

National Parks in Twitter: A German-speaking perspective

Maria Teresa Pellicer-Chenoll, Karolina Taczanowska, Pilar Serra-Añó, José Luis Toca-Herrera & Luis-Millán González

Keywords: *Twitter, German language, national park, protected area, mountains, social media, visitor monitoring, Covid-19*

Abstract

Mining data from social media platforms has become increasingly popular to explore aspects of human behaviour, including attitudes towards the natural environment or visiting protected areas. Most studies and analytical algorithms refer to digital content published in English. However, it is also useful to conduct research in other languages to complement existing international studies. Our main aim was to explore Twitter content on national parks, published between 2006 and July 2021, in German. The study also presents a differentiated analysis for tweets published in 2019 and 2020 on national parks and associated with the Covid-19 pandemic. The tweets came from German-speaking countries, but also other countries worldwide. The most frequently mentioned national parks were located mainly in mountain areas, yet terms, hashtags, emojis and topics directly relating to mountains were rare in comparison to other subjects. Tweets most frequently included words such as forest (*Wald*), holiday (*Urlaub*) and nature (*Natur*); messages related not only to the natural heritage and environmental protection but also to natural disasters. The Covid-19 pandemic and national parks were also a subject of discussion on Twitter, often accompanied by photographs or videos. As 85% of all the tweets studied were never retweeted, 92% never received a reply, and 74% were never assigned *likes*, we conclude that there is potential to improve (social media) communications by users interested in protected areas in mountainous regions.

Introduction

The managers of national parks (NPs) and the scientific community aim to understand the reasons underlying choices of travel destination, patterns of recreational use, human-environment interactions during visits to the NPs, as well as associated pre- and post-travel experiences. The methods traditionally used in recreation research consist of direct and indirect observational studies, on-site and online interviews and mail surveys, visitor tracking by GPS or mobile phone, or the use of administrative data such as tickets sold or entry permits issued (Bielanski et al. 2018; Cessford & Muhar 2003; Hartmann 1988).

In recent years, social networks have gained importance, providing alternative sources of data related to the use of protected areas (Di Minin et al. 2015; Ghermandi & Sinclair 2019; Sloan & Quan-Haase 2017; Teles da Mota & Pickering 2020; Toivonen et al. 2019), and the opportunity to obtain information at a minimal cost in time and resources (Li et al. 2019). Information from social media has been used, for example, to estimate visiting rates, spatial patterns of park use, visitor preferences, feelings and experiences, or to explore cultural ecosystem services (Wilkins et al. 2021).

Social media platforms, including Flickr, Twitter and Instagram, are commonly used as data sources (Tenkanen et al. 2017). Flickr and Instagram content is heavily image-based, while Twitter disseminates short text messages (tweets). Specific tools are required to obtain data published on these platforms (e.g. Application Programming Interfaces: APIs), and to analyse it (e.g. text-mining, computational statistics or machine learning) (Batrincea & Treleaven 2015) wikis, re-

ally simple syndication feeds, blogs, newsgroups, chat and news feeds. For completeness, it also includes introductions to social media scraping, storage, data cleaning and sentiment analysis. Although principally a review, the paper also provides a methodology and a critique of social media tools. Analyzing social media, in particular Twitter feeds for sentiment analysis, has become a major research and business activity due to the availability of web-based application programming interfaces (APIs). Such techniques are already used for analysis in the tourism sector (Bucur 2015; Giglio et al. 2020; Kalvet et al. 2020), in NP tourism in particular (Heikinheimo et al. 2018; Mangachena & Pickering 2021; Udyapuram & Gavirneni 2019).

Although social media allow communications in different languages, most research focuses on tweets published in English (Mangachena & Pickering 2021; Pickering & Norman 2020). In addition, much of the progress made on Natural Language Processing focuses on English. However, many tools which include other languages have recently been developed (Litvak & Vanetik 2019; Zierke n.d.).

From our point of view, accepting the axiom that English is the working language of all the citizens of the world is to accept that there is no diversity in the way people express their emotions, desires and concerns in other languages. Indeed, there is a population bias in the analyses conducted in English in non-English speaking countries: because not everybody knows or regularly uses English to communicate, a large part of the population is under-represented in the studies. It is thus useful to carry out research in languages other than English to complement existing international studies related to NPs (Teles da Mota &

Pickering 2021). In this study, we focus on German tweets on NPs, since this language is widely spoken in Europe (Data Europa EU 2012), German-speaking countries are among the leading EU economies (European Commission 2017), and the German-speaking population is among the most active nature-oriented tourists in the world (Starosta et al. 2019).

In this exploratory study of the German-language content on Twitter related to NPs, we address the following research questions (RQ):

RQ1: How many tweets in German are posted about NPs, and what words are used in them?

RQ2: Which are the most frequently mentioned NPs, and how do the users' interests in them vary over time?

RQ3: What emotions or ideas do the users express through emojis, and what people or organizations do they usually mention?

RQ4: What are the main topics of interest of German-speaking users posting tweets about NPs?

Due to the restrictions associated with the Covid pandemic, we also conducted a specific analysis of this topic between 2019 (pre-Covid) and 2020 (during the pandemic). Consequently, our final research question would be the following:

RQ5: How has the Covid-19 pandemic affected tweets about NPs posted in German?

Data and methods

Data retrieval and pre-processing

To build our database, we wrote a Matlab script (the simplest type of program file, which can be used for automating a series of commands) to perform queries through the Twitter search API v.2 (full-archive search).

The search focused on retrieving German-language tweets containing the term *national park* (*Nationalpark* in German) or variations of it (*Nationalparks*, *Nationalparke* and *Nationalparcken*). We also retrieved conversations in which the users expressed their ideas about the original tweet. The search was conducted on 9 July 2021. More than 200,000 original tweets were retrieved. After removing retweets and duplicates, using the unique tweet identifiers, the final corpus comprised 144,126 tweets.

The structure of the tweet along with associated data were stored in json format in order to allow the fields of interest of the study to be extracted. The files contained information on the tweet and its creator. Of more than 150 possible attributes, we used just 11. The standard recommendations used in similar studies (Jianqiang & Xiaolin 2017) were followed in preparing the tweet texts for further analysis. Each tweet was reduced to tokens (tokenization) – i. e. a string of characters representing a unit of text data (also known as a *unigram*) such as a word, number or email address.

The following actions were then performed on the tokens:

1. removing all hyperlinks (*http://url*), hashtags (*#hashtag*), emojis and username links (*@username*) in the tweets. The emojis and hashtags were stored for later analysis;
2. punctuation marks and special characters were removed;
3. all letters were converted to lowercase;
4. words that could add noise to the text and did not add content to the tweets (e.g. the German pronouns *der*, *die* and *das*) were removed using the German stopword list in Matlab's default text analytics toolbox;
5. the words were normalized using the Porter stemmer algorithm (German stemming; see Braschler & Ripplinger 2004) to reduce words (e.g. noun, adjective, verb or adverb) to their root forms.
6. finally, any words of fewer than 2 or more than 50 characters and those occurring only once in the corpus were also deleted.

Three bags of words were formed from the resulting tokens (unigrams: one token; bi-grams: two tokens in succession; tri-grams: three tokens in succession). The original documents (raw data) and associated fields were also stored for further analysis. The remaining fields containing text (e.g. username, user mentions, place...) were not pre-processed in any way.

Although strict duplicates were eliminated, a large number of tweets varied only slightly from others. For this reason, once the documents were cleaned, tweets with minor variations of the original text (e.g. presence or absence of urls) were removed. This identified the original tweets and not the number of tweets sent from one place to another. The IDs of the tweets used in our analysis can be found in Supplementary Material 1. To comply with Twitter's Terms of Service, we are publicly releasing the tweet IDs of the collected tweets only. The data are released for non-commercial research use. Users who wish to reuse our IDs can retrieve the original data using appropriate software (e.g. hydrator).

Descriptive analysis of tweets and national park dynamics over time

Tweet analysis began with a description of the pre-processed tweets and their date of creation. A count identified the number of tweets per day and quantified the tweets with the highest number of retweets, replies and likes over the period selected.

The user location field was used to find the approximate location of the Twitter users. This field is usually composed of the city and the country, separated by a comma (e.g. Hannover, Germany). The tweeter's geolocation was then identified using the nominatim 3.7.2 API, which uses the OpenStreetMap (OSM) dataset. Only when the city and country matched was the result accepted. 1,770 locations were geolocated in this way.

A frequency analysis of the main n-grams was also carried out. An n-gram is a sub-sequence of n ele-

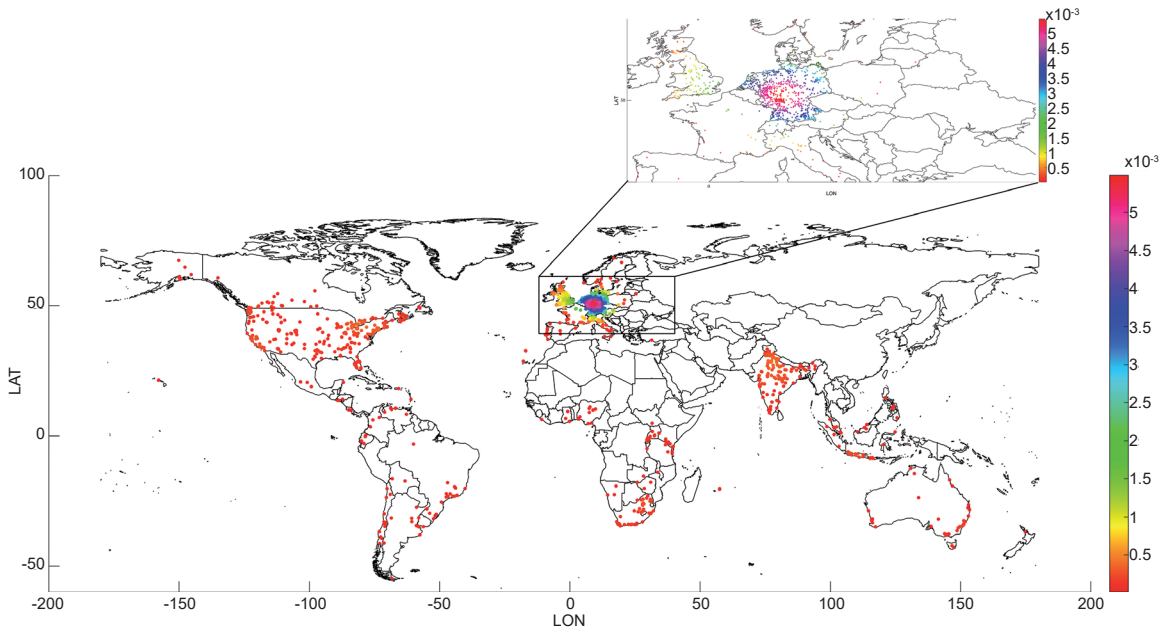


Figure 1 – Location reported by users ($n = 1,770$). The dots represent the geolocations of the Twitter users estimated through the user's location field. The colour bar shows the kernel density of each location.

ments of a given sequence of words. The frequency values of the main uni-grams, bi-grams and tri-grams were represented in word clouds. The bag of bi-grams was used to manually locate those NPs which were mentioned in the tweets a high number of times. To establish the growth and decline of the occurrence of the most-cited NPs throughout the study period, the NPs were represented in a heatmap. All the frequency values were normalized by means of a z-score.

The number of occurrences of words relating directly to Covid-19 that appeared during 2020 was also established.

Emojis

Emojis (images or pictograms that express ideas, emotions or feelings) were extracted from the text of the tweets and stored separately. A count was made of all of them, and those that were repeated most often were represented graphically (using the font Symbola.ttf for unicode symbols). We also looked for the most common meanings given by the users using the Full Emoji List, v13.0, which can be retrieved from the web unicode.org and emojiopedia.org. Emoticons (representations of facial expressions using keyboard characters such as punctuation marks) were not analysed, in order to facilitate text pre-processing (i. e. elimination of punctuation marks and special characters).

Most active users and building networks of user mentions

Certain users presented higher numbers of relevant outputs than others. The top producers were identified using the ID field and the total number of tweets from each of their individual accounts. The profiles of the top 20 producers were analysed.

Tweets often refer to other users or entities by including hyperlinks (e.g. hashtags, urls or usernames) or sources of information. We used the user mentions field to build a directed network in which the nodes were the users and the arcs were the number of times a user mentioned another user. Since most nodes in the network have only a few connections to other nodes, the subnetwork with the largest number of components (i. e. nodes) was represented.

The values of in-degree centrality (i. e. the number of times a node is mentioned by another node) were calculated, (Baek et al. 2022), as was pagerank, which assigns the relevance of a given node within the network, with higher values expressing higher relevance within the network (Borodin et al. 2005).

Topic modelling

To identify the topics that emerged from the corpus, a Latent Dirichlet Allocation (LDA) model was applied (Blei et al. 2003). This discovers underlying topics in a collection of documents and infers the word probabilities in the topics. For a more technical explanation of how the model works, see Büschken & Allenby (2016).

The LDA model fulfils a double function: i) it extracts the main topics that the users find interesting; ii) it serves as a method for selecting the tweets most closely related to certain topics of interest.

To carry out the analysis, the function *filllda.m* implemented in the Matlab text analytics toolbox was used on the pre-processed bag of words (uni-grams). In this case, the words *nationalpark*, *nationalparks*, *nationalparke* and *nationalparken* were removed to avoid distorting the results. Once the bag of words was ready, establishing the number of topics needed was

Table 1 – Retweets, replies and likes.

Total tweets Retweets (count)	66,380
never retweeted (%)	85.38
retweeted > 10 times (%)	0.55
mean retweeted (retweets/tweets)	0.46
Total tweets Replies (count)	18,743
never replied to (%)	92.15
replies > 10 replies (%)	0.07
mean replies (replies/tweets)	0.13
Total tweets Likes (count)	251,933
never liked (%)	73.45
like > 10 likes (%)	2.58
mean (likes/tweets)	1.74

the first step to ensure good cohesion in the resulting topics (González et al. 2021). After fixing the number at 30, a further calculation was used to select the most representative tweets on each topic. The tweets that contained a probability equal to or greater than 0.8 within a topic were selected. A qualitative analysis was carried out with these documents in order to examine them in more depth.

Results

General user data and tweets retrieved

After removing duplicates, a total of 144,126 tweets were retained. On average, the tweets comprised 131 characters. A total of 7,915 tweets included photos ($\approx 5.3\%$) or animated gifs ($>0.2\%$), while most videos were included in mentions of other entities (e.g. YouTube).

The tweets were posted by 64,929 users. Figure 1 shows the user distribution by country in which they reported being on their Twitter account. Most users

were in German-speaking countries, although a considerable number were scattered all over the world.

Statistical data on the number of retweets, replies and likes that the tweets obtained are presented in Table 1.

Hashtags, words and most frequently mentioned national parks in tweets

Figure 2 shows the main hashtags and n-grams posted over the study period. The actual words used as hashtags or uni-grams in the search were removed from the wordclouds to facilitate representation. The most frequent hashtags were #nationalparkservice (2,950), #travel (2,840), #nationalparktour (2,651), #schwarzwald (2,508) and #yosemite (2,505). Although the tweets were written in German, some users used English words in their hashtags, for example #photography (1,878).

In the uni-grams we found that the word *wald* (forest) was associated with the name of a NP (i.e. Bayerischer Wald NP), dominating the other words with a total of 12,938 occurrences. The words *urlaub* (holiday), *natur* (nature) and *neu* (new) also appeared frequently (9,514, 9,368 and 7,185 times respectively). For a detailed analysis, see Supplementary Material 2.

Words related to the environment or nature conservation were particularly important. Here we give those that appeared more than 1,000 times: *naturerb* (natural heritage; 2,392 occurrences); *schutz* (protection; 1,611); *naturschutz* (nature conservation; 1,547); *weltnaturerb* (world natural heritage; 1,144); *waldbrand* (forest fire; 1,431); *umwelt* (environment; 1,002). (Supplementary Material 3 contains several tweet quotations, illustrating the context of the messages).

Table 2 summarizes the frequency of NP names in the tweets. The top three are in Germany (Bayerischer

Table 2 – National parks (NP) that were named by Twitter users more than 1,000 times. US = United States of America

Name	URL	Counts	Country	Mountain area
Bayerischer Wald NP	https://www.nationalpark-bayerischer-wald.bayern.de/	5,860	Germany	yes
Hohe Tauern NP	https://hohetauern.at/de/	4,219	Austria	yes
Yosemite NP	https://www.nps.gov/yose/index.htm	3,968	US	yes
Schwarzwald NP	https://www.nationalpark-schwarzwald.de/de	3,331	Germany	yes
Harz NP	https://www.nationalpark-harz.de/de/startseite/	3,166	Germany	yes
Eifel NP	https://www.nationalpark-eifel.de/de/	3,030	Germany	yes
Krka NP	http://www.np-krka.hr/en/	2,487	Croatia	yes
Kruger NP	https://www.krugerpark.co.za/	1,855	South Africa	yes
Grand Teton NP	https://www.nps.gov/grte/index.htm	1,779	US	yes
Banff NP	https://www.pc.gc.ca/en/pn-np/ab/banff	1,736	Canada	yes
Zion NP	https://www.nps.gov/zion/index.htm	1,670	US	yes
Wattenmeer NP	https://www.nationalpark-wattenmeer.de/	1,602	Germany	no
Berchtesgaden NP	https://www.nationalpark-berchtesgaden.bayern.de/index.htm	1,568	Germany	yes
Sächsische Schweiz NP	https://www.nationalpark-saechsische-schweiz.de/	1,485	Germany	yes
Yellowstone NP	https://www.nps.gov/yell/index.htm	1,484	US	yes
Hainich NP	https://www.nationalpark-hainich.de/	1,387	Germany	yes
Jasmund NP	https://www.nationalpark-jasmund.de/	1,033	Germany	yes
Hunsrück-Hochwald NP	https://www.nationalpark-hunsrueck-hochwald.de/	1,027	Germany	yes
Jim Corbett NP	https://www.corbettnationalpark.in/	1,007	India	yes

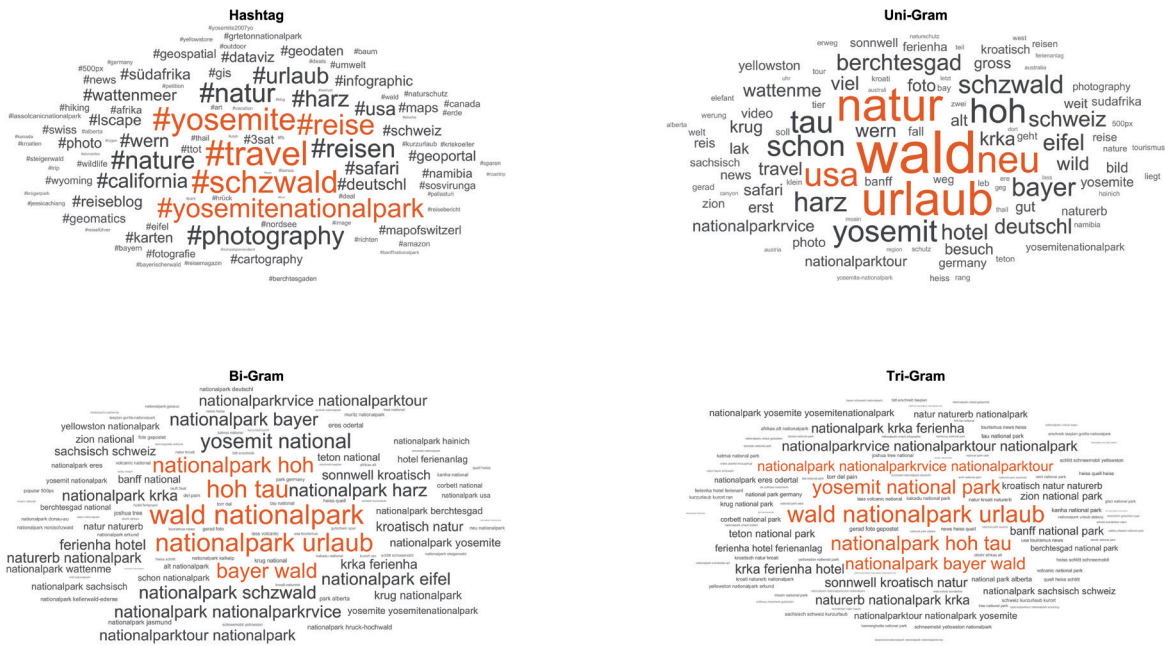


Figure 2 – Word clouds of the most frequent words and hashtags in tweets about national parks.

Wald NP), Austria (Hohe Tauern NP) and the USA (Yosemite NP). The users' interest in various NPs evolved over the years (Figure 3). Before 2011, Hohe Tauern, Yosemite, Eifel, Wattenmeer, Berchtesgaden and Hunsrück-Hochwald NPs were the most popular. In the last ten years, Bayerischer Wald, Hohe Tauern,

Schwarzwald, Harz and Hunsrück-Hochwald NPs had a larger number of tweets.

Interestingly, the most frequently mentioned NPs (>1,000 tweets) were mostly located in mountain areas (18 out of 19 NPs). Yet, tweet text relating to mountains generally was fairly rare in comparison

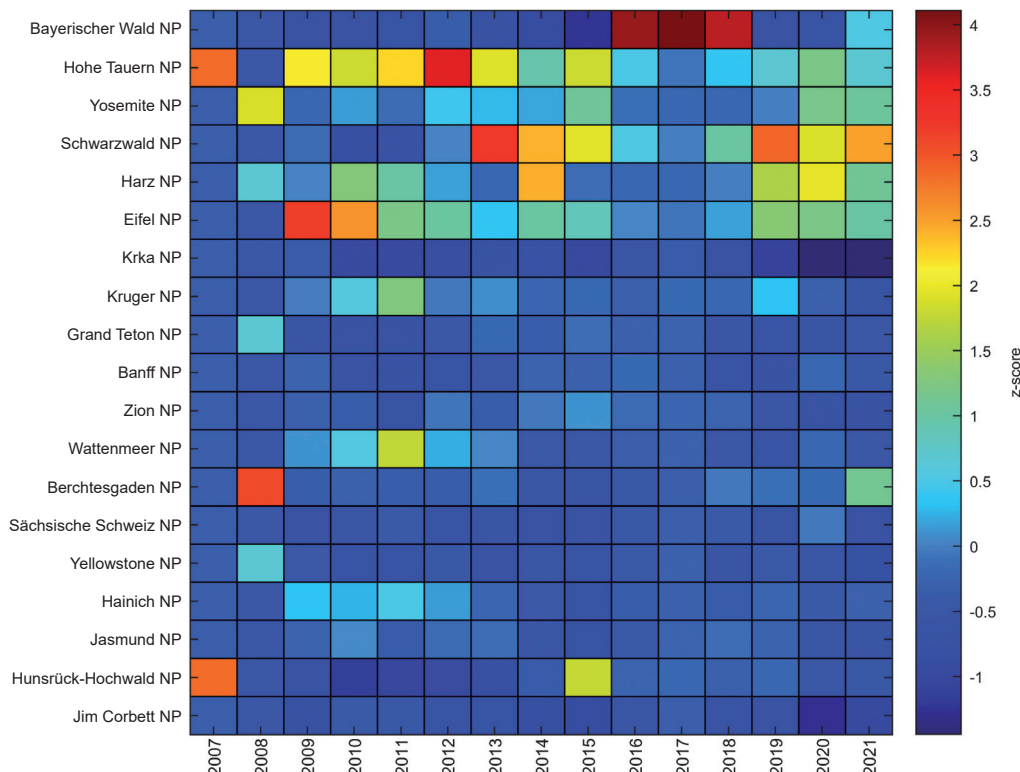


Figure 3 – Occurrence of national park (NP) names in tweets, 2007–2021 (national parks mentioned >1,000 times). The z-score is a measure of how many standard deviations you are away from the mean. 2006 was eliminated from this analysis because very few tweets were posted.

Table 3 – Emojis

Emoji (sentiment score [-1...+1])*	Count	Emoji (sentiment score [-1...+1])	Count
😊 Smiling Face with Heart-Eyes (0.678)	882	🌿 Herb (0.384)	165
📷 Camera (0.430)	803	🇺🇸 us Flag: United States (--)	159
🌳 Deciduous Tree (0.486)	506	🍁 Maple Leaf/Flag Canada (--)	152
❤️ Red Heart (0.746)	489	🇦🇹 at Flag: Austria (--)	146
🌲 Evergreen Tree (0.385)	480	🏔️ Snow-Capped Mountain (--)	141
😊 Smiling Face with Smiling Eyes (0.644)	404	🌊 Water Wave (0.500)	138
😉 Winking Face (0.463)	358	😊 Beaming Face Smiling Eyes (0.449)	138
📷 Camera with Flash (--)	355	🍂 Fallen Leaf (0.547)	130
👉 Backhand Index Pointing Right (0.390)	341	👌 OK Hand (0.563)	128
👍 Thumbs Up (0.521)	305	😊 Grinning Face (0.568)	125
➡️ Right Arrow (0.147)	282	🇨🇦 ca Flag: Canada (--)	123
😊 Face with Tears of Joy (0.221)	278	👣 Footprints (0.344)	121
😎 Smiling Face with Sunglasses (0.491)	268	📌 Round Pushpin (0.111)	118
☀️ Sun (0.465)	258	🍃 Leaf Fluttering in Wind (0.378)	116
🇩🇪 Flag: Germany (--)	254	☀️ Sun with Face (0.558)	114
🐘 Elephant (0.023)	243	🤔 Thinking Face (--)	111
💚 Green Heart (0.656)	209	➡️ Left Arrow (0.467)	105
🏞️ National Park (--)	193	🤗 Hugging Face (--)	105
❄️ Snowflake (0.506)	185	😱 Face Screaming in Fear (0.190)	102
♥️ Heart Suit (0.657)	182	💙 Blue Heart (0.730)	101
⛰️ Mountain (--)	182	😄 Grinning Face with Sweat (0.178)	99

* Sentiment scores are taken from Novak et al. (2015) and range from -1 to 1 , with -1 being the most negative possible sentiment and 1 the most positive. The textual meaning has been extracted from the emojiopedia database. (📷 Emojiopedia – 😊 Home of Emoji Meanings 🌿🍁🍂🏔️🌊👌👣📌🍃☀️🤔👉👍➡️💚🏞️❄️♥️⛰️ n.d.).

to other subjects. Frequencies of mountain-related hashtags were as follows: #mountain or #mountains: 318; #berge (mountains): 91; #bergtour (mountain tour): 26. Mountain-related uni-grams included *mountain* (1561) and *berg* (994). The full list of n-grams and hashtag frequencies can be found in Supplementary Material 2.

Finally, we describe the impact of the Covid-19 pandemic on the tweets in our corpus, considering the years 2019 (pre-Covid) and 2020 (during the pandemic) only. We found that tweet traffic decreased substantially (-10.6%) in 2020 (10,646 tweets) compared to 2019 (11,916). March 2020 had the fewest tweets in our historical series (507). In all tweets posted in 2020, only 447 (4.2%) contained our search words and also included the words *COVID**, *corona**, *SARS** and *pandemic**. The number of tweets posted each month before the outbreak of the pandemic (2019) and after (in 2020) can be seen in Supplementary Material 4.

Users' opinions changed throughout the pandemic. In February and March 2020, the comments referred mainly to the restrictions imposed on visiting NPs, e.g.:

"The #Tatra National Park will be closed from tomorrow until further notice due to the #Coronavirus..."

Several Twitter users highlighted the positive side of the pandemic for environmental conservation:

"One single, very positive effect I see with Corona, the chronic earth destroyers (tourists) no longer make a permanent pilgrimage across the globe for fun and if we now manage to turn Vienna-Schwechat into a national park, then the crisis really had a purpose!"

However, after the first strict lockdown many tweets contained messages about the large number of visitors to the NPs; other complaints related to overall nature conservation policies.

"#Corona brings #NationalPark #BlackForest visitor records @UmweltBW Franz Untersteller: "The National Park is booming, people enjoy and need recreation in #nature." https://t.co/EE7eKEHyj"

"The Tourist Wave. Leisure in the pandemic attracts masses of Germans to protected areas. Holidaymakers endanger rare animals, rangers are being threatened. Can the national parks counteract a collapse?++"

"Due to the Corona crisis, national parks lack important sources of revenue (visits, tours and school programmes) that are used to fulfil the mandate of #natureconservation, nature and environmental education and #research..."

After the restrictions were partially lifted in the summer months of 2020 and spring 2021, more tweets were posted concerning events and multimedia exhibitions. (Supplementary Material 3 gives a selection in the original language with English translations.)

Emojis

Users sometimes add emojis, emoticons and symbols to highlight their tweets. We found 1,039 different ones. Interestingly, the most-used one was "©" (Copyright) (1,498 times). The fact that it appears so frequently suggests that a large part of the tweet contents (mainly photographs) was copyrighted.

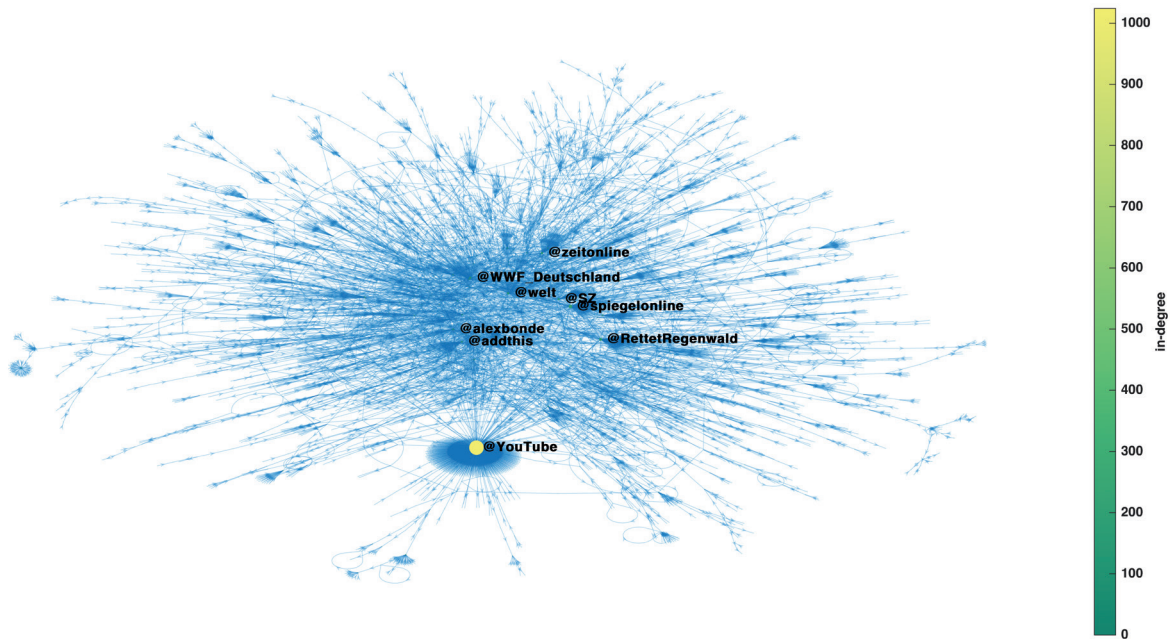


Figure 4 – Largest subnetwork of mentions appearing in tweets (8,519 nodes, 10,245 arcs). Only nodes receiving more than 60 mentions were tagged.

Our analysis focused only on emojis, not on emoticons. Table 3 shows the emojis that appeared in more than 99 tweets. Information on the meaning of each one and standard scores have been added (from Novak et al. 2015). Almost all of them refer to positive feelings. We have to go to the 70th position to find the first emoji with a negative sentiment score, 😏 Pouting Face (−0.173). Emojis related to the mountains were ranked in the middle of the frequency table (*Mountain* ranked 21st and *Snow-Capped Mountain* 26th).

Most active users and network of the most-mentioned user accounts in tweets

There were 90 users who posted more than 100 tweets over the years. The three most active accounts were @nlpschwarzwald (1,602), @NetBird (1,127) and @NpPartner (552). The first is a group of friends of the Schwarzwald NP, and the third is officially associated with a partner of the Bayerischen Wald. The @NetBird tweets were closely related to 3sat, a TV channel belonging to a consortium of German-speaking countries.

Remarkably, the top mentions were YouTube videos (in-degree, 1,772; pagerank, 5.31×10^{-2}), the daily spiegelonline (in-degree, 244; pagerank, 0.81×10^{-2}), and the non-profit association *Rettet den Regenwald e.V.* (in-degree, 197; pagerank, 0.66×10^{-2}). Other online newspapers and some other non-profit organizations appear less frequently. Alexander Bonde, general secretary of the German Federal Environment Foundation (DBU), appears among the most frequently mentioned. Figure 4 illustrates the relationships between Twitter user accounts.

Main tweet topics

Figure 5 shows the 30 largest groups of words that were most likely to appear together within a given topic. From left to right and top to bottom, they are ordered by probability of occurrence in the corpus (i. e. the words of each topic most likely to appear together in the total number of tweets collected). Supplementary Material 5 gives the 10 words associated with each topic and their probability of appearing together.

The first two topics show general words that logically fit in well with tweets on any subject. The rest of the topics relate mainly to the NPs mentioned in Table 2. Some new parks form a separate topic, although they did not appear so frequently – for example, Khao Sok NP in Thailand (topic 5) and Torres del Paine NP in Chile (topic 25). Perhaps most remarkable is the frequent presence of audiovisual content (topics 13, 17, 20, 27 and 29). We see a large number of words relating to photos and images of the parks taken by users, or to television programmes in which the parks were shown, for instance:

“*Gesäuse National Park, Styria Austria – Webvideo* <http://t.co/Hz3M0bvTVW> #national #park #nature #alps #mountains #styria #austria”

“*3sat.de aktuell: 14:05 Under the wings of the eagle - Kalkalpen National Park: After the great wildfire ...* <http://t.co/AHiQQCuWj3>”

Aspects of ecology and environmental conservation also appear in some topics. For example, we find demands for greater protection of NPs (topic 10), relating to Virunga NP and NPs in Ecuador, and the desire for a particular natural area to achieve NP status (topic 7).



Figure 5 – Main topics found. From left to right and from top to bottom, the topics most likely to appear in the corpus. The larger the word size, the higher the probability of appearing with the rest of the words in the topic.

“According to a recent #poll, 75 percent of citizens in the #Steigerwald region are in favour of a #national park. Approval is growing particularly strongly in the affected districts. #Forest #Nature #Preservation #Tourism”

There is also a group of tweets which celebrate certain anniversaries (topic 14), e.g.:

“Today, the Grand Canyon National Park turns 100 years old! 🎉😊 What is your favourite place in this beautiful park?”

Finally, some park names are associated commercially with the hospitality industry. These are tweets written as a form of advertising by certain tour operators, e.g.:

“Plitvice Lakes in Croatia: Tips for a holiday at the waterfalls: At the sight of the Plitvice Lakes National Park ...”

Selected tweets (original language + translations) related to the topics referred to above can be found in Supplementary Material 3.

Discussion and conclusions

In our study, we found a relatively small number of tweets about NPs compared to previous studies conducted in Finland (Heikinheimo et al. 2018), Nepal (Bhatt & Pickering 2021) and South Africa (Mangacheva & Pickering 2021). There are two possible explanations for this discrepancy: on the one hand, in earlier research, multilingual searches were used, and consequently the number of tweets was higher; on the other hand, the simplicity of our search strategy allowed us to reduce noise (i.e. unwanted results), but it may also have reduced the total number of relevant tweets.

Although the numbers of retweets and replies do not seem to be high, the number of likes is higher as a percentage. It should be noted that tweets containing images are more likely to be retweeted or to be rewarded with a like (Heikinheimo et al. 2018). There are many photographs, and links to videos were very common in our data.

The main Twitter users were from the former West Germany, in the Rheinland-Pfalz, Hessen and Baden-Württemberg regions. Like Scheffler et al. (2014), we found that Austria was under-represented in the number of tweets collected, despite the fact that one of its NPs was one of the most frequently mentioned in the data analysed. A high number of tweets were geolocated in southern England and northern France. This is understandable, since Great Britain is the non-German speaking country of choice for most Germans (≈142,000 people) living outside their country (German Federal Statistical Office n.d.). Nevertheless, it should be pointed out that our geolocation system is only an approximation, based on the addresses expressed in natural language by the users. The reader should therefore be cautious in the interpretation of the data, and should read the limitations at the end of this discussion.

The most frequently used uni-grams, bi-grams and tri-grams were words associated with the main NPs. This leads us to think that the contents of the tweets were mainly descriptive. Of the most frequently mentioned NPs, several are in Germany and only one in Austria. Our data agrees with a recent study by Sinclair et al. (2020), in which the most frequently mentioned parks in their study were also the most-visited ones. Factors that make a NP attractive (Puustinen et al.

2009; Siikamäki et al. 2015) were usually closely linked to their forests, scenery, biodiversity and the presence of water bodies. Most of the parks mentioned in our study stand out for their forests and/or their mountainous profile. In fact, the attraction of two of the top three parks (Yosemite NP and Hohe Tauern NP) is clearly based on their mountain scenery. The primeval forest located in a middle-high mountain range protected by the Bayerischer Wald NP is another example of an important nature-based attraction.

The German predilection for mountain parks was a preliminary finding, although it seems logical considering the easy accessibility of mountain protected areas in the main German-speaking countries. Surprisingly, however, terms, hashtags, emojis and topics directly related to mountains in general were comparatively rare in the Twitter content. More studies, using other research designs (e.g. quantitative surveys), would therefore be desirable to investigate actual appreciation of mountain environments and other environmental features. Several parks where water is the main feature (e.g. Wattenmeer NP in Germany, or Krka NP in Croatia) were also mentioned, while only one African park, a park known for its fauna (Kruger NP), appeared among the most frequently mentioned (Kruger et al. 2017).

Although most of the uni-grams referred to the names of NPs, others related to the conservation of ecosystems. It is important to note that German-speaking countries are pioneers in ecological claims and policies (Capra & Spretnak 1984), so this theme appears with some force. Interestingly, most of the tweets containing these words refer to NPs outside Central Europe. It seems that people in German-speaking countries perceive a greater danger to ecosystems when the parks are in countries with less strict regulations (Dahlberg et al. 2010) than those of the EU. However, there was also some criticism of local NPs.

Twitter is a perfect medium to express topics that people are concerned about. The analysis of Twitter users' opinions and moods has become a new decision-making tool for politicians and public assets managers (Segeber & Bennett 2011). Our data show that organizational accounts have a strong presence on Twitter: three of the most active accounts are institutional. It seems that organizations use Twitter for their official announcements or news. However, they could (or even should) also use it as a discussion forum where they can interact two-directionally with citizens (Feroz Khan et al. 2014).

We also quantified the most-used hashtags, and two of them obtained outstanding results. These were #nationalparkservice and #travel. The first refers to the US federal agency in charge of the management of NPs, national monuments and other protected sites (<https://www.nps.gov/index.htm>). Although the individual NPs usually have their own websites where the characteristics of the park are presented, the cen-

tralized management has been successful in attracting visitors, increasing revenue and improving employee satisfaction (Chung et al. 2010; Jones et al. 2017; Kranich et al. 1999). The second hashtag is a generic and stable one (Feng & Wang 2014) related to tourism (Park et al. 2016), showing the intention of Twitter users to visit parks far from home.

We also monitored the tweets on NPs and Covid-19 posted during 2020. Briefly, there are two large groups of tweets with opposing views. Some users show the positive side of the pandemic – for example, the fact that few people visited NPs during lockdown relieved pressure on the ecosystems. On the other hand, after lockdown was lifted, the higher numbers of visitors to NPs were perceived negatively, due to fear for the conservation of natural habitats.

There is already scientific evidence confirming Twitter users' perceptions of NPs during the pandemic (Bates et al. 2020; Miller-Rushing et al. 2021; Templeton et al. 2021), though in reality these perceptions were neither new nor unique to the pandemic (Gössling 1999; O'Reilly 1986). Although our data do not allow for an in-depth analysis of the consequences of the pandemic, both Twitter users' and experts' opinions indicate that the scenario created by the pandemic may be a good opportunity to rethink access to NPs in such a way as to improve conservation.

Our analysis also looked for the most common emojis that accompanied the main message, and the mentions of other sources of information in the text. Both mentions and emojis relate most frequently to photos and videos. YouTube and the © symbol associated with photos share the limelight. Although Twitter is based on sharing short text messages, users frequently used these texts to introduce audiovisual content. The intention behind a large proportion of the tweets about NPs was thus to show images of the parks; information, experiences and emotions associated with the NPs did not have great weight. Like Heikinheimo et al. (2018), we found that a substantial number of users shared content generated on other platforms, mainly YouTube (Pflugmacher et al. 2020).

Finally, we would like to highlight some important aspects of the topics identified. We used an LDA model to extract the most likely topics in our corpus, which delivered 30 topics. The topics included descriptions of the most important NPs, conservation of the environment, and the importance of the image as a form of expression. However, by grouping the topics together we were able to detect some issues more precisely.

Some users expressed concern about NPs in countries in which they perceived a lack of awareness about environmental conservation (e.g. Ecuador and Virunga NP). However, we also noted words associated with NPs more generally, and that parks often shared the same themes. The interpretation of the topics would go beyond the main objective of this article. Thus future studies should investigate the links between parks

(i.e. their shared topics) by using a more precise search strategy for specific NPs. Practical applications of these analyses range from the conservation and management of NPs to the promotion of tourist destinations attractive to the German-speaking market. Comparison with further languages other than English would also enrich and complement our results.

Our study has certain limitations that readers should bear in mind when interpreting our results. Our search strategy was rather restrictive (i.e. only the word nationalpark and its variations were used). This may have overestimated the relative importance of certain themes that focused on the description of parks. We also focused on the words that were most repeated over time. While this decision helped us to pinpoint the most common topics, it prevented us from identifying aspects that only appeared occasionally and which could be of public interest. Future studies should therefore focus on less frequently used terms to find minority opinions which could nevertheless provide useful knowledge. Of course, the 30 topics we found overlap to some extent, and this should be analysed in future work from a qualitative point of view. Researchers who would like to study our results in greater depth could usefully look for explanations for these associations or overlaps (e.g. topic 23 Berchtesgaden, Yellowstone and Joshua Tree parks), analysing the tweets containing these words to detect possible patterns that explain the coincidences. In addition, we believe that the analysis of just one social network could have limited, or biased, our conclusions. Future work should combine other social media networks that focus more on image-based messages (e.g. Instagram or Flickr) or GNSS/GPS-based Voluntary Geographic Information.

Our research, then, is a first attempt to characterize what Twitter users post in German about NPs. Most phrases found are associated with the names of the parks and with vacations. We found a clear interest in NPs in Central Europe and the United States. We also noticed that a large part of the content had multimedia links (e.g. photographs and videos). From our data, it can be deduced that the tweets seek to express verbally and to show places of special beauty. However, there is also increasing interest in aspects of ecology and environmental conservation. In future, analysis concerning the compatibility of multimedia content with conservation objectives would be desirable.

Several management implications can be derived from our study. The first is that Twitter in German-speaking societies is used by associations and institutions to post information, but there is little interaction between users. As posted information rarely elicits replies from other Twitter users, there is little enrichment of the original content by way of posting new opinions. Discussion of priority issues (e.g. environmental protection) should be encouraged in order to obtain the real opinions of the people who are interested in a given NP. The risks of the uncontrolled promotion

of protected areas and its negative consequences have recently been reported in the literature (Gretzel 2019; Silk et al. 2017), and it is therefore also important to balance promotion in social media and its effects on the physical environment. Digital information should be compatible with nature protection objectives. Finally, the German-speaking public on Twitter clearly prefers mountain NPs. Although this interest could potentially increase discussion related to environmental education, nature protection, geology, ecology, fauna and flora, physical activity, public health and wellbeing and other strategic objectives of the protected areas, curiously, the tweets focus mainly on the parks' scenic values. If the authorities were to repeatedly associate a NP's brand image with concepts that go beyond mass tourism, it might influence positively the ways in which the parks are used and enjoyed.

Acknowledgements

The authors would like to express their gratitude to Twitter for providing access to the data. This study was supported by a research grant from the University of Valencia, Spain (Ayudas para estancias de investigación: BEST/2021/095). Luis-Millán González gratefully acknowledges the support of the Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital under Project PROMETEO/2021/026.

References

- Emojiopedia—☺ Home of Emoji Meanings 🧑🏻 🏠 🌳 🌞 (n.d.). Available at: <https://emojipedia.org/> (accessed: 18/8/2021)
- Baek, E.C., R. Hyon, K. López, E.S. Finn, M.A. Porter & C. Parkinson 2022. In-degree centrality in a social network is linked to coordinated neural activity. *Nature communications* 13(1): 1–13. Doi: 10.1038/s41467-022-28432-3
- Bates, A.E., R.B. Primack, P. Moraga & C.M. Duarte 2020. COVID-19 pandemic and associated lockdown as a “Global Human Confinement Experiment” to investigate biodiversity conservation. *Biological Conservation* 248: 108665. Doi: 10.1016/j.biocon.2020.108665
- Batrinca, B. & P.C. Treleaven 2015. Social media analytics: A survey of techniques, tools and platforms. *Ai & Society* 30(1): 89–116. Doi: 10.1007/s00146-014-0549-4
- Bhatt, P. & C.M. Pickering 2021. Public Perceptions about Nepalese National Parks: A Global Twitter Discourse Analysis. *Society & Natural Resources* 34(6): 685–702. Doi: 10.1080/08941920.2021.1876193
- Bielański, M., K. Taczanowska, A. Muhar, P. Adamski, L.-M. González & Z. Witkowski 2018. Application of GPS tracking for monitoring spatially unconstrained outdoor recreational activities in protected areas – A case study of ski touring in the Tatra National Park, Poland. *Applied Geography* 96: 51–65. Doi: 10.1016/j.apgeog.2018.05.008

- Blei, D.M., A.Y. Ng & M.I. Jordan 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3(Jan): 993–1022.
- Borodin, A., G.O. Roberts, J.S. Rosenthal & P. Tsapras 2005. Link analysis ranking: Algorithms, theory, and experiments. *ACM Transactions on Internet Technology* 5(1): 231–297. Doi: 10.1145/1052934.1052942
- Braschler, M. & B. Ripplinger 2004. How Effective is Stemming and Decompounding for German Text Retrieval? *Information Retrieval* 7(3): 291–316. Doi: 10.1023/B:INRT.0000011208.60754.a1
- Bucur, C. 2015. Using Opinion Mining Techniques in Tourism. *Procedia Economics and Finance* 23: 1666–1673. Doi: 10.1016/S2212-5671(15)00471-2
- Büschken, J. & G.M. Allenby 2016. Sentence-Based Text Analysis for Customer Reviews. *Marketing Science*, 35(6): 953–975. Doi: 10.1287/mksc.2016.0993
- Capra, F. & C. Spretnak 1984. Green politics. *National Forum* 64(3): 21.
- Cessford, G. & A. Muhar 2003. Monitoring options for visitor numbers in national parks and natural areas. *Journal for Nature Conservation* 11(4): 240–250. Doi: 10.1078/1617-1381-00055
- Chung, J.Y., C.S. Jung, G.T. Kyle & J.F. Petrick 2010. Justice in the U.S. National Park Service: The Antecedents of Job Satisfaction. *Journal of Park and Recreation Administration* 28(3), Article 3. Available at: <https://js.sagamorepub.com/jpra/article/view/1371> (accessed: 18/05/2022)
- Dahlberg, A., R. Rohde & K. Sandell 2010. National Parks and Environmental Justice: Comparing Access Rights and Ideological Legacies in Three Countries. *Conservation and Society* 8(3): 209–224.
- Data Europa EU 2012. *Special Eurobarometer 386: Europeans and their Languages*. Available at: https://data.europa.eu/data/datasets/s1049_77_1_ebs386?locale=fr (accessed: 06/01/2022)
- Di Minin, E., H. Tenkanen & T. Toivonen 2015. Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science* 3: 63. Doi: 10.3389/fenvs.2015.00063
- Eagles, P.F.J., S.F. McCool & C.D. Haynes 2002. *Sustainable Tourism in Protected Areas: Guidelines for Planning and Management*. IUCN.
- European Commission 2017. *Statistical Annex of European Economy*. Available at: https://ec.europa.eu/economy_finance/db_indicators/statistical_annex/index_en.htm (accessed: 06/01/2022)
- Feng, W. & J. Wang 2014. We can learn your #hashtags: Connecting tweets to explicit topics. *2014 IEEE 30th International Conference on Data Engineering*: 856–867. Doi: 10.1109/ICDE.2014.6816706
- Feroz Khan, G., H. Young Yoon, J. Kim & H. Woo Park 2014. From e-government to social government: Twitter use by Korea's central government. *Online Information Review* 38(1): 95–113. Doi: 10.1108/OIR-09-2012-0162
- German Federal Statistical Office. n.d. *Deutsche mit Wohnsitz im europäischen Ausland 2020*. Available at: https://www.destatis.de/Europa/EN/Topic/Population-Labour-Social-Issues/Demography-migration/Deutsche_EUAusland.html (accessed: 23/8/2021)
- Ghermandi, A. & M. Sinclair 2019. Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change* 55: 36–47. Doi: 10.1016/j.gloenvcha.2019.02.003
- Giglio, S., F. Bertacchini, E. Bilotta & P. Pantano 2020. Machine learning and points of interest: Typical tourist Italian cities. *Current Issues in Tourism* 23(13): 1646–1658. Doi: 10.1080/13683500.2019.1637827
- González, L.M., J. Devís-Devís, M. Pellicer-Chenoll, M. Pans, A. Pardo-Ibañez, X. García-Massó & V. Pérez-Samaniego 2021. The impact of COVID-19 on sport in Twitter: A quantitative and qualitative content analysis. *International journal of environmental research and public health* 18(9): 4554. Doi: 10.3390/ijerph18094554
- Gössling, S. 1999. Ecotourism: A means to safeguard biodiversity and ecosystem functions? *Ecological Economics* 29(2): 303–320.
- Gretzel, U. 2019. *The role of social media in creating and addressing overtourism*. In: Dodds, R. & R. Buttler, *Overtourism - Issues, realities and solutions*. De Gruyter Studies in Tourism (Band 1). Doi: 10.1515/9783110607369-005
- Hartmann, R. 1988. Combining field methods in tourism research. *Annals of Tourism Research* 15(1), 88–105. Doi: 10.1016/0160-7383(88)90073-4
- Heikinheimo, V., H. Tenkanen, T. Hiiippala & T. Toivonen 2018. *Digital Imaginations of National Parks in Different Social Media: A Data Exploration*. Doi: 10.5281/zenodo.1472745
- IUCN 2016. Category II. IUCN. Available at: <https://www.iucn.org/theme/protected-areas/about/protected-areas-categories/category-ii-national-park> (accessed: 23/8/2021)
- Jianqiang, Z. & G. Xiaolin 2017. Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access* 5: 2870–2879.
- Jones, C., N. Shipley & S. Ul-Hasan 2017. Bringing Parks Back to the People: Revisiting the Dual Mandate and Core Values of the National Park Service. *The George Wright Forum* 34(1): 45–52.
- Kalvet, T., M. Olesk, M. Tiits & J. Raun 2020. Innovative Tools for Tourism and Cultural Tourism Impact Assessment. *Sustainability* 12(18): 7470. Doi: 10.3390/su12187470
- Krannich, R.S., B.W. Eisenhauer, D.R. Field, C. Pratt & A.E. Luloff 1999. Implications of the National Park Service Recreation Fee Demonstration Program for Park Operations and Management: Perceptions of NPS Managers. *Journal of Park and Recreation Administration* 17(3): Article 3. Available at: <https://js.sagamorepub.com/jpra/article/view/1632> (accessed: 18/05/2022)
- Kruger, M., A. Viljoen & M. Saayman 2017. Who visits the Kruger National Park, and why? Identifying target markets. *Journal of Travel & Tourism Marketing* 34(3): 312–340. Doi: 10.1080/10548408.2016.1156618

- Li, R., J. Crowe, D. Leifer, L. Zou & J. Schoof 2019. Beyond big data: Social media challenges and opportunities for understanding social perception of energy. *Energy Research & Social Science* 56: 101217. Doi: 10.1016/j.erss.2019.101217
- Litvak, M. & N. Vanetik 2019. *Multilingual Text Analysis: Challenges, Models, And Approaches*. World Scientific.
- Mangachena, J.R. & C.M. Pickering 2021. Implications of social media discourse for managing national parks in South Africa. *Journal of Environmental Management* 285: 112159. Doi: 10.1016/j.jenvman.2021.112159
- Miller-Rushing, A.J., N. Athearn, T. Blackford, C. Brigham, L. Cohen, R. Cole-Will, T. Edgar, E.R. Ellwood, N. Fisichelli & C.F. Pritz 2021. COVID-19 pandemic impacts on conservation research, management, and public engagement in US national parks. *Biological Conservation* 257: 109038.
- Novak, P.K., J. Smailović, B. Sluban & I. Mozetič 2015. Sentiment of Emojis. *PLOS ONE* 10(12): e0144296. Doi: 10.1371/journal.pone.0144296
- O'Reilly, A.M. 1986. Tourism carrying capacity: Concept and issues. *Tourism Management* 7(4): 254–258. Doi: 10.1016/0261-5177(86)90035-X
- Park, S.B., C.M. Ok & B.K. Chae 2016. Using Twitter Data for Cruise Tourism Marketing and Research. *Journal of Travel & Tourism Marketing* 33(6): 885–898. Doi: 10.1080/10548408.2015.1071688
- Pflugmacher, J., S. Escher, J. Reubold & T. Strufe 2020. The German-Speaking Twitter Community Reference Data Set. *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*: 1172–1177. Doi: 10.1109/INFOCOMWKSHPS50562.2020.9162728
- Pickering, C.M. & P. Norman 2020. Assessing discourses about controversial environmental management issues on social media: Tweeting about wild horses in a national park. *Journal of Environmental Management* 275: 111244. Doi: 10.1016/j.jenvman.2020.111244
- Puustinen, J., E. Pouta, M. Neuvonen & T. Sievänen 2009. Visits to national parks and the provision of natural and man-made recreation and tourism resources. *Journal of Ecotourism* 8(1): 18–31. Doi: 10.1080/14724040802283210
- Scheffler, T., J. Gontrum, M. Wegel & S. Wendler 2014. *Mapping German tweets to Geographic Regions*. Available at: <https://hildok.bsz-bw.de/frontdoor/index/index/docId/298> (accessed: 18/05/2022)
- Seegerberg, A. & W.L. Bennett 2011. Social Media and the Organization of Collective Action: Using Twitter to Explore the Ecologies of Two Climate Change Protests. *The Communication Review* 14(3): 197–215. Doi: 10.1080/10714421.2011.597250
- Siikamäki, P., K. Kangas, A. Paasivaara & S. Schroderus 2015. Biodiversity attracts visitors to national parks. *Biodiversity and Conservation* 24(10): 2521–2534. Doi: 10.1007/s10531-015-0941-5
- Silk M.J., S.L. Crowley, A.J. Woodhead & A. Nuno 2017. Considering connections between Hollywood and biodiversity conservation. *Conservation Biology* 32(3): 597–606. Doi: 10.1111/cobi.13030
- Sinclair, M., M. Mayer, M. Woltering & A. Ghermandi 2020. Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management* 263: 110418. Doi: 10.1016/j.jenvman.2020.110418
- Sloan, L. & A. Quan-Haase 2017. *The SAGE Handbook of Social Media Research Methods*. SAGE.
- Starosta, K., S. Budz & M. Krutwig 2019. The impact of German-speaking online media on tourist arrivals in popular tourist destinations for Europeans. *Applied Economics* 51(14): 1558–1573. Doi: 10.1080/00036846.2018.1527463
- Teles da Mota, V. & C. Pickering 2020. Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism* 30: 100295. Doi: 10.1016/j.jort.2020.100295
- Teles da Mota, V. & C. Pickering 2021. Geography of Discourse about a European Natural Park: Insights from a Multilingual Analysis of tweets. *Society & Natural Resources* 34(11): 1492–1509. Doi: 10.1080/08941920.2021.1971809
- Templeton, A.J., K. Goonan & A. Fyall 2021. COVID-19 and its impact on visitation and management at US national parks. *International Hospitality Review*. Doi: 10.1108/IHR-08-2020-0039
- Tenkanen, H., E. Di Minin, V. Heikinheimo, A. Hausmann, M. Herbst, L. Kajala & T. Toivonen 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports* 7(1): 17615. Doi: 10.1038/s41598-017-18007-4
- Toivonen, T., V. Heikinheimo, C. Fink, A. Hausmann, T. Hiippala, O. Järvi, H. Tenkanen & E. Di Minin 2019. Social media data for conservation science: A methodological overview. *Biological Conservation* 233: 298–315. Doi: 10.1016/j.biocon.2019.01.023
- Udyapuram, H.P. & S. Gavirneni 2019. Big Data Improves Visitor Experience at Local, State, and National Parks – Natural Language Processing Applied to Customer Feedback. In: Emrouznejad, A. & V. Charles (eds.), *Big Data for the Greater Good*: 171–181. Doi: 10.1007/978-3-319-93061-9_8
- Wilkins, E.J., S.A. Wood & J.W. Smith 2021. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environmental Management* 67(1): 120–132. Doi: 10.1007/s00267-020-01373-7
- Zierke, R. n.d. *Resources*. Available at: <https://www.inf.uni-hamburg.de/en/inst/ab/lt/resources.html> (accessed: 29/8/2021)

Authors

Maite Pellicer-Chenoll

is Assistant Professor in the Physical Education Department, University of Valencia (Spain). Her main research areas are human behaviour in outdoor environments, and the effects of physical activity on quality of life in society. She also has expertise in motor control in patients with musculoskeletal problems, and the influence of gender in the practice of sport and physical activity. Department of Physical Education and Sport, University of Valencia, Valencia, Spain.

Karolina Taczanowska – corresponding author

is a senior scientist at the Institute of Landscape Development, Recreation and Conservation Planning, University of Natural Resources and Life Sciences (BOKU) in Vienna, Austria. Her research focuses on sustainable tourism and outdoor recreation, with a specific interest in the spatial and temporal behaviour of visitors in natural environments. She is currently investigating the potential of big data resources (social media, VGI) to better understand human–environment relationships in outdoor recreation. Institute of Landscape Development, Recreation and Conservation Planning, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria. E-mail: karolina.taczanowska@boku.ac.at

Pilar Serra-Añó

is Professor at the Department of Physiotherapy, University of Valencia, Spain. She has extensive expertise in physical exercise as a therapeutic resource, and understanding environmental factors in promoting physical activity in the context of human health and wellbeing. Department of Physiotherapy, University of Valencia, Valencia, Spain

José L. Toca-Herrera

is Professor of biophysics and head of the Institute of Biophysics at the University of Natural Resources and Life Sciences (BOKU), Vienna. His research interests include the application of basic physical and mathematical concepts in interdisciplinary research fields. Institute of Biophysics, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Luis-Millán González

is an Associate Professor at the University of Valencia (Spain) and head of the Human Movement Analysis research unit. He currently teaches Physical Exercise and Health. His main line of research focuses on the quantification of physical activity in different populations and environments. Within this context, he has specialized in the analysis of large datasets (data mining) and textual data (text mining). Department of Physical Education and Sport, University of Valencia, Valencia, Spain.