African National Parks). (2017). *South African National Parks Annual Report 2016/17*. Retrieved from https://www.sanparks.org/assets/docs/general/annual-report-2017.pdf
Schwartz, A. J., Dodds, P. S., O'Neil-Dunne, J. P. M., Danforth, C. M., & Ricketts, T. H. (2019). Visitors to urban greenspace have higher sentiment and lower negativity on Twitter. *People and Nature*, *1*, 476–485. https://doi.org/10.1002/pan3.10045
Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *The Journal of Open Source Software*, *1*. https://doi.org/10.21105/joss.00037
Soga, M., Gaston, K., Yamaura, Y., Kurisu, K., Hanaki, K., Soga, M., … Hanaki, K. (2016). Both direct and vicarious experiences of nature affect children's willingness to conserve biodiversity. *International Journal of Environmental Research and Public Health*, *13*, 529. https://doi.org/10.3390/ijerph13060529
Tasci, A. D. A., Gartner, W. C., & Cavusgil, S. T. (2007). Conceptualization and operationalization of destination image. *Journal of Hospitality & Tourism Research*, *31*, 194–223. https://doi.org/10.1177/1096348006297290
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. *Scientific Reports*, *7*, 17615. https://doi.org/10.1038/s41598-017-18007-4
Terraube, J., Fernández-Llamazares, Á., & Cabeza, M. (2017). The role of protected areas in supporting human health: A call to broaden the assessment of conservation outcomes. *Current Opinion in Environmental Sustainability*, *25*, 50–58. https://doi.org/10.1016/j.cosust.2017.08.005
Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., … Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298–315. https://doi.org/10.1016/j.biocon.2019.01.023
van Zanten, B. T., van Berkel, D. B., Meetemeyer, R. K., Smith, J. W., Tieskens, K. F., & Vergurg, P. H. (2016). Continental scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 12974–12979. https://doi.org/10.1073/pnas.1614158113
Velarde, M. D., Fry, G., & Tveit, M. (2007). Health effects of viewing landscapes – Landscape types in environmental psychology. *Urban Forestry & Urban Greening*, *6*, 199–212. https://doi.org/10.1016/j.ufug.2007.07.001
Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, *20*, 1813–1831. https://doi.org/10.1177/1461444817707349
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
Willemen, L., Cottam, A. J., Drakou, E. G., & Burgess, N. D. (2015). Using social media to measure the contribution of red list species to the nature-based tourism potential of African protected areas. *PLoS ONE*, *10*, e0129785. https://doi.org/10.1371/journal.pone.0129785
Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspective*, *10*, 27–36. https://doi.org/10.1016/j.tmp.2014.01.001
Zhang, W., Goodale, E., & Chen, J. (2014). How contact with nature affects children's biophilia, biophobia and conservation attitude in China. *Biological Conservation*, *177*, 109–116. https://doi.org/10.1016/j.biocon.2014.06.011
Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*, *16*, S8. https://doi.org/10.1186/1471-2105-16-S13-S8

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hausmann A, Toivonen T, Fink C, et al. Understanding sentiment of national park visitors from social media data. *People Nat*. 2020;00:1–11. https://doi.org/10.1002/pan3.10130